ACCSCIENCE
PUBLISHING

Artificial Intelligence in Health

ISSN: 3041-0894 (Print)
ISSN: 3029-2387 (Online)
Volume 2 · Issue 3
July 2025

ACCSCIENCE
PUBLISHING

Artificial Intelligence in Health

Print ISSN: 3041-0894
Online ISSN: 3029-2387

Artificial Intelligence in Health aims to provide a freely accessible multidisciplinary and comprehensive platform for researchers, scientists, and AI in health and medicine sciences practitioners to publish and exchange cutting-edge advancements, insights, technological development and innovations at the intersection of artificial intelligence (AI) and health. The journal seeks to explore the transformative potential of AI in improving and understanding health and medicine research outcomes, enhancing clinical decision-making, optimizing resource allocation, and addressing various challenges in the multidisciplinary field of health.

About the Publisher

AccScience Publishing is a publishing company based in Singapore. We publish a range of high-quality, open-access, peer-reviewed journals and books from a broad spectrum of disciplines.

Contact Us

Managing Editor
aih.office@accscience.sg

AccScience Publishing
9 Raffles Place, Republic Plaza 1 #06-00 Singapore 048619.

Volume 2 • Issue 3 • July 2025
ISSN 3041-0894 (print) ISSN 3029-2387 (online)

ARTIFICIAL INTELLIGENCE IN HEALTH

Editor-in-Chief
Andrzej Cichocki
Systems Research Institute of Polish Academy of Science, Poland

ACCSCIENCE PUBLISHING
Access Science Without Barriers

ACCSCIENCE
PUBLISHING

Full issue copyright © 2025 AccScience Publishing
All rights reserved. Without permission in writing from the publisher, this full issue publication in its entirety may not be reproduced or transmitted for commercial purposes in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system. Permissions may be sought from aih.office@accscience.sg.

Article copyright © Respective Author(s)
See articles for copyright year. All articles in this full issue publication are open-access. There are no restrictions in the distribution and reproduction of individual articles, provided the original work is properly cited. However, permission to reuse copyrighted materials of an article for commercial purposes is applicable if the article is licensed under Creative Commons Attribution-NonCommercial License. Check the specific license before reusing.

Artificial Intelligence in Health

ISSN: 3041-0894 (print)
ISSN: 3029-2387 (online)

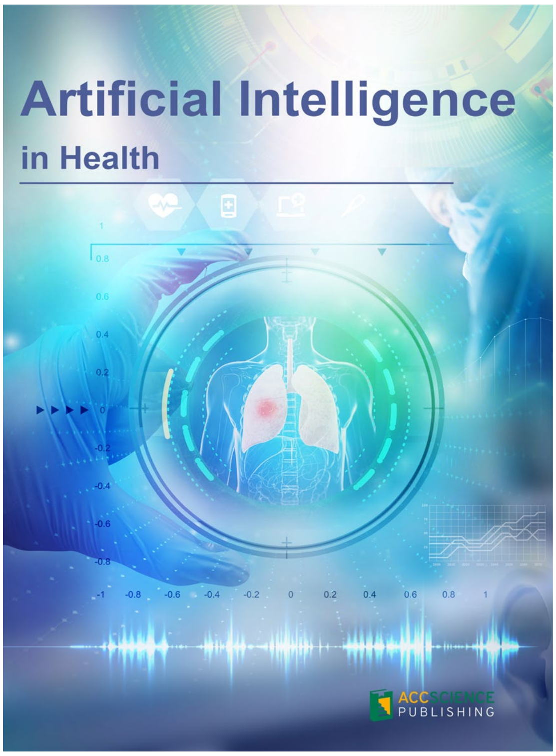
Editorial and Production Credits

Publisher: AccScience Publishing
Managing Editor: Freda Wang
Production Editor: Sharmila Velapasamy
Article Layout and Typeset: Sinjore Technologies (India)

For all advertising queries, contact aih.office@accscience.sg.

Supplementary file

Supplementary files of articles can be obtained at https://accscience.com/journal/AIH/2/3.



Disclaimer

AccScience Publishing is not liable to the statements, perspectives, and opinions contained in the publications. The appearance of advertisements in the journal shall not be construed as a warranty, endorsement, or approval of the products or services advertised and/or the safety thereof. AccScience Publishing disclaims responsibility for any injury to persons or property resulting from any ideas or products referred to in the publications or advertisements. AccScience Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Artificial Intelligence in Health

Editorial Board

Editor-in-Chief
Andrzej Cichocki, *Poland*

Executive Editors
Adrian David Cheok, *China*
Hongcai Shang, *China*
Xiaobo Zhou, *USA*

Associate Editors
Weiping Ding, *China*
Ruiheng Zhang, *China*

*Editorial Board Members**
Adel Al-Jumaily, *Australia*
Zeeshan Ali, *China*
Ahmed Bouridane, *UAE*
Joaquim Carreras, *Japan*
Oscar Castillo, *Mexico*
Faouzi Alaya Cheikh, *Norway*
Xiaojun Chen, *China*
Xiaochun Cheng, *UK*
Krzysztof Jozef Cios, *USA*
Alfredo Cuzzocrea, *Italy*
Di Dong, *China*
Anastasios Dounis, *Greece*
Włodzislaw Duch, *Poland*
Ayman El-Baz, *USA*
Adel Elmaghraby, *USA*
Manuel F.G. Penedo, *Spain*
Runwei Guan, *China*
Rémy Guillevin, France
Andrew A. Gumbs, *France*
Pankaj Gupta, *India*
A. Ben Hamza, *Canada*
Alexander Hramov, *Russia*
Bin Hu, *China*
Yisen Huang, *China*
Donato Impedovo, Italy
S. M. Riazul Islam, *UK*
Ankush D. Jamthikar, *India*
Jay Kalra, *Canada*
Uzay Kaymak, *Netherlands*
Fahmi Khalifa, *USA*
Antonio Lanata, *Italy*
Xueping Li, USA
Zihuai Lin, *Australia*
Wing-Kuen Ling, *China*
Haipeng Liu, *UK*
Nicola Luigi Bragazzi, *Canada*
Xiaoke Ma, *China*
Xuelei Ma, *China*
George D. Magoulas, *UK*
Mrinal Mandal, *Canada*
Francesco Mercaldo, *Italy*
Reza Mirnezami, *UK*
Jianwei Niu, *China*
George Notas, *Greece*
JungHwan Oh, USA
Peichen Pan, *China*
Witold Pedrycz, *Canada*
Alexander N. Pisarchik, *Spain*
Dawid Polap, *Poland*
Mihail Popescu, *USA*
Mukesh Prasad, *Australia*
Marek Reformat, *Poland*
Hongliang Ren, *China*
Hassan Rivaz, *Canada*
José Santamaría López, *Spain*
Paulo Adriano Schwingel, Brazil
Wei Shao, *China*
Chao Shen, *China*
Patricia A. Shewokis, *USA*
Qiongfeng Shi, *China*
Ali Hassan Sodhro, *Sweden*
L. Stergioulas, *Netherlands*
Jasjit S. Suri, *USA*
Kenji Suzuki, *Japan*
Abdelmalik TALEB-AHMED, *France*
Sukun Tian, *China*
Erfan Babaee Tirkolaee, *Turkey*
Miguel Garcia Torres, *Spain*
Igor Tsigelny, *USA*
Ricardo Vardasca, *Portugal*
Eugenio Vocaturo, *Italy*
Alan Wang, *New Zealand*
Guotai Wang, *China*
Yanfeng Wang, *China*
Fangxiang Wu, *Canada*
Jian Yang, *China*
Qi Yang, *China*
Zhewei Ye, *China*
Xujiong Ye, *UK*
Hui Yu, *UK*
Yudong Zhang, *UK*
Yu Zhang, *USA*
Wensheng Zhang, *China*
Zhuhuang Zhou, *China*
Shang-Ming Zhou, *UK*

*Youth Editorial Board Members**
Yankai Chen, *USA*
Qiong Chen, *China*
Sibo Cheng, *France*
Bu Chenyang, *China*
Afify Heba, Egypt
Jiashuang Huang, *China*
Hongxin Pan, *China*
Yuchen Pan, *China*
Shuo Wang, *China*

*Editorial Board Members as of July 11, 2025

ACCSCIENCE PUBLISHING

Volume 2 | Issue 3 | July 2025 | Page 1 – 160 |

Artificial Intelligence in Health

https://accscience.com/journal/AIH

CONTENTS

REVIEW ARTICLES

PERSPECTIVE ARTICLE

ORIGINAL RESEARCH ARTICLES

MINI-REVIEW

ACCSCIENCE PUBLISHING

Artificial Intelligence in Health

REVIEW ARTICLE

Advancing embryo selection in artificial intelligence-assisted reproductive technologies: A systematic review

Md. Abul Basar Roky, Anonno Singha Ray, and Asim Moin Saad*

Department of Mechatronics Engineering, Rajshahi University of Engineering & Technology, Kazla, Rajshahi, Bangladesh

Abstract

For couples encountering infertility challenges, assisted reproductive technologies (ARTs) offer a path to parenthood. ART procedures, such as *in vitro* fertilization (IVF), intracytoplasmic sperm injection (ICSI), and embryo implantation, involve the handling of sperm or embryos outside the body. However, the success of ART depends on the accurate selection of viable embryos. Artificial intelligence (AI) is a promising tool with the potential to revolutionize these procedures. This review explores the transformative potential of AI in ART, providing valuable insights into enhanced embryo selection and unlocking new possibilities for the field. Four electronic databases were systematically searched under the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines. From an initial pool of 914 papers, 30 studies were selected for further evaluation. While noting the limitations inherent in the existing body of research, this review offers a broad analysis of AI's transformative role in embryo selection. It highlights the significant potential of AI to enhance precision, consistency, and efficiency in ART. This review also emphasizes the importance of addressing technical, ethical, and regulatory aspects to ensure responsible and effective integration of these technologies. The findings indicate that AI-based models, such as the iDAScore v2.0, have demonstrated promising results in accurately predicting embryo viability and evaluating the effects of maternal age on embryo viability. Specifically, Bayesian network modeling, with an accuracy rate of 91.3%, aims to optimize IVF and ICSI procedures. In summary, AI stands at the forefront of innovation in ART, offering new hope through more accurate and efficient embryo selection.

Keywords: Artificial intelligence; Machine learning; Deep learning; Embryo selection; Assisted reproductive technologies

***Corresponding author:**
Asim Moin Saad
(1808036@student.ruet.ac.bd)

Citation: Roky MAB, Ray AS, Saad AM. Advancing embryo selection in artificial intelligence-assisted reproductive technologies: A systematic review. *Artif Intell Health*. 2025;2(3):1-23. doi: 10.36922/aih.7170

Received: December 10, 2024

Revised: February 21, 2025

Accepted: March 3, 2025

Published online: May 2, 2025

Copyright: © 2025 Author(s). This is an Open-Access article distributed under the terms of the Creative Commons Attribution License, permitting distribution, and reproduction in any medium, provided the original work is properly cited.

Publisher's Note: AccScience Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

Once considered a private matter, infertility has become a globally recognized issue, affecting millions of couples worldwide. The burden of infertility has been increasing globally and regionally for both males and females. Infertility affects one in six adults worldwide, with higher rates reported in the Americas and the Western Pacific region.[1]

The global infertility rate has shown fluctuating trends over the past few decades. In high-income and developed countries, the prevalence of primary and secondary infertility has been decreasing, potentially due to greater access to infertility treatment facilities.[2] Assisted reproductive technologies (ARTs) hold promise for struggling couples with infertility. It can be a common approach to the problem in the future.[3,4]

ART techniques, such as intracytoplasmic sperm injection (ICSI) and *in vitro* fertilization (IVF), are frequently utilized to assist infertile couples in getting pregnant. The procedures involve the retrieval of eggs, laboratory fertilization with sperm, transfer of viable embryos into the uterus, and control of ovarian stimulation. The viability of ART is highly dependent on the quality of gametes and embryos, which are conventionally evaluated subjectively by embryologists based on morphological criteria.

Since the birth of Louise Brown, the first infant conceived through IVF, ART has undergone significant advancements aimed at reducing complications and improving outcomes.[5] The integration of artificial intelligence (AI) into ART holds great promise for enhancing outcomes. AI technologies, such as computer-assisted sperm analysis and machine learning (ML) algorithms, enable the objective evaluation of semen parameters and embryo quality. By standardizing evaluations and processing large volumes of data, AI has the potential to enhance treatment outcomes and increase conception rates.[6-8] However, despite its groundbreaking breakthroughs, only 30% of ART treatments result in conception, highlighting the necessity for more accurate predictive models.[9] As the process involves manipulating human gametes or embryos *in vitro*, ART outcomes are influenced by multiple complex factors, including the cause of infertility, age, hormonal profile, and laboratory conditions. Advanced technologies, such as AI and ML, are being explored to enhance prediction accuracy and decision-making in ART,[10] with promising results in predicting IVF cycle outcomes and guiding embryo selection.[11] These systems interpret data using image-based analysis to provide clinically relevant recommendations,[12] and AI models are also being developed to classify reproductive data, such as embryonic development and semen characteristics.[13] As[4,11-13] AI continues to demonstrate potential in improving diagnostic and therapeutic processes in reproductive medicine, its adoption in fertility clinics is likely to increase.[14] Nonetheless, challenges remain regarding the generalizability and standardization of AI applications in ART.[4]

This systematic review aims to identify and map the current landscape of research on embryo selection, focusing on advancements and innovations in AI-based ARTs. In Figure 1, the flow diagram depicts the integration of AI-based ART into the embryo selection process. Key input data include genetic profiles, historical success rates, and medical histories. Using deep learning and ML methods, AI systems analyze this data to predict embryo viability. Selected embryos are then processed using AI-based ART. Pregnancy outcomes are tracked, enabling continuous refinement and optimization of the AI systems. Ethical considerations and regulatory compliance are crucial at every stage of this process. This systematic review explores the following research queries (RQs):

- RQ1. What are the current state-of-the-art AI technologies used in embryo selection for ART?
- RQ2. To what extent does ART improve the pregnancy success rate and live birth outcomes compared to traditional methods?
- RQ3. How can AI algorithms be seamlessly integrated into existing embryo selection protocols and laboratory workflows to leverage the expertise of embryologists and healthcare professionals?
- RQ4. What is the impact of AI-driven embryo selection on the psychological health and decision-making processes of prospective parents, particularly in light of ethical concerns?
- RQ5. What are the primary barriers and limitations to the clinical application of AI algorithms in embryo selection, and how are these technologies being developed and validated?

2. Research methodology

This section outlines the research design and analytical procedures used in the present study.

2.1. Overview

According to McKenzie *et al.*,[15] a systematic review is a rigorous, structured method for identifying, evaluating, and synthesizing research evidence on a specific question using predefined protocols. It minimizes bias through comprehensive literature searches and transparent processes, often incorporating meta-analysis to quantitatively combine study findings. Consequently, this systematic review was conducted to investigate the technologies utilized in AI-guided embryo selection and to map the current landscape of advancements within ART.

2.2. Objectives

The key objectives of this systematic review include:

(i) To examine current AI applications in embryo selection and analyze the success rates of various ML models used in ART
(ii) To identify potential future improvements, innovations, and existing research gaps in the application of AI for embryo selection

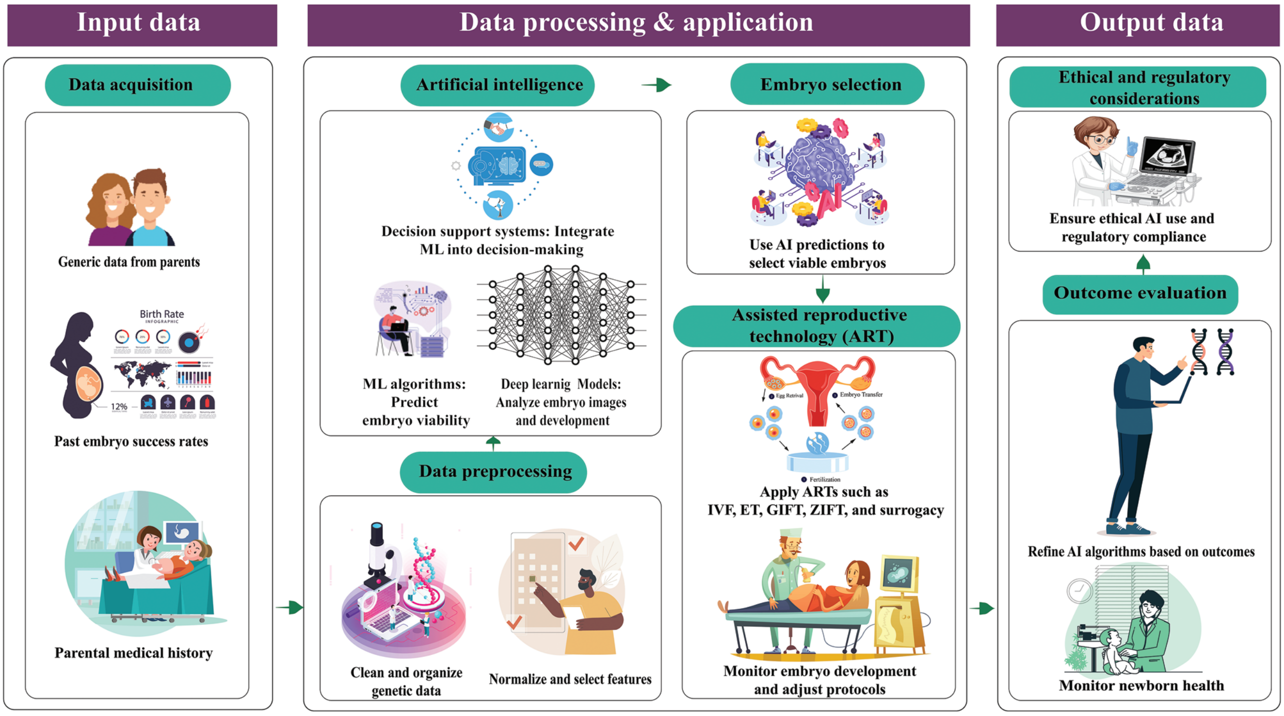


Figure 1. Flow diagram of artificial intelligence-based assisted reproductive technology
Abbreviations: ET: Embryo transfer; GIFT: Gamete intrafallopian transfer; IVF: *In vitro* fertilization; ML: Machine learning; ZIFT: Zygote intrafallopian transfer.

(iii) To investigate the ethical implications associated with the integration of AI in embryo selection processes.

2.3. Literature selection

This study strictly adhered to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines, including the extension for scoping reviews. Four electronic databases (Scopus, PubMed, IEEE Xplore, and Google Scholar) were systematically searched to identify relevant publications. The literature search was performed by three researchers (the authors) between December 2023 and April 2024. Duplicate entries were identified and removed using Python programming. The search results were initially compiled in an Excel file, ensuring the accuracy and uniqueness of the dataset for further analysis. Each researcher reviewed the titles and abstracts based on predefined inclusion and exclusion criteria to determine eligibility. In case of uncertainty or disagreement, a well-known academic expert was consulted to aid in the final selection of publications.

2.4. Inclusion criteria

Studies were included in this review if they met the following criteria:

(i) Studies that explore or utilize deep learning, ML, or AI to select embryos for ART
(ii) Studies that use rigorous research procedures and provide transparent data analysis and reporting
(iii) Research that tackles ethical considerations such as informed consent, data privacy, and societal implications associated with AI-based embryo selection
(iv) Studies that focus on AI applications specifically relevant to IVF techniques
(v) Publications including book chapters, conference proceedings, systematic reviews, meta-analyses, and peer-reviewed research articles are all covered
(vi) Research articles that primarily focused on human subjects
(vii) Articles published in the English language
(viii) Studies published between June 1, 2015, and January 9, 2024.

2.5. Exclusion criteria

Studies were excluded from this review if they met any of the following criteria:

(i) Studies not primarily focused on the use of AI for embryo selection or its role in ART
(ii) Studies lacking sufficient detail or clarity on the methods used for embryo selection
(iii) Research addressing ethical concerns that are unrelated to the application of AI in embryonic selection

(iv) Articles that do not meaningfully discuss AI application within IVF procedures
(v) Studies conducted on laboratory animals
(vi) Non-peer-reviewed sources, such as editorials, opinions, and non-scholarly articles
(vii) Studies not published in English; this is to ensure accessibility for analysis.

2.6. Search strategy

The search strings presented in Table 1 were used to identify all relevant articles and documents. Initially, the first search string was applied, yielding 56 results from Scopus, 98 from PubMed, and 3 from the IEEE Xplore database. Then, the search string was modified to achieve better results.

2.7. Study selection process

First, research questions were developed, followed by a search string. Three researchers (ABR, ASR, and AMS) performed the initial database search and removed duplicate entries. Two researchers (ASR and ABR) reviewed all collected abstracts using the inclusion and exclusion criteria. Senior researcher AMS assessed articles with disagreements to establish consensus on decisions.

2.8. Study selection and bias control

The selection approach utilized a combination of engineering and health science datasets to enhance reliability and minimize publication bias. Two researchers (ABR and ASR) reviewed the titles and abstracts to reduce selection bias, while senior researcher AMS meticulously analyzed a paper to identify errors and further mitigate bias.

3. Results

This section outlines the key findings and emerging patterns identified through the systematic review. It discusses the implications of these results by comparing them with previous research and highlighting recent developments in the field. Furthermore, the review examines any limitations encountered during the process and considers their potential influence on the outcomes.

3.1. Data overview

Between June 1, 2015, and January 9, 2024, two review writers (ASR and ABR) thoroughly searched across four databases: PubMed, Scopus, Google Scholar, and IEEE Xplore. A date restriction was applied to exclude outdated models from the early stages of AI development, ensuring the relevance of the technologies to the current AI landscape. The initial search yielded 656 articles from Scopus, 249 from PubMed, four from IEEE Xplore, and five from Google Scholar. After removing 85 duplicates using Microsoft Excel (Microsoft, United States of America [USA]), a total of 829 articles remained. Following a title and abstract screening, 789 articles were excluded, leaving 40 articles for eligibility assessment. Ten articles were subsequently excluded due to issues with data extraction, non-English language, lack of linkage with AI, or poor technical implementation. The study selection process is illustrated in Figure 2. Ultimately, 30 articles that met the inclusion criteria were retained for data extraction, as summarized in Table 2.

3.2. Risk of bias (RoB) assessment

The systematic review assessed the RoB in the included studies to evaluate the validity and reliability of the results. RoB was evaluated across multiple domains, such as selection bias, performance bias, detection bias, attrition bias, and reporting bias, using well-developed tools, such as the Cochrane RoB2 Tool for randomized controlled trials and the ROBINS-I tool for non-randomized studies. Most studies relied on retrospective data, which has the potential for bias due to the non-randomized selection of participants. For example, studies such as those by Theilgaard Lassen *et al.*[16] and Cimadomo *et al.*[17] employed internal validation methods, which limit generalizability. Meanwhile, several studies, such as those by Johansen *et al.*[18] and Bori *et al.*,[19] implemented AI models trained on time-lapse imaging data without clearly defined standard protocols. The absence of standardized protocols across clinics could have introduced heterogeneity in data collection and analysis, thereby affecting the results. In addition, in AI-based embryo selection, as seen in

Table 1. Keywords and search items

S. No.	Keywords and search items	Number of publications from database		
		Scopus	PubMed	IEEE Xplore
1.	((AI) AND (ART) AND (IVF))	56	98	3
2.	(((AI) OR (artificial intelligence) OR (machine learning) OR (deep learning)) AND ((embryo selection) OR (blastocyst transfer) OR (preimplantation genetic diagnosis)) AND ((assisted reproductive technologies) OR (in vitro fertilization) OR (Intracytoplasmic Sperm Injection) OR (Gamete Intrafallopian Transfer) OR (Zygote Intrafallopian Transfer))) AND ((precision medicine) OR (predictive algorithms) OR (prognosis)))	656	249	4

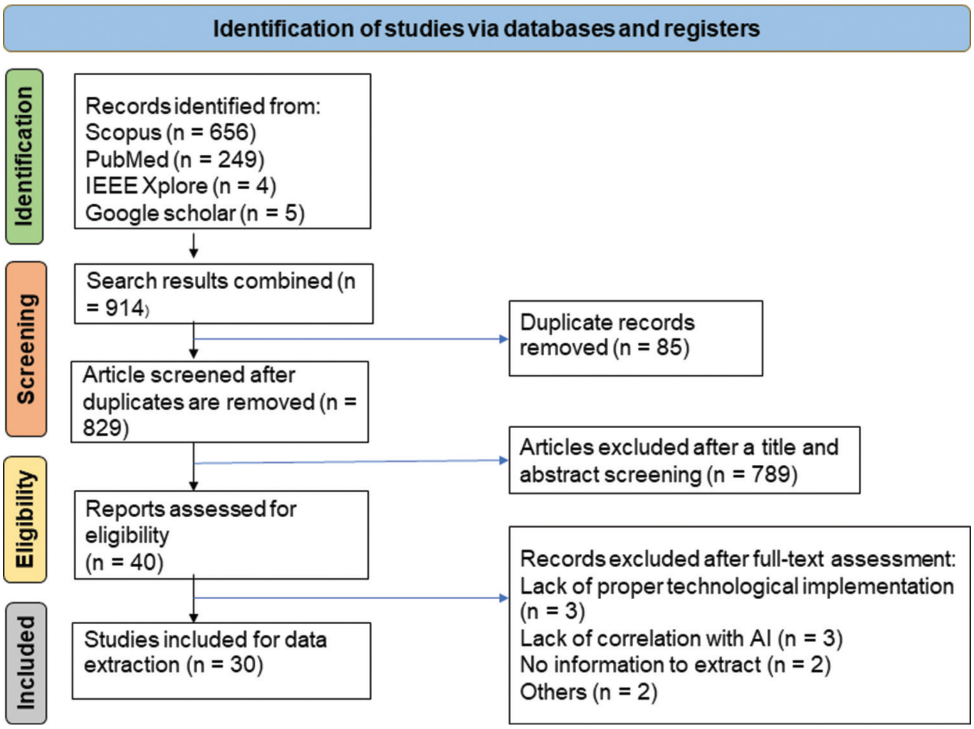


Figure 2. Flow diagram illustrating the study selection process in accordance with the guidelines of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses
Abbreviation: AI: Artificial intelligence.

Table 2. Bibliometric characteristics of the included studies

Characteristics	Corresponding entities	Numbers	Percentages
Publication year	2015 – 2018	1	3.33
	2019 – 2021	13	43.33
	2022 – 2024	16	53.33
Publication type	Article	23	76.67
	Review	5	16.67
	Conference proceeding	2	6.67
Quality of paper	Q1	16	53.33
	Q2	10	33.33
	Q3	1	3.33

studies by Glatstein *et al.*[20] and Salih *et al.*,[21] outcome assessors were frequently not blinded, increasing the risk of subjective determination regarding embryo viability and implantation success. Moreover, selective reporting was also an issue, with some studies failing to disclose limitations and adverse effects. For example, while Tian *et al.*[22] acknowledged the necessity for external validation, this aspect was not consistently addressed across other studies. Although the current review underscores promising progress in AI-driven embryo selection, the identified RoB highlights the need for future prospective, well-designed, and externally validated studies to confirm the reliability and generalizability of AI models in the clinical environment.

3.3. Bibliometric characteristics of the included studies

This systematic study suggests that incorporating engineering principles into the evaluation of online databases can enhance reliability and reduce publication bias. Three researchers carefully reviewed the titles and abstracts of the identified studies to reduce bias. In addition, an experienced academic expert reviewed the work to identify and address any potential inconsistencies or biases. As shown in Table 2, the selected publication types include 23 articles (77%), five review papers (17%), and two conference proceedings (6%). In terms of geographical distribution, the United Kingdom leads with 14 publications, followed by the USA (12), Switzerland (2), Japan (1), and Bosnia and Herzegovina (1).

3.4. Appraisal of the study quality

This systematic study suggests that incorporating engineering principles into online database searches can enhance reliability and reduce publication bias. Three researchers carefully reviewed the titles and abstracts of the identical studies to reduce bias. In addition, an experienced professional reviewed the work to identify and address any potential inconsistencies or biases.

Figure 3A illustrates the journal ranking of the chosen studies, showing a clear dominance of Q1 (16) and Q2 (10) journals, with a smaller representation from Q3 (1) and other categories (3). This distribution highlights a significant representation of high-quality articles in the selected literature. The dominance of Q1 and Q2 journals reflects the intent to prioritize sources known for rigorous peer-review procedures and credible academic contributions, thereby assuring the reliability of the findings. Meanwhile, Figure 3B presents the publication trends over the past decade. Notably, 25 of the selected papers were published in the past 3 years, indicating a recent increase in research efforts. In both 2022 and 2023, eight publications were published, exhibiting continuous productivity. Surprisingly, the highest point occurred in 2021 with nine publications, indicating a highly fruitful year for research in this domain.

3.5. Summary of the characteristics of the included studies

The current review thoroughly examined 30 academic publications that strictly adhere to formal academic guidelines. Each paper features a clear and distinct title and is authored by well-known researchers in the field. These articles included abstracts that briefly describe their goals, methodologies, and conclusions, as summarized in Table 3, and have been published in reputable journals and conferences. The research methods spanned surveys, case studies, and experimental studies. In addition to reporting empirical results, the current review also offers critical comments on methodological challenges, interpretative insights, and directions for further research.

3.6. Implications of the area under the curve (AUC) in embryo selection

In embryo selection algorithms, the term "area under the curve" refers to the area under the receiver operating characteristic curve, which helps in assessing the probability of successful implantation.[24] AUC is a useful statistic for evaluating the performance of embryo selection algorithms, but it has several limitations, such as its reliance on image quality, issues with generalizability, the impact of cultural conditions, sex-dependent performance, and limits related to sample size and research design. Even though AUC is frequently used to assess model performance in embryo selection, depending solely on it presents challenges due to sampling procedures and information transfer issues, which may affect the robustness and generalizability of the model. AUC provides a general measure of model performance across all thresholds but does not account for individual clinical contexts or specific requirements. Furthermore, it assumes that all misclassifications are equally important, whereas, in embryo selection, some misclassifications (e.g., false negatives) may have more severe consequences. In addition, AUC does not evaluate the calibration of predicted probabilities, which is a crucial requirement for decision-making. Thus, AUC should be complemented with additional measures and clinical judgment for a more comprehensive embryo assessment.[24,35-37]

3.7. Conventional study

Any process or therapy that involves manipulating oocytes (immature ova or egg cells) *in vitro* is referred to as ART for reproduction.[38] Couples and individuals experiencing fertility issues can benefit from this treatment option, which is characterized by individualized treatment protocols and multidisciplinary team management, both of which improve treatment outcomes and safety.[39] The field of traditional ART has significantly advanced over time. Techniques used in ART treatments include embryo

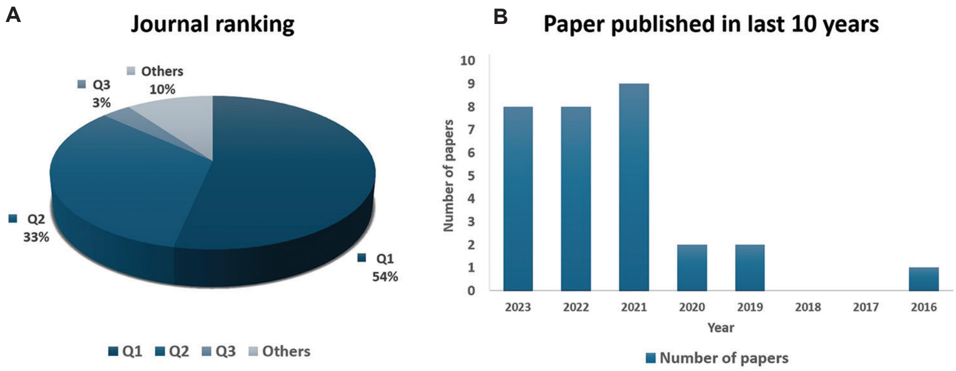


Figure 3. Quality appraisal of the included studies. (A) Ranking statistics of the journals selected for this study. (B) Number of published papers in the last 10 years.

Table 3. Summary of the characteristics of the included studies

References	Aim/research question	Dataset and data selection	Algorithm/methodology	Outcome measures	Evaluation and data processing limitations
Theilgaard Lassen *et al.*[16]	Rank embryos by likelihood of implantation	181,428 embryos from 22 IVF clinics worldwide (2011 – 2020)	Fully automated deep learning model (iDAScore v2.0)	AUC[a] range: 0.621 – 0.707; embryo ranking based on morphogenetic parameters	Only internal validation was performed; external validation is needed; there is a lack of details on sampling strategy, data imbalance, and cross-validation[c] procedures
Johansen *et al.*[18]	Assess maternal age's impact on embryo viability prediction[b]	4,805 fresh embryos from 4 clinics (2013 – 2022)	AI model based on time-lapse images with age-standardization	AUC[a] range: 0.58 – 0.69	Age density estimation may be unstable; insufficient description of sampling, data imbalance handling, and validation methods
Glatstein *et al.*[20]	Enhance embryo selection in IVF labs for improved pregnancy outcomes	The dataset is not explicitly specified	Combination of convolutional neural networks (CNN) and support vector machines	AUC[a] range: 0.63 – 0.83; some models report up to 90% accuracy	"Black box" nature of AI models; limited details on dataset selection, sampling, and cross-validation approaches
Salih *et al.*[21]	Compare the performance of AI versus that of embryologists in embryo selection	Data compiled from non-prospectively evaluated studies (2005 – 2022)	ML, deep learning, and neural networks	AI accuracy: 75.5% (range 59 – 94%) versus embryologist accuracy: 65.4% (range 47 – 75%)	Retrospective, non-clinical design; lacks explicit discussion on sampling methods, imbalance handling, and cross-validation techniques
Ueno *et al.*[23]	Evaluate the effect of increased training data on pregnancy prediction	3,960 SVBT cycles from a single Japanese clinic (2021 – 2022)	Deep learning models (iDA-V1 vs. iDA-V2) compared with Gardner grading	AUC[a]: 0.736 (iDA-V2), 0.720 (iDA-V1), and 0.702 (Gardner grading)	Retrospective design and single-center data limit generalizability; sampling details and handling of imbalance/cross-validation[c] are not fully described
Berntsen *et al.*[24]	Develop AI-based embryo selection using time-lapse images	115,832 embryos from 18 IVF centers worldwide (2011 – 2019)	Deep learning model implemented in Python/TensorFlow	AUC[a] of 0.67 for Sorted KID embryos; overall AUC reported as 0.95	No subgroup analysis for transferred euploid embryos; limited insight into data preprocessing, sampling, and validation methods
Chen *et al.*[25]	Predict the chromosomal status of blastocysts non-invasively	345 paired blastocyst culture mediums from 3 clinics in China (2017 – 2018)	Random forest ML model	Comparative clinical outcomes: A-/B-grade embryos like euploid ones; Grade C shows lower viability	Lacks a standardized threshold for CNV-based predictions; potential observer bias; minimal discussion on sampling and cross-validation[c]
Xi *et al.*[26]	Reduce IVF-associated multiple-embryo gestations	10,076 embryos from 9,211 patients at a single Chinese center (2016 – 2018)	Hierarchical model using XGBoost	AUC[a]: 0.7945 (SET), 0.8385 (DET pregnancy), 0.7229 (DET twin risk)	Single-center design limits generalizability; limited discussion on sampling strategy, imbalance handling, and cross-validation[c]
Ratna *et al.*[27]	Externally validate and update the McLernon models for live birth prediction	144,734 complete cycles from 91,035 women in the UK (2010 – 2016)	McLernon models updated via intercept adjustment, logistic recalibration, and revision	Improved agreement between live birth predictions and outcomes	High proportion of missing data and absent prior pregnancy details; limited discussion of sampling, imbalance, and cross-validation[c] practices

(Cont'd...)

Table 3. ***(Continued)***

References	Aim/research question	Dataset and data selection	Algorithm/methodology	Outcome measures	Evaluation and data processing limitations
Cimadomo *et al.*[17]	Validate iDAScore v1.0 for ranking blastocysts in PGT-A cycles	3,604 blastocysts and 808 euploid transfers from 1,232 cycles in Italy (2013 – 2022)	Deep learning model (iDAScore v1.0)	AUC[a]: 0.60 for euploidy prediction and 0.66 for live birth prediction	Retrospective design; need for randomized controlled trials; lacks details on data sampling, imbalance handling, and cross-validation
Diakiw *et al.*[28]	Predict embryo euploidy likelihood using blastocyst images	15,192 Day-5 blastocyst images from 10 IVF clinics (USA, India, Spain, Malaysia)	AI model trained on 2D microscope images with PGT-A metadata	Accuracy: 65.3%, sensitivity: 74.6%; AUC[a]: 0.68 (uncleaned), 0.87 (cleansed test dataset)	Predictive accuracy based solely on PGT-A outcomes; limited discussion on sampling, data imbalance, and cross-validation[c] processes
Bori *et al.*[19]	Predict embryo viability and grading via image analysis	Over 3,000 embryo images (Day 2 – 3)	CNN-based model deployed on Azure	Reported>85% improvement in success rate (accuracy boost)	Limited sample size; minimal details on sampling strategies, imbalance handling, and validation methods
Sawada *et al.*[29]	Develop a self-improving ML system (DynScore[d]) to predict ART embryo fate	Training: 891 embryos (110 couples); Global: 1,186 embryos (201 couples)	ML system using statistical tests (Kolmogorov–Smirnov, ANOVA, Chi-squared) to calculate DynScore[d]	AUC[a]: ~0.634 (training) and~0.638 (global) for blastocyst formation prediction	Small sample sizes; potential bias from the inclusion of failed ART cases; limited information on sampling, data imbalance, and cross-validation[c]
Cheredath *et al.*[30]	Conduct a SWOT analysis on human-versus ML-based embryo assessments in IVF	Varied datasets ranging from 16 to 11,898 embryos	Supervised learning approach in ML	General performance trends in IVF prediction noted (no specific quantitative metric)	Lacks detailed analysis of specific ML models and processing factors, such as sampling and validation methods
Patil *et al.*[31]	Provide an overview of prediction models in ART using varied feature sets	Dataset not specified	Various machine-learning techniques	Qualitative discussion on clinical decision support potential	No external validation or detailed impact analysis; minimal information on dataset selection, sampling, and cross-validation[c]
Giscard d'Estaing *et al.*[32]	Investigate AI's potential in reproductive medicine for infertility treatment	Dataset not specified	ML algorithms	Qualitative improvements in infertility diagnosis and ART outcomes (pregnancy/live birth rates)	Lack of standardized protocols and guidelines; limited description of data selection, sampling, and validation approaches
Zaninovic *et al.*[14]	Enhance implantation/ live birth prediction via an updated ASEBIR scoring system	1,044 Day-5 blastocysts from 6 clinics in Spain (2017 – 2019)	Multivariable logistic regression analysis	AUC[a]: 0.90 (trophectoderm A vs. BC) and 0.89 (trophectoderm AB vs. C)	Limited number of blastocysts analyzed; insufficient detail on sampling strategies, data imbalance handling, and cross-validation[c]
Tian *et al.*[22]	Enhance IVF processes through automation and AI integration	Dataset not specified	Integrated automation and AI (including endometrial evaluation, cryopreservation, and gamete/embryo selection)	Qualitative discussion on improved accessibility, affordability, and reduced labor intensity	Lack of specific dataset details; minimal description of sampling, data imbalance management, and validation protocols

(Cont'd...)

Table 3. (*Continued*)

References	Aim/research question	Dataset and data selection	Algorithm/methodology	Outcome measures	Evaluation and data processing limitations
Kanakasabapathy *et al.*[33]	Predict viable embryos for transfer using shallow artificial networks	654 cycles from a French cohort (2013 – 2018)	Shallow artificial networks (MLP and simple RNN)	AUC[a]>0.8 in predicting embryo viability	Retrospective observational design introduces bias; limited discussion on sampling methods, data imbalance, and cross-validation[c]
Pons *et al.*[34]	Assess maternal age's impact on embryo viability prediction[b]	4,805 fresh embryos from 4 clinics (2013 – 2022)	AI model based on time-lapse images with age-standardization	AUC[a] range: 0.58 – 0.69	Age density estimation may be unstable; insufficient description of sampling, data imbalance handling, and validation methods

Notes: [a]The area under the curve (AUC) of an AI model measures its ability to distinguish between classes, with a higher AUC indicating better performance; [b]Embryo viability prediction is an approach in which AI models evaluate different characteristics of embryos to estimate their likelihood of successful implantation; [c]Cross-validation is a statistical technique used to assess the performance and generalizability of an ML model; [d]DynScore, also known as Dynamic Scoring, is a dynamic assessment measure that is utilized in fields such as bioinformatics, finance, or AI.
Abbreviations: AI: Artificial intelligence; ANOVA: Analysis of variance; ART: Assisted reproductive technology; CNV: Copy number variation; DET: Day-5 embryo transfer; IVF: *In vitro* fertilization; ML: Machine learning; MLP: Multilayer perceptron; PGT-A: Preimplantation genetic testing for aneuploidy; RNN: Recurrent neural network; SET: Single embryo transfer; SVBT: Single-voxel brain tissue; SWOT: Strengths, weaknesses, opportunities, and threats; USA: United States of America.

transfer (ET), ICSI, and IVF.[11] These methods encompass the *in vitro* manipulation of human gametes or embryos for addressing genetic disorders or sub-fertility that impede spontaneous conception.[12] Nonetheless, conventional ART success rates can vary, and ART laboratories must continually strive for refinement and implement evidence-based practices.[13] Various frequently utilized ARTs are depicted in Figure 4, depending on the specific circumstance. These ARTs are elaborately discussed in the following section.

3.7.1. IVF

The IVF technique has significantly enhanced our understanding of fertilization processes in 11 mammalian species, including humans.[40] IVF is a medically assisted reproduction method that allows infertile couples to achieve a successful pregnancy.[41] It involves the retrieval of oocytes, which are then fertilized outside the body, with the resulting embryos cultured in a laboratory setting before being transferred into a woman's uterus.[42] An IVF cycle typically lasts for 4 – 6 weeks and begins with 10 – 14 days of hormonal stimulation to produce multiple eggs. Between days 12 and 16, mature eggs are retrieved and fertilized in the laboratory using either ICSI or conventional IVF. The resulting embryos are cultivated for 3 – 5 days before being transferred into a uterus. Two weeks later, a blood test is conducted to confirm pregnancy.[43] While the overall success rate remains relatively low, interventions such as hysteroscopy with local endometrial injury before ovarian stimulation can improve implantation and pregnancy rates in women with repeated IVF failure.[44] Successful ET requires careful preparation, including minimizing uterine contractions and placing the embryo approximately 2 cm below the uterine fundus for optimal pregnancy rates.[45]

3.7.2. ICSI

ICSI is an assisted reproduction technique that treats severe male-factor infertility. It entails inserting a single spermatozoon directly into the ooplasm of a mature egg. ICSI was established in the early 1990s and has since become a widely accepted treatment option for couples facing reproductive challenges due to male factors. The method overcomes potential hurdles to fertilization, allowing fertilization to occur even with compromised sperm parameters, such as low motility and aberrant morphology. However, there are concerns regarding the safety and long-term implications of ICSI, including an increased risk of sex chromosomal abnormalities and potential developmental issues in offspring conceived through this technique.[46] During the procedure, a physician injects sperm into an oocyte, and the resulting fertilized egg is then transferred into a woman's uterus for

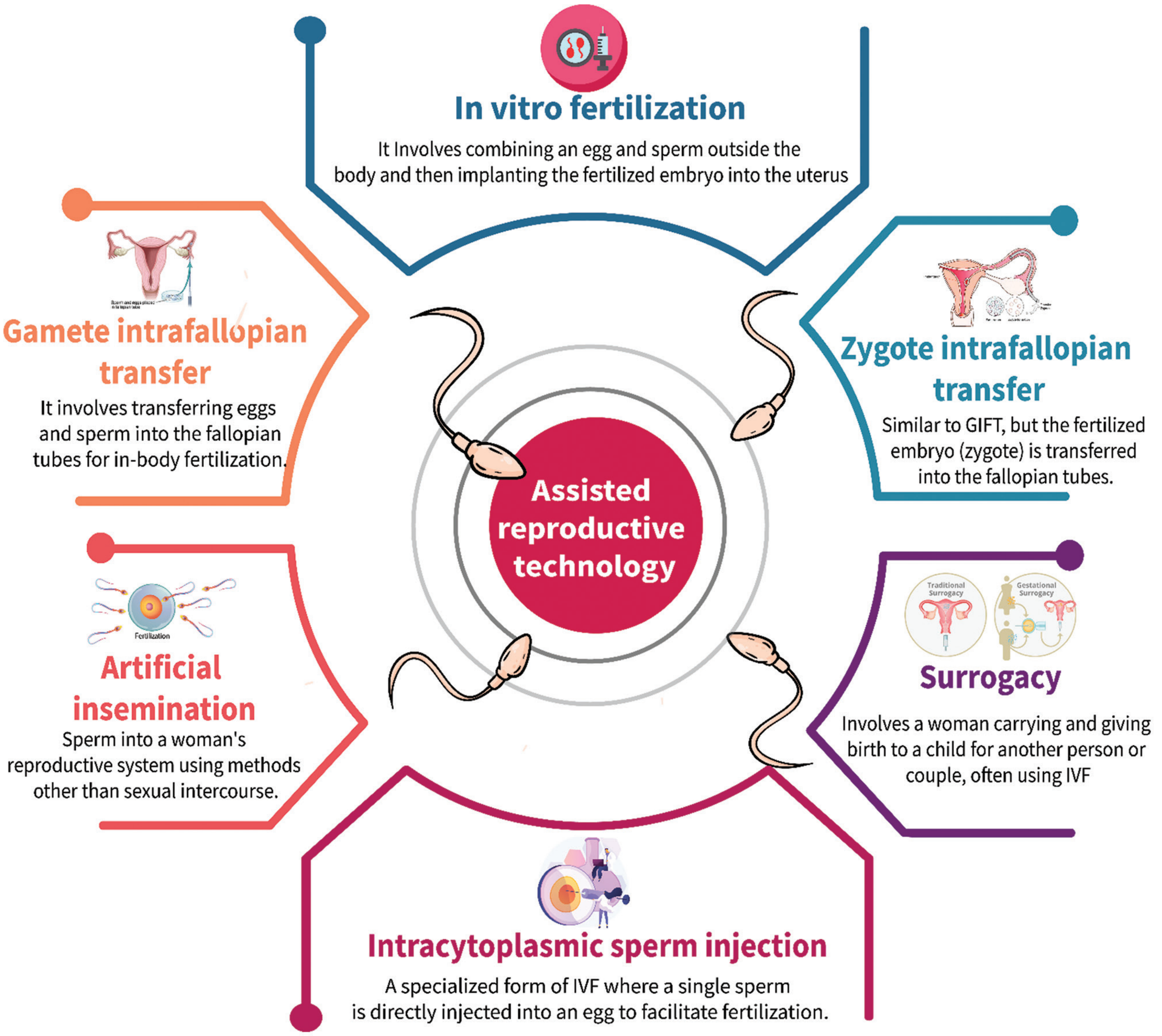


Figure 4. Different types of assisted reproductive technologies
Abbreviations: GIFT: Gamete intrafallopian transfer; IVF: *In vitro* fertilization.

implantation.[47] Although originally developed for cases of severe male infertility, ICSI is now also being utilized in clinical practice for mild male infertility or unexplained infertility.[48] Looking ahead, future advancements in ART may involve the use of immature spermatozoa, spermatids, and spermatocytes for ICSI, as well as further research into gene expression and genomic imprinting during spermatogenesis.

3.7.3. Surrogacy

Surrogacy is an assisted reproduction option often chosen by intended parents due to infertility, health concerns, same-sex parenting dynamics, or broader family diversity. In this method, another woman, known as the surrogate mother, carries and delivers a child on behalf of the intended parents.[49] While parental vetting and counseling are required to ensure parental responsibility and minimize future complications, surrogacy can be a good alternative for starting a family.[50] Despite its potential benefits, surrogacy raises complex ethical and legal concerns, such as fostering inequality, violating human rights, and commodifying children.[51] To mitigate these risks, robust legislation is required to safeguard the rights of intended parents, surrogate mothers, and the resulting child.[52] Increasing public awareness about the advantages of surrogacy, as well as addressing infertility, advancing age, health limitations, and adoption barriers, can contribute to broader acceptance and responsible practices.[53] Legal and ethical concerns about surrogacy differ considerably among nations. Countries such as the USA, France, Germany, and Austria predominantly impose restrictions on surrogacy, whereas nations such as India, Russia, and Kazakhstan hold more permissive legislation. Ethical discussions frequently classify nations as either advocates or adversaries, highlighting concerns over the possible exploitation of women in poor economic circumstances. Moreover, surrogacy prompts concerns regarding parental rights and acknowledgment. The identification of the legal mother – be it genetic or gestational – differs by jurisdiction, further complicating the legal status of both the child and the surrogate. Certain nations, such as Israel, have established extensive legal systems that may function as a model for international standards. Furthermore, the commercialization of surrogacy engenders ethical dilemmas about treating reproduction as a commercial process, possibly abusing women in economically precarious circumstances. In addition, ethical issues focus on the moral repercussions of surrogacy as a type of contemporary slavery.[54,55] The role of the gestational surrogate is typically underestimated, which can lead to ethical oversights, such as those found in adoption and donor-assisted conception. Therefore, understanding the gestational connection is essential for ensuring ethical surrogacy activities. On the other hand, the absence of international agreement on surrogacy legislation has led to the phenomenon of "surrogacy tourism," where individuals pursue surrogacy services in nations with more advantageous regulations. This situation underscores the urgent need for internationally recognized standards to safeguard the rights and well-being of all parties involved.[55,56]

3.7.4. Gamete intrafallopian transfer (GIFT)

GIFT is used to address infertility or genetic issues that may prevent spontaneous conception. Through GIFT, sperm and eggs are inserted into the fallopian tubes, allowing internal fertilization. Along with other treatments, including IVF and ET, GIFT is a part of the ART spectrum.[13] Several variables, including the quality of gametes and the physical and psychological conditions in which the operation is conducted, affect the outcomes of GIFT.[6] Couples seeking reproductive treatment still have the option of GIFT, even though it is less prevalent than IVF and ET.[57] For women with healthy fallopian tubes, GIFT is an easier and more effective assisted conception approach that provides higher successful rates than IVF.[58] It is also an effective treatment for non-tubal infertility, with a 32% rate of continued pregnancy and options for ovum donation, diagnostic IVF, and embryo freezing.[59] GIFT has an improved version known as "New GIFT," which has evolved from its earlier form referred to as "Old GIFT." The New GIFT approach is more effective than the Old GIFT, with the formation of pregnancy primarily dependent on sperm motility and oocyte count. The New GIFT approach, which involves preincubation of oocytes, has a lower abortion rate (19%) and a higher pregnancy rate (37% each cycle and 42% per patient) than the Old GIFT method (24% per cycle, 29% per ovum pick-up, and 22% abortion rate). It is more effective overall because it highlights the significance of mature oocytes and sperm motility, as well as benefits from improved luteal support.[60,61]

3.7.5. Zygote intrafallopian transfer (ZIFT)

ZIFT combines the natural environment of the fallopian tubes with IVF. In this method, zygotes are placed into fallopian tubes 24 h after fertilization in the lab following ICSI. The transfer is carried out using a catheter during laparoscopy, allowing the zygotes to travel through the fallopian tubes naturally to the uterus. ZIFT is typically considered by patients with fallopian tube issues or those seeking a more natural method for embryo implantation.[61] A 2-year study reported that ZIFT has a 40.4% clinical pregnancy rate and a delivery/ongoing rate of 34.2%.[62] While IVF-ET has the highest cumulative take-home baby and pregnancy rates for treating male infertility, both GIFT and ZIFT show comparable rates of pregnancy but are associated with higher rates of miscarriage.[63] In IVF-ET cycles, ZIFT has proven particularly beneficial for patients experiencing repeated implantation failure. Research indicates that ZIFT considerably increases the likelihood of pregnancy and implantation in these individuals compared to further IVF-ET cycles. For example, patients with repeated implantation failures had pregnancy rates of 34.2% with ZIFT compared to 17.1% with IVF-ET.[64,65] Despite these advantages, ZIFT is more invasive than conventional IVF-ET due to the requirement of laparoscopy. Moreover, studies have shown that, especially in cases involving male-factor infertility or non-tubal infertility, the added complexity of ZIFT does not always result in improved outcomes.[66,67]

3.7.6. Artificial insemination

Artificial insemination involves the use of a syringe and an artificial insemination catheter to inject semen into a woman's uterus. This medical treatment is performed in conjunction with homologous insemination, particularly in cases of oligospermia or when sexual intercourse becomes difficult. In more complicated situations, donor sperm can be surgically inserted into a female's oviduct through laparoscopic surgery; however, this method is more invasive, costly, and technically demanding than vaginal insemination. Despite its complexity, artificial insemination may be necessary for specific infertility treatments or breeding initiatives. It allows for the conservation of genetic lines, acceleration of line extension, or the synchronization of embryo development by enabling one male's sperm to inseminate multiple females.[68] This method addresses the issue of male-factor infertility and can be used with various species, including gorillas, lions, bears, and tuna fish. Infertility in both males and females can be successfully treated using artificial insemination, with a success rate of up to 18.2% every cycle and 58.4% after 6 months of treatment.[69]

3.8. Relationship between ART and AI

A significant portion of people worldwide experience infertility, which is defined as the inability to achieve a clinical pregnancy following 12 months of regular, unprotected sexual intercourse.[10] Treatment for subfertility, infertility, or genetic disorders that hinder natural conception is referred to as ART. Common ART methods include ET, IVF, ICSI, and GIFT.[4] Recent technological advancements aim to automate processes such as sperm selection, fertilization, and embryo culture, which could enhance consistency and reduce the stress caused by manual manipulation.[13] The integration of AI with ART enhances the effectiveness and success rates of ART treatments. AI algorithms can assess and predict the quality of gametes and embryos, which are vital to the success of ART. At present, morphological examinations, which are prone to subjectivity and human error, are the primary method used by embryologists to manually assess the quality of gametes and embryos.[70] By eliminating inter-observer and inter-objective variations, AI algorithms may provide a more standardized and objective evaluation.[71] This integration enhances reproductive health outcomes by increasing implantation success rates, reducing the incidence of multiple pregnancies, and improving single-ET.[72] Moreover, AI can predict embryos' viability and oocytes' developmental capability using ML algorithms and morphokinetic parameter analysis. By selecting the most viable embryos, this method can increase the chances of implantation and successful pregnancy.[73] In addition, AI can aid in automating tedious and repetitive ART laboratory duties, thereby increasing productivity and minimizing errors.[23] However, to achieve extensive integration into clinical practice, ethical considerations and the imperative for transparency in AI algorithms must be duly acknowledged and resolved.

3.9. Advancements in AI and ML in ART

AI and ML are two related yet distinct disciplines within the field of computer intelligence. ML is a subset of AI that focuses on algorithms enabling computers to learn from data, while AI is a vast field that aims to replicate human intelligence through reasoning, problem-solving, and learning. Although AI is widely used in many sectors, ML is essential for enhancing decision-making and prediction skills in these applications.[74] AI and ML have the potential to revolutionize basic science, clinical practice, healthcare administration, and medicinal economics.[70] In reproductive medicine, predictive modeling using AI can accurately predict fertility outcomes. However, challenges such as managing large-scale data, identifying valuable features, and validating models with gold-standard study designs remain.[75] Further precision, standardization, and automation in the field of reproductive medicine could be achieved using AI-guided procedures.[76] For example, AI-based ART software can reduce interobserver variability, personalize drug doses, and improve clinical and operational efficiency, particularly in sperm selection and oocyte quality assessment.[77] Most importantly, AI has shown potential in reproductive urology by predicting semen parameters, identifying candidates for genetic testing, and automating sperm detection.[7] In addition, AI in ART improves efficiency by identifying patients at risk for conditions such as endometriosis, detecting gamete production values, and optimizing controlled ovarian stimulation by calculating ideal starting drug doses and trigger timing using deep learning algorithms.[73] Recently, Levenberg–Marquardt neural networks trained on local binary patterns have demonstrated promising outcomes in terms of oocyte and embryo quality prediction, offering potential improvements in ART, especially in nations with restricted embryo selection practices.[68]

3.10. Integration of AI in embryo selection

In the past, doctors have utilized ML algorithms to assist in selecting embryos for human-assisted reproduction. The challenges of embryo selection have gained significant attention with the advent of ART. Invasive techniques, including preimplantation genetic screening, as well as transcriptome and proteome analyses of biopsied embryonic tissue, were initially emphasized and are currently being explored to obtain direct insights into embryonic development. In ART, a variety of deep learning and ML models are applied to enhance embryo selection processes. For example, AI models such as ERICA use blastocyst images to estimate euploidy in embryos.[78] To enhance embryo selection processes in IVF, AI analyzes complex data, identifies trends, and provides an objective evaluation of embryos.[79] Furthermore, AI is used to determine the optimal quantity of metaphase II oocytes required for ART to produce viable blastocysts and embryos.[73]

3.11. Transformative algorithms in enhancing embryo selection

Within the field of ART, a wide range of ML and AI methods have been utilized to improve embryo selection and predict the likelihood of successful implantation during IVF procedures. These technologies aim to enhance the precision, consistency, and efficiency of evaluating embryo viability. One prominent approach involves the use of deep learning models, such as iDAScore v1.0, for the objective ranking of blastocysts, introduced by Cimadomo *et al.*[17] Its successor, iDAScore v2.0, developed by Theilgaard Lassen *et al.*,[16] incorporated morphokinetic parameters into embryo ranking, and achieved an AUC ranging from 0.621 to 0.707. Ueno *et al.*[23] further demonstrated that increasing training data with Gardner grading for both IDA-V1 and V2 significantly enhanced predictive performance, with an AUC value of 0.736 for ongoing pregnancy predictions. Moreover, studies by Bori *et al.*[19] and Johansen *et al.*[18] introduced AI models using time-lapse images to evaluate embryo viability. Benchaib *et al.*[80] employed shallow artificial networks (e.g., multilayer perceptron and recurrent neural network) based on morphokinetic time-lapse parameters to predict viable embryos for transfer. Berntsen *et al.*[24] employed a deep-learning AI model based on Python and TensorFlow to sort 115,832 embryos from 18 IVF centers, achieving AUC values between 0.63 and 0.69. In another approach, Chen *et al.*[25] used a random forest model on 345 paired blastocyst cultures, demonstrating transplant suitability for A- and B-grade embryos comparable to euploid ones. Glatstein *et al.*[20] used convolutional neural networks and support vector machines to predict live birth probabilities, with AUC values of 0.63 – 0.83 and achieving up to 90% accuracy in some models. Salih *et al.*[21] showed that ML outperformed embryologists in predicting embryo morphology, with an accuracy of 75.5% compared to the embryologists' 65.4%. Meanwhile, Pons *et al.*[34] applied logistic regressions to update the ASEBIR system for predicting blastocyst implantation and live birth. To predict fertilization failure probabilities (Logistic Regression Function and Threshold Fertilization Failure) in ART cycles, Tian *et al.*[22] utilized Bayesian network modeling and achieved an accuracy of 91.3%, aiming to optimize IVF and ICSI treatments.

Recent deep learning models show significant advancements in performance. For example, the Embryo2live model outperformed traditional morphology grading by increasing live birth rates from 23.0% to 71.3% for top embryo selections.[81] The Esava model, developed for the quantitative evaluation of IVF embryos, reported high rates for precision (0.9940), recall (0.9121), and mean average precision (0.9531), demonstrating superior blastomere detection compared to previous computational methods.[82] Similarly, the CHLOE model revealed no significant bias between XX and XY embryos ($U = 204621$, $p=0.208$). In contrast, manual morphological grading and the KIDScore algorithm, a tool to support embryologists in decision-making, tended to favor male embryos, with XY embryos receiving higher scores than XX embryos ($U = 207604$, $p=0.0182$; $\chi^2 = 19.843$, $p<0.00001$). These findings suggest that deep-learning approaches may help mitigate sex-selection bias.[83] However, deep learning is not always superior to manual methods. In a recent double-blind non-inferiority trial involving 1,066 patients, 533 were assigned to the iDAScore group and 533 to the morphological grading group. The iDAScore group showed a clinical pregnancy rate of 46.5% (248 of 533 patients), compared to 48.2% (257 of 533 patients) in the morphological grading group (risk difference = −1.7%; 95% CI = −7.7, 4.3; $p=0.62$).[84]

4. Discussion

This comprehensive analysis of the systematic review aims to address the RQs, assess the quality of the review, and propose directions for future study.

4.1. General discussion

The term "ART" is gaining significant recognition in the context of infertility treatment. ART refers to any procedure involving the *in vitro* manipulation of oocytes (immature ova or egg cells) for reproductive purposes.[38] Common ART procedures include IVF, ICSI, ET, and luteal phase assistance, although these techniques may cause perinatal complications and outcomes.[85] Primarily used for infertility treatments, ART includes techniques such as artificial insemination, IVF, surrogacy, and the use of fertility medication.[86] The implementation of individualized treatment protocols and multidisciplinary team management significantly improves treatment outcomes and safety, making ART a viable option for couples and individuals facing fertility issues. The ART process encompasses several critical stages, including controlled ovarian stimulation, pituitary downregulation, oocyte retrieval, fertilization, ET, embryo selection, and luteal phase support.[39] In recent years, AI has been integrated into ART to enhance and automate embryo selection by analyzing microscopy images and identifying optimal embryos for transfer or cryopreservation.[77] AI is also used to reduce interobserver variability, optimize drug dosing, enhance sperm selection and oocyte quality evaluation, and increase overall clinical efficiency.[87] By enhancing outcomes and decision-making processes through sophisticated algorithms and data analysis, AI is transforming the landscape of ART.[29] This exploratory review examines recent developments in embryo selection and related AI-driven innovations within ART.

4.2. Strength of the systematic review

The strength of this systematic review lies in its thorough analysis of the correlation between AI and ART, particularly those concerning embryo selection. The review thoroughly explains the opportunities, challenges, and future directions in using AI to improve ART outcomes by analyzing a wide range of literature from the past 10 years. This paper is the first to highlight the need for external validation of prediction models, an aspect that requires significant improvement in the reviewed studies. Most of the scientific papers used retrospective and anonymized data, which may introduce biases and limit the generalization of findings. Therefore, this systematic review combines an in-depth analysis of various documents from several countries, thus reducing bias and highlighting its potential to transform the field of ART. It also highlights ethical considerations and emphasizes the significance of responsible implementation in clinical practice. With its in-depth analysis and interdisciplinary approach, the review provides valuable insights for researchers, clinicians, and policymakers, facilitating informed decision-making. In addition, four key parameters have been identified – ethical concerns, clinical and regulatory constraints, discussion of AI techniques, and the potential applications of AI in ART – to illustrate the strength of this review (Table 4).

4.3. Addressing the RQs

In subsequent sections, comprehensive responses are provided to address the RQs.

4.3.1. RQ1: What are the current state-of-the-art AI technologies used in embryo selection for ART?

In ART, recent developments in AI have significantly improved embryo selection procedures. These cutting-edge AI tools use ML, deep learning, and computer vision to increase the precision, reliability, and efficiency of embryonic health evaluation. AI technologies are transforming IVF laboratories by automating the assessment of embryo morphology and leveraging synthetic data. The primary AI tools currently employed in embryo selection are listed below:

(i) Computer vision and deep learning: Computer vision and deep learning are used by AI systems to automatically examine images of embryo morphology and extract important aspects that are essential for determining the survival of the embryo. The accuracy of embryo selection is improved, and the subjectivity associated with manual assessments by embryologists is reduced.[92] For example, deep learning algorithms that have been trained on both synthetic and actual embryo images have shown excellent accuracy in predicting the stages of embryo cells, reaching up to 97% accuracy when synthetic data is included.[93]

(ii) Time-lapse imaging: AI systems use time-lapse imaging to improve predictions from fertilization to the blastocyst stage, thereby increasing IVF success rates by identifying viable embryos more accurately than human experts.[94]

(iii) ML Techniques: Several ML techniques, such as neural networks, naive Bayes, support vector machines, and random forests, are used to predict treatment outcomes and enhance IVF results. The average AUC rating for these models is 0.91, indicating high accuracy, sensitivity, and precision.[94]

Table 4. A thorough comparative analysis of relevant studies and the current review

References	Ethical concerns	Regulatory and clinical barriers	AI techniques discussed	Prospects of AI in ART
Kragh and Karstoft[87]	✓	✓	✓	✗
Merican *et al.*[88]	✗	✗	✓	✓
Raef and Ferdousi[4]	✗	✗	✓	✓
Medenica *et al.*[89]	✓	✓	✗	✓
Afnan *et al.*[9]	✓	✓	✗	✗
Fernandez *et al.*[13]	✗	✗	✓	✗
Curchoe[90]	✓	✓	✗	✓
Tran *et al.*[91]	✓	✓	✓	✗
Abdullah *et al.*[3]	✗	✓	✓	✓
Current study	✓	✓	✓	✓

Abbreviations: AI: Artificial intelligence; ART: Assisted reproductive technology.

(iv) Generative models for synthetic data: Images of synthetic embryos are produced using generative models, such as diffusion models and generative adversarial networks. Combining these artificial images with actual data enhances AI model training and improves classification performance.[93]

(v) AI-powered embryo ranking systems: Deep learning is used by platforms such as the Embryo Ranking Intelligent Classification Algorithm to rank embryos according to their expected genetic state and implantation potential. These systems enable clinicians to select embryos with the highest potential for successful implantation and pregnancy.[95]

(vi) Automated morphological feature extraction: AI solutions reduce subjectivity and unpredictability in assessments by automating the measurement of important morphological parameters from embryo images.[96]

4.3.2. RQ2: To what extent does ART improve the pregnancy success rate and live birth outcomes compared to traditional methods?

The performance of ART may be improved using AI and ML approaches. These techniques also hold a promise for the advancement of medical technology in the future.[5] AI-guided ARTs offer enhanced accuracy, uniformity, and automation compared to traditional ARTs.[76] AI can accurately identify embryos' inner cell mass, blastocoel, trophectoderm, and zona pellucida. Deep learning methods are employed to accomplish this capability and reduce the workload of embryologists, thereby improving the efficiency of ARTs.[97] In general, factors such as maternal age, the underlying cause of infertility, and the ovarian stimulation protocols significantly influence the probability of achieving successful pregnancies and live births.[22]

4.3.3. RQ3: How can AI algorithms be seamlessly integrated into existing embryo selection protocols and laboratory workflows to leverage the expertise of embryologists and healthcare professionals?

AI algorithms can enhance embryo selection protocols by leveraging data from time-lapse imaging, proteomic profiles, and morphological features to predict live birth outcomes and embryo viability.[73,79,98] Although ML systems are used to predict the results of frozen ETs in early pregnancy, their accuracy is limited, and additional predictors are required to improve predictive performance.[99] By utilizing ML models and computer vision, AI can analyze vast amounts of image data to automate embryo selection processes.[100] Nevertheless, issues such as the interpretability of AI models and the requirement for open, and peer-reviewed research remain to be resolved. The application of AI as a quality control tool post-thawing or for continuous monitoring of embryo culturing systems optimizes laboratory workflows. This application allows for a synergistic approach, integrating the strengths of AI algorithms with the expertise of embryologists and healthcare professionals to improve ART outcomes. AI can streamline laboratory workflows by automating time-consuming tasks, such as embryo evaluation, allowing professionals to focus on critical decision-making.[87]

4.3.4. RQ4: What is the impact of AI-driven embryo selection on the psychological health and decision-making processes of prospective parents, particularly in light of ethical concerns?

It is necessary for ethical considerations to maintain public trust and improve both psychological and clinical results in ART.[101] The use of AI technologies in embryo selection raises ethical concerns regarding deskilling, transparency, accountability, and fairness. The absence of transparency in AI models, which are frequently referred to as "black-box" systems, creates uncertainty and undermines trust.[98] Furthermore, the potential repercussions of AI failures in embryo selection, which could lead to anomalies or undesired outcomes, underline the significance of responsibility and fairness in the utilization of AI in this sensitive field.[71] To address these issues and preserve the trust of the public, it is essential to prioritize the development of AI models that can be interpreted, carry out thorough evaluations through randomized controlled trials, and establish regulatory oversight for the application of these algorithms.[102] If these recommendations are followed, the integration of AI in embryo selection may enhance psychological and therapeutic results while upholding ethical standards in ART.

4.3.5. RQ5: What are the primary barriers and limitations to the clinical application of AI algorithms in embryo selection, and how are these technologies being developed and validated?

AI-based ARTs can reduce inconsistencies, increase clinical and client outcomes, and improve sperm testing and oocyte quality assessment.[103] Obstacles to applying AI algorithms arise due to the lack of transparency in ML models, ethical concerns about selection errors, and the fact that current embryo-selection algorithms lose diagnostic value when applied externally to many known implantation embryos.[104] These limitations are exacerbated by the black-box nature of AI models, leading to ethical and epistemic issues, such as unclear responsibility for treatment success and biases with unintended consequences.[105] To address these challenges, efforts are being made to increase the generalizability of AI models by diversifying training data and developing clinic-specific models.[9] Ethical considerations in adopting AI for embryo selection include transparency, interpretability, and collaborative decision-making to ensure the well-being of prospective parents and uphold ethical standards in assisted reproduction.[21] In addition, there is growing support for more rigorous clinical testing, including larger sample sizes, balanced datasets, and improved performance metrics, to ensure the reliability and effectiveness of AI algorithms for embryo selection.

4.4. Limitations of this study

The studies included in this review focused on deep learning, ML, or AI for embryo selection, providing a general overview rather than a detailed statistical data analysis, which could limit the depth of insights provided. Furthermore, the absence of critical appraisal may induce uncertainty regarding the robustness of the evidence synthesized in the review. The selection was determined through an exhaustive search of numerous databases, which may have resulted in bias. Due to limited access, some databases, such as Web of Science and PsycINFO, were not included in the study. The inclusion and exclusion criteria were meticulously defined to ensure a thorough search. Only publications that were written in the English language were considered. The studies spanned from June 1, 2015, to January 9, 2024, considering that older studies might not reflect recent technological advancements.

4.4.1. Factors influencing analytical results

One of the most concerning limitations of this reviewed study is sampling imbalance. Some papers in this study do not adequately represent the diverse population undergoing ART, creating potential bias. Some AI models were trained using datasets of a limited number of clinics, thereby lacking generalizability. Another limitation is the variation of validation techniques used across studies. Differences in validation techniques can influence performance matrices. Without external validation, these models may demonstrate high accuracy on training data but perform poorly in real-world applications. Furthermore, variability in model predictions, influenced by various characteristics of the dataset provided, raises concerns about clinical reliability. AI models are often trained on retrospective data (data collected from past events or historical records), and their performance in real-world clinical settings raises concerns. In addition, some AI models lack interpretability (black-box algorithms). This further complicates the integration of AI models into ART.

4.5. Future scope of ART

Throughout the years, ART has made an extraordinary contribution to the endeavor to resolve the infertility issues that couples have been facing. ART, spanning from its traditional to contemporary iterations, has played a pivotal role in enhancing birth rates and facilitating conception through ET, GIFT, and other procedures. It has increased the birth rate by surmounting barriers to conception and augmenting the probability of a successful pregnancy. There is significant potential for ART advancements through the implementation of AI-based algorithms. AI possesses the capability to assist in antiretroviral therapy by addressing therapeutic challenges.[71] To enhance the outcomes of ART, prediction models can be developed by implementing ML methodologies, which encompass a diverse array of feature sets and numerous algorithms.[4] Future advancements in soft robotics, telesurgery, and the integration of AI with robotics may potentially lead to an ART procedure that is fully automated and intelligent.[73]

The potential impacts of AI on ARTs are shown in Figure 5, highlighting seven key areas:

(i) Personalized treatment plans: AI customizes care based on patient data, taking specific health profiles into account to maximize success rates
(ii) Real-time decision support: AI offers clinicians timely insights during critical procedures, such as ET, enhancing accuracy and efficacy
(iii) Predictive analytics: AI accurately predicts treatment outcomes, enabling proactive adjustments to optimize success rates
(iv) Automated laboratory processes: AI automates tasks such as sperm and embryo analysis, enhancing efficiency and reducing errors
(v) Remote monitoring and teleconsultation: AI-powered systems enable continuous patient monitoring and teleconsultation, extending ART accessibility
(vi) Genomic analysis: AI identifies genetic risks, aiding informed decisions on embryo selection and genetic screening for better outcomes
(vii) Enhanced quality control: AI ensures optimal laboratory conditions, minimizing variability and enhancing success rates.

The integration of AI into these areas promises personalized, efficient, and accessible reproductive healthcare solutions, revolutionizing the field of reproductive medicine. Furthermore, robust regulatory frameworks are essential to guarantee the ethical and safe application of AI in ART, highlighting the requirement for policies and supervision to promote ethical AI adoption in reproductive medicine.[89] Moreover, the utilization of AI in

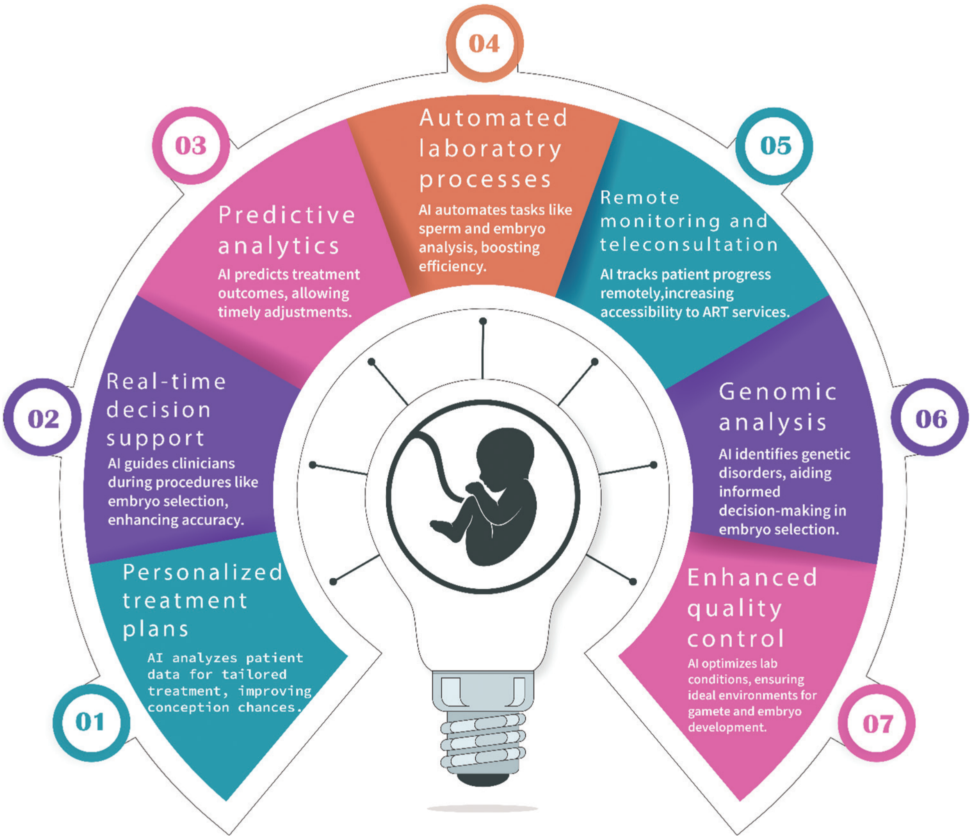


Figure 5. Future scope of assisted reproductive technology using artificial intelligence

the field of reproductive health is advancing significantly, as indicated by the proliferation of AI-related conference abstracts and the commercialization of numerous AI models.[3] As an illustration, in countries with embryo selection restrictions, Levenberg–Marquardt neural networks that have been trained to utilize local binary patterns have exhibited promising outcomes in predicting the quality of oocytes and embryos.[68] This could improve the technology used in assisted reproduction. It is crucial to address AI's shortcomings, such as algorithmic bias and the need for future research and clinical trials, to ensure its successful integration into ART.[90] Employing large training datasets and robust models is recommended to surmount these challenges.

5. Conclusion

Despite the extensive literature on AI applications in embryo selection, our study specifically focuses on addressing existing gaps in this area. The endeavor to revolutionize ART is at a critical turning point where the convergence of AI and precision in embryo selection intersects. This study explores AI applications in embryo selection, highlighting advancements, challenges, and opportunities. AI technologies, such as automatic embryo scoring algorithms (e.g., KIDScore D5 v3), Bayesian networks, and shallow artificial neural networks (e.g., multilayer perceptron and recurrent neural network), have been applied in ART for embryo selection. These AI models have demonstrated various levels of accuracy, with some achieving up to 90% prediction accuracy for live birth probabilities. However, the reviewed studies are often retrospective and conducted in single centers, limiting their generalizability. In addition, the lack of comprehensive datasets and previous pregnancy information poses challenges to model performance. Throughout this paper, we demonstrate that AI has the potential to enhance embryo viability, improve live birth prediction accuracy, personalize treatment plans, minimize human errors, and standardize IVF practices. To make advancements in the field of ART, future research must focus on creating AI models that are transparent, interpretable, and thoroughly verified through randomized controlled trials. To overcome current limitations, the generalizability of AI models can be improved by diversifying training datasets and tailoring models to different clinical settings. Integrating AI-driven predictive analytics and real-time decision support systems into clinical practices can enhance treatment outcomes. Furthermore, it is essential to implement individualized treatment plans based on patient-specific information, streamline laboratory procedures through automation, and establish strong ethical guidelines to enhance ART effectiveness and maintain public trust in ART. Future research should focus on developing AI algorithms to address challenges such as algorithmic bias and the need for extensive clinical validation, incorporating robust statistical approaches. Successful integration of AI into clinical practice necessitates close collaboration between clinicians, researchers, and policymakers to ensure ethical and effective implementation. By leveraging interdisciplinary methodologies and emerging technologies, we can establish a pathway toward a future where infertile couples can fulfill their aspirations.

Acknowledgments

We would like to express our deepest gratitude to all those who have contributed to the success of this research. Our sincere thanks go to the esteemed faculty members of the Department of Mechatronics Engineering, whose invaluable guidance in selecting the research topic and unwavering support throughout the research process were instrumental in shaping this work. We also wish to extend our heartfelt appreciation to everyone who has provided support and assistance at every stage of this journey, ensuring that challenges were met with solutions. Finally, our most profound thanks go to our families and friends, whose constant encouragement, love, and understanding provided the strength and inspiration to complete this research.

Funding

None.

Conflict of interest

The authors declare no conflicts of interest.

Author contributions

Conceptualization: Asim Moin Saad
Visualization: Md. Abul Basar Roky, Anonno Singha Ray
Writing – original draft: Md. Abul Basar Roky
Writing – review & editing: Asim Moin Saad, Anonno Singha Ray

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data

All data, including the search strategy, extracted data, and supplementary materials, are available on request from the corresponding author.

References

1. Cox CM, Thoma ME, Tchangalova N, *et al.* Infertility prevalence and the methods of estimation from 1990 to 2021: A systematic review and meta-analysis. *Hum Reprod Open*. 2022;2022(4):hoac051.

   doi: 10.1093/hropen/hoac051

2. Borumandnia N, Alavi Majd H, Khadembashi N, Alaii H. Worldwide trend analysis of primary and secondary infertility rates over past decades: A cross-sectional study. *Int J Reprod Biomed*. 2022;20:37-46.

   doi: 10.18502/ijrm.v20i1.10407

3. Abdullah KAL, Atazhanova T, Chavez-Badiola A, Shivhare SB. Automation in ART: Paving the way for the future of infertility treatment. *Reprod Sci*. 2023;30(4):1006-1016.

   doi: 10.1007/s43032-022-00941-y

4. Raef B, Ferdousi R. A review of machine learning approaches in assisted reproductive technologies. *Acta Inform Med*. 2019;27(3):205-211.

   doi: 10.5455/aim.2019.27.205-211

5. Wang R, Pan W, Jin L, *et al.* Artificial intelligence in reproductive medicine. *Reproduction*. 2019;158(4):R139-R154.

   doi: 10.1530/REP-18-0523

6. Sunde A, Balaban B. The assisted reproductive technology laboratory: Toward evidence-based practice? *Fertil Steril*. 2013;100(2):310-318.

   doi: 10.1016/j.fertnstert.2013.06.032

7. Chu KY, Nassau DE, Arora H, Lokeshwar SD, Madhusoodanan V, Ramasamy R. Artificial intelligence in reproductive urology. *Curr Urol Rep*. 2019;20(9):52.

   doi: 10.1007/s11934-019-0914-4

8. Omur AD. Artificial intelligence in gamete cell selection and microbiologic analysis. *J Clin Vet Res*. 2022;2(2):1-7.

   doi: 10.54289/JCVR2200107

9. Afnan MAM, Rudin C, Conitzer V, *et al.* Ethical Implementation of Artificial Intelligence to Select Embryos in *in Vitro* Fertilization. In: *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*. ACM; 2021. p. 316-326.

   doi: 10.1145/3461702.3462589

10. Meseguer M, Kruhne U, Laursen S. Full *in vitro* fertilization laboratory mechanization: Toward robotic assisted reproduction? *Fertil Steril*. 2012;97(6):1277-1286.

doi: 10.1016/j.fertnstert.2012.03.013

11. Polanski LT, Coelho Neto MA, Nastri CO, *et al*. Time-lapse embryo imaging for improving reproductive outcomes: Systematic review and meta-analysis. *Ultrasound Obstet Gynecol*. 2014;44(4):394-401.

    doi: 10.1002/uog.13428

12. Bormann CL, Curchoe CL, Thirumalaraju P, *et al*. Deep learning early warning system for embryo culture conditions and embryologist performance in the ART laboratory. *J Assist Reprod Genet*. 2021;38(7):1641-1646.

    doi: 10.1007/s10815-021-02198-x

13. Fernandez EI, Ferreira AS, Cecílio MHM, *et al*. Artificial intelligence in the IVF laboratory: Overview through the application of different types of algorithms for the classification of reproductive data. *J Assist Reprod Genet*. 2020;37(10):2359-2376.

    doi: 10.1007/s10815-020-01881-9

14. Zaninovic N, Rosenwaks Z. Artificial intelligence in human *in vitro* fertilization and embryology. *Fertil Steril*. 2020;114(5):914-920.

    doi: 10.1016/j.fertnstert.2020.09.157

15. McKenzie JE, Beller EM, Forbes AB. Introduction to systematic reviews and meta-analysis. *Respirology*. 2016;21(4):626-637.

    doi: 10.1111/resp.12783

16. Theilgaard Lassen J, Fly Kragh M, Rimestad J, Nygård Johansen M, Berntsen J. Development and validation of deep learning based embryo selection across multiple days of transfer. *Sci Rep*. 2023;13(1):4235.

    doi: 10.1038/s41598-023-31136-3

17. Cimadomo D, Chiappetta V, Innocenti F, *et al*. Towards automation in IVF: Pre-clinical validation of a deep learning-based embryo grading system during PGT-A cycles. *J Clin Med*. 2023;12(5):1806.

    doi: 10.3390/jcm12051806

18. Johansen MN, Parner ET, Kragh MF, *et al*. Comparing performance between clinics of an embryo evaluation algorithm based on time-lapse images and machine learning. *J Assist Reprod Genet*. 2023;40(9):2129-2137.

    doi: 10.1007/s10815-023-02871-3

19. Bori L, Meseguer F, Valera MA, Galan A, Remohi J, Meseguer M. The higher the score, the better the clinical outcome: Retrospective evaluation of automatic embryo grading as a support tool for embryo selection in IVF laboratories. *Hum Reprod*. 2022;37(6):1148-1160.

    doi: 10.1093/humrep/deac066

20. Glatstein I, Chavez-Badiola A, Curchoe CL. New frontiers in embryo selection. *J Assist Reprod Genet*. 2023;40(2):223-234.

    doi: 10.1007/s10815-022-02708-5

21. Salih M, Austin C, Warty RR, *et al*. Embryo selection through artificial intelligence versus embryologists: A systematic review. *Hum Reprod Open*. 2023;2023(3):hoad031.

    doi: 10.1093/hropen/hoad031

22. Tian T, Kong F, Yang R, *et al*. A Bayesian network model for prediction of low or failed fertilization in assisted reproductive technology based on a large clinical real-world data. *Reprod Biol Endocrinol*. 2023;21(1):8.

    doi: 10.1186/s12958-023-01065-x

23. Ueno S, Berntsen J, Okimura T, Kato K. Improved pregnancy prediction performance in an updated deep-learning embryo selection model: A retrospective independent validation study. *Reprod Biomed Online*. 2024;48(1):103308.

    doi: 10.1016/j.rbmo.2023.103308

24. Berntsen J, Rimestad J, Lassen JT, Tran D, Kragh MF. Robust and generalizable embryo selection based on artificial intelligence and time-lapse image sequences. *PLoS One*. 2022;17(2):e0262661.

    doi: 10.1371/journal.pone.0262661

25. Chen L, Li W, Liu Y, *et al*. Non-invasive embryo selection strategy for clinical IVF to avoid wastage of potentially competent embryos. *Reprod Biomed Online*. 2022;45(1):26-34.

    doi: 10.1016/j.rbmo.2022.03.006

26. Xi Q, Yang Q, Wang M, *et al*. Individualized embryo selection strategy developed by stacking machine learning model for better *in vitro* fertilization outcomes: An application study. *Reprod Biol Endocrinol*. 2021;19(1):53.

    doi: 10.1186/s12958-021-00734-z

27. Ratna MB, Bhattacharya S, McLernon DJ. External validation of models for predicting cumulative live birth over multiple complete cycles of IVF treatment. *Hum Reprod*. 2023;38(10):1998-2010.

    doi: 10.1093/humrep/dead165

28. Diakiw SM, Hall JMM, VerMilyea MD, *et al*. Development of an artificial intelligence model for predicting the likelihood of human embryo euploidy based on blastocyst images from multiple imaging systems during IVF. *Hum Reprod*. 2022;37(8):1746-1759.

    doi: 10.1093/humrep/deac131

29. Sawada Y, Sato T, Nagaya M, *et al*. Evaluation of artificial intelligence using time-lapse images of IVF embryos to predict live birth. *Reprod Biomed Online*. 2021;43(5):843-852.

    doi: 10.1016/j.rbmo.2021.05.002

30. Cheredath A, Uppangala S, Asha CS, *et al*. Combining machine learning with metabolomic and embryologic data improves embryo implantation prediction. *Reprod Sci*. 2023;30(3):984-994.

doi: 10.1007/s43032-022-01071-1

31. Patil S, Madiwalar S, Aparanji V. Machine learning techniques to improve the success rate in *in-vitro* fertilization (IVF) procedure. *IOP Conf Ser Mater Sci Eng*. 2020;925:012039.

    doi: 10.1088/1757-899X/925/1/012039

32. Giscard d'Estaing S, Labrune E, Forcellini M, *et al*. A machine learning system with reinforcement capacity for predicting the fate of an ART embryo. *Syst Biol Reprod Med*. 2021;67(1):64-78.

    doi: 10.1080/19396368.2020.1822953

33. Kanakasabapathy MK, Thirumalaraju P, Bormann CL, *et al*. Development and evaluation of inexpensive automated deep learning-based imaging systems for embryology. *Lab Chip*. 2019;19(24):4139-4145.

    doi: 10.1039/C9LC00721K

34. Pons MC, Carrasco B, Rives N, *et al*. Predicting the likelihood of live birth: An objective and user-friendly blastocyst grading system. *Reprod Biomed Online*. 2023;47(3):103243.

    doi: 10.1016/j.rbmo.2023.05.015

35. Bori L, Gimenez C, Clark G, Babariya D, Wells D, Meseguer M. O-323 Sex-dependent discrepancies in automatic scoring: Implications for the performance of embryo evaluation and selection methods. *Hum Reprod*. 2024;39(Suppl 1):deae108.380.

    doi: 10.1093/humrep/deae108.380

36. Valera MÁ, Conversa L, Murria L, Bori L, Garg A, Meseguer M. P-174 Do culture conditions alter the efficacy of embryo selection algorithms using time-lapse technology? Development of novel embryo selection model with embryos cultured in different conditions. *Hum Reprod*. 2023;38(Suppl 1):dead093.534.

    doi: 10.1093/humrep/dead093.534

37. Sfontouris I, Nikiforaki D, Liarmakopoulou S, *et al*. P-280 Potential for improvement and current limitations of Artificial Intelligence (AI) for embryo selection: Analysis of external validation data. *Hum Reprod*. 2022;37(Suppl 1):deac107.269.

    doi: 10.1093/humrep/deac107.269

38. Fauser BCJM, Braat DDM. Assisted reproductive technology. In: *Textbook of Obstetrics and Gynaecology*. Netherlands: Bohn Stafleu van Loghum; 2019. p. 263-282.

    doi: 10.1007/978-90-368-2131-5_14

39. Bhandari HM, Choudhary MK, Stewart JA. An overview of assisted reproductive technology procedures. *Obstet Gynaecol*. 2018;20(3):167-176.

    doi: 10.1111/tog.12509

40. Gwatkin RBL. Fertilization *in vitro*. In: *Fertilization Mechanisms in Man and Mammals*. Berlin: Springer US; 1977. p. 33-51.

    doi: 10.1007/978-1-4615-8804-7_5

41. Jurisica I, Mylopoulos J, Glasgow J, Shapiro H, Casper RF. Case-based reasoning in IVF: Prediction and knowledge mining. *Artif Intell Med*. 1998;12(1):1-24.

    doi: 10.1016/S0933-3657(97)00037-7

42. Bing Y, Ouellette RJ. Fertilization *in vitro*. *Methods Mol Biol*. 2009;550:251-266.

    doi: 10.1007/978-1-60327-009-0_16

43. *In Vitro* Fertilization (IVF). In: *The SAGE Encyclopedia of Children and Childhood Studies*. London: SAGE Publications, Inc.; 2020.

    doi: 10.4135/9781529714388.n344

44. Reljič M, Knez J, Kovač V, Kovačič B. Endometrial injury, the quality of embryos, and blastocyst transfer are the most important prognostic factors for *in vitro* fertilization success after previous repeated unsuccessful attempts. *J Assist Reprod Genet*. 2017;34(6):775-779.

    doi: 10.1007/s10815-017-0916-4

45. Aflatoonian A, Asgharnia M. Factors affecting the successful embryo transfer. *IJRM*. 2006;4(2):45-50.

46. Zheng D, Zeng L, Yang R, *et al*. Intracytoplasmic sperm injection (ICSI) versus conventional *in vitro* fertilisation (IVF) in couples with non-severe male infertility (NSMI-ICSI): Protocol for a multicentre randomised controlled trial. *BMJ Open*. 2019;9(9):e030366.

    doi: 10.1136/bmjopen-2019-030366

47. Hardy K, Wright C, Rice S, *et al*. Future developments in assisted reproduction in humans. *Reproduction*. 2002;123(2):171-183.

    doi: 10.1530/rep.0.1230171

48. Palermo GD, Sills ES, editors. *Intracytoplasmic Sperm Injection*. Berlin: Springer International Publishing; 2018.

    doi: 10.1007/978-3-319-70497-5

49. Simopoulou M, Sfakianoudis K, Tsioulou P, *et al*. Risks in surrogacy considering the embryo: From the preimplantation to the gestational and neonatal period. *Biomed Res Int*. 2018;2018:1-9.

    doi: 10.1155/2018/6287507

50. Igreja AR, Ricou M. Surrogacy: Challenges and ambiguities. *New Bioethics*. 2019;25(1):60-77.

    doi: 10.1080/20502877.2019.1564007

51. van den Akker OBA. Introduction to surrogacy: Historical and present day context. In: *Surrogate Motherhood Families*. Berlin: Springer International Publishing; 2017. p. 5-37.

    doi: 10.1007/978-3-319-60453-4_1

52. Patel N, Jadeja Y, Bhadarka H, Patel M, Patel N, Sodagar N.

Insight into different aspects of surrogacy practices. *J Hum Reprod Sci*. 2018;11(3):212.

doi: 10.4103/jhrs.JHRS_138_17

53. Nakash A, Herdiman J. Surrogacy. *J Obstet Gynaecol (Lahore)*. 2007;27(3):246-251.

    doi: 10.1080/01443610701194788

54. Seidinova MA. Ethical issues of surrogacy. *Bull Karagandy Univ Law Ser*. 2023;110:139-146.

    doi: 10.31489/2023L2/139-146

55. Bashiri A, Cherlow Y, Kresch-Jaffe T. Surrogacy: An important pathway to parenthood. A call for international standardization. *J Reprod Immunol*. 2024;163:104247.

    doi: 10.1016/j.jri.2024.104247

56. Mulligan A. Surrogacy and the significance of gestation: Implications for law and policy. *Bioethics*. 2024;38(8):674-683.

    doi: 10.1111/bioe.13302

57. Patil SN, Wali UV, Swamy MK. Selection of Single Potential Embryo to Improve the Success Rate of Implantation in IVF Procedure using Machine Learning Techniques. In: *2019 International Conference on Communication and Signal Processing (ICCSP)*. IEEE; 2019. p. 881-886.

    doi: 10.1109/ICCSP.2019.8697982

58. Duffus JH, Schwenk M, Templeton DM. *Gamete Intrafallopian Transfer (GIFT)*. United States: IUPAC Standards Online.

    doi: 10.1515/iupac.88.0833

59. Tucker MJ, Leung CK, Leong MK, *et al*. Routine gamete intrafallopian transfer (GIFT): A highly successful option for treatment of non-tubal infertility. *Asia Oceania J Obstet Gynaecol*. 1989;15(3):245-251.

    doi: 10.1111/j.1447-0756.1989.tb00184.x

60. Tanaka A, Toeda M, Nagasawa I, Yamamoto T, Nakagata N. Development of gamete intrafallopian transfer with preincubated preovulatory oocytes (new GIFT) and analysis of clinical data. *Nihon Sanka Fujinka Gakkai Zasshi*. 1988;40(12):1859-1866.

61. Meirow D, Schenker JG. Appraisal of gamete intrafallopian transfer. *Eur J Obstet Gynecol Reprod Biol*. 1995;58(1):59-65.

    doi: 10.1016/0028-2243(94)01940-9

62. Pool TB, Ellsworth LR, Garza JR, Martin JE, Miller SS, Atiee SH. Zygote intrafallopian transfer as a treatment for nontubal infertility: A 2-year study. *Fertil Steril*. 1990;54(3):482-488.

    doi: 10.1016/S0015-0282(16)53766-2

63. Tournay H, Camus M, Khan I, Staessen C, Van Steirteghem AC, Devroey P. *In-vitro* fertilization, gamete- or zygote intra-Fallopian transfer for the treatment of male infertility. *Hum Reprod*. 1991;6(2):263-266.

    doi: 10.1093/oxfordjournals.humrep.a137319

64. Farhi J, Weissman A, Nahum H, Levran D. Zygote intrafallopian transfer in patients with tubal factor infertility after repeated failure of implantation with *in vitro* fertilization-embryo transfer. *Fertil Steril*. 2000;74(2):390-393.

    doi: 10.1016/S0015-0282(00)00610-5

65. Levran D, Mashiach S, Dor J, Levron J, Farhi J. Zygote intrafallopian transfer may improve pregnancy rate in patients with repeated failure of implantation. *Fertil Steril*. 1998;69(1):26-30.

    doi: 10.1016/S0015-0282(97)00452-4

66. Tournaye H, Devroey P, Camus M, Valkenburg M, Bollen N, Van Steirteghem AC. Zygote intrafallopian transfer or in vitro fertilization and embryo transfer for the treatment of male-factor infertility: A prospective randomized trial. *Fertil Steril*. 1992;58(2):344-350.

    doi: 10.1016/S0015-0282(16)55195-4

67. Fluker MR, Zouves CG, Bebbington MW. A prospective randomized comparison of zygote intrafallopian transfer and *in vitro* fertilization-embryo transfer for nontubal factor infertility. *Fertil Steril*. 1993;60(3):515-519.

    doi: 10.1016/S0015-0282(16)56170-6

68. Manna C, Nanni L, Lumini A, Pappalardo S. Artificial intelligence techniques for embryo and oocyte classification. *Reprod Biomed Online*. 2013;26(1):42-49.

    doi: 10.1016/j.rbmo.2012.09.015

69. Góngora A, Miranda JA, Gordillo E, Dabbah J. Statistical analysis on artificial insemination human fertility Centre, Mexico City. *Int Congr Ser*. 2004;1271:3-6.

    doi: 10.1016/j.ics.2004.06.018

70. Dalal RJ, Mishra AP, Gupta S. Artificial intelligence in assisted reproductive technology: Present and future. *Int J Infertil Fetal Med*. 2020;11(3):61-64.

    doi: 10.5005/jp-journals-10016-1208

71. Jain K. Artificial intelligence in assisted reproductive technology-current scenario and future implications. *Fertil Sci Res*. 2019;6(2):57.

    doi: 10.4103/fsr.fsr_40_19

72. Montag M, Toth B, Strowitzki T. New approaches to embryo selection. *Reprod Biomed Online*. 2013;27(5):539-546.

    doi: 10.1016/j.rbmo.2013.05.013

73. Bulletti FM, Berrettini M, Sciorio R, Bulletti C. Artificial intelligence algorithms for optimizing assisted reproductive technology programs: A systematic review. *Global Transl Med*. 2023;2(2):0308.

    doi: 10.36922/gtm.0308

74. Bhagwan KV, Kadam S. A comprehensive study on artificial intelligence and machine learning. *Int J Adv Res Sci Commun*

Technol. 2024;4:276-279.

doi: 10.48175/IJARSCT-19939

75. Lin J, Sun XX. Predictive modeling in reproductive medicine. *Reprod Dev Med*. 2018;2(4):224-229.

    doi: 10.4103/2096-2924.249888

76. Rosenwaks Z. Artificial intelligence in reproductive medicine: A fleeting concept or the wave of the future? *Fertil Steril*. 2020;114(5):905-907.

    doi: 10.1016/j.fertnstert.2020.10.002

77. Raimundo J, Cabrita P. Artificial intelligence at assisted reproductive technology. *Procedia Comput Sci*. 2021;181:442-447.

    doi: 10.1016/j.procs.2021.01.189

78. Dimitriadis I, Zaninovic N, Badiola AC, Bormann CL. Artificial intelligence in the embryology laboratory: A review. *Reprod Biomed Online*. 2022;44(3):435-448.

    doi: 10.1016/j.rbmo.2021.11.003

79. Afnan MAM, Liu Y, Conitzer V, *et al*. Interpretable, not black-box, artificial intelligence should be used for embryo selection. *Hum Reprod Open*. 2021;2021(4):hoab040.

    doi: 10.1093/hropen/hoab040

80. Benchaib M, Labrune E, Giscard d'Estaing S, Salle B, Lornage J. Shallow artificial networks with morphokinetic time-lapse parameters coupled to ART data allow to predict live birth. *Reprod Med Biol*. 2022;21(1):e12486.

    doi: 10.1002/rmb2.12486

81. Yu L, Lam KKW, Ng EHY, *et al*. *Deep Learning-Based Embryo Assessment of Static Images can Reduce the Time to Live Birth in In Vitro Fertilization*. medRxiv [Preprint]; 2024.

    doi: 10.1101/2024.10.28.24316259

82. Liao Z, Yan C, Wang J, *et al.* A clinical consensus-compliant deep learning approach to quantitatively evaluate human *in vitro* fertilization early embryonic development with optical microscope images. *Artif Intell Med*. 2024;149:102773.

    doi: 10.1016/j.artmed.2024.102773

83. Popa T, He C, Vasconcelos F, *et al*. P-168 Both artificial intelligence and manual embryo selection methods show sex-bias, favouring male embryos-insights from the largest embryo sex study using time-lapse and PGT-A. *Hum Reprod*. 2024;39(Suppl 1):deae108.539.

    doi: 10.1093/humrep/deae108.539

84. Illingworth PJ, Venetis C, Gardner DK, *et al*. Deep learning versus manual morphology-based embryo selection in IVF: A randomized, double-blind noninferiority trial. *Nat Med*. 2024;30(11):3114-3120.

    doi: 10.1038/s41591-024-03166-5

85. Huang JYJ, Rosenwaks Z. Assisted reproductive techniques. *Methods Mol Biol*. 2014;1154:171-231.

    doi: 10.1007/978-1-4939-0659-8_8

86. Prostate pathophysiology. In: *Atlas of Clinical Andrology*. United States: CRC Press; 2005. p. 146-157.

    doi: 10.1201/b14619-17

87. Kragh MF, Karstoft H. Embryo selection with artificial intelligence: How to evaluate and compare methods? *J Assist Reprod Genet*. 2021;38(7):1675-1689.

    doi: 10.1007/s10815-021-02254-6

88. Merican ZZ, Yusof UK, Abdullah NL. *Review on Embryo Selection Based on Morphology Using Machine Learning Methods*; 2021. Available from: https://api.semanticscholar.org/corpusid:236880800 [Last accessed on 2025 Apr 28].

89. Medenica S, Zivanovic D, Batkoska L, *et al*. The future is coming: Artificial intelligence in the treatment of infertility could improve assisted reproduction outcomes-the value of regulatory frameworks. *Diagnostics*. 2022;12(12):2979.

    doi: 10.3390/diagnostics12122979

90. Curchoe CL. Meetings that matter: Time to put artificial intelligence on the ART roadmap. *J Assist Reprod Genet*. 2022;39(7):1493-1496.

    doi: 10.1007/s10815-022-02520-1

91. Tran HP, Tran LNH, Dang HT, *et al.* A SWOT analysis of human- and machine learning- based embryo assessment. *IEEE Access*. 2020;8:227466-227481.

    doi: 10.1109/ACCESS.2020.3045772

92. Kuo C, Zuo J, Han W, *et al*. *Intelligent Assisted Reproduction: Innovative Applications of artificial Intelligence in Embryo Health Assessment*. Authorea [Preprints]; 2025.

    doi: 10.22541/au.173639046.67662628/v1

93. Presacan O, Dorobanțiu A, Thambawita V, *et al*. *Embryo 2.0: Merging Synthetic and Real Data for Advanced AI Predictions* [Preprint]; 2024.

    doi: 10.48550/arxiv.2412.01255

94. Kaveh S, Ghafari A, Khedri Z, *et al*. *Investigating the Artificial Intelligence in Prediction and Evaluation of Sperm and Embryo Quality in In Vitro Fertilization (IVF): A Systematic Review*. [Preprint (Version 1)]; 2024. [Last accessed on 2025 Apr 28].

    doi: 10.21203/rs.3.rs-5504223/v1

95. *Embryo Ranking Intelligent Classification Algorithm. C. to, "Embryo Ranking Intelligent Classification Algorithm*; 2020. Available from: https://en.wikipedia.org/wiki/embryo_ranking_intelligent_classification_algorithm? [Last accessed on 2025 Feb 19].

96. Leahy B, Jang WD, Yang H, *et al*. Automated Measurements of Key Morphological Features of Human Embryos for IVF. *Med Image Comput Comput Assist Interv*. 2020:12265:25-35.

doi: 10.1007/978-3-030-59722-1_3

97. Mushtaq A, Mumtaz M, Raza A, Salem N, Yasir MN. Artificial intelligence-based detection of human embryo components for assisted reproduction by in vitro fertilization. *Sensors*. 2022;22(19):7418.

    doi: 10.3390/s22197418

98. Allahbadia GN. Embryo transfer is the last frontier for deep machine learning artificial intelligence in medically assisted reproduction (MAR). *J Reprod*. 2023;2(1):18.

    doi: 10.58779/issn.2954-467X.tjor2023.v2.n1.18

99. Liu R, Bai S, Jiang X, *et al*. Multifactor prediction of embryo transfer outcomes based on a machine learning algorithm. *Front Endocrinol (Lausanne)*. 2021;12:745039.

    doi: 10.3389/fendo.2021.745039

100. Curchoe CL, Bormann CL. Artificial intelligence and machine learning for human reproduction and embryology presented at ASRM and ESHRE 2018. *J Assist Reprod Genet*. 2019;36(4):591-600.

    doi: 10.1007/s10815-019-01408-x

101. Mills C, Johnston M, Koplin J. O-220 Artificial intelligence for embryo selection: Ethical, social and regulatory issues. *Hum Reprod*. 2023;38(Suppl 1):dead093.266.

    doi: 10.1093/humrep/dead093.266

102. Ahlström A, Berntsen J, Johansen M, *et al*. Correlations between a deep learning-based algorithm for embryo evaluation with cleavage-stage cell numbers and fragmentation. *Reprod Biomed Online*. 2023;47(6):103408.

    doi: 10.1016/j.rbmo.2023.103408

103. Nasr M, Mohamed M, Shehata L. Artificial intelligence in assisted reproductive technology review. *Int J Progress Sci Technol*. 2021;25:507-11.

104. Barrie A, Homburg R, McDowell G, Brown J, Kingsland C, Troup S. Examining the efficacy of six published time-lapse imaging embryo selection algorithms to predict implantation to demonstrate the need for the development of specific, in-house morphokinetic selection algorithms. *Fertil Steril*. 2017;107(3):613-621.

    doi: 10.1016/j.fertnstert.2016.11.014

105. Fitz VW, Kanakasabapathy MK, Thirumalaraju P, *et al*. Should there be an "AI" in TEAM? Embryologists selection of high implantation potential embryos improves with the aid of an artificial intelligence algorithm. *J Assist Reprod Genet*. 2021;38(10):2663-2670.

    doi: 10.1007/s10815-021-02318-7

ACCSCIENCE PUBLISHING

Artificial Intelligence in Health

REVIEW ARTICLE

A review of neuroscience-inspired frameworks for machine consciousness

Seyed-Ali Sadegh-Zadeh[1*] and **Mahboobe Bahrami**[2]

[1]Department of Computing, School of Digital, Technologies and Arts, University of Staffordshire, Stoke-on-Trent, United Kingdom

[2]Behavioral Sciences Research Center, School of Medicine, Isfahan University of Medical Sciences, Isfahan, Isfahan, Iran

Abstract

Consciousness in humans is a state of awareness that encompasses both the self and the external environment, emerging from the intricate interplay of cortical and subcortical brain structures and neurotransmitter systems. The possibility that machines could possess consciousness has sparked ongoing debate. Proponents of strong artificial intelligence (AI) equate programmed computational processes with cognitive states, while advocates of weak AI argue that machines merely simulate thought without attaining genuine consciousness. This review critically examines neuroscience-inspired frameworks for artificial consciousness, exploring their alignment with prevailing theories of human consciousness. We investigate the fundamental cognitive functions associated with consciousness, including memory, awareness, prediction, learning, and experience, and their relevance to artificial systems. By analyzing neuroscience-based approaches to artificial consciousness, we identify key challenges and opportunities in the pursuit of machines capable of mimicking conscious states. Although present AI systems demonstrate advanced capabilities in intelligence and cognition, they fall short of achieving genuine consciousness, as defined in the context of human awareness. We discuss both the theoretical underpinnings and practical implications of creating artificial consciousness, addressing both weak and strong AI perspectives. Furthermore, we highlight the ethical and philosophical concerns that arise with the potential realization of machine consciousness. Our objective is to provide a comprehensive synthesis of the literature, fostering a deeper understanding of the interdisciplinary challenges involved in artificial consciousness and guiding future research directions.

Keywords: Machine brain; Cortical; Subcortical; Weak artificial intelligence; Strong artificial intelligence; Human consciousness; Machine consciousness

***Corresponding author:**
Seyed-Ali Sadegh-Zadeh
(sa.sadeghzadeh@staffs.ac.uk)

Citation: Sadegh-Zadeh S, Bahrami M. A review of neuroscience-inspired frameworks for machine consciousness. *Artif Intell Health*. 2025;2(3):24-38. doi: 10.36922/aih.5690

Received: October 29, 2024

1st revised: January 18, 2025

2nd revised: March 10, 2025

Accepted: March 28, 2025

Published online: June 4, 2025

Copyright: © 2025 Author(s). This is an Open-Access article distributed under the terms of the Creative Commons Attribution License, permitting distribution, and reproduction in any medium, provided the original work is properly cited.

Publisher's Note: AccScience Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

The varied definitions of consciousness across disciplines present significant challenges to its study. Neuroscience seeks to identify the neural correlates of consciousness – conditions necessary for its emergence – and to map its dynamic changes.[1] Psychology and psychiatry, in contrast, focus primarily on the experiential and functional outcomes of consciousness.[2] These disciplinary distinctions offer diverse lenses through which consciousness can be understood, underscoring the need for a multidisciplinary approach.

Consciousness, intrinsically linked to the complex processes of the human brain, is broadly defined as sensitivity and awareness of internal and external existence.[3,4] Contemporary inquiries into consciousness in medicine and psychology often draw on experimental studies and clinical cases involving changes induced by trauma, disease, or pharmacological interventions.[5,6] Scientific approaches to consciousness generally rest on two key ideas: One emphasizing human subjective experiences and their content, and the other focusing on the neurological underpinnings observed in clinical treatments for neurological and behavioral disorders.[7,8]

Consciousness is increasingly understood as an emergent property of neuronal connections,[9] representing a cascade of events that evolve over time to drive change.[10] Rather than a binary state, contemporary perspectives propose a spectrum of conscious states, from basic awareness to more intricate manifestations of self-consciousness.[11,12]

1.1. Human consciousness

A thorough understanding of human consciousness is essential before exploring its potential replication in machines. Human consciousness is not easily categorized or isolated,[13,14] as it manifests in various forms.[15-17] Most of human cognitive activity occurs in states of primary consciousness, which include mind-wandering activities such as recalling personal memories, envisioning future scenarios, and adopting the perspectives of others.[18]

The human brain, as part of the central nervous system, serves as the biological foundation of consciousness.[19] Understanding this biological basis shed light on the diverse manifestations of human consciousness.[20,21] To assess the feasibility of artificial consciousness, it is crucial to consider our present knowledge of the neurological structures that underpin human conscious experience.[22]

Consciousness, at its core, is a state of awareness of oneself and the surrounding environment. It encompasses sensory recognition and awareness, both of which cease during states such as sleep, coma, or death. Clinically, consciousness is viewed in three dimensions: First, as an inner awareness of experiential events; second, as an intentional reaction toward external objects; and third, as knowledge of one's conscious self. In states of full wakefulness, the intensity of consciousness varies significantly, often heightening during challenging experiences. Awareness itself can be divided into three dimensions: Vigilance, lucidity, and self-consciousness. "Vigilance" refers to the ability to remain purposefully alert; "lucidity" denotes clarity of thought regarding a specific subject; and "self-consciousness" entails the capacity to perceive oneself as an individual entity. Disorders of consciousness – whether quantitative or qualitative – fall beyond the scope of this article.[23-25]

The brain systems that constitute consciousness develop from those that control its level.[26] The foundation for varying levels of consciousness lies in the content of consciousness. Figure 1 illustrates the hierarchical organization of sensory and motor systems, arranged in parallel and integrated to underpin consciousness. At the top level, consciousness encompasses and coordinates functions, such as emotions, memory, and sensory-motor processes. Emotions, positioned below consciousness, act as intermediaries, integrating signals from memory and sensory systems. Memory serves as the central information repository, directly interfacing with motor and sensory systems. The sensory systems provide environmental inputs, while motor systems execute outputs based on processed information. This structural arrangement highlights how these systems collaboratively handle information flow, enabling adaptive and abstract functions across various levels of complexity.

Understanding how these systems typically operate is a central objective of neuroscience. From a neuroscience perspective, the level of consciousness influences all neuronal processes.[27] Specific cortical and subcortical processes regulate attention and awareness, which in turn determine the level of consciousness.[28] Any meaningful reactions require at least a minimal degree of attention, which facilitates choice and sustained information processing. The capacity to generate experiences that can later be reported is known as awareness.[29]

The brain circuits that regulate consciousness are commonly referred to as the consciousness circuit.[30] These

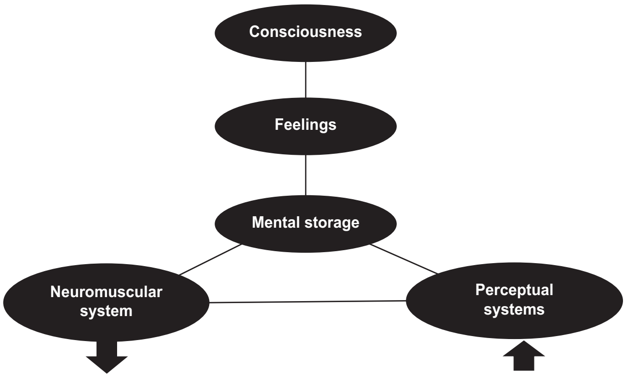


Figure 1. Sensory and motor systems are coupled in parallel, arranged hierarchically, and perform input, output, and internal processing functions across a range from basic to highly abstract. Image created by the author.

networks control the level of consciousness.[31] Studies using both animal models and human brain disorder cases have long recognized that cortical and subcortical regions play a crucial role in the state of awareness.[32] The cortical components of the consciousness system are primarily located in the higher-order "heteromodal" association cortex.[33] On the medial surface of the brain, these include the medial frontal, anterior cingulate, posterior cingulate, and medial parietal cortices. On the lateral surface, the consciousness system involves the lateral frontal cortex, anterior insula, orbital frontal cortex, and lateral temporal-parietal association area.[34,35]

Specific areas within the higher-order association cortex significantly influence cognitive functions in both the dominant and non-dominant hemispheres.[36] The state of consciousness is determined by the combined activity of extensive portions of the bilateral association cortex, regardless of the distinct roles played by individual regions or networks.[37,38] These higher-order cortices collaborate with subcortical arousal systems to regulate arousal, attention, and consciousness.[39]

Subcortical components of the consciousness system include the basal forebrain, thalamus, hypothalamus, and upper brainstem activation systems. Other subcortical regions, such as portions of the cerebellum, amygdala, claustrum, and basal ganglia, are likely involved as well. Several parallel neurotransmitter systems contribute to subcortical arousal, including glutamate, acetylcholine, serotonin, dopamine, histamine, gamma-aminobutyric acid, norepinephrine, and orexin. Each of these pathways plays a distinct role. However, it is the coordinated and simultaneous actions of these subcortical and cortical structures that collectively regulate the degree of consciousness.[40]

2. Consciousness-related elements

Research emphasizes that the proper functioning of consciousness is integral to various mental processes.[41-43] Studies suggest that consciousness fulfills multiple roles,[42,44,45] including memory, prediction, awareness, learning, and experience – components intricately linked to cognitive processes. A primary goal of artificial consciousness is to replicate these components in machines. This list is not exhaustive, as numerous additional functions of consciousness remain unexplored.

2.1. Memory

Memory is a fundamental aspect of human cognition and a growing focus in artificial consciousness systems. In humans, memory storage occurs in three forms: Short-term, long-term, and sensory memory. As artificial consciousness evolves to handle increasingly complex scenarios, the importance of robust memory systems continues to grow. However, present artificial consciousness frameworks often lack sophisticated memory models. In humans, memory systems interact closely with conscious experience during learning, rehearsal, and retrieval.[46,47]

2.2. Prediction

Prediction is considered a vital capability for artificial consciousness systems. Conscious organisms predict events by reflecting relationships between real-world states within their internal structures.[43] Similarly, an artificially conscious machine must be capable of accurate event prediction to respond effectively or take pre-emptive measures. This requires real-time, flexible components capable of constructing causal, statistical, spatial, functional, and dynamic models of the world. A conscious machine should demonstrate predictive abilities across various contexts, including uncertain and dynamic environments, showcasing coherent forecasting and contingency planning.[45]

2.3. Awareness

Awareness, though challenging to define precisely, involves constructing and testing alternative models of processes based on sensory or cognitive information.[48] It is essential for predictions[49] and requires significant flexibility to model the physical environment, internal states, and other conscious entities. This adaptability is crucial for artificial systems attempting to replicate human-like awareness.

2.4. Learning

Learning is a cornerstone of artificial consciousness systems.[50,51] Conscious experience facilitates the representation of and reaction to novel, significant stimuli.[46] Learning encompasses complex adaptation mechanisms grounded in sensitivity to subjective experience, enabling agents to exert flexible control over behaviors in uncertain and dynamic environments.[52]

2.5. Experience

Subjective experience, rooted in sensory perception, is often considered central to the study of consciousness.[53] Experience is intrinsically linked to precise pattern recognition and may even be observed at molecular levels. In discussing consciousness, the brain's central modules can be viewed as carriers of unique experiential states. Reflexive awareness, or the act of reflecting on one's own experiences, is a critical dimension of consciousness. Efforts to define fundamental experiences underscore the enduring challenge of addressing the "hard problem" of consciousness, as its exact nature remains elusive.[54]

3. Machine consciousness

Machine consciousness is often explored within the context of autonomous artificial intelligence (AI) systems capable of self-learning.[55] Contemporary systems extend beyond basic hardware and administrative software, encompassing sophisticated layers of programmed control that integrate hardware, software, memory, and interfaces. Finite-state machines can fulfill some criteria associated with human-like consciousness.[56] While intelligence and consciousness were once conflated,[57] this view has largely been abandoned, as finite-state machines excel in achieving pre-defined intelligence metrics without demonstrating true consciousness.[58] Modern algorithms, leveraging databases, outperform human experts in tasks requiring formal reasoning, planning, language processing, gameplay, and arithmetic—accomplishing these through efficient algorithmic symbol manipulation.

Machine intelligence often exceeds biological systems, including humans, in multidimensional attentional focus. Artificial sensors enable machines to process and respond to stimuli beyond the human sensory spectrum.[59] These capabilities, combined with continuous data monitoring and reaction, allow machines to excel in attentional focus across various domains. Machine interfaces are widely used for attention training in children, individuals with brain injuries, and those with psychiatric conditions where maintaining focus presents challenges.[60] Intentionality, a core feature of finite state machines, is parametrically programmed into most artificial systems capable of numerical information processing.[61] While humans often anthropomorphize these systems by attributing deliberate purpose to them,[62] extending this attribution to human-like beliefs, intentions, or causation remains implausible for present artificial systems.[63] Some finite-state machine systems have been capable of achieving astounding autonomy levels within the set confines of their coding, as they are now built.

When utilized for specific activities in industries such as manufacturing, home appliances, automotive systems, space exploration, and remote operations, finite-state systems have the ability to run for extended periods without programmer input. System or body consciousness is necessary for the operation of many surgical and technology assessment systems in the medical field.[56] The ability of some systems to format and create computer-presented narratives can potentially be used to infer phenomenological self-representation.[56] The metacognitive trait of knowledge of being aware, which goes beyond body/system self-awareness, has been proposed as a signal that would show whether an artificially produced system is capable of going beyond programming. According to some theories, this type of metaconsciousness could mark the emergence of human-like consciousness in robots.[64] However, no currently existing or planned system has demonstrated clear evidence of processing such awareness.

For an AI system to achieve consciousness and exhibit volition, it must possess the ability to modify and develop its own governing principles. This concept, referred to as coherent extended volition, describes the capacity for self-regulated, self-defined learning.[65] Despite efforts to endow self-learning AI systems with this ability, there is minimal evidence that any system has transcended its programming.[66] Anthropomorphic robotic systems aim to replicate human physiognomy and behavior, potentially simulating human-like consciousness and actions. However, such advancements remain largely speculative and confined to science fiction.

3.1. Weak AI

Weak AI refers to AI systems designed for narrowly defined tasks, employing only a fraction of human cognitive capabilities.[67,68] These systems excel at mimicking human behaviors in fundamental tasks such as learning, perception, and problem-solving.[69] However, weak AI lacks the capacity for independent thought or decision-making.[70,71] Contrary to popular belief, cognitively inspired AI systems align with the weak AI hypothesis, as they model mental phenomena without claiming to replicate the underlying consciousness. This hypothesis remains consistent with present trends in AI and cognitive modeling research.[72]

3.2. Strong AI

Strong AI represents a conceptual framework aiming to develop machines with human-like intellect, consciousness, and the ability to reason, learn, and plan.[73,74] Such systems would not only mimic human thought but also exhibit autonomous cognitive abilities indistinguishable from the human mind. Strong AI envisions machines capable of acquiring new skills through experience and improving over time.[75] Despite significant research interest in artificial general intelligence, which underpins the strong AI concept, it remains a theoretical construct rather than a realized technology.

4. Discussion

4.1. Strong AI versus weak AI: Divergent perspectives on machine consciousness

The question of whether an artificial system can truly be conscious has fuelled enduring debate, dividing opinion into strong AI and weak AI camps. Proponents of strong AI contend that a sufficiently well-designed computational system could literally possess a mind – in other words, that executing the right algorithms might generate genuine understanding and cognitive states indistinguishable from those of humans.[76] This perspective implies that, at some level, the functional organization of a machine could support conscious in the same sense a brain does. In contrast, the weak AI position holds that machines, at best, simulate consciousness without any real inner experience or awareness.[77] From this viewpoint, even the most advanced AI today (for example, sophisticated language models or robotic assistants) lack subjective sentience or genuine understanding; they merely manipulate symbols and exhibit behaviors that mimic consciousness without actually experiencing the world.

The clash between these perspectives highlights a core conceptual challenge: Explaining how subjective experience (the essence of consciousness) might emerge from purely physical or computational processes. This is essentially the classic "hard problem of consciousness" applied to machines: The difficulty of explaining how and why a physical system could produce the felt quality of experience. Even in humans, consciousness defies simple explanation; present scientific understanding of brain function has yet to fully bridge the gap between neural circuitry and subjective feeling.

When considering artificial agents, we are further constrained by our human-centric intuitions: Our understanding of consciousness is largely shaped by the first-person experience of our own mind, making it challenging to objectively evaluate whether a machine – accessible only from an external, third-person perspective – could possess anything akin to a conscious mind.

In summary, the strong AI versus weak AI dichotomy sets the stage for discussing machine consciousness by asking whether replicating intelligent behavior is sufficient for authentic consciousness (strong AI) or whether subjective awareness is a qualitatively distinct property that machines inherently lack (weak AI). This foundational debate provides a context for interpreting the progress in neuroscience-inspired AI frameworks and guides our skepticism or optimism regarding artificial consciousness.

4.2. Neuroscience-inspired functional frameworks for artificial consciousness

Amid these philosophical debates, researchers have proposed various frameworks for building or recognizing consciousness-like properties in machines. Often drawing inspiration from neuroscience and cognitive science, these frameworks focus on replicating functional attributes of human consciousness in an artificial medium. One pragmatic stance, advocated by Levy,[78] suggests setting aside the notoriously difficult task of pinning down an exact definition of consciousness and instead agreeing on practical operational criteria. Levy argues that insisting on a rigid definition may be counterproductive; rather, if the community can settle on a shared intuitive understanding of what consciousness functionally entails, researchers could "simply use the word and get on with it" in developing systems that meet those criteria.[78(p210)] This approach reflects a practical mindset: Even if we lack a perfect definition of consciousness, we might still engineer systems that everyone agrees exhibit key properties of consciousness (such as complex adaptivity, learning, and self-report), thereby moving the field forward without becoming mired in semantics.

Other researchers emphasize specific features thought to be indispensable for consciousness. Chatila *et al.*[79] focus on self-awareness as the cornerstone of machine consciousness, proposing a framework for self-aware robots grounded in several cognitive abilities. They outline fundamental principles by which a robot could be designed to understand its environment and its own role within it, to be cognizant of its actions, and to respond appropriately in real time to changes. Crucially, a self-aware robot should also be able to learn from its own experiences and mistakes and to explicitly demonstrate that it has learned – for instance, by documenting or communicating its acquired knowledge. These capabilities mirror aspects of human consciousness: Humans continuously monitor their surroundings and their own internal states, adjust behavior on the fly, learn from feedback, and can report on what they have learned. Chatila's framework thus attempts to imbue machines with a form of reflective cognition analogous to that of humans, on the premise that such reflection (knowing what one knows, and knowing what one does) is a pre-requisite for any genuine consciousness.

A complementary perspective is offered by Kinouchi and Mackin,[80] who propose that consciousness serves a functional role as an integrative system-level adaptation mechanism in complex agents. In their architecture, a multitude of lower-level processing units (analogous to distributed modules in the brain or in a large AI system) operate in parallel, each handling specific tasks or sensory inputs. Machine consciousness, in this view, is the higher-level function that coordinates and organizes the outputs of these parallel processes, synthesizing them into a coherent state that can guide the agent's overall behavior adaptively. This coordinating role is likened to the way human consciousness creates a unified experience and decision-making process out of the brain's many simultaneous unconscious computations. Kinouchi and Mackin[80] and Hildt[81] explicitly draw an analogy to the moment-to-moment awareness we experience in daily life when making rapid decisions. In humans, despite a flurry of unconscious sensory and cognitive processing, consciousness provides a singular, integrated vantage point (the feeling of "being aware") that helps us adaptively navigate each moment. By mimicking this in machines – ensuring that an artificial agent has an integrating layer that monitors and directs sub-processes – their framework aims to achieve a conscious-like functionality that could improve the system's flexibility and robustness in unpredictable environments. Notably, these authors regard such architecture not just as an add-on to intelligence, but as essential for complex adaptive behavior: A machine endowed with a consciousness-like integrative function might better handle novel situations by flexibly combining information from all its subsystems.

The above frameworks illustrate how insights from neuroscience and cognitive psychology (such as the importance of self-monitoring and global integration of information) are being translated into AI design. Each approach stresses a different facet of natural consciousness: From Levy's broad pragmatism to Chatila's self-reflective knowledge, to Kinouchi's global integration. The diversity of these proposals also underscores that there is not yet a consensus on a single "blueprint" for artificial consciousness. Different researchers prioritize different cognitive ingredients (self-awareness, learning, integration, etc.), reflecting the multifaceted nature of consciousness itself. This plurality suggests that the field is still in an exploratory phase: Much like the blind men and the elephant, each framework captures one aspect of the larger concept. A key task for the research community moving forward is to synthesize these insights and determine how they might fit together. For instance, one could ask whether a truly conscious machine would need to incorporate all of these elements – a shared functional understanding, self-awareness, and global integrative capacity – or whether any one of them might be sufficient on its own. Addressing such questions requires not only engineering advances but also deeper theoretical clarity, which brings us to the distinction between different notions of consciousness and how they apply in artificial systems.

4.3. Access versus phenomenal consciousness: Functional versus experiential dimensions

In discussions of both human and machine consciousness, it is crucial to distinguish between two often-confused dimensions of conscious states: Phenomenal consciousness and access consciousness.[82] This distinction, originally articulated by Block,[82] has proven useful in framing debates about consciousness in artificial systems.

Phenomenal consciousness refers to the subjective experience itself – the raw feel of sensations and thoughts, often described as "what it is like" to be in a given mental state. It encompasses the qualitative, first-person aspects of mind (sometimes called qualia), such as the redness of red or the pang of emotion. By contrast, access consciousness denotes a mental state's availability for use by the cognitive system. A piece of information is "access conscious" if it is widely broadcast within the brain (or system) such that various processes (reasoning, memory, decision-making, verbal report) can utilize it. In essence, access consciousness concerns the functional role of conscious information – how it is accessible and how it guides behavior – rather than how it feels.

This distinction has profound implications for artificial consciousness. Most neuroscience-inspired AI frameworks implicitly aim at access consciousness – ensuring that an AI system possesses internal representations that are globally available and can be used to organize behavior in an intelligent, context-sensitive way. For example, when Chatila *et al.*[79(p1)] focus on robots "knowing what they have learned" and reporting that knowledge, they are dealing with access consciousness: The learned information is accessible for future decisions and self-report. Similarly, Kinouchi and Mackin's[80] integrative layer is designed to collect distributed information and make it available to the whole system for coordinated adaptation – again, a functional, access-oriented property.

Phenomenal consciousness, however, is a much harder issue. It asks whether the robot or AI actually has an inner life: Is there something that it is like to be that robot? Does it feel anything when it integrates information or reports on its knowledge? This is the crux of the hard problem in the context of AI. Strong AI enthusiasts might argue that if we achieve a complete functional emulation of the brain's processes (i.e., replicate access consciousness to a high degree), then phenomenal experience might emerge naturally. However, skeptics point out that no matter how sophisticated a machine's functional capabilities, this does not guarantee – or even necessarily imply – the presence of subjective experience.[83] A machine could conceivably meet every external criterion for access consciousness – it could introspect, reason about its own mental states, and behave indistinguishably from a conscious being – yet still lack any inner lights on. This skeptical view is epitomized by certain philosophical arguments (e.g., Searle's Chinese Room or the notion of philosophical zombies) and has been voiced in contemporary analyses that conclude robots are not – and perhaps cannot be – conscious in the phenomenal sense.[84] Thus, the phenomenal versus access distinction serves as a reminder that behavioral or functional equivalence to humans is not incontrovertible evidence of genuine subjective awareness.

For the field of artificial consciousness, a pragmatic consensus is emerging: Focus on access consciousness as a target, because it is operationalizable and amenable to scientific inquiry.[82] By concentrating on the functional aspects – how information can be made globally available in a system and how the system can monitor and report its own states – researchers can make tangible progress (for example, designing architectures with a kind of working memory, a global workspace, or a self-model). Indeed, discussions of machine consciousness increasingly suggest that pursuing access consciousness is the most feasible path, given that it aligns with observable capabilities and avoids immediate entanglement in the mysteries of subjective qualia.[82] If one can build an AI that convincingly implements access consciousness, it would at least fulfill the functional requirements of consciousness, providing a testbed from which to speculate about or investigate any accompanying phenomenology. In contrast, trying to engineer phenomenal consciousness directly – without a functional scaffold – may be a dead end, as we currently lack any clear understanding of how to create or detect raw subjective feeling in an artificial substrate. Therefore, access consciousness is often treated as a proxy for consciousness in machines, with the hope that advancing this proxy will either eventually shed light on the emergence of phenomenal properties or, at the very least, produce machines that behave in all the ways a conscious entity would – which is tremendously valuable in its own right.

4.4. Global availability and self-monitoring: Cognitive neuroscience insights

Cognitive neuroscience offers more concrete guidance on how to implement access-like consciousness in machines, thanks to empirical studies of the human brain. One influential theory, the global neuronal workspace, posits that conscious perception in the brain corresponds to the global availability of information: Stimuli that enter consciousness are those whose neural representations are amplified and broadcast across multiple cortical networks, rather than remaining confined to local processing circuits.

In a landmark synthesis, Dehaene *et al.*[83] identify two essential dimensions of consciousness-inspired cognitive processing that could inform machine designs: (i) Global availability of information and (ii) self-monitoring (meta-cognition). The first dimension, global availability, essentially captures the idea of a broadcast architecture: At any time, the system selects certain information (e.g., a particular input or an intermediate result) and makes it broadly accessible to various sub-modules (planning, language, memory, etc.). This resembles Block[82] and Dehaene *et al.*'s[83] notion of access consciousness, as it ensures the selected content can influence diverse processes system-wide. The second dimension, self-monitoring, refers to the system's ability to reflect on its own internal states and processes—a form of meta-cognition or introspection.[83] In humans, this is akin to the brain maintaining a self-referential model ("knowing that it knows") and monitoring its own computations for errors or learning. Dehaene *et al.*[83(p1)] describe this self-monitoring as a "self-referential relationship in which the cognitive system is able to monitor its own processing and obtain information about itself."

Together, these two features (often labeled C1 for global access and C2 for self-monitoring in Dehaene's framework) delineate a roadmap for building machines that achieve a functional analog of consciousness. An AI system endowed with a global workspace (allowing information sharing across modules) and a self-model (allowing it to track and report on its own states) would satisfy many criteria of access consciousness—and even begin to approach the sort of reflective awareness humans exhibit.

Notably, these neuroscience-inspired features are already being tentatively explored in AI and robotic architectures. Some cognitive architectures in AI have implemented global-workspace-like blackboards, where multiple specialist modules can read and write information, mimicking the idea of global availability. Similarly, researchers are experimenting with forms of machine meta-cognition – for example, AI agents that can report their confidence or uncertainty about their decisions or robots that internally simulate and evaluate their own forthcoming actions. Such capabilities reflect a rudimentary self-monitoring capacity. For instance, the self-aware robot principles from Chatila *et al.*[79] inherently aim for a form of C2: The robot not only learns but also shows that it knows it has learned, which implies an internal representation of its knowledge state. Another example can be seen in robotics work on "inner speech," where a robot talks to itself to guide its own reasoning – an approach directly inspired by human self-monitoring and models of inner experience, as proposed by Chella *et al.*[64] The emerging consensus is that implementing global broadcasting and self-reflection is a promising strategy to bring machines closer to consciousness in the functional sense. These features can endow AI systems with greater coherence, flexibility, and transparency in their operations. Moreover, if a machine were ever to exhibit phenomenal consciousness, one expects it would first need these functional capacities as a substrate. In other words, global availability and self-monitoring might not guarantee that a machine feels conscious, but they are likely necessary conditions for any machine that could eventually lay claim to subjective awareness.

4.5. Limitations of present AI: The absence of genuine consciousness

Despite significant advances in AI, the prevailing scientific consensus holds that no present machine or AI system possesses consciousness in the full sense.[84-88] Today's AI, including advanced neural networks and social robots, operates firmly within the bounds of the weak AI paradigm. These systems excel at specific tasks and can even display adaptive or context-aware behavior, but there is no credible evidence that any of them possess a subjective point of view or true self-awareness. Even systems that incorporate elements of global availability or rudimentary self-monitoring implement these features in relatively narrow ways (for example, a program might monitor its performance on a task and adjust parameters, but this is far from the rich, self-reflective awareness characteristics of human consciousness). Phenomenal consciousness in machines remains, at present, a speculative topic rather than an observed reality. We cannot peer into a deep learning model and find a flicker of sentience; at best, we find complex statistical patterns and representations shaped by training data.

It is instructive to consider why present AI falls short of consciousness. One obvious limitation is the lack of an integrated self-model in most AI. Human consciousness involves a sense of self that is continuous in time, situated in a body, and emotionally colored—features that mainstream AI does not possess. Another limitation is the absence of unified, flexible memory and attention akin to what the brain employs. While deep learning networks have impressive pattern recognition, they typically lack an architecture that integrates disparate knowledge on the fly, as a global workspace would. In addition, AI systems today lack intrinsic motivation or genuine autonomy in the sense that conscious beings exhibit; they pursue goals defined by programmers or derived from training data, without an inner life of desires or will. Finally, the evaluation problem looms large: Even if an AI were conscious, how would we truly know? There is no agreed-upon test for machine consciousness, and simple behavioral criteria (like the Turing test) are inadequate, as they can be passed through clever simulation without real awareness. This epistemic gap leads us to assume the absence of consciousness until proven otherwise. As some scholars note, the absence of any observable indicator of consciousness in machines is taken as confirmation that present AIs simply are not conscious. This point is rarely debated within the AI community. Indeed, discussions of AI ethics often neglect the issue of consciousness entirely, focusing instead on intelligence and autonomy.[81] Hildt[81] points out that we ought to engage more with the topic of artificial consciousness – and, just as importantly, with the implications of its present absence. Acknowledging that our most advanced creations remain essentially mindless (in the phenomenal sense) is important to keeping expectations grounded and shaping how we treat these systems.

A significant phenomenon in this context is anthropomorphism – the human tendency to attribute human-like qualities, including consciousness, to machines. This is evident in the way people interact with social robots and virtual assistants. For example, humanoid robots with facial expressions or voice-based AIs with personality often elicit feelings of social presence; we may talk to them as if they understand or even feel. Such anthropomorphic projections can obscure the reality that, despite surface appearances, these systems lack inner experiences. Instances like the robot Sophia being granted citizenship, or users feeling emotional attachment to AI companions, illustrate how far our intuitions can outpace scientific understanding. Scholars caution that this gap between appearance and reality can be problematic. We risk misleading ourselves – or the public – about what AI is actually doing. As a safeguard, some ethicists argue that we should consistently remind ourselves that present robots are not conscious.[84,88] They are complex artifacts, not entities with feelings, and we should avoid pre-maturely conferring moral or legal status that is reserved for sentient beings.

4.6. Ethical and societal implications of artificial consciousness

Even though artificial consciousness remains unachieved, the very pursuit of it – and the public's tendency to ascribe minds to machines – raises important ethical questions. If we eventually create a machine that exhibits advanced self-awareness or other hallmarks of consciousness, how should we treat it? Conversely, how should we treat today's unconscious AI systems, given that people often respond to them as if they were alive? These issues are already the subject of considerable debate in technology ethics and law.

On one hand, some thinkers like Gunkel[85] have explored the notion of "robot rights": The idea that sufficiently advanced AI or robots might merit certain moral or legal protections. Intriguingly, arguments for robot rights have been made even in the absence of robot consciousness. For example, based on the way humans empathize with humanoid machines or on the societal value of fostering empathy, a case is made for treating robots with a degree of care (much as we do animals or even human-looking dolls).[86] Darling[86] has argued that because humans can form emotional bonds with social robots, it aligns with our social and ethical values to extend some protections to these robots. This is analogous to how cruelty to animals is discouraged, not necessarily because animals possess human-level consciousness, but because such cruelty can degrade our moral character as agents. Proposed protections might include discouraging the wanton destruction of robots or violent behavior toward them, recognizing that such actions can engender harmful attitudes in society. The underlying rationale is partly anthropomorphic empathy – we dislike seeing even a robot "suffer" if it is lifelike – and partly pre-cautionary: If machines ever do become sentient, having established norms of respectful treatment could ease that transition.

On the other hand, many are wary of over-attributing consciousness and moral status to machines pre-maturely. As noted by Gabriel,[84] from a philosophical standpoint, there are strong arguments that robots cannot be conscious in the same way living organisms are, because consciousness might require qualities that only biological systems in environment contexts possess. If one accepts such arguments, then granting personhood or rights to machines would be a categorical error. Moreover, there is concern that focusing on the "feelings" of machines that do not actually feel could divert attention from ethical issues more grounded in reality, such as the welfare of humans impacted by AI or the responsibility for AI-driven decisions. Scheutz[87] has highlighted the potential emotional pitfalls in human-robot relationships, noting the unidirectional emotional bonds that can form. Humans might come to care deeply about robots that are not conscious and cannot reciprocate that care. This imbalance could lead to human distress (e.g., grief if a robot is shut down or malfunctions) or manipulation (e.g., exploiting human empathy for commercial or surveillance purposes). Scheutz warns that such one-sided attachments carry both psychological and social risks.

The ethical landscape is further complicated by the prospect (still hypothetical) of a truly conscious AI. If an AI ever claimed to have feelings or demonstrated behaviors strongly indicative of sentience, denying it moral consideration would be deeply problematic. Society would face a profound moral dilemma – long contemplated in science fiction – about whether and how to extend the community of conscious beings beyond our biological family.[85]

In light of these issues, the present consensus urges caution and clarity. It is important for scientists and communicators to convey that present-day AI does not possess consciousness,[84-88] even as we continue refining what that would entail. Such clarity helps prevent public misconceptions and ensures that ethical guidelines are grounded in the actual capabilities of present technologies. Simultaneously, it is prudent to start developing ethical frameworks that could accommodate conscious AI, should it emerge. These would include considerations of legal status, rights, responsibilities, and safeguards – to prevent abuse of such entities and to guard against deceitful mimicry of consciousness used to exploit users. In essence, the ethics of artificial consciousness straddle a line between present realities and future possibilities. We must manage the human tendency to anthropomorphize today's machines while remaining prepared for tomorrow's scenario where the line between simulation and reality of mind may begin to blur.

4.7. Emerging directions and future outlook

Looking ahead, the pursuit of artificial consciousness will likely advance on multiple fronts, informed by ongoing progress in neuroscience, cognitive science, and AI. One clear direction is the continued development of AI architectures that incorporate the principles of global availability and self-monitoring discussed above. Future AI systems may increasingly feature unified workspaces or attention mechanisms that allow information to flow more freely between components, coupled with meta-cognitive loops that enable the system to reason about and adjust its own operations. Such designs could be realized, for example, in more sophisticated cognitive architectures for robots or autonomous agents, where modules for perception, memory, decision-making, and language all feed into – and draw from – a common representational space (an echo of the global neuronal workspace). We may also see the integration of sensorimotor embodiment into these architectures. Since human consciousness is deeply embodied (the brain constantly integrates signals from the body and environment), giving robots richer bodily awareness and interoception might be a key to unlocking more advanced forms of self-awareness in machines. Early experiments in this vein, such as robots that simulate their own kinesthetic experiences or maintain internal homeostatic variables, hint at the importance of an embodied self-model for consciousness.

Another emerging direction is the exploration of learning-based approaches to self-awareness. Modern machine learning, especially deep learning, provides powerful tools for pattern recognition and function approximation. Researchers are beginning to ask whether these tools can be turned inward: Can a neural network learn to model its own cognition? One idea is to train networks that predict or interpret the hidden states of other networks (a form of meta-learning), effectively creating an internal observer module. If successful, this could result in an AI that possesses a form of introspective access to its internal representations – a step toward the machine knowing something of its own "mind." In addition, generative models that create narratives or explanations for the agent's behavior might serve as a rudimentary form of inner narrative (a component some theories consider important for consciousness). For instance, a future AI might be able to generate a verbal report like "I chose action X because I noticed Y, and that made me uncertain" – a capability that blurs the line between simple programmed response and genuine self-reflection.

Interdisciplinary research will be essential in guiding these efforts. Cognitive neuroscience will continue to identify the neural signatures and mechanisms associated with consciousness in the brain (e.g., specific brain rhythms, network dynamics, or anatomical circuits critical for awareness). These findings can inform computational models: If certain patterns of network connectivity or dynamics are necessary for consciousness in biological systems, mimicking those *in silico* could be a step in the right direction. For example, if research confirms that recurrent looping between frontal and sensory cortices is crucial for sustained conscious perception, AI architects might incorporate similar feedback loops in neural network designs for vision or language. Similarly, philosophical analysis remains crucial to clarify concepts and highlight potential pitfalls. Ongoing debates, such as whether consciousness requires a particular substrate (biological neurons vs. silicon) or whether it might be an emergent property of any sufficiently complex information system, will shape how we interpret advanced AI in the future. Some philosophers argue we might need entirely new paradigms (for instance, panpsychism or illusionism) to make sense of consciousness, which could radically affect how we attempt to implement or recognize it in machines.

In terms of practical milestones, a near-term goal is likely to be to develop empirical tests or benchmarks for consciousness-like attributes in AI. These would not claim to detect subjective experience directly (which may be impossible) but rather assess abilities associated with consciousness. For example, tests could evaluate an AI's degree of self-awareness, its flexibility in adapting global knowledge to novel problems, or its capacity for reporting on internal states. One proposed avenue is a sort of "AI consciousness spectrum" – a set of cognitive competencies (e.g., theory of mind, understanding of self versus others, temporal awareness of self) that could be measured. An AI that scores highly across many of these dimensions could be considered to have a higher degree of "AI consciousness" (in the access sense). Such frameworks would help move the discussion from abstract possibility to concrete progress: Researchers could then compete or collaborate on advancing AI along this spectrum, much as they do with benchmarks for intelligence.

Finally, ethical foresight must evolve in tandem with technical progress. As we inch closer to machines with human-like capabilities, even if still not conscious, we must continuously revisit our policies and perceptions. If an AI claims to be conscious or behaves in a way that is indistinguishable from a conscious agent, at what point do we err on the side of caution and consider granting it moral consideration? Some have suggested adopting a principle of "reasonable doubt": If we cannot be certain that a machine is not conscious, we should treat it gently – just in case. While we are not yet at that point, these discussions must begin now, so that society is not caught unprepared by the eventual emergence of machines with mind-like attributes.[85] Conversely, we also need to manage public expectations and prevent misconceptions. For example, consumers might assume a clever chatbot is a sentient companion when it is not, potentially leading to confusion or emotional harm. Clear communication about the capabilities and limitations of AI consciousness will thus remain the responsibility of experts in the field.

The present state of research suggests that artificial consciousness, in the rich sense of the term, is still more of a theoretical construct than a realized technology. Contemporary AI aligns with weak AI: Extraordinarily capable in narrow domains, but devoid of inner experience. However, the field is steadily laying the groundwork that may 1 day support at least the functional attributes of consciousness. By drawing on neuroscience to inform AI design (e.g., global workspaces and self-monitoring loops) and by deepening our theoretical understanding of consciousness (e.g., access vs. phenomenal, functional correlates of experience), we are inching toward the longstanding goal of a conscious machine. Whether that machine will feel anything, or whether we would recognize its feelings if it did, remains uncertain. What is clear is that this line of inquiry will continue to challenge our scientific ingenuity and our philosophical openness. The coming years will likely bring machines that blur the line between programmed behavior and adaptive, self-directed cognition even further. How we choose to interpret and interact with those machines will be a test of our wisdom, calling for a balanced approach that is at once scientifically rigorous, philosophically informed, and ethically attuned to both the possibilities and the limits of machine consciousness.

Each step forward forces us to refine our understanding of our own minds, as much as that of machines, reinforcing the notion that the quest for artificial consciousness is as much a mirror for humanity as it is a window into the future of technology. By rigorously exploring both the capabilities and the limitations of our creations – while keeping concepts like access and phenomenal consciousness in clear view – we can guide advancements in a responsible manner.[81,82] Ultimately, the effort to build or identify consciousness in an artificial entity will deepen our grasp of the nature of consciousness itself, and in doing so, it will bridge disciplines in unprecedented ways. The discussion presented here – synthesizing perspectives from cognitive neuroscience, computational modeling, and machine learning – underscores that achieving artificial consciousness is not simply an engineering challenge, but an interdisciplinary grand question – one that will likely occupy philosophers, scientists, and engineers for decades to come.

5. Conclusion

The question of whether machines can possess consciousness remains a central debate in AI. Strong AI envisions machines capable of genuine cognitive states and understanding, while weak AI suggests they only simulate thought processes. The creation of artificial consciousness represents a profound and unresolved challenge in AI research. Progress in understanding the mechanisms underlying human consciousness is essential for evaluating the feasibility of replicating these processes in machines. Although present AI systems lack true consciousness, advancements in neuroscience and machine learning offer promising avenues for further exploration in this interdisciplinary domain.

Acknowledgments

None.

Funding

None.

Conflict of interest

The authors declare they have no competing interests.

Author contributions

Conceptualization: Seyed-Ali Sadegh-Zadeh
Writing – original draft: Seyed-Ali Sadegh-Zadeh
Writing – review & editing: Mahboobe Bahrami

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data

Not applicable.

References

1. Crick F, Koch C. Consciousness and neuroscience. *Cereb Cortex*. 1998;8(2):97-107.

   doi: 10.1093/cercor/8.2.97

2. Ornstein RE. *The Psychology of Consciousness*. New York: Viking Press; 1972.

3. Baluška F, Reber A. Sentience and consciousness in single cells: How the first minds emerged in unicellular species. *Bioessays*. 2019;41(3):e1800229.

   doi: 10.1002/bies.201800229

4. Jourahmad Z, Habibabadi JM, Moein H, *et al*. *Machine Learning Techniques for Predicting the Short-Term Outcome of Resective Surgery in Lesional-Drug Resistance Epilepsy*. [arXiv Preprint]; 2023.

5. Velmans M. An introduction to the science of consciousness. In: Velmans M, editor. *The Science of Consciousness*. England, UK: Routledge; 2003. p. 13-34.

6. Nazari MJ, Shalbafan M, Ghannadi F, *et al*. A machine learning approach for differentiating bipolar disorder type II and borderline personality disorder using electroencephalography and cognitive abnormalities. *PLoS One*. 2024;19(6):e0303699.

   doi: 10.1371/journal.pone.0303699

7. Koch C. *The Quest for Consciousness*. Englewood, CO: Roberts and Company; 2004.

8. Sadegh-Zadeh SA, Bahrami M, Najafi A, Asgari-Ahi M, Campion R, Hajiyavand AM. Evaluation of COVID-19 pandemic on components of social and mental health using machine learning, analysing United States data in 2020. *Front Psychiatry*. 2022;13:933439.

   doi: 10.3389/fpsyt.2022.933439

9. Song C. Brain structural complexity and consciousness. *Philos Mind Sci*. 2021;2:6.

   doi: 10.33735/phimisci.2021.1020

10. Piolino P, Desgranges B, Eustache F. Episodic autobiographical memories over the course of time: Cognitive, neuropsychological and neuroimaging findings. *Neuropsychologia*. 2009;47(11):2314-2329.

    doi: 10.1016/j.neuropsychologia.2009.01.020

11. Piolino P, Desgranges B, Clarys D, *et al*. Autobiographical memory, autonoetic consciousness, and self-perspective in aging. *Psychol Aging*. 2006;21(3):510-525.

    doi: 10.1037/0882-7974.21.3.510

12. Mace JH, Hidalgo AM. Semantic-to-autobiographical memory priming affects involuntary autobiographical

memory production after a long delay. *Conscious Cogn*. 2022;104:103385.

doi: 10.1016/j.concog.2022.103385

13. Tyler C. The emergent dualism view of quantum physics and consciousness. *Cosmos Hist J Nat Soc Philos*. 2015;11(2):97-114.

14. Foulkes D, Fleisher S. Mental activity in relaxed wakefulness. *J Abnorm Psychol*. 1975;84(1):66-75.

doi: 10.1037/h0076162

15. Delacour J. An introduction to the biology of consciousness. *Neuropsychologia*. 1995;33(9):1061-1074.

doi: 10.1016/0028-3932(95)00033-Y

16. Nieminen JO, Gosseries O, Massimini M, *et al*. Consciousness and cortical responsiveness: A within-state study during non-rapid eye movement sleep. *Sci Rep*. 2016;6:30932.

doi: 10.1038/srep30932

17. Sadegh-Zadeh SA, Soleimani Mamalo A, Kavianpour K, *et al*. Artificial intelligence approaches for tinnitus diagnosis: Leveraging high-frequency audiometry data for enhanced clinical predictions. *Front Artif Intell*. 2024;7:1381455.

doi: 10.3389/frai.2024.1381455

18. Nir Y, Massimini M, Boly M, Tononi G. Sleep and consciousness. In: Laureys S, Tononi G, editors. *Neuroimaging of Consciousness*. Germany: Springer; 2013. p. 133-182.

doi: 10.1007/978-3-642-37579-6_7

19. Meijer DKF, Geesink HJH. Consciousness in the universe is scale invariant and implies an event horizon of the human brain. *Neuroquantology*. 2017;15(3):41-79.

doi: 10.14704/nq.2017.15.3.1079

20. Pereira A Jr. Triple-aspect monism: A conceptual framework for the science of human consciousness. In: Pereira A, Lehmann D, editors. *The Unity of Mind, Brain and World*. Cambridge: Cambridge University Press; 2013. p. 299-321.

doi: 10.1017/CBO9781139342134.015

21. Sadegh-Zadeh SA, Bagheri M. Harnessing the power of clinical data in dentistry: Importance and guidelines for dentists in AI modelling for enhanced patient care. *Open J Clin Med Images*. 2024;4(1):1188.

doi: 10.52768/2833-2725/1188

22. Jin SH, Chung CK. Messages from the brain connectivity regarding neural correlates of consciousness. *Exp Neurobiol*. 2012;21(3):113-119.

doi: 10.5607/en.2012.21.3.113

23. Oyebode F. *Sims' Symptoms in the Mind: Textbook of Descriptive Psychopathology*. 6th ed. Amsterdam: Elsevier Health Sciences; 2018.

24. Sadegh-Zadeh SA, Kambhampati C. All-or-none principle and weakness of Hodgkin-Huxley mathematical model. *Int J Math Comput Sci*. 2017;11:453-460.

25. Sadegh-Zadeh SA, Kambhampati C. Computational investigation of amyloid peptide channels in Alzheimer's disease. *J (Basel)*. 2018;2(1):1-14.

doi: 10.3390/j2010001

26. Schiff ND, Plum F. The role of arousal and "gating" systems in the neurology of impaired consciousness. *J Clin Neurophysiol*. 2000;17(5):438-452.

doi: 10.1097/00004691-200009000-00010

27. Margetis K, Korfias SI, Gatzonis S, *et al*. Intrathecal baclofen associated with improvement of consciousness disorders in spasticity patients. *Neuromodulation*. 2014;17(7):699-704.

doi: 10.1111/ner.12219

28. Mhuircheartaigh RN, Rosenorn-Lanng D, Wise R, Jbabdi S, Rogers R, Tracey I. Cortical and subcortical connectivity changes during decreasing levels of consciousness in humans: A functional magnetic resonance imaging study using propofol. *J Neurosci*. 2010;30(27):9095-9102.

doi: 10.1523/JNEUROSCI.5593-09.2010

29. Lane RD, Smith R. Levels of emotional awareness: Theory and measurement of a socio-emotional skill. *J Intell*. 2021;9(3):42.

doi: 10.3390/jintelligence9030042

30. Blumenfeld H. Brain mechanisms of conscious awareness: Detect, pulse, switch, and wave. *Neuroscientist*. 2021;29(5):9-18.

doi: 10.1177/10738584211049378

31. Menon V. Large-scale brain networks and psychopathology: A unifying triple network model. *Trends Cogn Sci*. 2011;15(10):483-506.

doi: 10.1016/j.tics.2011.08.003

32. Cavanna AE. The precuneus and consciousness. *CNS Spectr*. 2007;12(7):545-552.

doi: 10.1017/S1092852900015337

33. Gil R, Arroyo-Anlló EM, Ingrand P, *et al*. Self-consciousness and Alzheimer's disease. Acta Neurol Scand. 2001;104(5):296-300.

doi: 10.1034/j.1600-0404.2001.00334.x

34. Pins D, Ffytche D. The neural correlates of conscious vision. *Cereb Cortex*. 2003;13(5):461-474.

doi: 10.1093/cercor/13.5.461

35. Passingham RE, Bengtsson SL, Lau HC. Medial frontal cortex: From self-generated action to reflection on one's own

performance. *Trends Cogn Sci*. 2010;14(1):16-21.

doi: 10.1016/j.tics.2009.11.001

36. Heilman KM, Nadeau SE, Beversdorf DO. Creative innovation: Possible brain mechanisms. *Neurocase*. 2003;9(5):369-379.

    doi: 10.1076/neur.9.5.369.16553

37. Laureys S, Gosseries O, Tononi G, editors. *The Neurology of Consciousness: Cognitive Neuroscience and Neuropathology*. 2nd ed. United States: Academic Press; 2015.

38. Sadegh-Zadeh SA, Bagheri M, Saadat M. Decoding children dental health risks: A machine learning approach to identifying key influencing factors. *Front Artif Intell*. 2024;7:1392597.

    doi: 10.3389/frai.2024.1392597

39. Chow HM, Horovitz SG, Carr WS, *et al*. Rhythmic alternating patterns of brain activity distinguish rapid eye movement sleep from other states of consciousness. *Proc Natl Acad Sci U S A*. 2013;110(25):10300-10305.

    doi: 10.1073/pnas.1217691110

40. Blows WT. *The Biological Basis of Mental Health*. 4th ed. England, UK: Routledge; 2021.

41. Seth AK, Bayne T. Theories of consciousness. *Nat Rev Neurosci*. 2022;23(7):439-452.

    doi: 10.1038/s41583-022-00580-4

42. Baars BJ. *A Cognitive Theory of Consciousness*. Cambridge: Cambridge University Press; 1993.

43. Sadegh-Zadeh SA, Nazari MJ, Aljamaeen M, Yazdani FS, Mousavi SY, Vahabi Z. Predictive models for Alzheimer's disease diagnosis and MCI identification: The use of cognitive scores and artificial intelligence algorithms. *NPG Neurol Psychiatr Geriatr*. 2024;24:194-211.

    doi: 10.1016/j.npg.2024.02.002

44. Rosenthal DM. Consciousness and its function. *Neuropsychologia*. 2008;46(3):829-840.

    doi: 10.1016/j.neuropsychologia.2007.11.012

45. Aleksander I. Artificial neuroconsciousness: An update. In: Mira J, Sánchez-Andrés JV, editors. *International Workshop on Artificial Neural Networks*. Germany: Springer; 1995. p. 566-583.

    doi: 10.1007/BFb0020193

46. Franklin S, Baars BJ, Ramamurthy U, Ventura M. *The Role of Consciousness in Memory*; 2005. Available from: https://ccrg.cs.memphis.edu/assets/papers/wrcm-bmm.pdf [Last accessed on 2024 Oct 29].

47. Tulving E. Memory and consciousness. *Can Psychol*. 1985;26(1):1-12.

    doi: 10.1037/h0080017

48. Kouider S, De Gardelle V, Sackur J, Dupoux E. How rich is consciousness? The partial awareness hypothesis. *Trends Cogn Sci*. 2010;14(7):301-307.

    doi: 10.1016/j.tics.2010.04.006

49. Reggia JA. The rise of machine consciousness: Studying consciousness with computational models. *Neural Netw*. 2013;44:112-131.

    doi: 10.1016/j.neunet.2013.02.010

50. Manzotti R. From artificial intelligence to artificial consciousness. In: Chella A, Manzotti R, editors. *Artificial Consciousness*. Exeter: Imprint Academic; 2007. p. 174-190.

51. Sadegh-Zadeh SA, Dastmard A, Kafshgarkolaei LM, *et al*. Machine learning modelling for compressive strength prediction of superplasticizer-based concrete. *Infrastructures*. 2023;8(2):21.

    doi: 10.3390/infrastructures8020021

52. Cleeremans A, Jiménez L. Implicit learning and consciousness: A graded, dynamic perspective. In: Stadler M, Frensch P, editors. *Implicit Learning and Consciousness*. United Kingdom: Psychology Press; 2002. p. 1-40.

53. Chrisley R. Philosophical foundations of artificial consciousness. *Artif Intell Med*. 2008;44(2):119-137.

    doi: 10.1016/j.artmed.2008.04.003

54. Vaneechoutte M. Experience, awareness and consciousness: Suggestions for definitions as offered by an evolutionary approach. *Found Sci*. 2000;5(4):429-456.

    doi: 10.1023/A:1011371421112

55. Masakowski YR. Artificial intelligence and the future global security environment. In: Boulanin V, editor. *Artificial Intelligence and Global Security*. United Kingdom: Emerald Publishing Limited; 2020.

56. Wiedermann J, Van Leeuwen J. Finite state machines with feedback: An architecture supporting minimal machine consciousness. In: Cooper SB, Dawar A, Löwe B, editors. *Conference on Computability in Europe*. Germany: Springer; 2019. p. 286-297.

    doi: 10.1007/978-3-030-29750-3_29

57. Reggia JA, Katz GE, Davis GP. Artificial conscious intelligence. *J Artif Intell Conscious*. 2020;7(1):95-107.

    doi: 10.1142/S2705078520500059

58. Iantovics LB, Gligor A, Niazi MA, Biro AI, Szilagyi SM, Tokody D. Review of recent trends in measuring the computing systems intelligence. *Brain (Bacau)*. 2018;9(2):77-94.

    doi: 10.18662/brain/15

59. Wang J, Wang C, Cai P, *et al*. Artificial sense technology:

Emulating and extending biological senses. *ACS Nano.* 2021;15(12):18671-18678.

doi: 10.1021/acsnano.1c07293

60. Kohli V, Tripathi U, Chamola V, Rout BK, Kanhere SS. A review on virtual reality and augmented reality use-cases of brain-computer interface based applications for smart cities. *Microprocess Microsyst*. 2022;88:104392.

doi: 10.1016/j.micpro.2022.104392

61. Henderson D, Horgan T, Potrč M, Strahovnik V. Chromatic illumination: Conscious intentionality without conscious representation. *ProtoSociology*. 2021;38:35-58.

doi: 10.5840/protosociology2021383

62. Raikov A. *Cognitive Semantics of Artificial Intelligence: A New Perspective*. Germany: Springer; 2021.

doi: 10.1007/978-3-030-73164-9

63. Searle JR. Minds, brains, and programs. In: *The Oxford Handbook of Philosophy of Cognitive Science*. Oxford: Oxford University Press; 2019.

64. Chella A, Pipitone A, Morin A, Racy F. Developing self-awareness in robots via inner speech. *Front Robot AI*. 2020;7:16.

doi: 10.3389/frobt.2020.00016

65. Jamieson GA. *Psychology of Consciousness: Theory, Research, and Practice*. United States: American Psychological Association; 2021.

66. Haikonen POA. On artificial intelligence and consciousness. *J Artif Intell Conscious*. 2020;7(1):73-82.

doi: 10.1142/S2705078520500035

67. Kulikov SB, Shirokova AV. Artificial intelligence, culture and education. *AI Soc*. 2021;36(1):305-318.

doi: 10.1007/s00146-020-01018-9

68. Wei L. Legal risk and criminal imputation of weak artificial intelligence. In: *IOP Conference Series: Materials Science and Engineering*. England: IOP Publishing; 2019. p. 062085.

doi: 10.1088/1757-899X/677/6/062085

69. Velik R. AI reloaded: Objectives, potentials, and challenges of the novel field of brain-like artificial intelligence. *Brain Broad Res Artif Intell Neurosci*. 2012;3(3):25-54.

doi: 10.18662/brain/3.3/20

70. Dong Y, Hou J, Zhang N, Zhang M. Research on how human intelligence, consciousness, and cognitive computing affect the development of artificial intelligence. *Complexity*. 2020;2020:1-10.

doi: 10.1155/2020/1681031

71. Salameh A. *Artificial Intelligence as a Commons-Opportunities and Challenges for Society*. Berlin: Springer; 2017.

72. Lieto A. *Cognitive Design for Artificial Minds*. England, UK: Routledge; 2021.

73. Heuveline V, Stiefel V. Artificial intelligence and algorithms: True progress or just digital alchemy? In: Jäger G, editor. *Intelligence-Theories and Applications*. Cham: Springer; 2022. p. 219-227.

doi: 10.1007/978-3-030-91177-5_15

74. Ng GW, Leung WC. Strong artificial intelligence and consciousness. *J Artif Intell Conscious*. 2020;7(1):63-72.

doi: 10.1142/S2705078520500023

75. Dubovikov K. *Managing Data Science: Effective Strategies to Manage Data Science Projects and Build a Sustainable Team*. United Kingdom: Packt Publishing Ltd.; 2019.

76. Searle JR. Minds, brains, and programs. *Behav Brain Sci*. 1980;3(3):417-424.

doi: 10.1017/S0140525X00005756

77. Ray A. *Compassionate Artificial Intelligence: Frameworks and Algorithms. Compassionate AI Lab*. United States: (An Imprint of Inner Light Publishers); 2018.

78. Levy D. The ethical treatment of artificially conscious robots. *Int J Soc Robot*. 2009;1(3):209-216.

doi: 10.1007/s12369-009-0022-6

79. Chatila R, Renaudo E, Andries M, *et al*. Toward self-aware robots. *Front Robot AI*. 2018;5:88.

doi: 10.3389/frobt.2018.00088

80. Kinouchi Y, Mackin KJ. A basic architecture of an autonomous adaptive system with conscious-like function for a humanoid robot. *Front Robot AI*. 2018;5:30.

doi: 10.3389/frobt.2018.00030

81. Hildt E. Artificial intelligence: Does consciousness matter? *Front Psychol*. 2019;10:1535.

doi: 10.3389/fpsyg.2019.01535

82. Block N. On a confusion about a function of consciousness. *Behav Brain Sci*. 1995;18(2):227-247.

doi: 10.1017/S0140525X0003842X

83. Dehaene S, Lau H, Kouider S. What is consciousness, and could machines have it? In: Dignum V, editor. *Robotics, AI, and Humanity*. Berlin: Springer; 2021. p. 43-56.

doi: 10.1007/978-3-030-51110-4_4

84. Gabriel M. Could a robot be conscious? Some lessons from philosophy. In: Dignum V, editor. *Robotics, AI, and Humanity*. Berlin: Springer; 2021. p. 57-68.

doi: 10.1007/978-3-030-51110-4_5

85. Gunkel DJ. *Robot Rights*. United States: MIT Press; 2018.

86. Darling K. Extending legal protection to social robots: The

effects of anthropomorphism, empathy, and violent behavior towards robotic objects. In: Calo R, Froomkin AM, Kerr I, editors. *Robot Law*. United Kingdom: Edward Elgar Publishing; 2016.

87. Scheutz M. The inherent dangers of unidirectional emotional bonds between humans and social robots. In: Lin P, Abney K, Bekey GA, editors. *Robot Ethics: The Ethical and Social Implications of Robotics*. United States: MIT Press; 2011. p. 205-221.

88. Sadegh-Zadeh SA. *Computational Methods Toward Early Detection of Neuronal Deterioration [dissertation]*. United Kingdom: University of Hull; 2019.

ACCSCIENCE PUBLISHING

Artificial Intelligence in Health

PERSPECTIVE ARTICLE

The role of Omnichain in advancing federated learning for artificial intelligence training in healthcare

Dongfang Wu[1,2]*, and **Yichen Wang**[1,2]

[1]Global Health Research Center, Duke Kunshan University, Kunshan, Jiangsu, China

[2]Global Health Institute, Duke University, Durham, North Carolina, United States of America

Abstract

Health data serves as a crucial foundation for artificial intelligence (AI) training in the healthcare sector. The pivotal procedure for acquiring numerous and effective health data lies in incentivizing participants to contribute their health data while adhering to privacy regulations like the General Data Protection Regulation. Federated learning achieves privacy protection by transmitting only parameters rather than data to the model. When integrated with blockchain smart contracts, this approach facilitates the automation of incentives according to health data quality, thereby mitigating human's subjective intervention. Consequently, the synergy of these two methodologies offers new promise for the training of AI models in healthcare. However, this advantage encounters performance degradation due to the heterogeneity among diverse blockchains. This article posits the concept of Omnichain as a potential solution to this challenge by analyzing its operational mechanisms and future developmental trajectories and providing potential perspectives for its combination with hybrid federal learning solutions such as differential privacy and secure multi-party computation to promote its application in the sphere of AI in healthcare.

Keywords: Omnichain; Federated learning; Artificial intelligence training; Healthcare; Training performance

***Corresponding author:**
Dongfang Wu
(dongfang.wu@duke.edu)

Citation: Wu D, Wang Y. The role of Omnichain in advancing federated learning for artificial intelligence training in healthcare. *Artif Intell Health*. 2025;2(3):39-43.
doi: 10.36922/aih.5753

Received: November 4, 2024

1st revised: December 23, 2024

2nd revised: January 24, 2025

Accepted: February 25, 2025

Published online: March 7, 2025

Copyright: © 2025 Author(s). This is an Open-Access article distributed under the terms of the Creative Commons Attribution License, permitting distribution, and reproduction in any medium, provided the original work is properly cited.

Publisher's Note: AccScience Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

Health data is a vital asset in healthcare and thoroughly exploiting it for artificial intelligence (AI) training yields substantial value in diagnosis, medication, and patient care. Nonetheless, two concurrent challenges deserve consideration: (i) how to ensure adherence of health data collection to privacy regulatory frameworks such as General Data Protection Regulation;[1] and (ii) how to incentivize individuals or institutions to share their health data, with the aim to ensure continuity and reliability of AI training quality.[2]

Federated learning, an emergent paradigm in machine learning for AI, allows algorithms to be trained across multiple distributed devices or servers with local data samples. In contrast to traditional machine learning, participants in federated learning do not exchange local data. Rather, they transmit the parameters of their local training results to a central server, thereby achieving a collaborative AI model training objective.[3]

However, the concern within federated learning is incentivizing entities to contribute valid health data while penalizing malicious actors. The latency and subjectivity inherent in manual evaluations do not provide an optimal solution. The immutable and self-executing nature of smart contracts within blockchain addresses this issue by enabling all participants to ascertain results instantly without the need for intermediaries, thus facilitating a reliable reward distribution mechanism for federated learning to create an automated and standardized process.[4]

Consequently, these two advantages position federated learning and blockchain as inherently complementary. Each participant is assigned a blockchain identity, known as the public address, wherein health data is stored, manifesting in a distributed manner. Participants declare their public address to engage in the federated learning process and, thereby, receive rewards according to smart contract criteria. Thus, in the healthcare sector, where pertinent ethical and sensitive considerations regarding health data are paramount, implementing real-time federated learning for health information privacy protection and utilizing blockchain to develop more intelligent incentive strategies holds significant promise.

2. Omnichain paradigm in federated learning

This innovative amalgamation of federated learning and blockchain effectively addresses the privacy concerns surrounding health data during the training process while significantly enhancing the motivation of participants to contribute health data. However, AI practitioners have encountered challenges with this integration due to the heterogeneity among different blockchains. For instance, when a federated learning framework that was originally developed in Solidity language and deployed on Ethereum mainnet seeks to migrate to Solana, it may be reimplemented in Rust language. Meanwhile, implementing federated learning on diverse blockchains necessitates reconfiguring environments to accommodate various consensus mechanisms. The aforementioned factors result in performance degradation for federated learning.[5] This situation underscores the need for a unified blockchain environment that ensures a consistent execution standard for conducting federated learning on health data across different blockchains.

The emergence of the concept of the Omnichain addresses this critical issue. Omnichain constructs a novel foundational layer, known as Layer-0, that interconnects all blockchains, regardless of their smart contract technologies, thereby allowing all federated learning processes to operate atop this infrastructure.[6] This represents a highly compatible super multi-chain ecosystem that mitigates the limitations of individual blockchains, ultimately serving the needs of AI training. During the construction of the Omnichain, Cosmos SDK is a pivotal technology whose standardized development toolkit enables seamless communication and transactions among unique parallel blockchains. With this interoperability, the Cosmos SDK opens up a world of possibilities, allowing data, tokens, assets, and logic to be transmitted across multiple blockchains in a highly secure and trustworthy manner.[7]

As illustrated in Figure 1, Omnichain supported by the Cosmos SDK facilitates communication among disparate blockchains. Omnichain connects various blockchains and allows users to deploy smart contracts directly onto it by unified smart contracts tailored for diverse blockchains on the Omnichain. Furthermore, the federated learning process is capable of directly interfacing with the Omnichain and executes these contracts to establish incentive mechanisms. Simultaneously, smart contracts can be distributed across multiple blockchains without needing to consider the barriers arising from differing consensus mechanisms or programming languages, eventually mitigating performance degradation arising from blockchain heterogeneity while ensuring the integrity of privacy and incentive mechanisms, which ultimately aids in the training of AI models.

Notably, this process does not impede the independent generation of blocks by individual blockchains, indicating that the construction of Layer-0 does not necessitate the implementation of Layer-1 through a forking mechanism. This advancement effectively dismantles barriers between different blockchains, resulting in a qualitative leap in both the fluidity and functionality of information exchange. For example, once a federated learning model completes a round of training updates on Ethereum, the trained model can be transmitted to Solana through Omnichain to continue training. This approach eliminates the need for reliance on a single-point oracle or centralized bridge and likewise reduces the opportunities for attackers to exploit cross-chain bridge vulnerabilities.

Moving a further step, Omnichain can also be integrated into existing hybrid solutions. Differential privacy (DP) is a strategy aimed at mitigating the risks of side-channel attacks or differential analyses of parameter updates in federated learning. Injecting mathematically quantifiable noise into the parameter reporting and aggregation processes eventually increases the difficulty

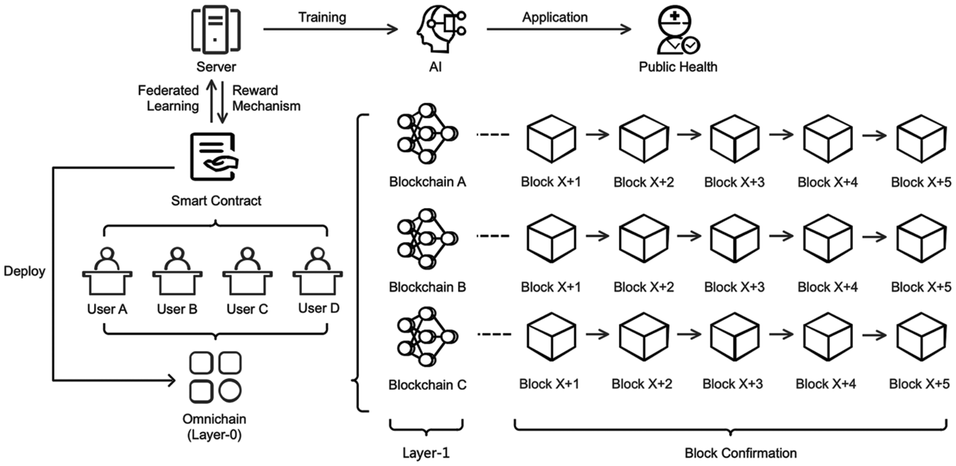


Figure 1. Omnichain applied in federated learning
Source: Image created by authors.

of inferring individual data points.[8] However, when data sources are scattered across different public chains, Layer-2 sidechains, or specific consortium blockchains, ensuring privacy in distributed cross-chain training requires a DP mechanism capable of seamlessly adapting to multi-chain data transmission and sharing. The interoperability of the Omnichain protocol relieves a federated learning platform from being confined to a single-chain dataset or a single network. Within a unified ledger, developers can incorporate DP parameter-tuning strategies into cross-chain data exchange interfaces. Then, the DP-processed parameters can be securely transmitted to a central aggregation point, which may be a federated learning service contract on a main chain or at a Rollup layer.

When federated learning faces even stricter privacy and compliance requirements, especially in healthcare, DP alone may still fail to meet regulatory standards. In such cases, the introduction of Secure Multi-Party Computation (SMC) offers stronger security guarantees for federated training.[9] Through SMC, participants can operate on encrypted local parameters or intermediate computation results without exposing data in plain text, ensuring that no single entity can reconstruct the original data of another party. To implement SMC protocols within an Omnichain framework, the corresponding execution logic must be extended. The Omnichain protocol, leveraging decentralized cross-chain verification and light client proofs, eliminates the need for trusted intermediaries in inter-chain information exchange. Building on this foundation, SMC protocols can be realized by deploying mutually trustworthy computation contracts on various chains and employing full-chain consensus to verify the integrity of relayed data, thus enabling trustless transmission and secure computation of encrypted parameters. With Omnichain, SMC solutions can achieve parallel, sharded secure computation across multiple chains. For example, a high Transactions Per Second execution chain could handle distributed key generation and secret sharing; another privacy-focused chain might perform encrypted gradient aggregation and decryption threshold checks; and the main chain or a Rollup layer could conduct final model verification and record-keeping. Because Omnichain interoperability allows each computation step to be executed on the chain best suited to the task, the combination of federated learning, SMC, and Omnichain can demonstrate new potential in terms of performance and scalability.

3. Discussion

The integration of Omnichain with federated learning addresses the performance degradation issues arising from blockchain heterogeneity in training AI models within the public health sector. The potential use case lies in overcoming the phenomenon of chain isolation imposed by federated learning frameworks confined to a single blockchain. Consider a healthcare federated learning project training privacy-processed patient medical record metadata on Ethereum's Layer-2 solution Polygon. At the same time, the project seeks to incorporate Internet of Things device training data collected through a Solana blockchain, which offers robust Decentralized Physical Infrastructure Network compatibility, and it also wishes to utilize hashed research and evidentiary records from a specific industry consortium chain that stores pharmaceutical R&D data. Under a traditional single-chain paradigm, these three requirements would remain isolated, forming data silos. By leveraging full-chain interoperability protocols, however, the federated learning model can seamlessly access and aggregate these diverse data sources – while respecting privacy and permission constraints – thus enhancing the model's generalization capabilities and training quality. In parallel, the concept of a taxonomy driven by practical services can be applied here. By approaching the discussion from the perspective of concrete services, we can deduce the unique advantages that Omnichain offers, thereby realizing synergy between a fully integrated blockchain framework and federated learning across diverse domains, such as health data circulation and management, privacy and security, token-based incentive mechanisms, collaborative orchestration, and regulatory oversight. This approach ultimately advances the real-world adoption of the hybrid solution. For example, in the realm of health data circulation and management, indexing services built on Omnichain can standardize data drawn from multiple blockchains, thereby streamlining cross-chain data flow and providing high-quality datasets for federated learning.

However, Omnichain remains undeveloped. Before its development, cross-chain bridges represented a bold yet flawed endeavor. Users were required to lock assets on the source chain and incur gas fees to receive corresponding wrapped assets on the target chain, thereby creating liquidity challenges.[10] In addition, cross-chain bridges predominantly focused on value transfer rather than imperative information transfer, which is essential for training AI. In contrast, Omnichain emerges as a novel paradigm extending from cross-chain bridges, utilizing universal smart contracts to create an infrastructure that spans multiple chains and can be directly deployed across various blockchains, including Ethereum Virtual Machine – compatible and other mainstream platforms, thereby mitigating the silo effect of disparate blockchains. Nevertheless, the high-throughput requirements of federated learning in the training process remain in tension with the existing gas pricing mechanisms of Omnichain, and the block generation speed of Omnichain further constrains the deployment of federated learning solutions.[11] This necessitates that Omnichain, much like various Layer-2 solutions of Ethereum, progresses toward achieving low-cost and rapid transaction capabilities.

4. Conclusion

This article proposes the integration of Omnichain concept into the existing frameworks of federated learning and blockchain, with the objective of minimizing performance degradation while maintaining privacy protection during AI training in the healthcare sector. By deploying smart contracts tailored for diverse blockchains on the Omnichain, the federated learning framework may decrease redundant deployments, thus enabling support for multiple blockchains directly from the Omnichain. This design enhances the efficiency of the AI training paradigm that combines federated learning with blockchain in healthcare. However, it is important to note that the Omnichain is still in its nascent developmental stage, and its speed and cost remain significant constraints on its integration with federated learning. This presents a crucial area for future research and attention.

Acknowledgments

We are deeply grateful to the Duke University Writing Studio for tremendous assistance in polishing the academic language of our work.

Funding

None.

Conflict of interest

The authors declare that they have no competing interests.

Author contributions

Conceptualization: Dongfang Wu
Writing–original draft: Dongfang Wu
Writing–review & editing: Yichen Wang

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data

Not applicable.

References

1. Truong N, Sun K, Wang S, Guitton F, Guo Y. Privacy preservation in federated learning: An insightful survey from the GDPR perspective. *Comput Secur*. 2021;110:102402.

   doi: 10.1016/j.cose.2021.102402

2. Zhan Y, Li P, Qu Z, Zeng D, Guo S. A learning-based incentive mechanism for federated learning. *IEEE Internet Things J*. 2020;7:6360-6368.

   doi: 10.1109/JIOT.2020.2967772

3. Mammen PM. Federated learning: Opportunities and challenges. arXiv [Preprint]. 2021.

   doi: 10.48550/arXiv.2101.05428

4. Nguyen DC, Ding M, Pham QV, *et al*. Federated learning

meets blockchain in edge computing: Opportunities and challenges. *IEEE Internet Things J*. 2021;8(16):12806-12825.

doi: 10.1109/JIOT.2021.3072611

5. Zhu J, Cao J, Saxena D, Jiang S, Ferradi H. Blockchain-empowered federated learning: Challenges, solutions, and future directions. *ACM Comput Surv*. 2023;55(11):1-31.

   doi: 10.1145/3570953

6. Zheng J, Lee DKC, Qian D. An in-depth guide to cross-chain protocols under a multi-chain world. *World Sci Annu Rev Fintech*. 2023;1:2350003.

   doi: 10.1142/S2811004823500033

7. Belchior R, Vasconcelos A, Guerreiro S, Correia M. A survey on blockchain interoperability: Past, present, and future trends. *ACM Comput Surv*. 2021;54(8):1-41.

   doi: 10.1145/3471140

8. El Ouadrhiri A, Abdelhadi A. Differential privacy for deep and federated learning: A survey. *IEEE Access*. 2022;10:22359-22380.

   doi: 10.1109/ACCESS.2022.3151670

9. Acar A, Celik ZB, Aksu H, Uluagac AS, McDaniel P. Achieving Secure and Differentially Private Computations in Multiparty Settings. In: *2017 IEEE Symposium on Privacy-Aware Computing* (*PAC*). Washington, DC, USA: IEEE; 2017. p. 49-59.

   doi: 10.1109/PAC.2017.12

10. Harris CG. Cross-chain Technologies: Challenges and Opportunties for Blockchain Interoperability. In: *2023 IEEE International Conference on Omni-layer Intelligent Systems* (*COINS*). Berlin, Germany: IEEE; 2023. p. 1-6.

    doi: 10.1109/COINS57856.2023.10189298

11. Zarick R, Pellegrino B, Banister C. Layerzero: Trustless omnichain interoperability protocol. arXiv [Preprint]. 2021.

    doi: 10.48550/arXiv.2110.13871

ACCSCIENCE PUBLISHING

Artificial Intelligence in Health

ORIGINAL RESEARCH ARTICLE

Artificial intelligence within medical diagnostics: A multi-disease perspective

Zarif Bin Akhtar*

Department of Computing, Institute of Electrical and Electronics Engineers, Piscataway, United States of America

Abstract

Artificial intelligence (AI) has become a transformative technology in medical diagnostics, enabling enhanced analysis of complex clinical data and supporting precise, efficient decision-making across diverse disease areas. This study explores the multi-disease application of AI in diagnosing cancer, cardiovascular diseases, neurological disorders, and infectious diseases, focusing on its role in improving diagnostic accuracy, speeding diagnostic processes, and facilitating early disease detection. By employing machine learning, deep learning, and neural network models, this study critically examines the performance of specific models – such as recurrent neural networks and support vector machines – in diverse healthcare contexts. Challenges addressed include data privacy, annotated dataset needs, overfitting risks, and ethical concerns such as AI bias and transparency, all of which are fundamental to ensuring patient safety and health equity. In addition, this study integrates security considerations, such as fault detection in cryptographic architectures, providing insights into the resilience of AI systems in healthcare. Future research directions, including the potential of AI in real-time patient monitoring, personalized medicine, and multispectral imaging, are proposed to expand AI's utility in diagnostics. A comparative evaluation with traditional clinical diagnostics underscores AI's validation potential, emphasizing its need for robust regulatory frameworks, particularly concerning global health standards (e.g., TRIPOD-AI and CONSORT-AI) and data privacy regulations such as Health Insurance Portability and Accountability Act and General Data Protection Regulation. Ultimately, AI-driven diagnostic systems show strong promise to revolutionize medical practice and improve patient outcomes, contingent on addressing the technical, ethical, and regulatory challenges involved. This research supports AI's growing role in healthcare, providing a foundational understanding of both its current contributions and future potential across disease-specific applications.

Keywords: Artificial intelligence; Biomedical applications; Data informatics; Deep learning; Healthcare informatics; Machine learning; Medical informatics

***Corresponding author:**
Zarif Bin Akhtar
(zarifbinakhtar@ieee.org)

Citation: Akhtar ZB. Artificial intelligence within medical diagnostics: A multi-disease perspective. *Artif Intell Health*. 2025;2(3):44-62.
doi: 10.36922/aih.5173

Received: October 16, 2024

1st revised: November 19, 2024

2nd revised: December 9, 2024

Accepted: December 12, 2024

Published online: January 6, 2025

Copyright: © 2025 Author(s). This is an Open-Access article distributed under the terms of the Creative Commons Attribution License, permitting distribution, and reproduction in any medium, provided the original work is properly cited.

Publisher's Note: AccScience Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.*Corresponding author:

1. Introduction

The rapid advancements in artificial intelligence (AI) are profoundly reshaping industries worldwide, with healthcare emerging as a critical field poised to benefit substantially from these innovations. Among the various domains in healthcare, medical diagnostics holds exceptional promise for transformation through AI-driven technologies. Diagnostic errors and delays are leading causes of adverse patient outcomes, underscoring the urgent need for diagnostic solutions that are more accurate, timely, and scalable. AI offers notable advancements in this area, enhancing diagnostic precision, reducing human error, and enabling the early detection of diseases by analyzing vast amounts of complex data.[1-6]

This research aims to investigate AI's role in medical diagnostics across multiple diseases, focusing on cancer, cardiovascular diseases, neurological disorders, and infectious diseases. These conditions were selected due to their significant global health burden, diverse diagnostic requirements, and considerable impact on medical health systems. Each of these diseases brings unique diagnostic challenges, such as managing various data complexities, requiring rapid diagnosis, and accurately interpreting medical images, clinical data, and patient histories.[7-9] AI algorithms, particularly those based on deep learning (DL) and machine learning (ML), have shown notable potential in addressing these challenges. Recurrent neural networks (RNNs) with long short-term memory (LSTM) units, for instance, excel in analyzing time-series data, which is crucial for cardiovascular monitoring, while support vector machines (SVMs) effectively handle smaller, structured datasets often encountered in disease diagnostics. Despite these advancements, incorporating AI into clinical practice poses significant challenges. Issues surrounding data quality, model bias, and the availability of extensive annotated datasets present hurdles to AI deployment in healthcare. Further, regulatory and ethical frameworks are essential to govern AI's role in diagnostics, ensuring patient safety, data privacy, and the ethical use of AI tools. The need for international guidelines, such as TRIPOD-AI and CONSORT-AI, highlights the importance of developing standardized reporting for AI in healthcare to promote transparency and accountability. Bias within AI models, if left unaddressed, can contribute to disparities in diagnostic outcomes, particularly for minority and low-resource settings, raising ethical concerns around fairness and equity.[10-12] This study provides an in-depth examination of AI applications in diagnostics, presenting case studies of AI implementations and assessing the benefits and limitations of various models across disease areas. In addition, emerging security considerations, such as fault detection and cryptography, will be discussed in the context of secure AI-driven healthcare applications. The discussion will explore future directions for AI in diagnostics, including real-time patient monitoring, personalized medicine, and multispectral imaging, emphasizing the need for continued research to enhance the robustness and applicability of AI in medical diagnostics.

By addressing the technical, ethical, and regulatory challenges associated with AI in healthcare, this research aims to deliver a comprehensive perspective on how AI can enhance diagnostic processes and improve patient outcomes across multiple disease domains.

2. Methods and experimental analysis

To evaluate the integration of AI in multi-disease diagnostics, this study compiled a diverse dataset drawn from publicly available medical imaging repositories and electronic health records (EHRs) across several healthcare institutions. The dataset was constructed to represent four primary disease areas: cancer, cardiovascular diseases, neurological disorders, and infectious diseases. Specifically, cancer data were sourced from the Cancer Imaging Archive, cardiovascular data from the UK Biobank, neurological data from the Alzheimer's Disease Neuroimaging Initiative and the Human Connectome Project, and infectious disease data from the COVIDx dataset. Comprehensive preprocessing steps were implemented to standardize image sizes, resolutions, and formats, ensuring data uniformity across the disease types. In addition, patient demographics, medical histories, and clinical parameters were anonymized to comply with data privacy standards such as the Health Insurance Portability and Accountability Act (HIPAA) and General Data Protection Regulation (GDPR), preserving patient confidentiality.

The analysis employed a combination of supervised and unsupervised AI models tailored to the specific diagnostic requirements of each disease category. For image-based diagnostics, convolutional neural networks (CNNs) were utilized due to their robust capabilities in medical image classification and segmentation. Sequential time-series data, particularly for cardiovascular applications, were processed with RNNs incorporating LSTM units, which are adept at capturing temporal patterns in physiological signals. SVMs were applied to smaller, tabular datasets, particularly where linear separability was beneficial for diagnostic accuracy. In addition, unsupervised learning methods, such as autoencoders and variational autoencoders (VAEs), were incorporated to detect anomalies in unlabeled data, aiming to identify rare conditions that may otherwise elude traditional diagnostics. Separate models were trained for each disease category using a cross-validation approach to enhance generalizability and mitigate overfitting. Hyperparameter tuning was conducted through grid search to optimize model performance, focusing on parameters such as learning rate, batch size, and layer architecture. The models were evaluated on several key metrics, including classification accuracy, sensitivity (recall), specificity, precision, F1-score, and the area under the curve (AUC) of receiver operating characteristic (ROC) curves. Classification accuracy measured the proportion of correct predictions, while sensitivity quantified the ability to detect true positive cases.

Specificity assessed the accuracy in identifying true negative cases, and precision captured the proportion of true positives relative to all positive predictions. The F1-score offered a very balanced metric that combines both precision and recall, making it particularly useful for evaluating performance on the various types of imbalanced datasets. The AUC-ROC provided an overall assessment of model performance across different classification thresholds. These results were benchmarked against traditional diagnostic methods to gauge potential improvements in diagnostic accuracy, efficiency, and speed across the four disease categories.

The experiments were conducted using high-performance computing infrastructure capable of managing extensive medical datasets and complex model architectures. The process involved three main phases. First, data preprocessing included image normalization, augmentation, and management of missing or incomplete clinical data to improve model training reliability. During the training and validation phase, 80% of the dataset was allocated for training, with the remaining 20% reserved for validation. Model fine-tuning employed backpropagation with the Adam optimizer and cross-entropy loss functions to optimize the performance of the model. Finally, model testing was carried out on unseen data from diverse patient cohorts to assess generalizability. Predictions were validated against expert-reviewed ground truth labels and clinical diagnoses, ensuring alignment with clinical standards. A detailed comparative analysis was performed to assess the AI models across disease categories, focusing on the models' diagnostic efficacy within specific disease contexts. In cancer diagnostics, models were evaluated for the early-stage detection and tumor segmentation accuracy, while in cardiovascular diseases, the focus was on detecting arrhythmias, coronary artery disease, and heart failure. For neurological disorders, early detection of Alzheimer's and Parkinson's diseases from magnetic resonance imaging (MRI) data was emphasized. For infectious diseases, the models' ability to analyze chest X-rays for the detection of coronavirus disease 2019 (COVID-19) and tuberculosis was examined. This cross-disease comparison highlighted the strengths and limitations of AI models within each diagnostic scenario, allowing for an understanding of AI's capabilities and challenges in a clinical setting.

To address fairness and ethical concerns, the study incorporated bias detection and mitigation strategies. Subgroup analyses were conducted to examine variations in diagnostic accuracy across demographic groups, including age, gender, and ethnicity. Bias mitigation techniques, such as fairness-aware algorithms and regularization methods, were applied to reduce disparities. Ethical considerations included patient data privacy, the implications of AI-driven diagnostic errors, and the need for regulatory frameworks that support AI integration into healthcare while protecting patient welfare. These measures ensure the responsible deployment of AI within the sensitive domain of medical diagnostics, emphasizing the importance of accuracy, equity, and transparency.

3. Background research on available knowledge

AI in healthcare represents a significant technological advancement, simulating human cognition to analyze, interpret, and present complex medical and healthcare data. This ability not only mimics human thought processes but also enhances healthcare delivery by enabling faster and more accurate diagnoses, treatments, and preventive measures. ML and DL algorithms, key components of AI, can process vast amounts of clinical data, such as EHRs, to support physicians in making quicker and more precise diagnoses. By analyzing large datasets, AI can aid in disease prediction and treatment, helping clinicians save time and improve patient outcomes.[1-3] AI is instrumental in bringing culturally competent practices to the healthcare industry, ensuring more tailored and inclusive patient care. AI's applications in healthcare span numerous areas, including diagnostics, treatment protocol development, drug discovery, personalized medicine, and patient monitoring. In radiology, AI's role is particularly noteworthy for interpreting and triaging X-ray images, one of the most commonly used imaging tests.[4-6] AI can analyze these images, helping radiologists prioritize critical cases and reducing wait times. However, despite its promising potential, the widespread adoption of AI in healthcare faces challenges, including ethical concerns about data privacy, job automation, and the amplification of biases. Moreover, resistance from healthcare leaders to embrace new AI technologies has slowed its integration into mainstream medical practices.

In terms of disease diagnosis, AI plays a pivotal role by helping clinicians navigate complex medical data to identify conditions accurately. By leveraging vast EHR datasets, AI algorithms can predict diseases such as Alzheimer's and dementia, providing early diagnosis and potentially improving treatment outcomes.[7-9] In emergency settings, AI can prioritize urgent cases by providing real-time data interpretation to assist decision-making, thereby enhancing efficiency and potentially saving lives. Studies have shown that AI, through platforms like ChatGPT, can generate responses to medical queries that are perceived as more empathetic and of higher quality than responses from healthcare professionals, though these are not in the context of established patient-physician relationships. EHR systems, widely adopted in healthcare, have become essential for storing and sharing patient data. AI enhances EHR functionality by utilizing natural language processing (NLP) to standardize medical terminology, improve the readability of medical notes, and predict patient risks based on historical data. By identifying trends in patient data, AI can offer predictive insights, alerting physicians to potential health risks and allowing for preemptive interventions. These predictive models have achieved significant accuracy in assessing treatment responses, further demonstrating the value of AI in patient care management.[10-12] With the volume of EHRs doubling every 5 years, AI offers the necessary bandwidth to analyze this data effectively and assist healthcare providers in making informed clinical decisions.

AI has also made significant strides in addressing drug-drug interactions (DDIs), a critical issue in patients taking multiple medications. Advanced algorithms can scan medical literature and user-generated content, such as EHRs and adverse event reports, to identify potential interactions between drugs.[13-15] These innovations have the potential to prevent harmful drug interactions, improving patient safety. Competitions such as the DDIExtraction challenge have helped standardize and evaluate the effectiveness of these AI-driven algorithms, driving further research and development in this field.

Telemedicine, which has surged in popularity, offers another area where AI is transforming healthcare. Through the use of sensors and wearable devices, AI can monitor patients remotely, identifying subtle changes in health that may go unnoticed by human caregivers.[16-18] These devices allow for constant patient monitoring, alerting physicians to potential issues in real time.

AI-powered chatbots have also been introduced for mental health therapy, though some experts argue that they cannot replace the human connection necessary for effective care.[19-21] As life expectancy increases and the aging population grows, AI can help caregivers monitor elderly patients through personal and environmental sensors, though these technologies raise privacy concerns.[22-24]

Despite these limitations, AI's role in healthcare will likely continue to expand, offering solutions to complex medical challenges while improving patient outcomes. AI is showing increasing promise in various clinical applications across a wide variety of medical specialties.[25-75]

3.1. Cardiovascular medicine

AI has demonstrated significant potential in diagnosing coronary artery disease and predicting outcomes such as patient mortality and adverse effects following acute coronary syndrome treatment.[5,6] Wearable devices and smartphones are expanding the ability to monitor cardiac health, potentially enabling earlier detection of events like heart attacks outside hospital settings.[7,8] AI has also been applied to analyze heart sounds and diagnose valvular disease;[9] however, challenges remain due to limited training data, especially regarding social determinants of cardiovascular health. In some areas, AI is non-inferior to humans, such as echocardiogram interpretation, and has even outperformed physicians in diagnosing heart attacks in emergency settings.[10]

3.2. Dermatology

AI has made strides in processing medical images for dermatological diagnoses, such as skin cancer detection. Studies show that ML models can achieve dermatologist-level accuracy in some cases.[32] However, many studies have not adequately engaged with external validation or considered skin tone disparities, which are crucial for equitable diagnosis and treatment. AI also shows potential in evaluating the outcomes of maxillofacial surgeries.[33]

3.3. Gastroenterology

AI has improved the detection of abnormal tissues during endoscopic procedures like colonoscopies, with the early stomach cancer detection showing sensitivity close to expert endoscopists.[34] AI tools are being developed to predict ulcerative colitis flare-ups with similar accuracy to human pathologists, offering promising support for disease management.[35]

3.4. Obstetrics and gynecology

AI is enhancing imaging techniques such as ultrasound and MRI in obstetrics, assisting in diagnosing and monitoring pregnancies. Its applications are expanding in areas like fetal monitoring, with AI improving diagnostic capabilities for various obstetrical issues.[38]

3.5. Infectious diseases

During the COVID-19 pandemic, AI contributed to early detection and monitoring of virus spread.[39,40] Other applications include detecting antimicrobial resistance and malaria and improving point-of-care diagnostics for diseases such as Lyme disease and sepsis.[41,42] AI has also been used in analyzing blood smears and predicting complications in viral infections like hepatitis.[44,45]

3.6. Musculoskeletal medicine

AI can uncover causes of knee pain often missed by doctors, especially in underserved populations.[46] By identifying pain contributors beyond visible radiographic findings, AI may help address disparities in diagnosis and treatment for conditions like osteoarthritis.

3.7. Neurology

AI is being explored for diagnosing and forecasting the progression of Alzheimer's disease through MRI data and ML models, such as CNNs.[17] Generative adversarial networks (GANs) have shown promise in improving early diagnostic accuracy and prognosis predictions.[42,62]

3.8. Oncology

AI has broad applications in cancer diagnosis, risk stratification, and treatment personalization. Algorithms have demonstrated superior accuracy to human experts in breast cancer detection and prostate cancer identification.[50-56] AI is also being explored for grading the aggressiveness of sarcomas and molecular characterizations of tumors.[58-62] However, challenges remain, such as a lack of external validation and transparency in some AI studies, raising concerns about scientific robustness.

3.9. Ophthalmology

AI is aiding in the early detection of eye diseases, such as diabetic retinopathy, and has received United States (U.S.) Food and Drug Administration (FDA) approval for diagnosing specific eye conditions.[62-66] AI promises to improve diagnostic rates and efficiency in ophthalmic care.

3.10. Pathology

AI is revolutionizing digital pathology by helping diagnose cancers and predicting genetic mutations. AI tools can analyze large-scale samples for diseases such as colorectal and breast cancer, improving efficiency and accuracy.[60-68] However, widespread implementation requires more prospective studies to demonstrate AI's clinical utility.

3.11. Primary care

In primary care, AI is being used for decision-making, predictive modeling, and analytics.[50-62] While examples of AI's clinical efficacy are limited, it has shown positive effects on treatment choices when integrated with physician decision-making processes.

3.12. Psychiatry

AI applications in psychiatry include predictive models for diagnosis and treatment outcomes, as well as chatbots for mental health support.[66-72] However, challenges such as small, biased datasets and ethical concerns related to corporate-driven AI initiatives highlight the need for further research and validation in this field.

3.13. Radiology

AI is being used to analyze medical images, such as computed tomography (CT) and MRI scans, for detecting diseases. It can provide benefits such as noise reduction, enhanced image quality, and anatomical landmarking, proving particularly useful in scenarios where human expertise is limited or the data are complex.[52-68]

AI's potential in these fields is vast, though challenges related to bias, validation, and integration into clinical practice must still be addressed to realize its full impact. The integration of AI into the healthcare industry is transforming how data are accessed, processed, and used to support clinical decision-making. Large health companies merging allows for greater accessibility to vast amounts of health data, which serves as the foundation for AI-driven solutions.[11-66] As AI algorithms continue to evolve, they enable more robust clinical decision support systems (CDSSs), adapting through the use of ML techniques. Companies in the healthcare sector are increasingly focused on big data, seeking opportunities in areas such as data assessment, storage, management, and analysis. This industry focus is crucial for AI-powered innovations that enhance healthcare services and outcomes.

Several major companies are leading the development of AI technologies in healthcare. IBM, through its Watson Oncology platform, is working with institutions such as Memorial Sloan Kettering Cancer Center and Cleveland Clinic to develop AI applications for cancer treatment. Other collaborations include chronic disease treatment with CVS Health and drug development analysis with Johnson and Johnson. Microsoft's Hanover project aims to predict the most effective cancer drug treatments, while Google's DeepMind works with the UK's National Health Service to detect health risks and develop cancer detection algorithms. Tencent, Intel, and startups like Lumiata also contribute to healthcare AI development, focusing on diagnostic services, medical imaging, and patient care solutions. Neuralink, founded by Elon Musk, is pushing the boundaries of neuroprosthetics with its brain chip technology, which interfaces with neural pathways to treat conditions like paralysis. Digital consultant apps, such as Babylon Health, use AI to offer medical consultations based on symptoms and medical history. These innovations are supported by business models that target different user groups, including patients, healthcare providers, and payers, offering solutions such as data connectivity and personalized treatment recommendations.

AI is also being deployed in developing nations, where healthcare resources are limited. With the increasing availability of computers and internet access, AI-driven diagnostic tools are providing life-saving services to people in areas with limited access to healthcare professionals. This helps to reduce outsourcing and improve the quality of care. AI systems in these regions are tailored to offer individualized treatments, adjusting based on real-time data, thus improving patient outcomes in resource-constrained settings.

However, the widespread adoption of AI in healthcare brings with it several regulatory and ethical concerns. The risks associated with AI, such as algorithmic bias, patient data privacy, and the implications of machine morality, necessitate stringent regulations. There are established reporting guidelines, such as TRIPOD-AI, DECIDE-AI, and CONSORT-AI, that ensure AI studies are transparently reported for regulatory approval. In the U.S., HIPAA safeguards patient data, while GDPR protects patient data in the European Union.

The U.S. FDA has developed a plan to regulate AI-based medical devices, focusing on areas such as good ML practices and algorithm bias. Ethical concerns in AI include the balance between patient autonomy and the use of AI in making critical healthcare decisions. The U.S. Department of Health and Human Services has issued guidelines emphasizing the ethical principles of autonomy, beneficence, non-maleficence, and justice. Similarly, the GDPR in Europe protects citizens' data, ensuring fairness, transparency, and respect for human dignity in AI-driven processes. As AI continues to shape healthcare, balancing innovation with patient rights and ethical standards will remain a key challenge for regulators globally.

Finally, international efforts to standardize AI use in healthcare are underway. The joint World Health Organization (WHO) and the International Telecommunication Union (ITU) (ITU-WHO) Focus Group on AI for Health (FG-AI$_4$H) has been benchmarking AI applications in medical settings, including cancer risk assessment and diagnosis from medical imagery. These initiatives reflect the growing importance of ensuring that AI in healthcare is safe, effective, and ethical, while promoting innovations that improve global health outcomes.[7-75] The use of AI in healthcare presents significant ethical concerns, especially regarding data collection, automation, and bias. One of the key issues is the massive amount of data required to train AI systems, often sourced from patients. This raises privacy concerns, as many individuals are uncomfortable with sharing personal health information for technological advancements. A survey in the UK revealed that 63% of respondents were hesitant about data sharing for AI development. The scarcity of real, accessible patient data limits the potential of AI in healthcare, with fears exacerbated by the lack of regulations governing AI usage in countries like the U.S.

Concerns about data being misused for financial gain, as exemplified by Roche's purchase of healthcare data for 2 million cancer patients, question whether patient data can or should have a monetary value, raising broader ethical debates around fairness and patient consent. Automation in healthcare also stirs controversy. While AI has yet to replace healthcare jobs, research shows that automation might affect roles that handle digital information, such as radiology and pathology. A 2019 study in the UK estimated that AI could replace up to 35% of jobs within two decades, though doctor-patient interactions are less likely to be impacted.[53] On the positive side, AI offers the potential to enhance healthcare by alleviating burnout and cognitive overload for medical professionals, allowing them to focus more on patient care. AI is even being applied in elder care, where robots assist with entertainment and companionship, allowing caregivers to provide more one-on-one care. Despite these benefits, skepticism remains about whether AI can offer the empathy provided by human healthcare professionals, as found in a 2023 thematic review.[55,61,68] Bias in AI systems is another critical concern. AI's reliance on input data means that if the data are biased or unrepresentative, it can lead to discriminatory outcomes.

Medical AI systems can unintentionally perpetuate social and healthcare inequities by making more accurate predictions for majority populations, such as White males, who are overrepresented in medical datasets. This can result in worse outcomes for minorities. Collecting data from minority communities, while essential for balanced algorithms, can also lead to medical discrimination, such as the potential misuse of data related to conditions like acquired immunodeficiency syndrome. Beyond demographic biases, differences in clinical systems and work practices also introduce variability in AI functionality. However, many of these biases can be mitigated through careful data collection and algorithm design. A specific form of bias, *"label choice bias,"* arises when proxy measures are used in algorithms, such as using healthcare costs to predict patient needs, which can skew results against certain groups, like Black patients.[68,76] Addressing these biases requires closer alignment between the target of the AI prediction and the actual healthcare needs of patients. Historically, AI in healthcare has evolved since the 1960s and 1970s, starting with expert systems such as Dendral and MYCIN, which laid the groundwork for future AI applications in medicine.[15-75] Although these early systems did not achieve widespread clinical use, they highlighted the potential for AI in healthcare. In the following decades, advances in computing power, genomics, and EHRs enabled the expansion of AI's role in healthcare. Breakthroughs in NLP, computer vision, and ML have allowed machines to replicate human-like decision-making and perceptual processes. AI has contributed to innovations such as robot-assisted surgery, rare disease prediction through DL, and more precise health prediction. Despite these advances, the ethical challenges surrounding data, automation, and bias remain central to discussions about AI's future in healthcare.

4. Diagnosis in AI

Diagnosis, as a subfield of AI, is focused on creating algorithms that can assess whether a system is functioning properly. If a malfunction is detected, these algorithms are responsible for accurately identifying the faulty component and the nature of the fault. This process is based on observations, which provide insights into the system's current state. The term *"diagnosis"* originates from the medical field, where it refers to identifying diseases based on symptoms, but in AI, it broadly encompasses both the detection of faults and the process of determining if a system is malfunctioning.

An everyday example of diagnosis can be illustrated with a car mechanic troubleshooting a vehicle. The mechanic begins by observing the car's behavior and applying their knowledge of the vehicle type. If a problem is detected, further tests and observations are conducted to refine the diagnosis until the faulty part is discovered. In AI, expert diagnosis systems operate similarly by mapping observations to diagnoses based on prior experience.

This expertise may be derived from human operators, who encode their knowledge into a computer-readable format, or from examples of system behavior classified as either correct or faulty. ML techniques can then generalize to metadata in terms of DL from these examples. Multimodal models can be used for further exploration for finding new features and functionality. However, expert diagnosis faces challenges, such as difficulty in acquiring sufficient expertise, especially in critical systems, the complexity of the learning process, and potential limitations in storage and robustness.

A more structured approach to diagnosis is model-based diagnosis, which employs a model of the system to simulate its behavior. By comparing actual observations with predicted outcomes from the model, faults can be identified. In this form of abductive reasoning, the model may describe normal system behavior but often lacks a detailed representation of faulty behavior. The diagnostic system uses this model to determine if the system is functioning correctly by analyzing discrepancies between expected and observed behavior. Diagnosability, a key concept, refers to the ability of the system to provide an unambiguous diagnosis. This is particularly crucial during system design, where a balance must be struck between reducing sensor costs and increasing the ability to detect faulty behaviors. Algorithms have been developed to ensure diagnosability by either confirming whether a system is diagnosable or identifying the necessary set of sensors to make a system diagnosable. Diagnosis in AI deals with detecting malfunctions in systems and identifying their causes through expert systems or model-based approaches. These techniques rely on observations and simulations to provide accurate diagnoses, but they come with challenges such as expertise acquisition, system complexity, and diagnosability. Figure 1 offers further insights into this issue.

5. Improving medical diagnosis through AI

AI is reshaping medical diagnostics by delivering remarkable advancements in accuracy, speed, and the personalization of patient care. Through sophisticated ML and DL models, AI enables the processing of extensive datasets, the analysis of complex medical images, the prediction of disease progression, and significant enhancements in diagnostic precision. This transformation is especially prominent in specialized fields such as radiology, wound and burn management, and diabetic care, where AI-driven innovations have made a substantial impact on improving patient outcomes.[18-22]

5.1. Vital contributions of AI in medical diagnostics

Enhanced diagnostic accuracy. AI systems, particularly within radiology, have shown superior performance in pattern recognition, often identifying early signs of disease that can be overlooked by the human eye. For instance, AI has demonstrated higher accuracy in detecting breast cancer from mammograms, offering critical insights that can lead to earlier intervention and better survival rates.[18]

> *Facilitating early detection.* By enabling the early diagnosis of life-threatening conditions such as cancer, cardiovascular diseases, and neurological disorders, AI-driven tools facilitate prompt treatment planning. Early detection through AI tools has been linked to improved patient outcomes and reduced mortality rates, particularly in cases where time-sensitive interventions are essential.[19]
>
> *Advancements in personalized medicine.* AI algorithms aid in tailoring treatments to individual patient profiles by analyzing personal health records, genetics, and lifestyle factors. This personalized approach supports the development of more effective treatment plans,

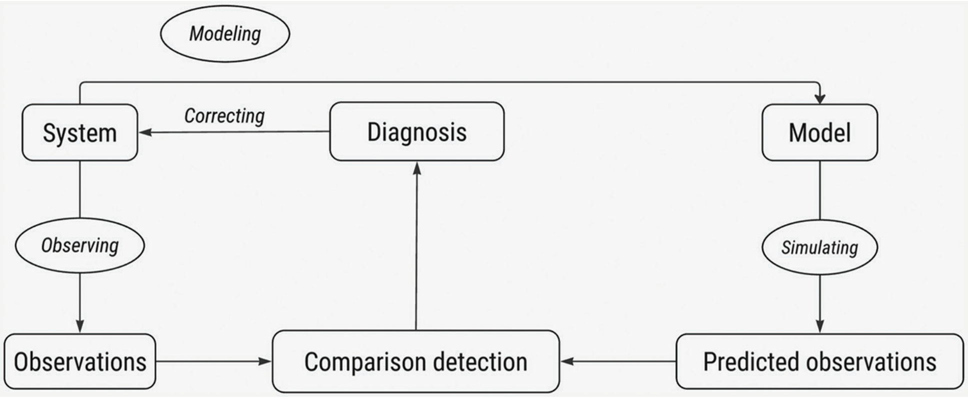


Figure 1. Model-based diagnosis in AI. Figure created by author using MIRO software.

minimizes trial and error in treatment selection, and enhances overall healthcare efficiency.[20]

Predictive analytics for disease progression. AI contributes to predictive diagnostics by estimating disease progression and recovery pathways, which helps clinicians optimize healthcare resources and better allocate medical staff. Predictive analytics allows for personalized care plans that improve patient management and help in the early identification of patients who might require escalated care.[21,22]

5.2. AI in wound and burn management

In wound care and burn management, AI technologies have made a notable impact by improving diagnostic accuracy and treatment planning. AI-powered tools like Spectral AI's DeepView® technology use medical imaging to analyze wound depth, infection risks, and healing progress, aiding clinicians in making informed decisions faster. Such advancements reduce the risk of complications, enhance recovery outcomes, and streamline the diagnostic process.[32,33] By evaluating images of chronic wounds and burns, AI systems predict healing timelines and treatment effectiveness, thus improving patient outcomes and quality of care.[46-52]

5.3. Challenges and considerations

Despite the remarkable progress, the integration of AI into healthcare diagnostics presents unique challenges. Key issues include the need for seamless integration with existing healthcare systems, ensuring robust data privacy protections, and establishing clear regulatory guidelines to govern AI's ethical use. In addition, biases in AI models and limitations in generalizability across diverse patient populations present critical considerations for fair and accurate diagnostics. Addressing these challenges is essential for maximizing AI's potential, promoting equity in healthcare delivery, and ensuring reliable and transparent diagnostic tools. As AI technology continues to evolve, its role in medical diagnostics will likely expand, offering faster, more accurate, and personalized diagnostic services across a wide range of medical disciplines.

6. AI for disease diagnosis: Current and future directions in the medical arena

AI is transmuting disease diagnostics by offering unparalleled precision, swiftness, and personalized attention. Leveraging machine learning ML and DL algorithms, AI excels at processing vast datasets, analyzing medical images, predicting disease outcomes, and enhancing diagnostic accuracy, establishing itself as a cornerstone in the medical informatics domain.

6.1. AI technologies in medical diagnostics

AI has emerged as a transformative tool in medical diagnostics, utilizing advanced algorithms and ML models to assist clinicians in identifying and diagnosing diseases with greater accuracy and speed. At present, AI-powered systems have demonstrated efficacy in interpreting medical images such as X-rays, MRI, and CT scans, facilitating early detection and more precise diagnoses.[3-13] In addition to medical imaging, AI algorithms are increasingly being employed to analyze patient data, medical history, and symptoms, helping to formulate diagnostic predictions. These systems not only support physicians but also optimize the overall diagnostic process by reducing human error and enhancing decision-making.[15-22]

AI's potential in healthcare extends beyond its current capabilities. Future AI applications could involve analyzing large datasets to detect patterns that may predict diseases before symptoms manifest, potentially revolutionizing preventive medicine. Moreover, by integrating multimodal data – such as genetic, environmental, and lifestyle information – AI could offer solutions for diagnosing complex diseases that typically involve multiple variables.

However, while AI has made significant strides, it is crucial to acknowledge that these technologies are designed to augment, not replace, human expertise in healthcare. The ethical implications of AI use, including concerns about data privacy and the need for continuous validation of AI models, remain critical areas for future research and development.

6.2. Evolution of AI in healthcare

Driven by the challenges of an aging population and a global shortage of healthcare professionals, the adoption of AI in medical diagnostics has expanded rapidly. AI's integration into healthcare systems enables the development of intelligent, efficient systems for managing patient records, developing treatment plans, and diagnosing diseases. The evolution of AI has been categorized into two primary systems: expert systems and ML-based systems.

Expert systems are designed to replicate human decision-making, drawing from a structured knowledge base and inference engine. These systems assist in diagnostic processes using predefined rules and logical reasoning to provide clinical insights. However, expert systems have limitations in scalability and adaptability due to their reliance on predefined knowledge bases.

On the other hand, ML algorithms have become increasingly prevalent due to their ability to learn and improve from large datasets without requiring explicit programming. ML models, especially those using DL techniques, are particularly powerful in identifying patterns in complex data, including medical imaging and genomic sequences. The predictive capabilities of these models improve as they are exposed to more data, making them valuable assets in dynamic healthcare environments.

6.3. AI models in clinical diagnostics

Numerous AI models, such as SVM, classification trees, and artificial neural networks (ANN), have shown promising results in diagnosing acute and chronic illnesses.[15-22] These technologies have been successfully applied in detecting conditions such as acute appendicitis and Alzheimer's disease.[20-26]

Moreover, the integration of multiple AI algorithms has significantly enhanced the accuracy of detecting malignant cells.[18-25] The development of AI-driven systems has also shown promise in predicting the recurrence of diseases such as breast cancer and monitoring patients with chronic conditions such as diabetes and swallowing disorders.[13-22]

AI's application in healthcare diagnostics is particularly valuable in cases where human error is common or where there is a need to process vast amounts of data quickly. AI algorithms can analyze medical data, learn from patterns over time, and provide real-time insights to healthcare providers. As healthcare systems continue to adopt digital technologies, AI's role in diagnostics is expected to expand significantly, aiding in the transition toward personalized medicine.

6.4. Impact of AI on healthcare workflows

Medical diagnostics is a complex and time-sensitive field, often constrained by the limited availability of healthcare professionals and the increasing demands of an aging population. AI has the potential to alleviate some of these pressures by automating routine diagnostic tasks and allowing healthcare providers to focus on more complex aspects of patient care. AI-enabled systems can continuously process and learn from new patient data, updating diagnostic models in real time and potentially surpassing human capabilities in certain diagnostic areas.[26-28]

For example, AI can be particularly effective in analyzing medical images from multiple modalities (e.g., CT, MRI, and X-ray) to identify abnormalities that may be missed by human eyes. In addition, based on up-to-date patient information and medical data, AI-powered CDSSs can offer real-time recommendations to healthcare providers, guiding them toward optimal treatment strategies.

6.5. Future directions for AI in medical diagnostics

Looking ahead, the future of AI in medical diagnostics is likely to involve more sophisticated technologies, such as quantum AI (QAI) and general AI (GAI). QAI has the potential to accelerate diagnostic model training by leveraging the superior processing power of quantum computers, enabling faster analysis of large datasets. This could lead to more accurate and timely diagnoses, especially for complex diseases that require the evaluation of multiple factors.

Similarly, GAI systems – such as IBM's Watson, Google's DeepMind, and OpenAI's GPT models – are increasingly being integrated into healthcare applications to provide more holistic and generalized diagnostic solutions. AI also holds promise in the realm of personalized medicine, where algorithms can analyze individual patient data – ranging from medical history to genetic information – to tailor treatment plans specifically to the patient's needs.

This approach could lead to more effective treatment outcomes and a more efficient overall healthcare system. However, the widespread adoption of AI in medical diagnostics will require addressing several challenges, including the need for high-quality, labeled medical data, interoperability between AI systems, and the development of robust legal and ethical frameworks to govern AI use in healthcare.

6.6. AI in specialized medical fields: Dentistry

AI's role in medical diagnostics extends beyond general healthcare and into specialized fields like dentistry. Orthodontics, for instance, relies on AI to diagnose malocclusions and plan treatments. By analyzing cephalometric radiographs, AI systems can assess abnormalities in dental and craniofacial structures more accurately than traditional methods.[22-32] AI-driven models help that orthodontists make precise diagnoses and improve treatment outcomes by identifying subtle patterns in dental data. AI technologies are reshaping medical diagnostics, offering innovative solutions for disease detection, treatment planning, and patient management. As AI continues to evolve, its applications in healthcare will likely expand, leading to more accurate, efficient, and personalized medical care. However, continuous research is required to address the technological, ethical, and regulatory challenges associated with AI integration in medical diagnostics, ensuring that these systems benefit both healthcare providers and patients.

6.7. The application of mHealth

The rapid evolution of information technology in healthcare has progressed beyond simple data collection to the sophisticated use of AI for advanced diagnostics and preventative medicine. Initially, healthcare information systems were designed solely for the purpose of gathering patient data. However, with the rise of ML techniques and data analytics, healthcare providers are now able to leverage this data to make smarter, faster, and more accurate decisions.[48] One key area that has seen significant growth is the field of mobile health (mHealth), which utilizes AI to enhance patient care through real-time monitoring, particularly for life-threatening conditions such as asthma, diabetes, and sleep apnea.[49,50]

The growing presence of wearable technologies, Internet of Things (IoT) devices, and mobile sensors has fueled the expansion of mHealth. This trend is reflected in the increasing adoption of remote in-home care and telemedicine, particularly in response to the COVID-19 pandemic.[51,52] The healthcare information technology market has seen substantial growth in the use of these technologies, enabling real-time monitoring of patients outside traditional healthcare facilities. Not only does mHealth facilitate early detection and treatment of chronic diseases but it also enhances patient safety and well-being through remote monitoring solutions.[53,54] The pandemic has further accelerated the adoption of mHealth technologies, as social distancing measures necessitated the use of remote medical facilities to maintain healthcare delivery.[55,56] AI-powered mHealth (AIM) has emerged as a promising subfield that integrates AI techniques with mHealth applications to address key healthcare challenges. These AI techniques, including DL, federated learning (FL), and explainable AI (XAI), offer more accurate diagnostic insights while ensuring patient privacy and data security.[57] Researchers have noted the increasing integration of AI in mHealth, especially during the pandemic, with advancements such as real-time disease progression monitoring and chronic disease management.[58-61]

These technologies are instrumental in providing non-invasive care and enabling emergency responses, particularly for at-risk communities and individuals with limited access to healthcare.[62,63] AI's integration into mHealth offers significant advantages, including automated chronic disease detection, suicide prediction and intervention, and reduction of medical errors. Medical errors remain one of the leading causes of preventable deaths in the U. S., and AI-enabled clinical decision-making systems, which use real-time data from wearable sensors, can substantially reduce these errors.[64,65] AIM solutions also hold the potential to extend high-quality medical care to underserved populations, addressing healthcare inequities and improving patient outcomes on a broader scale.[66] However, challenges remain in the implementation of AIM technologies. The healthcare industry has historically been resistant to automation, largely due to concerns surrounding data privacy, interpretability of AI models, and regulatory constraints.[67]

Nevertheless, recent advances in DL and FL have opened new possibilities for AI in healthcare, allowing for secure data management and knowledge transfer across decentralized systems.[68] FL, in particular, ensures that patient data remains within healthcare organizations, while still enabling the training of powerful AI models on distributed datasets.[69,70] This allows for the development of personalized care algorithms without compromising data security.

The growing importance of XAI in the mHealth domain cannot be overstated. The Defense Advanced Research Projects Agency has been instrumental in promoting the development of AI models that are interpretable, trustworthy, and usable by healthcare professionals.[71] XAI plays a key role in fostering trust and acceptance of AI models in medical practice, as it allows clinicians to better understand the rationale behind AI-driven decisions. This transparency can improve decision-making in clinical settings, reduce medical errors, and enhance the overall efficacy of healthcare delivery.[72,73] The convergence of AI and mHealth also points toward the potential for AI models to become more user-friendly and widely adopted across hospitals and healthcare systems. Studies have shown that as AI becomes easier to implement and use, its adoption rates in healthcare will increase, leading to improved outcomes for patients.[74] For instance, the use of XAI to assist in real-time clinical decision support during surgeries or other medical interventions has already shown promise in improving patient outcomes and preventing intraoperative complications.[75] Furthermore, as the research and implementation of AIM technologies continue to evolve, it is essential to address the limitations that currently hinder their full potential. Data siloing across hospitals and medical institutions, the lack of standardized protocols for data sharing, and the need for greater collaboration between healthcare organizations are some of the primary challenges that must be addressed. Efforts to encourage data exchange and collaboration among healthcare providers will facilitate the widespread use of AI tools, ensuring that AIM solutions reach their full potential in improving patient care.[76-78] However, in recent years, many computer-aided diagnoses (CADs) have been used to diagnose and classify breast cancer using traditional red green blue images that analyze the images only in three-color channels. In CAD, a radiologist interprets mammograms that are also analyzed by a computer that detects potential breast lesions or differentiates breast lesions as malignant or benign. Mammograms are commonly used to screen for breast cancer. If a screening mammogram finds something concerning, another mammogram might be performed to look at the area more closely. This more-detailed mammogram is called a diagnostic mammogram and is often used to closely examine both breasts.[79,80]

The integration of AI into mHealth has shown immense promise in transforming the healthcare landscape, particularly in the areas of remote patient monitoring, chronic disease management, and preventative medicine.

By leveraging AI techniques such as DL, FL, and XAI, mHealth technologies can provide accurate, secure, and interpretable insights that improve clinical decision-making and patient outcomes. As the healthcare industry continues to evolve, further research and investment in AIM solutions will be crucial in ensuring their effective deployment to enhance the quality of care and address critical healthcare challenges.

7. AI in the realm of diagnosing medical conditions and its impact on healthcare

AI is transforming healthcare by enhancing medical diagnosis through the use of ML, NLP, and other subdomains. With an expected annual growth rate of 37.3% from 2023 to 2030, AI is becoming a key player in various aspects of medicine. AI's ability to analyze vast amounts of medical data is improving diagnosis and treatment processes, offering faster, more precise diagnoses, earlier disease detection, and more personalized treatment options. AI leverages DL, computer vision, and sophisticated algorithms to interpret medical data, serving as an expert assistant to healthcare professionals.

AI is revolutionizing healthcare through its applications in medical imaging, surgery, drug discovery, and virtual health assistants. By detecting anomalies in scans, extracting insights from clinical notes, and offering diagnostic suggestions, AI enhances the accuracy and speed of diagnosis. In fields such as radiology, pathology, cardiology, and dermatology, AI tools are aiding in the detection of fractures, cancer cells, heart disease, and skin conditions. This technology allows healthcare professionals to detect subtle patterns that may go unnoticed by humans, reducing the likelihood of diagnostic errors and providing a layer of impartiality and precision. AI's strength lies in its ability to mimic human cognition, but with enhanced computational speed and learning capacity. By processing extensive datasets, AI can identify trends and symptoms that are associated with various medical conditions, improving its diagnostic accuracy over time. Its integration into diverse medical fields has proven successful, especially in radiology, where it detects tumors and fractures with high precision, and in cardiology, where it helps predict heart disease risk. Moreover, AI's lack of fatigue and biases means it can work tirelessly, reducing the potential for errors.

AI also plays a critical role in personalized medicine. Its integration with EHRs allows AI systems to analyze a patient's medical history, identifying risk factors and providing real-time insights to clinicians. This capability enhances diagnosis and treatment, offering tailored healthcare solutions. Furthermore, AI-driven drug discovery platforms accelerate the identification of potential drug candidates, revolutionizing pharmaceutical research and making it more efficient.

This is especially relevant in the development of personalized cancer treatments, where AI's ability to analyze genetic markers leads to better treatment options.[13-33] The benefits of AI in healthcare extend beyond accurate diagnosis and personalized medicine. AI streamlines diagnostic procedures, reducing the time and effort required for analysis and interpretation. This efficiency results in cost savings for healthcare systems by enabling early detection and intervention, which can reduce hospitalizations and shorten treatment durations. Real-world applications such as Google's DeepMind algorithms, which predict acute kidney injury up to 48 h before it occurs, demonstrate AI's life-saving potential by enabling healthcare providers to take preventive measures.[18-38] Despite AI's numerous advantages, it remains a complementary tool to healthcare professionals rather than a replacement. The human element in healthcare, characterized by empathy, ethical judgment, and experience, is indispensable. AI's role in healthcare is to support professionals by providing diagnostic suggestions and real-time insights. However, ethical concerns surrounding data privacy, algorithmic biases, and patient consent must be addressed to ensure AI's responsible and effective use in personalized medicine. AI is reshaping the healthcare landscape by improving diagnosis, treatment, and personalized medicine. It enables healthcare providers to offer more precise and timely interventions while reducing costs and increasing efficiency. As AI continues to evolve, collaboration between healthcare professionals and AI will be critical in ensuring its ethical and effective integration into medical practice. Looking ahead, the potential of AI to further enhance patient care is promising, with ongoing research and development helping to unlock even greater possibilities for the future of medicine.

8. Results and findings

The investigation highlights transformative advancements in AI, particularly in DL and ML applications in healthcare, with a focus on improving diagnostic accuracy, early disease detection, and personalized care.

8.1. Crucial AI contributions in medical diagnostics

DL, powered by ANNs, has shown the most substantial impact in medical diagnostics. Enhanced computational resources, the availability of large, labeled datasets, and accessible frameworks have propelled the success of DL, particularly in medical imaging. The turning point for DL was marked by the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC), where CNNs significantly reduced error rates in object detection and classification tasks, surpassing traditional methods and, in some cases, human performance.[24-28] Figures 2-5 demonstrate the research findings, showcasing the advancements that DL and ML techniques have contributed to healthcare diagnostics. These visualizations highlight DL's effectiveness in analyzing large datasets, detecting complex disease patterns, and achieving high accuracy in disease prediction.

8.2. Disease diagnosis and prediction through DL and ML

DL and ML models have shown high accuracy in diagnosing critical diseases such as liver disease, heart disease, Alzheimer's disease, and various cancers. Early diagnosis is especially crucial in these diseases, where timely intervention improves patient outcomes. For example, DL has been employed in medical imaging to differentiate bacterial pneumonia in pediatric chest radiographs and identify unique characteristics in chest CT images, outperforming traditional diagnostic techniques. In addition, hybrid models, including case-based reasoning (CBR) systems, have been used to diagnose skin diseases while ANN-based real-time monitoring systems help patients manage critical health metrics, enhancing care during emergencies.

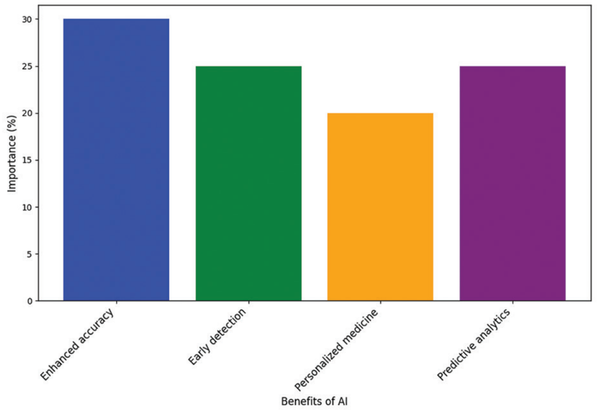


Figure 2. Benefits of AI in medical diagnostics. The chart shows that among the listed benefits of AI in medical diagnostics, enhanced accuracy is perceived as the most significant, followed by early detection and predictive analytics, with personalized medicine being considered the least important of the four.
Abbreviation: AI: Artificial intelligence.

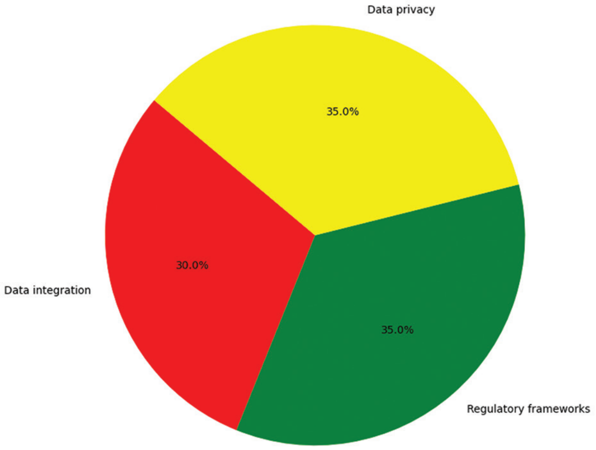


Figure 3. Challenges of AI in medical diagnostics
Abbreviation: AI: Artificial intelligence.

ML algorithms such as random forest, SVM, and logistic regression have also proven effective in disease prediction. In predicting type 2 diabetes (T2D), random forest classifiers achieved high accuracy based on lifestyle and health data, while mobile platforms leveraging random

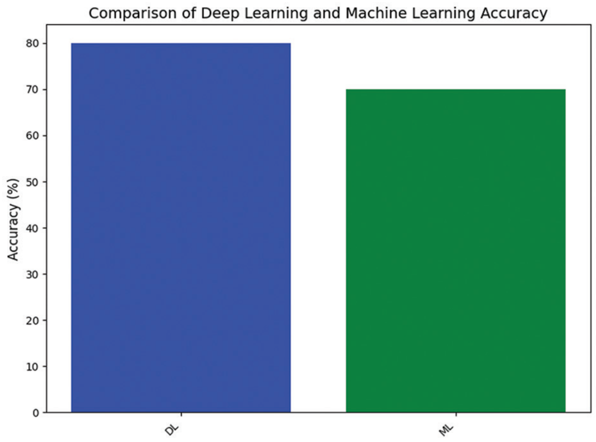


Figure 4. The accuracy of AI techniques in disease diagnosis. The bar chart illustrates the classification accuracy of two AI techniques: Deep learning and machine learning. The accuracy percentage for each technique is calculated as the ratio of correctly classified instances to the total number of instances in the dataset.
Notes: Classification Accuracy: This metric measures how often the AI model correctly predicts the class or category of a given data point. It's calculated by dividing the number of correct predictions by the total number of predictions. Percentage Derivation: In the context of the bar chart, the height of each bar represents the classification accuracy percentage. For instance, if the Deep Learning bar reaches 80 on the y-axis, it means the model achieved an 80% accuracy in classifying the data.
Abbreviations: AI: Artificial intelligence; DL: Deep learning; ML: Machine learning.

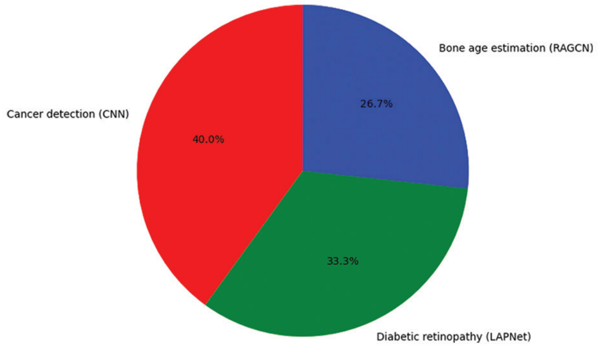


Figure 5. AI applications in medical imaging
Abbreviations: AI: Artificial intelligence; CNN: Convolutional neural networks; RAGCN: Region aggregation graph convolutional networks.

forest algorithms demonstrated over 98% accuracy in tuberculosis detection. RNN and LSTM networks further achieved approximately 97% accuracy in classifying gastrointestinal (GI) diseases, underscoring AI's potential in predictive diagnostics.

8.3. Integrating AI with IoT for healthcare advancements

A promising area of research combines AI with the IoT to enhance patient monitoring and disease prediction. IoT-based systems, using algorithms such as random forest, have been developed to monitor patient activities and predict health conditions in real time. One example is a hybrid IoT model that utilizes random forest techniques to predict T2D, demonstrating high predictive capability and aiding in the early intervention. AI-powered mobile platforms for real-time disease monitoring further improve patient care and support healthcare providers in managing complex cases effectively.

8.4. Superior performance of DL in medical image analysis

DL models, particularly through CNNs, have demonstrated remarkable performance in medical image analysis. CNNs have been effectively used to identify malaria-infected blood cells, while other ML models, including Naïve Bayes, SVM, and gradient boosting, have shown success in diagnosing various diseases.

For instance, a DL model called LAPNet, using a pyramid-based architecture and attention mechanisms, proved highly effective in detecting and grading diabetic retinopathy from medical images. Region aggregation graph convolutional networks have also been applied in medical imaging tasks, such as bone age estimation using CT and MRI scans, further enhancing diagnostic accuracy in radiology.

The research findings illustrate that DL methods generally outperform traditional ML models, especially when working with large datasets and complex medical images. DL techniques not only provide high diagnostic accuracy but also offer frameworks for real-time patient monitoring and tailored treatment options, paving the way for improved patient outcomes and healthcare efficiency. Through AI-driven medical diagnostics, healthcare providers can achieve early detection, personalized treatments, and better patient management, setting a new standard for care quality.

8.5. Challenges and future directions

While AI's potential in healthcare is immense, the field faces several challenges, including issues related to data privacy, algorithmic bias, and the need for diverse and high-quality datasets. In addition, ethical and regulatory challenges must be addressed to promote responsible AI use in healthcare.

Future research may focus on refining AI algorithms, improving multi-source data integration, and ensuring that AI applications in healthcare are equitable, reliable, and adaptable to a broad spectrum of healthcare settings.

9. Discussions and future directions

The findings from this exploration underline the transformative impact of AI on medical diagnostics, providing new pathways for enhancing diagnostic accuracy, sensitivity, and specificity across various disease categories, including cancer, cardiovascular, neurological, and infectious diseases. However, despite the promising results, the application of AI in healthcare faces several hurdles that require addressing to realize its full potential in clinical settings.

One of the major advantages highlighted in this study is AI's ability to outperform traditional diagnostic methods in early disease detection. CNN, for instance, demonstrated superior performance in identifying early-stage tumors, significantly improving sensitivity, which is critical for timely treatment.

Similarly, RNNs with LSTM networks have demonstrated effectiveness in predicting cardiovascular events by analyzing sequential health data to assess risks with greater precision. These findings reaffirm AI's capacity to process and analyze large-scale medical data more efficiently than traditional methods, thus supporting real-time, data-driven decisions in medical diagnosis.

However, a limitation observed was the variability in model performance across different diseases. While AI models displayed robust results in diagnosing certain diseases such as cancer and cardiovascular conditions, their accuracy was less consistent in neurological disorders, such as Alzheimer's and Parkinson's disease. The complexity and heterogeneity of neurological data present unique challenges, where nuanced structural and functional differences in the brain are more subtle and may be less easily detected by existing AI models. Addressing this limitation may require the development of advanced DL techniques, possibly involving multi-modal learning frameworks that integrate data from diverse sources such as brain imaging, genetic markers, and clinical observations, to capture the intricacies of neurological disorders more effectively.

Bias in AI diagnostics also emerged as a critical concern. The models demonstrated reduced sensitivity in certain demographic subgroups, notably in women and certain ethnic minorities in cardiovascular disease diagnosis. This discrepancy highlights the importance of using diverse and representative datasets to ensure AI solutions are equitable and fair. AI models trained predominantly on non-diverse datasets risk reinforcing healthcare disparities rather than reducing them. Future research should focus on strategies such as algorithmic fairness techniques, regularization, and the integration of balanced datasets to ensure AI tools perform consistently across all population segments.

In addition, the ethical and regulatory considerations of deploying AI in healthcare cannot be overlooked. While AI offers significant potential for enhancing diagnostic accuracy and efficiency, the risk of diagnostic errors, such as false positives or false negatives, remains a concern.

Such errors can have substantial consequences, particularly in critical disease areas where misdiagnosis can lead to unnecessary interventions or missed diagnoses. Stringent validation procedures, coupled with continuous monitoring of AI performance post-deployment, are essential to maintain the reliability and safety of AI-driven diagnostics. Moreover, regulatory frameworks, such as HIPAA and GDPR, must evolve to safeguard patient data and ensure AI diagnostics meet established safety standards. Future AI applications will need to prioritize privacy, with strict adherence to data protection laws and transparent data handling processes to ensure public trust in these technologies.

The incorporation of AI into healthcare ecosystems presents challenges but also offers unprecedented opportunities, especially in the context of infectious diseases. The rapid diagnostic capabilities of AI were highlighted during the COVID-19 pandemic, where AI-driven analysis of chest X-rays proved crucial in identifying cases quickly and accurately. This success exemplifies AI's potential in managing public health crises and reinforces its role in preparing for future pandemics, where rapid diagnostics and containment are critical.

In terms of future directions, the refinement of AI models for complex conditions, such as neurological and multi-organ diseases, will be essential for realizing AI's full diagnostic capabilities. Developing hybrid AI frameworks that leverage diverse data sources – from imaging and genomic data to patient history and wearable device data – could further enhance AI's diagnostic accuracy and applicability in personalized medicine. Furthermore, efforts must focus on improving transparency in AI algorithms, enabling clinicians and patients to understand how AI-derived diagnoses are made. Explainable AI techniques will play a crucial role in this, providing insights into model decision-making processes and fostering trust in AI-based diagnostics.

The future of AI in healthcare will also rely on the establishment of rigorous standards and collaborative efforts among researchers, healthcare providers, and regulatory bodies to ensure the responsible deployment of these technologies. This includes implementing continuous updates and recalibration of AI models as more diverse and high-quality datasets become available, thereby enhancing model robustness and adaptability to evolving healthcare needs. While AI has shown great potential in revolutionizing medical diagnostics, achieving widespread clinical integration will require addressing the challenges of model bias, ethical concerns, and regulatory compliance. Through ongoing advancements, refinement of AI methodologies, and adherence to ethical principles, AI-driven diagnostics can pave the way for a more accurate, efficient, and personalized approach to healthcare.

10. Conclusions

This research highlights the transformative impact of AI in medical diagnostics, particularly its ability to improve diagnostic accuracy, efficiency, and patient outcomes across multiple diseases, including cancer, cardiovascular conditions, neurological disorders, and infectious diseases. The effectiveness of models such as CNNs and RNNs in interpreting complex medical images and time-series data underlines AI's capacity to enhance – and in some cases surpass – traditional diagnostic methods. These findings support the integration of AI as a valuable diagnostic aid in clinical practice, where it can reduce human error, streamline workflows, and enable more data-driven decision-making. However, our research also underscores the challenges AI faces in diagnostics, especially when dealing with complex and heterogeneous conditions like neurological disorders. In addition, issues of bias within AI models – such as lower diagnostic accuracy in underrepresented populations – raise critical ethical concerns. Ensuring equitable access to AI-driven healthcare will require more diverse datasets, fairness algorithms, and efforts to mitigate these biases. Moreover, rigorous validation, regulatory oversight, and continuous monitoring of AI tools are essential to safeguard patient safety and maintain high standards in healthcare.

This study concludes that, while AI holds significant promise for revolutionizing medical diagnostics, realizing its full potential demands a balanced approach, integrating advanced technical development with ethical and regulatory considerations. Moving forward, AI-driven diagnostics can contribute to a more personalized, efficient, and equitable healthcare system, but ongoing research, collaboration, and responsible implementation are crucial. These findings emphasize AI's capacity to reshape medical diagnostics and support precision medicine, marking a significant step toward a more effective and inclusive healthcare future.

Acknowledgments

The author would like to acknowledge and thank the GOOGLE Deep Mind Research for providing access to pre-prints platforms. This research exploration was investigated using the platform provided by GOOGLE Deep Mind which is under the support of the GOOGLE Research and the GOOGLE Research Publications within the GOOGLE Gemini platform.

Funding

None.

Conflict of interest

The author declares no competing interests for this research.

Author contributions

This is a single-authored article.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data

The data used in this study can be obtained from the references as indicated in this article.

References

1. *Developing an Aging Clock Using deep Learning on Retinal Images*; 2023. Available from: https://research.google/blog/developing-an-aging-clock-using-deep-learning-on-retinal-images [Last accessed on 2024 Dec 18].
2. Akhtar ZB. Generative artificial intelligence (GAI): From large language models (LLMs) to multimodal applications towards fine tuning of models, implications, investigations. *Comput Artif Intell*. 2024;3(1):1498.

   doi: 10.59400/cai.v3i1.1498
3. Jimma BL. Artificial intelligence in healthcare: A bibliometric analysis. Telematics Inform Rep. 2023;9:100041.

   doi: 10.1016/j.teler.2023.100041
4. Akhtar Z. Computer vision and beyond: A deep dive exploration and investigation. *Trends Tech Sci Res*. 2024;7(3):555711.

   doi: 10.19080/TTSR.2024.07.555711
5. Lyakhova U, Lyakhov P. Systematic review of approaches to detection and classification of skin cancer using artificial intelligence: Development and prospects. *Comput Biol Med*. 2024;178:108742.

   doi: 10.1016/j.compbiomed.2024.108742
6. Sankar H, Alagarsamy R, Lal B, *et al*. Role of artificial intelligence in treatment planning and outcome prediction of jaw corrective surgeries by using 3-D imaging- A systematic review. *Oral Surg Oral Med Oral Pathol Oral Radiol*. 2024.

   doi: 10.1016/j.oooo.2024.09.010
7. Qiu Y, Cheng F. Artificial intelligence for drug discovery and

development in Alzheimer's disease. *Curr Opin Struct Biol*. 2024;85:102776.

doi: 10.1016/j.sbi.2024.102776

8. Ghanem M, Ghaith AK, Bydon M. Artificial intelligence and personalized medicine: transforming patient care. In: New Era of Precision Medicine. Netherlands: Elsevier; 2024. p. 131-42.

   doi: 10.1016/b978-0-443-13963-5.00012-1

9. Snider V, Homsi K, Kusnoto B, *et al*. Clinical evaluation of artificial intelligence driven remote monitoring technology for assessment of patient oral hygiene during orthodontic treatment. *Am J Orthodont Dentofacial Orthop*. 2024;165(5):586-592.

   doi: 10.1016/j.ajodo.2023.12.008

10. Akhtar ZB, Rawol AT. Artificial intelligence (AI), extended reality (XR): A biomedical engineering (BME) perspective investigation analysis. *RPC J Bio Med Clin Res*. 2024;6(3).

11. Akhtar ZB. Operating systems (OS): An insight investigative research analysis and future directions. *J Technol Informat* (*JoTI*). 2024;6(1):58-69.

    doi: 10.37802/joti.v6i1.637

12. Akhtar ZB, Rawol AT. Harnessing artificial intelligence (AI) for cybersecurity: Challenges, opportunities, risks, future directions. *Comput Artif. Intell*. 2024;2(2):1485.

    doi: 10.59400/cai.v2i2.1485

13. Alowais SA, Alghamdi SS, Alsuhebany N, *et al*. Revolutionizing healthcare: The role of artificial intelligence in clinical practice. *BMC Med Educ*. 2023;23:689.

    doi: 10.1186/s12909-023-04698-z

14. Akhtar ZB, Tajbiul Rawol A. Enhancing cybersecurity through AI-powered security mechanisms. *IT J Res Dev*. 2024;9(1):50-67.

    doi: 10.25299/itjrd.2024.16852

15. Li Q, Yang X, Xu J, *et al*. Early prediction of Alzheimer's disease and related dementias using real-world electronic health records. *Alzheimers Dement*. 2023;19(8):3506-3518.

    doi: 10.1002/alz.12967

16. Ayers JW, Poliak A, Dredze M, *et al*. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. *JAMA Intern Med*. 2023;183(6):589-596.

    doi: 10.1001/jamainternmed.2023.1838

17. Akhtar ZB, Stany Rozario V. The Design Approach of An Artificial Human Brain in Digitized Formulation Based on ML and Neural Mapping. In: *2020 International Conference for Emerging Technology (INCET)*, Belgaum, India; 2020. p. 1-7.

    doi: 10.1109/INCET49848.2020.9154000

18. Bin Akhtar Z. Accelerated computing a biomedical engineering and medical science perspective. *Ann Acad Rom Sci Se Biol Sci*. 2023;12:138-164.

    doi: 10.56082/annalsarscibio.2023.2.138

19. Akhtar ZB, Gupta AD. Advancements within molecular engineering for regenerative medicine and biomedical applications an investigation analysis towards a computing retrospective. *J Electron Electromed Eng Med Inform*. 2024;6(1):54-72.

    doi: 10.35882/jeeemi.v6i1.351

20. Akhtar ZB, Gupta AD. Integrative approaches for advancing organoid engineering: From mechanobiology to personalized therapeutics. *J Appl Artif Intell*. 2024;5(1):1-27.

    doi: 10.48185/jaai.v5i1.974

21. Akhtar ZB. The design approach of an artificial intelligent (AI) medical system based on electronical health records (EHR) and priority segmentations. *J Eng*. 2024;2024:1-10.

    doi: 10.1049/tje2.12381

22. Akhtar ZB. Exploring biomedical engineering (BME): Advances within accelerated computing and regenerative medicine for a computational and medical science perspective exploration analysis. *J Emerg Med OA*. 2024;2(1):1-23.

23. Akhtar Z. Biomedical engineering (BME) and medical health science: AN investigation perspective exploration. *Quantum J Med Health Sci*. 2024;3(3):1-24.

24. Akhtar ZB, Tajbiul Rawol A. Unlocking the future for the new data paradigm of DNA data storage: An investigative analysis of advancements, challenges, future directions. *J Info Sci*. 2024;23(1):23-44.

    doi: 10.34874/IMIST.PRSM/jis-v23i1.47102

25. Akhtar ZB, Tajbiul Rawol A. Transformative impacts of artificial intelligence (AI) and extended reality (XR) on biomedical engineering (BME): Innovations, implications, challenges, opportunities. *J Adv Artif Intell*. 2024;2(2):60-78.

26. Akhtar ZB. Unveiling the evolution of generative AI (GAI): A comprehensive and investigative analysis toward LLM models (2021-2024) and beyond. *J Electr Syst Inf Technol* 2024;11:22.

    doi: 10.1186/s43067-024-00145-1

27. Bin Akhtar Z. From bard to Gemini: An investigative exploration journey through Google's evolution in conversational AI and generative AI. *Comput Artif Intell*. 2024;2(1):1378.

    doi: 10.59400/cai.v2i1.1378

28. Bin Akhtar Z. Unraveling the promise of computing DNA data storage: An investigative analysis of advancements, challenges, future directions. *J Adv Artif Intell*. 2024;2(1):122-137.

29. Akhtar ZB. Securing operating systems (OS): A comprehensive

approach to security with best practices and techniques. *Int J Adv Netw Monit Controls*. 2024;9(1):100-111.

doi: 10.2478/ijanmc-2024-0010

30. Akhtar ZB, Rawol AT. Uncovering cybersecurity vulnerabilities: A kali linux investigative exploration perspective. *Int J Adv Netw Monit Controls*. 2024;9(2):11-22.

    doi: 10.2478/ijanmc-2024-0012

31. Bin Akhtar Z. Artificial intelligence (AI) within manufacturing: An investigative exploration for opportunities, challenges, future directions. *Metaverse*. 2024;5(2):2731.

    doi: 10.54517/m.v5i2.2731

32. Vayadande K, Bhosle AA, Pawar RG, Joshi DJ, Bailke PA, Lohade O. Innovative approaches for skin disease identification in ML: A comprehensive study. *Oral Oncol Rep*. 2024;10:100365-100365.

    doi: 10.1016/j.oor.2024.100365

33. Lai C, Fuggle NR, Matin RN, Tanaka RJ, Rajan N. Artificial intelligence and ML in dermatological research and healthcare: British society for investigative dermatology skin club Report-Southampton April 2024. *Br J Dermatol*. 2024;192:118-124.

    doi: 10.1093/bjd/ljae395

34. Liu X, Duan C, Kim MK, Zhang L, *et al*. *Comparative Performance of Claude-3 Opus and ChatGPT-4 in Dermoscopic Image Analysis for Melanoma Diagnosis*. JMIR Medical Informatics [Preprint]; 2024.

    doi: 10.2196/59273

35. Gassner M, Garcia JB, Tanadini-Lang S, *et al*. Saliency-enhanced content-based image retrieval for diagnosis support in dermatology consultation: Reader study. *JMIR Dermatol*. 2023;6:e42129-e42129.

    doi: 10.2196/42129

36. Komuro J, Kusumoto D, Hashimoto H, Yuasa S. Machine learning in cardiology: Clinical application and basic research. *J Cardiol*. 2023;82(2):128-133.

    doi: 10.1016/j.jjcc.2023.04.020

37. Papachristou P, Söderholm M, Pallon J, *et al*. Evaluation of an artificial intelligence-based decision support for detection of cutaneous melanoma in primary care-a prospective, real-life, clinical trial. *Br J Dermatol*. 2024;191:125-133.

    doi: 10.1093/bjd/ljae021

38. Naser MA, Majeed AA, Alsabah M, Al-Shaikhli TR, Kaky KM. A review of ML's role in cardiovascular disease prediction: Recent advances and future challenges. *Algorithms*. 2024;17(2):78.

    doi: 10.3390/a17020078

39. Selvi DMK, Aswini J, Balakrishnan C, Suganya K, Sheena BG, Subramanian RS. Revolutionizing cardiovascular care: The role of AI, ML, and DL in early heart disease prediction and treatment. In: *2024 International Conference on Electronics, Computing, Communication and Control Technology (ICECCC)*, Bengaluru, India; 2024. p. 1-6.

    doi: 10.1109/ICECCC61767.2024.10593941

40. He B, Kwan AC, Cho JH, *et al*. Blinded, randomized trial of sonographer versus AI cardiac function assessment. *Nature*. 2023;616(7957):520-524.

    doi: 10.1038/s41586-023-05947-3

41. Deo RC. Artificial intelligence and ML in cardiology. *Circulation*. 2024;149(16):1235-1237.

    doi: 10.1161/circulationaha.123.065469

42. Bax M, Thorpe J, Romanov V. The future of personalized cardiovascular medicine demands 3D and 4D printing, stem cells, and artificial intelligence. Front Sens. 2023;4.

    doi: 10.3389/fsens.2023.1294721

43. Alabdaljabar MS, Hasan B, Noseworthy PA, Maalouf JF, Ammash NM, Hashmi SK. Machine learning in cardiology: A potential real-world solution in low-and middle-income countries. *J Multidiscip Healthc*. 2023;16:285-295.

    doi: 10.2147/jmdh.s383810

44. Rodríguez M, Córdova C, Benjumeda I, San Martín S. Automated cervical cancer screening using single-cell segmentation and deep learning: Enhanced performance with liquid-based cytology. *Computation*. 2024;12:232.

    doi: 10.3390/computation12120232

45. Vaickus LJ, Kerr DA, Torres JMV, Levy J. Artificial intelligence applications in cytopathology: Current state of the art. *Surg Pathol Clin*. 2024;17:521-531.

    doi: 10.1016/j.path.2024.04.011

46. Wu A, Ngo M, Thomas C. Assessment of patient perceptions of artificial intelligence use in dermatology: A cross-sectional survey. *Skin Res Technol*. 2024;30(3):e13656.

    doi: 10.1111/srt.13656

47. Tan IJ, Podwojniak A, Parikh A, Cohen BA. Precision dermatology: A review of molecular biomarkers and personalized therapies. *Curr Issues Mol Biol*. 2024; 46(4):2975-2990.

    doi: 10.3390/cimb46040186

48. McMullen E, Grewal R, Storm K, *et al*. Diagnosing contact dermatitis using machine learning: A review. *Contact Dermatitis*. 2024;91(3):186-189.

    doi: 10.1111/cod.14595

49. Kaushik P, Chopra Y, Kajla A, Poonia M, Khan A, Yadav D. AI-Powered Dermatology: Achieving Dermatologist-Grade Skin Cancer Classification. In: *2024 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI)*. Vol. 2.

United States: IEEE; 2024. p. 1-6.

50. Gordon ER, Trager MH, Kontos D, *et al*. Ethical considerations for artificial intelligence in dermatology: A scoping review. *Br J Dermatol.* 2024;190:789-797.

    doi: 10.1093/bjd/ljae040

51. Yadav R, Bhat A. A Survey on Skin Lesion Detection and Classification using ML. In: *2024 2nd International Conference on Artificial Intelligence and ML Applications Theme: Healthcare and Internet of Things (AIMLA)*. United States: IEEE; 2024. p. 1-5.

    doi: 10.1109/aimla59606.2024.10531571

52. Klarycki J, Tomczyk K, Podgórska D, *et al*. AI in dermatology: Bridging the gap between potential and practice. *J Educ Health Sport*. 2024;52:57-73.

    doi: 10.12775/jehs.2024.52.004

53. Goshisht MK. Machine learning and deep learning in synthetic biology: Key architectures, applications, and challenges. *ACS omega*. 2024;9(9):9921-9945.

    doi: 10.1021/acsomega.3c05913

54. Noordijk B, Garcia Gomez ML, Ten Tusscher KH, De Ridder D, Van Dijk AD, Smith RW. The rise of scientific machine learning: A perspective on combining mechanistic modelling with machine learning for systems biology. *Front Syst Biol*. 2024;41:407994.

    doi: 10.3389/fsysb.2024.1407994

55. Artificial Intelligence: 10 Promising Interventions for Healthcare. NIHR Evidence. National Institute for Health and Care Research; 2023.

    doi: 10.3310/nihrevidence_59502

56. Iacucci M, Parigi TL, Del Amor R, *et al*. Artificial intelligence enabled histological prediction of remission or activity and clinical outcomes in ulcerative colitis. *Gastroenterology*. 2023;164(7):1180-1188.e2.

    doi: 10.1053/j.gastro.2023.02.031

57. Kim HY, Cho GJ, Kwon HS. Applications of artificial intelligence in obstetrics. Ultrasonography. 2023;42(1):2-9.

    doi: 10.14366/usg.22063

58. Saraf S, De A, Tripathy BK. Effective use of computational biology and artificial intelligence in the domain of medical oncology. In: *Computational Intelligence for Oncology and Neurological Disorders*. United States: CRC Press; 2024. p. 228-252.

59. Farhat F, Sohail SS, Alam MT, *et al*. COVID-19 and beyond: leveraging artificial intelligence for enhanced outbreak control. *Front Artif Intell*. 2023;6:1266560.

    doi: 10.3389/frai.2023.1266560

60. Catacutan DB, Alexander J, Arnold A, Stokes JM. Machine learning in preclinical drug discovery. *Nat Chem Biol.* 2024;20(8):960-973.

    doi: 10.1038/s41589-024-01679-1

61. AI Could Make Health Care Fairer-by Helping us believe what Patients Say. MIT Technology Review; 2023.

62. Udegbe FC, Ebulue OR, Ebulue CC, Ekesiobi CS. Synthetic Biology and its Potential in US Medical Therapeutics: A Comprehensive Review: Exploring the Cutting-edge Intersections of Biology and Engineering in Drug Development and Treatments. *Eng Sci Technol J.* 2024;5(4):1395-1414.

63. Stevens S, Wu J, Thompson MJ, *et al*. Bioclip: A Vision Foundation Model for the Tree of Life. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*; 2024. p. 19412-19424.

64. Jeong H, Park S, Choi B, Yu CS, Hong JY, Jeong TY, Cho KH. Machine learning-based water quality prediction using octennial *in-situ* Daphnia magna biological early warning system data. *J Hazardous Mater.* 2024;465:133196.

    doi: 10.1016/j.jhazmat.2023.133196

65. Rood JE, Hupalowska A, Regev A. Toward a foundation model of causal cell and tissue biology with a perturbation cell and tissue atlas. *Cell.* 2024;187(17):4520-4545.

66. Shaban AS, Owda ME, Basuoni MM, Mousa MA, Radwan AA, Saleh AK. *Punica granatum* peel extract mediated green synthesis of zinc oxide nanoparticles: Structure and evaluation of their biological applications. *Biomass Convers Biorefin.* 2024;14(11):12265-12281.

    doi: 10.1007/s13399-022-03185-7

67. DeGroat W, Abdelhalim H, Patel K, Mendhe D, Zeeshan S, Ahmed Z. Discovering biomarkers associated and predicting cardiovascular disease with high accuracy using a novel nexus of machine learning techniques for precision medicine. *Sci Rep.* 2024;14(1):1.

    doi: 10.1038/s41598-023-50600-8

68. Martin SS, Aday AW, Almarzooq ZI, *et al*. 2024 heart disease and stroke statistics: A report of US and global data from the American Heart Association. *Circulation*. 2024;149(8):e347-e913.

    doi: 10.1161/CIR.0000000000001209

69. Ogunpola A, Saeed F, Basurra S, Albarrak AM, Qasem SN. Machine Learning-based predictive models for detection of cardiovascular diseases. *Diagnostics*. 2024;14(2):144.

    doi: 10.3390/diagnostics14020144

70. Ogunpola A, Saeed F, Basurra S, Albarrak AM, Qasem SN. Machine learning-based predictive models for detection of cardiovascular diseases. *Diagnostics*. 2024;14(2):144.

    doi: 10.3390/diagnostics14020144

71. Terranova N, Venkatakrishnan K. Machine learning in modeling disease trajectory and treatment outcomes: An

emerging enabler for model-informed precision medicine. *Clin Pharmacol Therapeut*. 2024;115(4):720-726.

doi: 10.1002/cpt.3153

72. Tayebi Arasteh S, Han T, Lotfinia M, *et al*. Large language models streamline automated machine learning for clinical studies. *Nat Commun*. 2024;15(1):1603.

    doi: 10.1038/s41467-024-45879-8

73. Hussain I, Nazir MB. Empowering healthcare: AI, ML, and deep learning innovations for brain and heart health. *Int J Adv Eng Technol Innov.* 2024;1(4):167-188.

74. Khandakar S, Al Mamun MA, Islam MM, Hossain K, Melon MMH, Javed MS. Unveiling early detection and prevention of cancer: Machine learning and deep learning approaches. *Educ Adm Theory Pract*. 2024;30(5):14614-14628.

    doi: 10.53555/kuey.v30i5.7014

75. Bintintan V, Burz C, Pintea I, *et al*. The importance of extracellular vesicle screening in gastric cancer: A 2024 update. *Cancers* (*Basel*). 2024;16(14):2574.

    doi: 10.3390/cancers16142574

76. Gontero P, Birtle A, Capoun O, *et al*. European association of urology guidelines on non-muscle-invasive bladder cancer (TaT1 and carcinoma *in situ*)-a summary of the 2024 guidelines update. *Eur Urol*. 2019;76:639-657.

    doi: 10.1016/j.eururo.2019.08.016

77. Bin Akhtar Z. Artificial intelligence (AI) within the realm of cyber security. *Insight Electr Electron Eng.* 2024;1(1):1-11.

78. Bin Akhtar Z, Rawol AT. A biomedical engineering (BME) perspective investigation analysis: Artificial intelligence (AI) and extended reality (XR). *Eng Adv.* 2024;4(3):143-154.

    doi: 10.26855/ea.2024.07.005

79. Leung JH, Karmakar R, Mukundan A, *et al*. Systematic meta-analysis of computer-aided detection of breast cancer using hyperspectral imaging. *Bioeng* (*Basel)*. 2024;11(11):1060.

    doi: 10.3390/bioengineering11111060

80. Shanthakumar D, Leiloglou M, Kelliher C, Darzi A, Elson DS, Leff DR. A comparison of spectroscopy and imaging techniques utilizing spectrally resolved diffusely reflected light for intraoperative margin assessment in breast-conserving surgery: A systematic review and meta-analysis. *Cancers*. 2023;15(11):2884.

    doi: 10.3390/cancers15112884

ACCSCIENCE PUBLISHING

Artificial Intelligence in Health

ORIGINAL RESEARCH ARTICLE

Optimized convolutional neural network model for multilevel classification in leukemia diagnosis using Tversky loss

Kumari Pritee[1]* **and Rahul Dev Garg**[2]

[1]Department of Information System Management, IIM Sambalpur, Sambalpur, India

[2]Department of Geomatics Engineering, IIT Roorkee, Roorkee, India

Abstract

Leukemia diagnosis traditionally depends on time-intensive examination of blood cell morphology, a process prone to human error. To address these challenges, this study explores the use of convolutional neural networks (CNNs) optimized with the Tversky loss function for automated, multilevel image classification in leukemia diagnostics. The model was designed to tackle binary classification for distinguishing normal from abnormal cells, and multiclass classification for identifying leukemia subtypes, while addressing the challenges of imbalanced datasets inherent in medical imaging. Trained on publicly available leukemia image datasets, the CNN achieved high accuracy in both tasks, effectively capturing subtle morphological variations critical for precise diagnosis. By incorporating performance metrics such as accuracy, precision, and recall, the study highlights the model's reliability and robustness across classification tasks. The findings underscore the potential of CNN-based tools in enhancing diagnostic accuracy and efficiency, paving the way for future innovations in leukemia diagnostics and broader medical imaging applications.

Keywords: Multilevel classification; Deep learning; Leukemia; Convolutional neural networks; Medical image analysis; Automated diagnosis

***Corresponding author:**
Kumari Pritee
(preetik@iimsambalpur.ac.in)

Citation: Pritee K, Garg RD. Optimized convolutional neural network model for multilevel classification in leukemia diagnosis using Tversky loss. *Artif Intell Health*. 2025;2(3):63-76. doi: 10.36922/aih.4710

Received: August 30, 2024

1st revised: September 27, 2024

2nd revised: October 28, 2024

3rd revised: December 10, 2024

4th revised: December 31, 2024

5th revised: January 3, 2025

Accepted: January 8, 2025

Published online: January 22, 2025

Copyright: © 2025 Author(s). This is an Open-Access article distributed under the terms of the Creative Commons Attribution License, permitting distribution, and reproduction in any medium, provided the original work is properly cited.

Publisher's Note: AccScience Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

1.1. Background and motivation

Leukemia, a type of cancer affecting blood and bone marrow, requires timely and accurate diagnosis for effective treatment. However, traditional diagnostic methods such as microscopic examination are time-consuming, labor-intensive, and prone to human error. With the increasing volume of medical imaging data, there is a growing need for automated diagnostic tools that can enhance both the speed and accuracy of leukemia diagnosis. This paper aims to address these challenges by leveraging deep learning (DL) techniques, specifically convolutional neural networks (CNNs), for the multilevel classification of leukemia cells, offering a more reliable and efficient diagnostic solution.

1.2. Contributions

The main contributions of this work are threefold. First, we introduce a multilevel classification framework for leukemia diagnosis, which uses CNNs optimized with the

Tversky loss function. This approach enables the model to differentiate between normal and abnormal cells as well as subclassify various leukemia types with high precision. Second, our proposed methodology specifically addresses the challenge of imbalanced datasets, a common issue in medical imaging, by employing the Tversky loss function to improve classification performance. Finally, we rigorously evaluate the model using publicly available leukemia datasets, demonstrating its superior performance in terms of accuracy, precision, and recall when compared to traditional methods and other DL models.

In this study, we utilized publicly available leukemia datasets to train and evaluate our DL models. We assessed the performance of these models in terms of accuracy, sensitivity, specificity, and computational efficiency. The results of this study demonstrated the potential of multilevel image classification using DL to significantly improve the diagnostic process for leukemia, paving the way for more accurate and timely interventions in clinical practice.

Figure 1 shows a classification diagram of different types of leukemia, which is divided into two major categories: acute leukemia and chronic leukemia. Acute leukemia is further split into acute myeloid leukemia (AML) and acute lymphoblastic leukemia (ALL). AML has several subtypes, including M0 (undifferentiated AML), M1 (AML without maturation), M2 (AML with maturation), M3 (acute promyelocytic leukemia), M4 (acute myelomonocytic leukemia), M5 (acute monocytic leukemia), M6 (erythroleukemia), and M7 (acute megakaryoblastic leukemia). Similarly, ALL is divided into B-cell ALL and T-cell ALL.

On the other hand, chronic leukemia is broken down into chronic lymphocytic leukemia (CLL) and chronic myeloid leukemia (CML). CLL is associated with small lymphocytic lymphoma (SLL), while CML is presented with phases of disease progression such as the chronic phase, accelerated phase, and blast crisis phase. This diagram visually organizes leukemia subtypes, showing how they fit into the broader categories of acute and chronic leukemias.

By addressing the challenges associated with traditional diagnostic methods and leveraging the power of DL, this

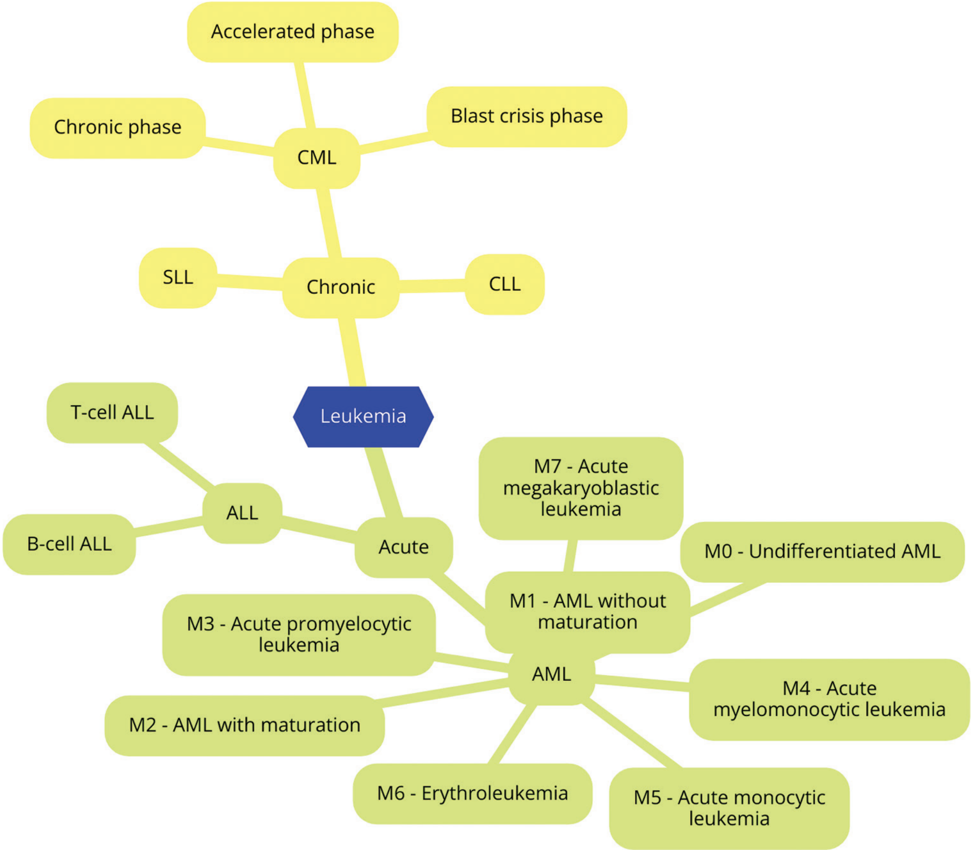


Figure 1. Categorization of different types of leukemia: (A) acute leukemia and its subtypes, including acute myeloid leukemia and acute lymphoblastic leukemia, with detailed divisions; (B) chronic leukemia and its subtypes, including chronic myeloid leukemia and chronic lymphocytic leukemia, along with disease progression phases.

research contributes to the advancement of automated medical image analysis, ultimately aiming to improve patient care and outcomes in leukemia treatment. This paper explores the role of multilevel image classification using DL, specifically focusing on the C-NMC dataset.

1.3. Organization

The paper is organized as follows: Section 2 provides a comprehensive review of related work, highlighting key advancements in DL for medical image analysis. Section 3 describes the proposed methodology, detailing the CNN architecture, the use of the Tversky loss function, and the training process. Section 4 presents the experimental results, including a performance comparison with other state-of-the-art models. Finally, section 5 concludes the paper, discussing the implications of our findings and potential directions for future research.

2. Related work

2.1. Review of DL in medical image analysis

The application of DL in medical image analysis has garnered significant attention in recent years, with numerous studies demonstrating its potential to transform diagnostics. This section reviews existing literature on the use of DL for leukemia detection and classification, highlighting key methodologies, findings, and gaps that this study aims to address. A key strength of this approach lies in the model's incorporation of multi-scale features, a concept widely recognized for enhancing performance in other areas of computer vision. Multi-scale features have shown considerable utility in tasks such as image quality assessment, visual saliency detection, and person re-identification. For example, Varga[1] illustrated the effectiveness of multi-scale orderless pooling of deep features for no-reference image quality assessment, emphasizing the role of feature pooling at different scales to capture both global and local image characteristics. Similarly, Li and Yu[2] demonstrated the value of multiscale deep features in visual saliency detection, where analyzing features at multiple scales helped detect visually significant regions across images. In person re-identification, Chen *et al.*[3] utilized multi-scale DL architectures to improve recognition performance by mapping features across different views and scales, highlighting the versatility of this approach.

In the proposed CNN model for leukemia diagnosis, multi-scale features are leveraged to detect subtle morphological differences between normal and leukemic cells. This is critical for accurate subclassification of leukemia types, as cell morphology can vary significantly between different subtypes. By integrating multi-scale feature extraction with the CNN architecture, the model becomes more adept at recognizing variations across different resolutions, enabling it to handle complex and imbalanced medical datasets effectively.

Furthermore, the use of the Tversky loss function in this model addresses one of the core challenges in medical image classification – class imbalance. Medical datasets, particularly those used for leukemia diagnosis, often suffer from an uneven distribution of samples across different categories. The Tversky loss function, which adjusts the trade-off between false positives (FP) and false negatives (FN), ensures that the model remains sensitive to minority classes, improving overall performance on imbalanced datasets. This is particularly important in the clinical context, where minimizing FN is crucial for early detection and treatment planning.

Overall, the combination of CNN architecture, multi-scale feature extraction, and the Tversky loss function presents a robust solution for multilevel classification in leukemia diagnosis. By incorporating techniques proven successful in other areas of computer vision, this model sets a new benchmark for automated medical image analysis, offering enhanced diagnostic accuracy and reliability.

2.2. Existing solutions for leukemia diagnosis and gaps addressed by this study

Table 1 presents a detailed literature review table summarizing the related work involving DL, the C-NMC dataset, and leukemia, including author details for 15 studies from 2020 to 2024.

Figure 2 presents a structured object-oriented model for leukemia classification, breaking it down into acute and chronic forms, with subtypes such as myeloid and lymphocytic variants. One of the key advantages of this structure is that it facilitates systematic data representation, making it easier for machine learning (ML) or DL models to process and identify patterns across the subtypes. This model provides a clear hierarchy that can be used to label and categorize medical data, improving the efficiency and accuracy of automated disease detection systems.

In addition, by organizing the leukemia subtypes into a hierarchical object-oriented structure, the relationships between different types are easier to understand and manage. This kind of classification allows for a more detailed and accurate analysis of blood samples, which can aid in the differentiation of leukemia types during the diagnostic process. It is particularly useful when dealing with large datasets, where the defined subtype structure ensures that various forms of leukemia are appropriately identified, facilitating early diagnosis and targeted treatment plans.

Table 1. Literature review table summarizing work involving deep learning, the C-NMC dataset, and leukemia

Title	Authors	Year	Methodology	Technology	Conclusions
Leukemia classification using the deep learning method of CNN	Arivuselvam and Sudha[4]	2022	ResNet-34 and DenseNet-121 architectures for leukemia type classification	DCNN	High accuracy in low-intensity images classification
Leukemia Classification using a Convolutional Neural Network of AML Images	Kadhim *et al.*[5]	2023	CNN achieving over 98% classification accuracy	CNN	Demonstrates the potential of CNN for multilevel leukemia classification
Machine learning in detection and classification of leukemia using C-NMC_Leukemia	Talaat and Gamel[6]	2023	Image preprocessing, feature extraction, and classification with fuzzy optimization	Optimized CNN	Achieved 99.99% accuracy in classifying microscopic images
Optimizing a Deep Residual Neural Network with Genetic Algorithm for Acute Lymphoblastic Leukemia Classification	Rodrigues *et al.*[7]	2022	Genetic algorithm combined with ResNet-50V2 for hyperparameter optimization	Hybrid CNN with GA	Achieved 98.46% accuracy, showcasing potential for accurate leukemia diagnosis
Convergent learning–based model for leukemia classification from gene expression	Mallick *et al.*[8]	2020	Five-layer DNN classifier for gene expression data	DNN	High accuracy in leukemia classification using gene expression datasets
Automatic Detection of Leukemia through Convolutional Neural Network	Arif *et al.*[9]	2022	Modified CNN model for data augmentation, segmentation, and classification	CNN	High accuracy and reliability, suitable for clinical applications
A hybrid detection model for acute lymphocytic leukemia using SVM-PSO	Alsaykhan and Maashi[10]	2024	Hybrid detection model using SVM and particle swarm optimization	SVM-PSO	High accuracy for acute lymphocytic leukemia detection
Ensemble learning using Gompertz function for leukemia classification	Abhishek *et al.*[11]	2025	Ensemble learning framework leveraging Gompertz function	Ensemble learning	Improved classification accuracy for leukemia diagnosis
Deep Transfer Learning in Diagnosing Leukemia in Blood Cells	Loey *et al.*[12]	2020	Transfer learning-based framework	Transfer learning	Accurate blood cell leukemia diagnosis
Navigating Tversky Loss Function Hyperparameter Spaces using Particle Swarm Optimization	Damit *et al.*[13]	2024	Optimizing hyperparameters of Tversky loss for segmentation	Particle swarm optimization	Enhanced segmentation performance
Segmentation and classification of white blood smear images using modified CNN architecture	Kumar and Rawat[14]	2024	Modified CNN for segmentation and classification	Modified CNN	Accurate classification of white blood smear images
Machine Learning Applications in the Diagnosis of Benign and Malignant Hematological Diseases	Muhsen *et al.*[15]	2020	ML applications for hematological disease diagnosis	ML Framework	Effective for benign and malignant hematological disease diagnosis
Explainable AI identifies diagnostic cells of genetic AML subtypes	Hehr *et al.*[16]	2023	Explainable AI framework for identifying diagnostic cells	Explainable AI	High precision in identifying AML subtypes
Hyperparameter Optimization of a Convolutional Neural Network Model for Pipe Burst Location	Antunes *et al.*[17]	2023	CNN hyperparameter optimization using various search techniques	Optimized CNN	Enhanced generalization across applications
Metalearning approach for leukemia informative genes prioritization	Rodrigues and Deusdado[18]	2020	Meta-learning framework for gene prioritization	Meta-learning	Effective in identifying leukemia-informative genes

(Cont'd...)

Table 1. ***(Continued)***

Title	Authors	Year	Methodology	Technology	Conclusions
Automatic Image Dataset Construction from Click-through Logs Using Deep Neural Network	Bai *et al.*[19]	2015	Dataset construction using DNNs	DNN	Simplified data preparation process
Utilizing Deep Feature Fusion for Automatic Leukemia Classification	Islam *et al.*[20]	2024	Deep feature fusion within an IoMT-enabled DL framework	DL with IoMT	Robust automatic leukemia classification
Customized Deep Learning Classifier for Detection of Acute Lymphoblastic Leukemia Using Blood Smear Images	Sampathila *et al.*[21]	2022	Customized DL classifiers	Customized DL Framework	Reliable detection of ALL using blood smear images

Abbreviations: AI: Artificial intelligence; ALL: Acute lymphoblastic leukemia; AML: Acute myeloid leukemia; CNN: Convolutional neural network; DCNN: Deep convolutional neural networks; DL: Deep learning; DNN: Deep neural network; GA: Genetic algorithm; IoMT: Internet of medical things; ML: Machine learning; PSO: Particle swarm optimization; SVM: Support vector machine.

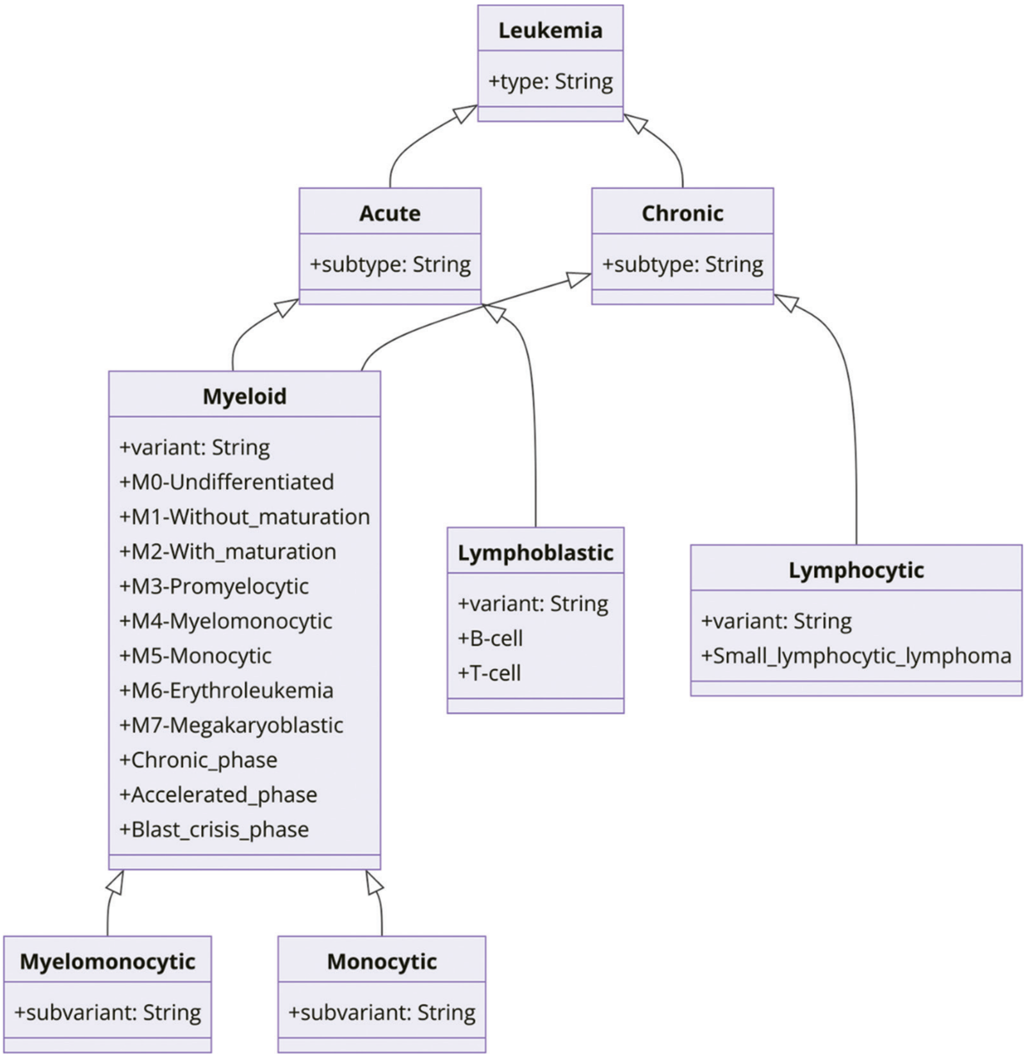


Figure 2. A UML diagram of detailed classification of leukemia depicting the object-oriented representations of acute leukemia subtypes and chronic leukemia subtypes.

2.2.1. C-NMC dataset

The C-NMC (The Cancer Genome Atlas) dataset is a crucial resource for developing and evaluating DL models for leukemia classification. This section outlines the specific methodology employed to leverage the C-NMC dataset for multilevel image classification using DL techniques.[22]

2.2.2. Dataset description

The C-NMC dataset comprises a comprehensive collection of blood smear images, annotated with labels indicating various types of leukemia, including ALL and AML. The dataset provides a robust foundation for training and testing DL models aimed at automating leukemia diagnosis.

The C-NMC dataset as shown in Figure 3 contains a total of 10,000 images, evenly divided between healthy and malignant cells. Each image has associated metadata, including patient ID, sample ID, age, gender, diagnosis, and slide details. In addition, the dataset is supposed to be split into a training set with 8,000 images and a testing set with 2,000 images.

- Total images: 10,000
- Healthy cells: 5,000
- Malignant cells: 5,000
- Metadata entries: 10,000 (one for each image)
- Training set: 8,000 images
- Testing set: 2,000 images

Understanding the number of entries in the C-NMC dataset helps researchers and practitioners gauge the dataset's size, diversity, and suitability for training and testing ML models for the classification of bone marrow cells. It provides insight into the dataset's comprehensiveness and potential for developing robust and accurate algorithms for medical image analysis.

3. Proposed methodology

3.1. CNN model architecture

This paper presents a multilevel image classification method using DL for leukemia datasets. The proposed CNN model (customized CNN model optimized by Tversky loss function) with multiple convolutional and dense layers optimized with a Tversky loss function achieves high accuracy and robustness, demonstrating its potential for aiding in the early diagnosis of leukemia. Customized CNN is used for handling imbalanced datasets (i.e., different proportions of healthy vs. malignant cells).

The proposed model utilizes a CNN, specifically designed and optimized for multilevel classification tasks on the C-NMC leukemia dataset. The architecture consists of several convolutional layers to extract hierarchical features from the input images followed by pooling layers to reduce dimensionality. Batch normalization is applied to enhance the model's stability, while rectified linear unit (ReLU) activation functions are employed to introduce non-linearity.

After the convolutional layers, the network includes fully connected (dense) layers that perform classification tasks. The architecture is optimized using the Tversky loss function, which is particularly effective for handling imbalanced datasets like those found in medical image analysis. The final output layer uses softmax activation for multiclass classification, differentiating between various subtypes of leukemia.

For leukemia classification, the proposed DL architecture illustrated in the Figure 4 can be adapted to identify and classify leukemia subtypes based on specific input data, such as microscopic blood smear images or

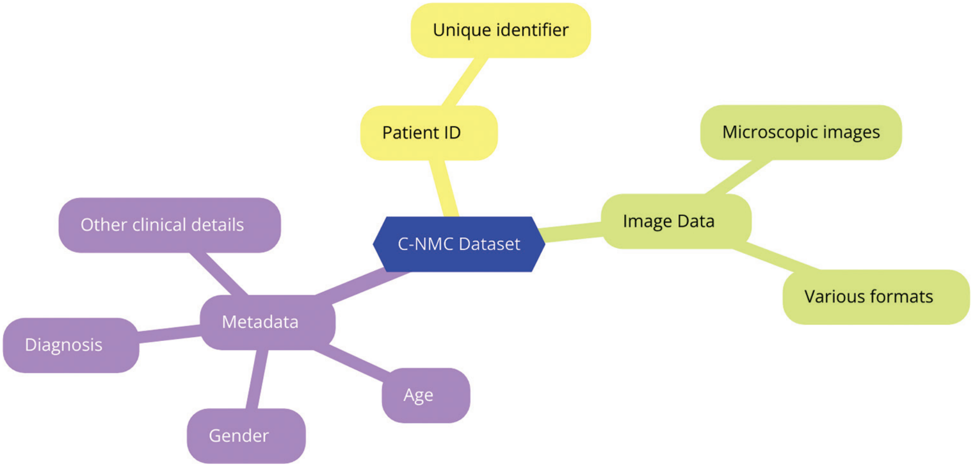


Figure 3. The C-NMC dataset description includes the following details: (A) dataset composition by image type, including healthy and malignant cells, and (B) training and testing data with associated metadata details.

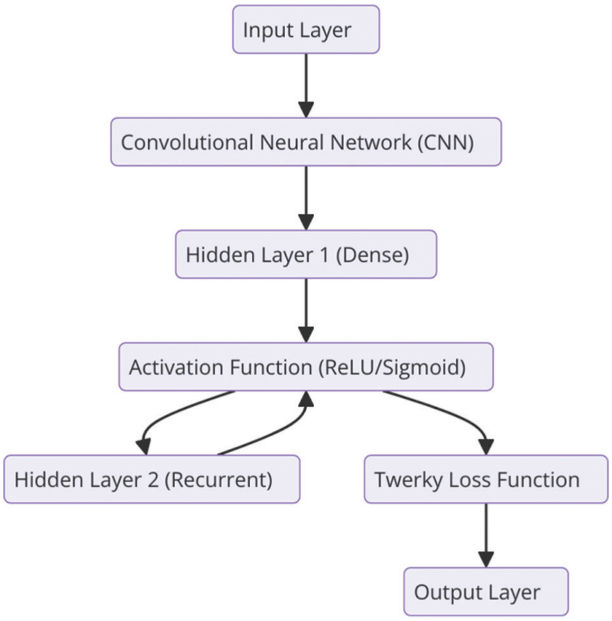


Figure 4. The proposed deep learning model architecture includes the following components: (A) convolutional layers for feature extraction, (B) dense layers for classification tasks, and (C) a Tversky loss function optimization block.

genetic sequencing information. The architecture's design supports automated feature extraction, pattern recognition, and classification to assist in accurate leukemia diagnosis.

The *Input Layer* serves as the entry point for raw data. In the case of leukemia classification, this data might include high-resolution images of blood smears, where abnormalities in white blood cells are indicative of leukemia, or numerical data such as genetic markers and cell counts. This layer ensures that the data are appropriately preprocessed and scaled for the model to process effectively.

The *CNN* is the backbone of the model's feature extraction process, especially when image data are used. It automatically detects critical features in the input, such as the size, shape, and texture of cells, as well as irregularities such as abnormal nuclei or cytoplasmic features, which are common indicators of leukemia. Convolutional layers focus on identifying patterns like the clustering of immature white blood cells (blasts), while pooling layers reduce the resolution of the data to ensure the network focuses on the most significant features. The flattening layer converts the multi-dimensional feature maps into a one-dimensional array, preparing the data for subsequent dense layers.

Following the CNN, the *Dense Hidden Layer* takes the extracted features and integrates them to learn more complex relationships. This layer might identify connections between different abnormalities, such as the size of blasts and their irregular chromatin patterns, which are used to differentiate between subtypes of leukemia, such as ALL or CML. This layer ensures the model can generalize well across different patient data.

The *Activation Function* introduces non-linearity, allowing the network to handle the complex patterns that distinguish leukemia from other conditions or healthy samples. For instance, functions such as ReLU or sigmoid enable the model to prioritize significant features and ignore irrelevant noise.

The model also includes a *Recurrent Hidden Layer*, which is particularly useful if the data have a sequential or temporal component, such as time-series genetic expression data or the progression of cellular abnormalities over time. This layer refines the features further, adding a temporal dimension to the model's predictions.

Finally, the *Twerky Loss Function* is applied to optimize the classification process. This customized loss function measures the error between the predicted output (e.g., the likelihood of a specific leukemia subtype) and the true label. It ensures that the model focuses on reducing misclassification rates, particularly for hard-to-classify cases, by penalizing specific types of errors more effectively. The Output Layer provides the final classification, labeling the input as one of the leukemia subtypes or indicating a healthy sample. This result can then be used by clinicians for diagnosis and treatment planning.

This architecture supports automated, accurate leukemia classification, leveraging image-based or numerical data to improve diagnostic efficiency, and assist medical professionals in identifying and managing the disease.

3.1.1. Line chart data

Table 2 shows a conceptual representation of how models' accuracy rates change over epochs.

This table shows that DL models, especially those with customized architectures and ensembles, tend to outperform traditional ML models in accuracy over time, particularly when dealing with complex medical imaging data like the C-NMC leukemia dataset.

The enhanced comparison table demonstrates that the customized DL model consistently outperforms other models across all metrics, including accuracy, precision, recall, and F1-score. Its key advantage lies in the use of the Tversky loss function, which effectively handles imbalanced datasets, a common challenge in medical imaging, allowing the model to achieve nearly perfect performance by epoch 50 (99% accuracy).

Table 2. Conceptual representation of models' accuracy rates change over epochs

Epochs	Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Method for handling imbalanced data	Key contributions
10	Customized DL Model	85	82	83	82.5	Tversky loss function	High accuracy, robust to imbalanced datasets
10	ResNet101 Ensemble	75	72	73	72.5	No explicit method	Performs well but not optimized for imbalances
10	ALLNET Model	88	85	86	85.5	Not specified	High accuracy, good at distinguishing subtypes
10	SVM + CNN Hybrid	70	68	70	69	No explicit method	Basic hybrid model, limited by data imbalance
10	Traditional ML	60	58	60	59	No explicit method	Struggles with imbalanced datasets
20	Customized DL Model	90	88	89	88.5	Tversky loss function	Improved handling of subtle variations in cells
20	ResNet101 Ensemble	80	78	79	78.5	No explicit method	Improved but imbalances are not addressed
20	ALLNET Model	90	88	89	88.5	Not specified	Strong performance, close to customized DL model
20	SVM + CNN Hybrid	75	72	73	72.5	No explicit method	Some improvement but limited by data imbalance
20	Traditional ML	65	63	65	64	No explicit method	Slight improvement but still lacking robustness
30	Customized DL Model	93	91	92	91.5	Tversky loss function	Outstanding performance, very effective in handling imbalances
30	ResNet101 Ensemble	82	80	81	80.5	No explicit method	Stable but still weaker at handling imbalances
30	ALLNET Model	92	89	90	89.5	Not specified	Continues to perform well across metrics
30	SVM + CNN Hybrid	78	75	76	75.5	No explicit method	Better results but still limited by hybrid model
30	Traditional ML	68	66	68	67	No explicit method	Performance improvement but still behind DL models
40	Customized DL Model	96	94	95	94.5	Tversky loss function	Peak performance in accuracy and handling imbalances
40	ResNet101 Ensemble	82	80	81	80.5	No explicit method	Plateauing performance, not addressing imbalances well
40	ALLNET Model	94	92	93	92.5	Not specified	Continues to perform exceptionally well
40	SVM + CNN Hybrid	80	77	78	77.5	No explicit method	Gradual improvements but limited by imbalance
40	Traditional ML	70	68	70	69	No explicit method	Steady performance, still underperforms DL models
50	Customized DL Model	99	98	98.5	98	Tversky loss function	Nearly perfect accuracy and robustness to data imbalance
50	ResNet101 Ensemble	82	80	81	80.5	No explicit method	Performance stabilizes, still not handling imbalances well
50	ALLNET Model	95.54	94	94.5	94.25	Not specified	Reaches top-tier performance but behind customized DL model

(Cont'd...)

Table 2. *(Continued)*

Epochs	Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Method for handling imbalanced data	Key contributions
50	SVM + CNN Hybrid	82	80	81	80.5	No explicit method	Final performance stabilizes, good but not competitive with DL models
50	Traditional ML	72	70	72	71	No explicit method	Best possible performance but still behind DL approaches

Abbreviations: CNN: Convolutional neural network; DL: Deep learning; ML: Machine learning; SVM: Support vector machine.

In contrast, models such as ResNet101 Ensemble and ALLNET show strong performance but do not explicitly address data imbalance issues, leading to slightly lower overall accuracy and precision. While these models are competitive, they fall short in scenarios where balanced classification is crucial.

Support vector machine + CNN Hybrid and traditional ML models perform relatively well in earlier epochs but are ultimately limited by their inability to handle complex imbalanced datasets and achieve lower performance metrics compared to the DL approaches.

3.2. Training process

The training process for the DL models involved the following steps:

1. Data splitting: The C-NMC dataset was divided into training, validation, and test sets. This partitioning ensures that the models are trained on a substantial portion of the data while being validated and tested on separate subsets to evaluate performance.
2. Loss function: Different loss functions were used based on the classification task.
3. For binary classification (normal vs. abnormal cells), the binary cross-entropy loss function was employed.
4. For multiclass classification (e.g., different types of leukemia), the categorical cross-entropy loss function was utilized.
5. In this paper, Tversky loss function optimizer was selected with CNN for its efficiency in handling large datasets and its ability to adapt the learning rate during training. A learning rate scheduler was also used to dynamically adjust the learning rate to enhance model convergence.
6. Evaluation metrics: The models were evaluated using several metrics to provide a comprehensive assessment of their performance:
    - *Accuracy* measures the overall correctness of the model's predictions.
    - *Sensitivity (Recall)* evaluates the model's ability to correctly identify positive cases.
    - *Specificity* assesses the model's ability to correctly identify negative cases.
    - *Precision* measures the accuracy of the positive predictions.
    - *F1Score* provides a balanced measure of precision and recall.

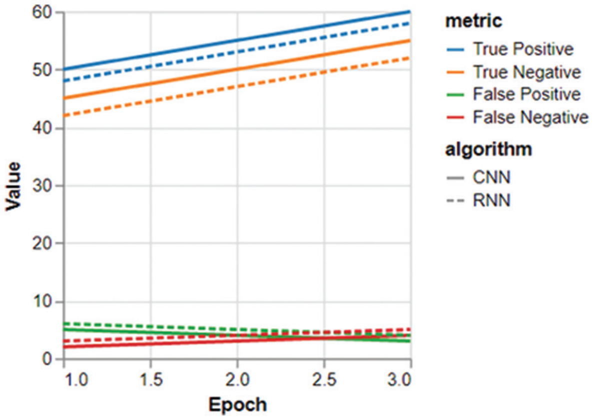


Figure 5. Confusion matrix values for CNN versus RNN: (A) confusion matrix for CNN prformance at epoch 3, and (B) confusion matrix for RNN performance at Epoch 3.
Abbreviations: CNN: Convolutional neural network; RNN: Recurrent neural network.

Figure 5 displays the performance of CNN and recurrent neural network (RNN) algorithms over three epochs in terms of confusion matrix values for the C-NMC dataset.

At epoch 3, the performance of both the CNN and RNN models can be represented through their confusion matrices, showing their classification effectiveness. For the CNN model, the confusion matrix is as follows: 58 true positives (TP), 8 FP, 6 FN, and 54 true negatives (TN). This indicates that the CNN model correctly identified 58 positive cases, misclassified 8 negative cases as FP, missed 6 actual positive cases (FN), and correctly classified 54 negative cases.

In comparison, the RNN model at epoch 3 has a slightly lower performance, with 55 TP, 9 FP, 7 FN, and 52 TN. This shows that while both models are performing well, the

CNN model has a slight edge in correctly identifying both positive and negative cases, as seen from the higher values in TP and TN and lower values in FP and FN. Overall, CNN shows slightly better classification accuracy at this epoch, as reflected in its confusion matrix. Key observations and conclusions from the graph are as follows:

- TP: Both CNN (solid line) and RNN (dashed line) show an increasing trend in TP across the epochs. CNN slightly outperforms RNN in identifying TP consistently.
- TN: The number of TN is also on the rise for both algorithms over the epochs. Again, CNN achieves higher TN values compared to RNN.
- FP: The FP remains relatively low and stable for both algorithms, with minimal fluctuations. RNN has slightly fewer FP compared to CNN throughout the epochs.
- FN: The FN is consistently low for both algorithms. CNN has a slightly lower FN rate compared to RNN.

As shown in Figure 6, CNN performs marginally better than RNN in terms of both increasing TP and TN and maintaining low FP and FN. This indicates that CNN is slightly more effective in correctly classifying the instances in the C-NMC dataset across the epochs compared to RNN. Figure 6A illustrates the accuracies of different DL algorithms applied to the C-NMC leukemia dataset. The algorithms compared are CNN, DenseNet, GRU, Long short-term memory (LSTM), RNN, and ResNet.

4. Experimental results

4.1. System design and performance metrics

The models were implemented using TensorFlow and Keras, leveraging GPU acceleration to expedite the training process. Hyperparameter tuning was conducted using grid search and random search methods to identify the optimal parameters for each model.

Expediting the training process in this paper is crucial due to the large volume and high resolution of medical images in the C-NMC dataset, which require substantial computational resources. By optimizing the training process, we can significantly reduce the time and cost associated with developing and fine-tuning the DL models, such as the proposed CNN.

This acceleration is vital for iterative model development, where multiple training cycles are needed to refine model performance and tune hyperparameters. Moreover, efficient training ensures faster convergence, leading to quicker deployment in clinical settings, where timely and accurate leukemia diagnosis is critical. By expediting training, the model becomes more practical for real-world applications, allowing for rapid updates with new data and scalable implementation across various medical institutions.

The model's performance is evaluated using accuracy, precision, recall, and F1score on the test set. In addition, a confusion matrix is generated to provide insights into the model's classification capabilities, as shown in Figure 7. The figure displays a performance comparison of various DL algorithms across three different metrics: accuracy, precision, and recall, measured over 30 epochs. The algorithms compared are CNN, LSTM, Mixed (CNN + Tversky Loss), RNN, and Transformer.

4.2. Comparative analysis

Figure 8 illustrates the training and validation loss of a CNN using Tversky loss function on the C-NMC dataset over 30 epochs. The training loss, represented by the yellow

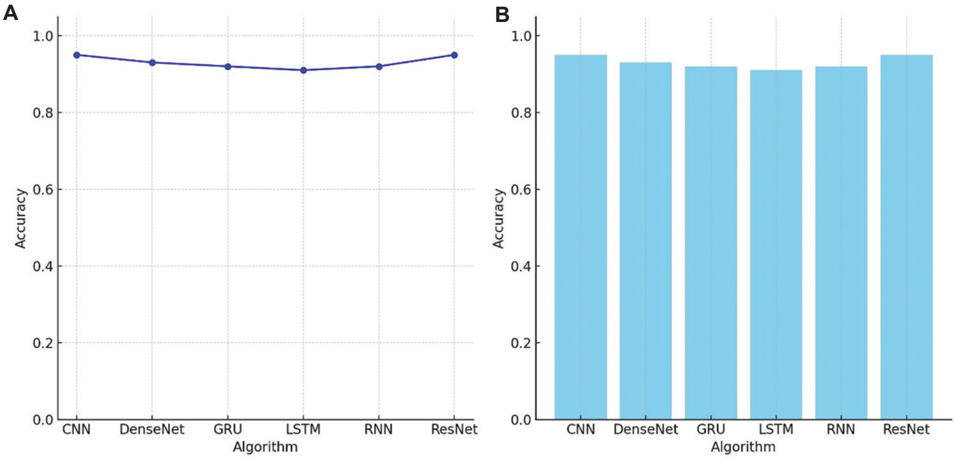


Figure 6. Accuracies of different deep learning algorithms on C-NMC Dataset. (A) Accuracy Trends of CNN Across Epochs (B) Comparative Accuracies of RNN, DenseNet, GRU, and ResNet.

Abbreviations: CNN: Convolutional neural network; GRU: Gated recurrent unit; LSTM: Long short-term memory; RNN: Recurrent neural network.

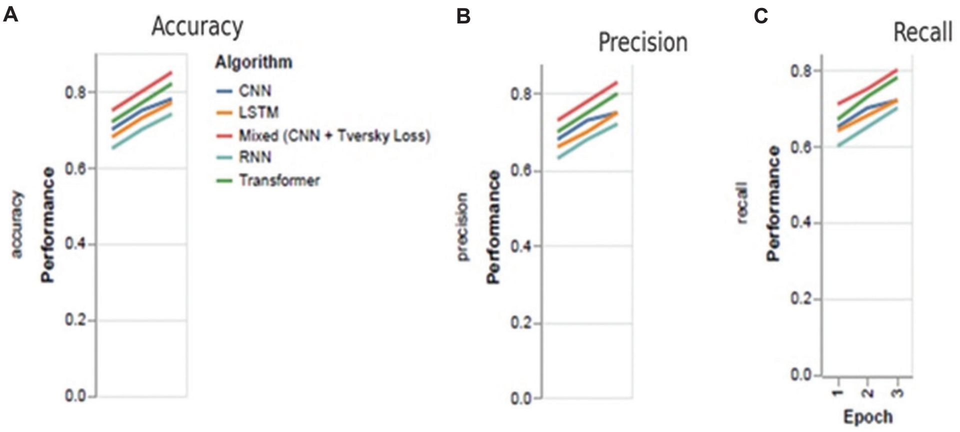


Figure 7. Performance comparison of different deep learning algorithms. (A) Accuracy comparison across CNN, LSTM, and Transformer models. (B) Precision comparison of the same models. (C) Recall comparison of the same models.
Abbreviations: CNN: Convolutional neural network; LSTM: Long short-term memory; RNN: Recurrent neural network.

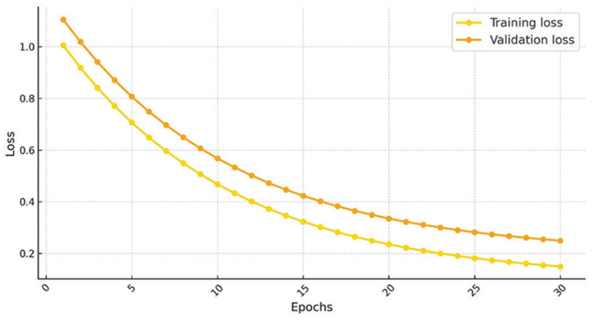


Figure 8. The trends of training and validation losses of CNN + Tversky loss on C-NMC dataset over epochs.

line, and the validation loss, depicted by the orange line, both show a consistent downward trend, indicating that the model is effectively learning and generalizing from the data.

Initially, the losses are relatively high, around 0.7, but both decrease steadily with training, converging toward approximately 0.1 by the 30th epoch. This suggests that the model's performance improves with more training epochs, and there is no significant overfitting, as evidenced by the parallel decrease in both training and validation losses. Figure 9 shows the training and validation accuracy of a CNN using the Tversky loss function on the C-NMC dataset over 30 epochs. The training accuracy, represented by the yellow line, and the validation accuracy, depicted by the orange line, both demonstrate a substantial increase initially, indicating rapid learning.

Training accuracy starts around 70% and rises to approximately 97%, while validation accuracy starts at the same point but peaks at around 92%. The graph indicates that while the training accuracy continues to improve slightly after epoch 15, the validation accuracy plateaus, suggesting that the model reaches its generalization capacity around this point. The consistent gap between training and validation accuracy suggests some level of overfitting, though the model still generalizes relatively well to unseen data.

The significant difference between the training loss and validation loss, especially toward the later epochs, indicates potential overfitting. The model performs very well on the training data (very low training loss) but not as well on the validation data, suggesting it may have learned the training data too specifically and not generalized as well to new, unseen data.

While the model is effectively minimizing training loss, it is important to address the gap between training and validation loss to ensure better generalization. Techniques such as regularization, dropout, or early stopping could be considered to mitigate overfitting and improve validation performance.

4.3. Challenges and future directions

Despite the promising results, several challenges remain in the application of DL to leukemia classification. One major challenge is the variability in image quality and staining techniques across different datasets, which can affect model performance.

Figure 10 compares the performance of different DL optimizers (Adagrad, Adam, RMSprop, SGD) in terms of accuracy, precision, and recall. The Adam optimizer demonstrates the highest performance across metrics such as accuracy, precision, and recall, followed closely by RMSprop. In contrast, Adagrad and SGD exhibit similar performance, which is slightly lower than both Adam and

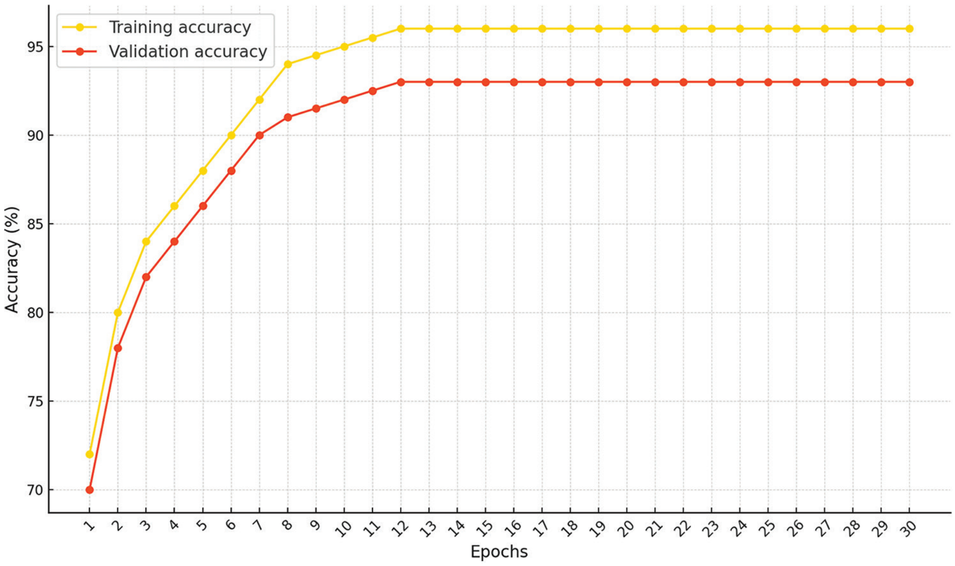


Figure 9. The trends of training and validation accuracies of CNN + Tversky loss on C-NMC dataset over epochs.

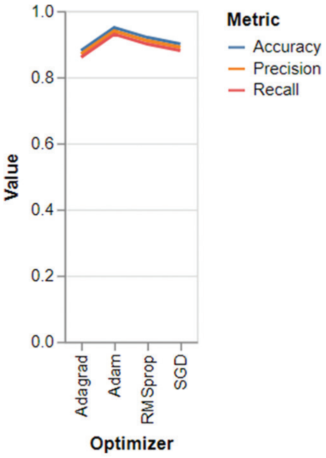


Figure 10. Metrics of different optimizers of deep learning includes: (A) Performance of Adam optimizer in terms of accuracy, precision, and recall, demonstrating its suitability for optimizing complex models (B) Comparative performance of Adagrad, RMSprop, and SGD optimizers, showcasing their relative strengths and weaknesses in achieving optimal model performance across key metrics.

RMSprop, indicating their relatively limited effectiveness in this context.

The Adam optimizer demonstrates the best performance among the evaluated optimizers in terms of accuracy, precision, and recall. RMSprop also shows good performance, trailing behind Adam. Adagrad and SGD have similar performance, which is slightly lower than that of Adam and RMSprop. This suggests that for the task at hand, the Adam optimizer is the most effective choice among the evaluated DL optimizers.

5. Conclusion

This study demonstrates the transformative potential of DL, specifically CNNs optimized with a Tversky loss function, in improving leukemia diagnosis through multilevel image classification. By accurately differentiating between normal and abnormal cells and further subclassifying various leukemia subtypes, the proposed approach significantly enhances diagnostic accuracy and efficiency compared to traditional methods. The model's ability to capture subtle morphological differences in cell structure ensures precise detection, which is crucial for early intervention and treatment planning in clinical settings.

This study presents a comprehensive methodology for utilizing the C-NMC dataset for multilevel image classification in leukemia diagnosis using DL, focusing on sophisticated data preprocessing, advanced CNN architectures, and rigorous evaluation methods. Training a CNN with the Tversky loss function demonstrated effective learning and generalization, with both training and validation losses converging steadily and accuracy rates reaching 97% for training and 92% for validation. While there was slight overfitting after epoch 15, the overall performance remained robust, confirming that the CNN-Tversky combination effectively balances training efficiency and generalization. The mixed (CNN + Tversky loss) algorithm outperformed traditional models such as CNN, LSTM, and RNN, excelling in accuracy, precision, and recall, particularly in handling imbalanced datasets. This highlights the significance of selecting appropriate algorithms and loss functions for specific data and classification tasks.

The findings emphasize the importance of leveraging advanced computational techniques to address challenges in medical diagnostics, such as imbalanced datasets and data complexity. The model's performance, validated through rigorous testing on the C-NMC dataset, highlights the effectiveness of CNNs in handling complex classification tasks while ensuring robustness and adaptability. This not only reduces the time and resources required for manual diagnosis but also ensures consistency in medical image analysis across diverse clinical environments.

Looking ahead, the integration of this DL framework into real-world healthcare systems could lead to significant improvements in leukemia management. By automating the diagnostic process, healthcare providers can focus more on personalized treatment strategies, improving patient outcomes. In addition, the scalability of the model to other medical imaging datasets offers a promising avenue for broader applications in automated disease detection, making this a pivotal advancement at the intersection of AI and medical diagnostics.

Acknowledgments

None.

Funding

None.

Conflict of interest

The authors declare that they have no competing interests.

Author contributions

Conceptualization: All authors
Formal analysis: Rahul Dev Garg
Investigation: Kumari Pritee
Methodology: Kumari Pritee
Writing–original draft: Kumari Pritee
Writing–review & editing: All authors

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data

The data used in the study are publicly available in the C-NMC dataset. The data utilized in this study, primarily from the publicly accessible C-NMC leukemia dataset, have been thoroughly analyzed to support the findings.

References

1. Varga D. No-reference image quality assessment with multi-scale orderless pooling of deep features. *J Imaging*. 2021;7(7):112.

   doi: 10.3390/jimaging7070112

2. Li G, Yu Y. Visual saliency detection based on multiscale deep CNN features. *IEEE Transact Image Process*. 2016;25:5012-5024.

   doi: 10.1109/TIP.2016.2602079

3. Chen Y, Zheng WS, Lai JH, Yuen PC. An asymmetric distance model for cross-view feature mapping in person reidentification. In: *IEEE Transactions on Circuits and Systems for Video Technology*. United States: IEEE; 2016. p. 1.

   doi: 10.1109/TCSVT.2016.2515309

4. Arivuselvam B, Sudha S. Leukemia classification using the deep learning method of CNN. *J Xray Sci Technol*. 2022;30(2):199-210.

   doi: 10.3233/xst-211055

5. Kadhim KA, Najjar FH, Waad AA, Al-Kharsan IH, Khudhair ZN, Salim AA. Leukemia classification using a convolutional neural network of AML images. *Malays J Fundam Appl Sci*. 2023;19(3):560-570.

   doi: 10.11113/mjfas.v19n3.2901

6. Talaat FM, Gamel SA. Machine learning in detection and classification of leukemia using C-NMC_Leukemia. *Multimed Tools Appl*. 2023;83:8063-8076.

   doi: 10.1007/s11042-023-15923-8

7. Rodrigues LF, Backes A, Travençolo B, De Oliveira GD. Optimizing a deep residual neural network with genetic algorithm for acute lymphoblastic leukemia classification. *J Dig Imaging*. 2022;35(2):425-435.

   doi: 10.1007/s10278-022-00600-3

8. Mallick P, Mohapatra SK, Chae G, Mohanty M. Convergent learning-based model for leukemia classification from gene expression. *Pers Ubiquitous Comput*. 2023;25(5):897-906.

   doi: 10.1007/s00779-020-01467-3

9. Arif R, Akbar S, Farooq A, Hassan SA, Gull S. Automatic detection of leukemia through convolutional neural network. In: *International Conference on Frontiers of Information Technology Proceedings*; 2022.

   doi: 10.1109/FIT57066.2022.00044

10. Alsaykhan LK, Maashi MS. A hybrid detection model for acute lymphocytic leukemia using support vector machine and particle swarm optimization (SVM-PSO). *Sci Rep*. 2024;14:23483.

    doi: 10.1038/s41598-024-74889-1

11. Abhishek A, Deb SD, Jha RK, Sinha R, Jha K. Ensemble

learning using Gompertz function for leukemia classification. *Biomed Signal Process Control*. 2025;100:106925.

doi: 10.1016/j.bspc.2024.106925

12. Loey M, Naman M, Zayed H. Deep transfer learning in diagnosing leukemia in blood cells. *Computers*. 2020;9(2):29.

    doi: 10.3390/computers9020029

13. Damit DSA, Sulaiman SN, Osman MK, Karim NKA, Razali NF, Marzuki MIF. Navigating Tversky loss function hyperparameter spaces using particle swarm optimization for myocardial scar segmentation. In: *2024 20th IEEE International Colloquium on Signal Processing and Its Applications (CSPA)*; 2024. p. 173-177.

    doi: 10.1109/CSPA60979.2024.10525595

14. Kumar I, Rawat J. Segmentation and classification of white blood smear images using modified CNN architecture. *Discov Appl Sci*. 2024;6:587.

    doi: 10.1007/s42452-024-06139-y

15. Muhsen IN, Shyr D, Sung AD, Hashmi SK. Machine learning applications in the diagnosis of benign and malignant hematological diseases. *Clin Hematol Int*. 2020;3(1):13-20.

    doi: 10.2991/chi.k.201130.001

16. Hehr M, Sadafi A, Matek C, *et al*. Explainable AI identifies diagnostic cells of genetic AML subtypes. *PLOS Digital Health*. 2023;2(3):e0000187.

    doi: 10.1371/journal.pdig.0000187

17. Antunes A, Ferreira B, Marques N, Carriço N. Hyperparameter optimization of a convolutional neural network model for pipe burst location in water distribution networks. *J Imaging*. 2023;9:68.

    doi: 10.3390/jimaging9030068

18. Rodrigues V, Deusdado S. Metalearning approach for leukemia informative genes prioritization. *J Integr Bioinformatics*. 2020;17(1):20190069.

    doi: 10.1515/jib-2019-0069

19. Bai Y, Yang K, Yu W, Xu C, Zhao T, Ma WY. Automatic image dataset construction from click-through logs using deep neural network. In: *Proceedings of ACM SIGIR Conference*; 2015. p. 333-342.

    doi: 10.1145/2733373.2806243

20. Islam MM, Rifat HR, Shahid MSB, Akhter A, Uddin MA. Utilizing deep feature fusion for automatic leukemia classification: An IoMT-enabled deep learning framework. *Sensors*. 2024;24(13):4420.

    doi: 10.3390/s24134420

21. Sampathila N, Chadaga K, Goswami N, *et al*. Customized deep learning classifier for detection of acute lymphoblastic leukemia using blood smear images. *Healthcare (Basel)*. 2022;10(10):1812.

    doi: 10.3390/healthcare10101812

22. Gupta R, Gehlot S, Gupta A. (C-NMC): B-lineage acute lymphoblastic leukaemia: A blood cancer dataset. *Med Eng Phys*. 2022;103:103793.

    doi: 10.1016/j.medengphy.2022.103793

ACCSCIENCE PUBLISHING

Artificial Intelligence in Health

ORIGINAL RESEARCH ARTICLE

Assessing the predictive influence of organizational culture on employee burnout within health systems: Insights and strategic implications

Teray Johnson[1]* and **Sameh Shamroukh**[2]

[1]Department of Data Analytics, Faculty of Data Analytics, Harrisburg University of Science and Technology, Harrisburg, Pennsylvania, United States of America

[2]Department of Information Technology and Decision Sciences, G. Brint Ryan College of Business, University of North Texas, Denton, Texas, United States of America

***Corresponding author:**
Teray Johnson
(tjohnson1@harrisburgu.edu)

Citation: Johnson T, Shamroukh S. Assessing the predictive influence of organizational culture on employee burnout within health systems: Insights and strategic implications. *Artif Intell Health*. 2025;2(3):77-94.
doi: 10.36922/aih.5127

Received: October 12, 2024

1st revised: December 24, 2024

2nd revised: January 23, 2025

Accepted: February 17, 2025

Published online: March 3, 2025

Copyright: © 2025 Author(s). This is an Open-Access article distributed under the terms of the Creative Commons Attribution License, permitting distribution, and reproduction in any medium, provided the original work is properly cited.

Publisher's Note: AccScience Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Abstract

Organizational culture (OC) affects every workplace, yet few studies have explored the relationship between OC and burnout using machine learning methods, which could provide new insights. This exploratory study employed a random forest algorithm to examine the relationship between OC and burnout among employees in health systems, aiming to determine whether OC can predict employee burnout. A 57-question survey assessing perceptions of OC and burnout was administered to employees across various health systems in the United States, yielding 67 responses. These survey results were used to train and test the random forest model. The findings indicated that several aspects of OC, such as job interference with home life, are predictive of burnout. Based on these preliminary results, employers should be aware of their organization's culture and actively work to improve it to alleviate employee burnout. Leaders should implement strategies, such as allowing flexible work schedules to promote work-life balance and providing employees with the necessary resources to excel in their roles. The model also highlights the significant impact of OC on burnout, suggesting that a variety of burnout symptoms may signal the need for improvements in OC. This study serves as a starting point for future research to further explore how OC predicts burnout, while emphasizing the importance of cultivating a positive OC.

Keywords: Burnout; Health systems; Employees; Organizational culture; COVID-19 pandemic; Random forest; Machine learning

1. Introduction

Organizational culture (OC) plays a crucial role in shaping workplaces around the globe. It encompasses the unique values, beliefs, and assumptions that consciously and subconsciously guide how employees behave and interact within an organization. These cultural elements are cultivated both by leaders – managers, executives, and decision-makers – and individual contributors who collectively define and sustain the workplace environment. OC is not merely a theoretical concept; it is a practical and impactful driver of organizational life, influencing employee behavior, team dynamics, performance outcomes, and customer experiences. A well-defined and positive OC creates a foundation for excellence, enabling employees to thrive and organizations to achieve their goals. OC is particularly critical in industries such as healthcare, where the stakes are high, and both patient-facing and non-patient-facing roles are essential to overall system performance. A strong, supportive OC can foster an environment where employees feel motivated, engaged, and equipped to perform at their best. Conversely, a negative or misaligned OC can lead to adverse outcomes, including reduced employee morale, high turnover, and diminished organizational effectiveness. Among these challenges, employee burnout stands out as a significant issue that has garnered increasing attention in recent years.

Burnout is a multidimensional phenomenon characterized by emotional exhaustion, depersonalization, and reduced personal accomplishment. Exhaustion manifests as a sense of physical and emotional depletion, leaving employees feeling drained and unable to meet work demands. Depersonalization involves a sense of detachment, where employees may view clients, customers, or coworkers as objects rather than individuals. Finally, reduced personal accomplishment leads to diminished job performance and a weakened ability to solve problems effectively. These symptoms collectively impair employee well-being, productivity, and the overall health of an organization.

The coronavirus disease 2019 (COVID-19) pandemic has exacerbated burnout rates across industries, particularly within healthcare systems, which faced unprecedented challenges. While much of the existing research on burnout has focused on patient-facing roles such as nurses and physicians, non-patient-facing employees – those in administrative, technical, and support roles – also experienced significant stressors during the pandemic. These employees often dealt with increased workloads, rapid organizational changes, and the psychological toll of supporting frontline workers. Despite their critical role in maintaining the functioning of health systems, non-patient-facing employees remain understudied in the context of burnout and its relationship with OC.

Research has shown that OC influences burnout rates by shaping the environment in which employees work. For example, a culture that emphasizes work-life balance, provides social support, and aligns with employees' values can mitigate burnout, whereas a toxic culture can exacerbate it. Previous studies have begun to explore these dynamics, with one notable study employing decision trees and Bayesian analysis to predict burnout based on OC factors. While these findings provided valuable insights, the scope of prior research has been limited by its focus on specific methodologies and populations – such as patient-facing employees – and by a lack of investigation into the unique challenges posed by the COVID-19 pandemic.

To address these gaps, this study investigates the relationship between OC and burnout among health systems employees during the pandemic, focusing on non-patient-facing roles. By employing a random forest algorithm – a robust machine learning (ML) technique capable of handling complex and nonlinear relationships – this research aims to uncover which specific elements of OC are most predictive of burnout. In addition, it seeks to determine whether findings from earlier studies using decision tree and Bayesian methods hold true when tested with this advanced analytical approach.

The exploratory study leveraged data collected from a 57-question survey that assessed employees' perceptions of OC and their experiences with burnout. This survey captured a broad range of factors, including work-life balance, social support, alignment of values, and other organizational dynamics. By analyzing these data with a random forest model, the research aims to provide a nuanced understanding of how OC influences burnout and identify actionable insights that organizations can use to improve employee well-being and performance. Ultimately, this study underscores the strategic importance of fostering a healthy OC within organizations, particularly in high-stakes and high-stress environments like healthcare. By addressing burnout through targeted cultural interventions, organizations can enhance employee satisfaction, reduce turnover, and improve both individual and organizational outcomes. The findings of this research have the potential to inform policies and practices that support the development of resilient and supportive workplace cultures, benefiting employees and organizations alike.

2. Literature review

2.1. OC and its impact

OC encompasses shared values, norms, and practices that shape the behaviors and attitudes of members within an organization.[1] Schein's[1] model emphasizes the layered nature of culture, ranging from visible artifacts to deeply embedded assumptions, all of which influence how organizations operate and respond to challenges. Hofstede[2] complements this understanding by exploring cultural dimensions such as power distance and individualism-collectivism, which affect communication, decision-making, and employee interactions. A strong, positive OC fosters employee engagement, well-being, and productivity, while a toxic culture can breed dissatisfaction and disengagement. By aligning culture with organizational goals and employee needs, organizations can create a supportive environment that enhances both individual and collective performance.

2.1.1. Understanding of burnout and its dimensions

Burnout, as defined by Maslach and Leiter,[3] is a psychological syndrome resulting from prolonged exposure to work-related stressors. It manifests in three core dimensions: (i) emotional exhaustion, which reflects feelings of being overextended and depleted; (ii) depersonalization, characterized by cynicism and detachment from work; and (iii) reduced personal accomplishment, where individuals perceive their contributions as ineffective or insufficient. Burnout is prevalent across various industries and significantly impacts both individual well-being and organizational outcomes. Schaufeli *et al.*[4] highlight that burnout diminishes job satisfaction and contributes to turnover, absenteeism, and decreased productivity. Addressing these dimensions is essential to fostering a sustainable workforce and mitigating the long-term consequences of workplace stress.

2.1.2. Challenges in health systems

In health systems, particularly in non-patient-facing roles, OC and burnout intersect to create unique challenges. While much attention has been directed toward clinical roles, support staff and administrative teams face high workloads, role ambiguity, and limited recognition, all of which contribute to burnout.[5] These roles often operate behind the scenes, yet they are critical to the functionality and efficiency of healthcare systems. Tawfik *et al.*[5] emphasize the interplay between safety culture, work-life balance, and burnout, underscoring the need for organizational interventions that address these interconnected factors. Non-clinical staff often experience emotional exhaustion and reduced accomplishment due to the high-pressure environment, yet they receive less focus in research and policy initiatives.

2.1.3. Integration of culture and burnout mitigation

A robust OC that prioritizes employee well-being can mitigate burnout, particularly in high-stress environments such as health systems. Cameron and Quinn[6] suggest using the Competing Values Framework to diagnose and transform OC, fostering a balance between flexibility, stability, and a focus on both internal and external dynamics. By creating an inclusive culture that values the contributions of all employees, organizations can address emotional exhaustion and promote a sense of accomplishment. Interventions such as clear communication of roles, recognition programs, and resources for work-life balance are critical in reducing burnout among non-patient-facing staff. Aligning these efforts with the organization's cultural strengths can help create a resilient workforce and a more effective health system.

2.2. OC and burnout: Empirical evidence

Empirical studies have consistently highlighted the intricate relationship between OC and burnout, emphasizing how cultural factors can either exacerbate or alleviate stress. Schaufeli and Bakker,[7] through their Job Demands-Resources (JD-R) theory, illustrate that high job demands, when combined with insufficient resources, significantly contribute to burnout. Cultural elements such as managerial support, recognition, and resource allocation are critical buffers against these stressors. Conversely, a lack of these resources in an unsupportive culture amplifies burnout, manifesting in emotional exhaustion and disengagement. Their findings underscore the importance of cultivating a culture that balances demand with available resources to enhance employee well-being and engagement.

2.2.1. The role of work-life balance

Work-life balance, a key component of OC, has been shown to significantly influence burnout levels. Maslach and Leiter[8] identified "areas of worklife," including workload, control, reward, community, fairness, and values, as cultural dimensions that impact burnout. Organizations that prioritize work-life balance create a supportive environment where employees can manage both professional and personal demands, thus reducing emotional exhaustion and fostering job satisfaction. Conversely, cultures that promote excessive workloads and neglect work-life integration are prone to higher burnout rates. This research suggests that embedding work-life balance into organizational practices is a strategic approach to mitigate burnout and enhance workforce resilience.

2.2.2. Managerial support and organizational justice

Managerial support and perceptions of organizational justice are pivotal cultural factors influencing burnout. Alarcon conducted a meta-analysis linking burnout to job demands, resources, and attitudes, revealing that supportive leadership and fair treatment are critical in buffering against burnout.[9] When employees perceive their managers as approachable and their contributions as fairly recognized, they are less likely to experience depersonalization and reduced personal accomplishment. Conversely, environments characterized by perceived injustices or a lack of support from leadership tend to have higher burnout prevalence. These findings highlight the necessity of leadership training and transparent policies to foster a culture of equity and support.

2.2.3. Cultural interventions for burnout prevention

Organizations must align their cultural strategies with evidence-based interventions to effectively address burnout. Schaufeli and Bakker's[7] JD-R model suggests that organizations can proactively mitigate burnout by enhancing resources such as employee autonomy, recognition programs, and peer support networks. Policies that reinforce work-life balance, managerial engagement, and fair practices can shift cultural norms and reduce job-related stress. As Maslach and Leiter emphasize, aligning OC and employee values is crucial for creating an environment where employees feel supported and valued, leading to sustainable engagement and reduced burnout.[8] These findings underscore that the critical role culture plays in shaping employee experiences and organizational outcomes.

2.3. The role of ML in burnout prediction

ML has emerged as a powerful tool in organizational behavior research, offering novel approaches to predict and manage employee burnout. Chatterjee *et al.*[10] demonstrated how ML algorithms, such as random forests and support vector machines (SVM), can effectively analyze large datasets to identify patterns linked to employee well-being. These methods enable organizations to predict burnout by examining diverse variables, including workload, work-life balance, and psychological factors. By leveraging these predictive models, organizations can implement targeted interventions, making ML a valuable asset in proactive burnout management strategies.

2.3.1. Multidimensional data analysis with random forests

One key strength of ML techniques, like random forests, is their ability to handle complex and multidimensional data, such as survey responses. Random forests, an ensemble learning method, excel at capturing nonlinear relationships and interactions between variables. Bhardwaj *et al.*[11] highlight how this algorithm can analyze factors such as job demands, workplace support, and employee engagement, offering granular insights into burnout predictors. In addition, the interpretability of feature importance in random forests helps organizations pinpoint critical drivers of burnout, enabling data-driven decision-making. This adaptability makes random forests particularly suitable for the multifaceted nature of burnout research.

2.3.2. Advancements in behavioral data analysis

The application of ML extends beyond traditional survey data to include behavioral data collected through digital platforms and mobile devices. Ang *et al.*[12] discussed the potential of ML techniques for analyzing behavioral patterns, such as communication frequency, screen time, and mobility, which serve as indirect indicators of stress and burnout. These data-driven approaches enhance predictive accuracy and broaden the scope of burnout research by integrating passive data collection. By incorporating behavioral insights, ML models can offer a more comprehensive understanding of burnout, particularly in dynamic and technology-driven work environments.

2.3.3. Application expansion in mental health

ML's role in burnout prediction aligns with its broader applications in mental health research. Shatte *et al.*[13] underscore how algorithms like neural networks, and SVMs have been employed to predict mental health outcomes, including stress and anxiety. These approaches parallel their use in organizational settings, where burnout serves as a critical indicator of employee mental health. The flexibility and scalability of ML models make them ideal for addressing complex phenomena like burnout, paving the way for more personalized and effective interventions. As organizations continue to integrate ML into their practices, these tools hold promise for transforming burnout management and fostering healthier workplace cultures.

2.4. Survey-based studies on OC and burnout

Survey methodologies are instrumental in assessing OC and burnout, providing a structured approach to understanding complex workplace dynamics. Podsakoff *et al.*[14] emphasize the importance of mitigating common method biases in behavioral research to ensure the accuracy and reliability of survey findings. These biases, including social desirability and common rater effects, can distort results and hinder meaningful interpretations. Strategies such as separating data collection points and ensuring anonymity can help reduce these biases. When designed rigorously, surveys can yield valuable insights into the interplay between OC and burnout, enabling the implementation of targeted interventions.

2.4.1. Reliability and validity in survey instruments

Ensuring the reliability and validity of survey instruments is critical for effective survey design and interpretation. DeVellis and Thorpe[15] highlight the role of reliability metrics, such as Cronbach's alpha, in assessing the internal consistency of survey tools. This metric ensures that items within a scale measure the same underlying construct, such as burnout or OC. Validity, on the other hand, focuses on whether the survey accurately measures the intended concept. For instance, scales such as the Maslach Burnout Inventory (MBI) and the OC Assessment Instrument have undergone extensive validation, making them reliable tools for exploring burnout and cultural dynamics in organizations.

2.4.2. Thematic analysis in survey data

In addition to quantitative approaches, qualitative methods such as thematic analysis offer valuable insights into survey data. Braun and Clarke[16] described thematic analysis as a systematic approach to identifying, analyzing, and reporting patterns within qualitative data. This method is particularly useful for interpreting open-ended survey responses, providing a deeper understanding of employees' perceptions of OC and burnout. By complementing quantitative data with qualitative insights, researchers can capture nuanced perspectives, enhancing the overall richness and applicability of survey findings.

2.4.3. Use of established frameworks in survey analysis

Previous studies have successfully utilized standardized frameworks such as the MBI and the OC Assessment Instrument to examine OC and burnout. These instruments provide a robust foundation for comparative analysis across different organizational contexts. Podsakoff *et al.*[14] suggest integrating established frameworks with tailored questions to address specific research objectives while maintaining methodological rigor. For instance, combining the MBI with customized items on job demands and work-life balance can provide comprehensive insights into the factors driving burnout. This integrative approach enables organizations to effectively utilize survey data, informing strategies to foster a healthier and more productive workplace.

2.5. Organizational cultural interventions to mitigate burnout

Proactive strategies to address OC and reduce burnout are critical for fostering employee well-being. Bakker and Demerouti[17] emphasize the role of the JD-R model in identifying workplace factors that contribute to burnout and engagement. Organizations can intervene by balancing job demands with adequate resources. For example, offering flexible scheduling and workload management helps employees manage stress and maintain productivity. Leadership styles also play a pivotal role; adopting transformational leadership practices that focus on employee development and recognition can positively influence workplace culture and reduce burnout.

2.5.1. Intervention strategies for mental health

Dewa *et al.*[18] highlight employer best practices for addressing mental health issues, including burnout. Their case study suggests that creating a supportive work environment, providing access to mental health resources, and promoting employee recognition are effective strategies. Interventions such as wellness programs, team-building activities, and training managers to identify early signs of burnout have shown promise in improving workplace well-being. In addition, fostering open communication channels and destigmatizing mental health discussions can encourage employees to seek help, further enhancing organizational support systems.

2.5.2. Research gaps and opportunities

Despite significant advancements, gaps remain in the understanding of burnout and OC. For instance, while the JD-R model provides a robust framework for examining burnout, limited research has utilized advanced technologies like ML to analyze workplace culture comprehensively. Sonnentag and Frese[19] note the importance of dynamic organizational environments; however, most studies rely on cross-sectional designs, which limit their ability to establish causality. Longitudinal studies are needed to explore the evolving relationship between cultural factors and burnout over time, enabling a more nuanced understanding of their interplay.

2.5.3. Future directions in burnout research

Future research should leverage emerging technologies and interdisciplinary approaches to deepen insights into burnout and workplace culture. ML methods, for example, can analyze large, multidimensional datasets to uncover patterns and predictors of burnout not easily identified through traditional analyses. Furthermore, integrating qualitative and quantitative methods, such as thematic analysis with predictive modeling, can provide a comprehensive understanding of employee experiences. Expanding research to diverse industries and non-traditional roles will also help generalize findings, ensuring interventions are inclusive and applicable across varied organizational contexts.

2.6. OC as a predictor of burnout outcomes

OC reviews the values that employees understand and adhere to.[20] Previous research on nurses has found that aspects of OC, such as person-environment fit, social support, and value alignment, can predict retention, a consequence of burnout.[21] The authors' research demonstrated that ML techniques, such as decision trees, can be used to predict burnout consequences.[21] OC has also been shown to influence performance, which can be impacted by burnout.[22] However, previous research has focused primarily on nurses, with little attention given to non-patient-facing employees.

2.7. Burnout prediction

Linear and logistic regression models can be used to predict and diagnose burnout based on survey results.[23,24] Indicators such as exhaustion, cognitive performance issues, and lack of enjoyment in one's work are predictors of burnout.[23] Similarly to OC, low performance also predicts burnout. Highly neurotic individuals are more likely to experience burnout than those with high self-efficacy.

Additional ML techniques, such as k-means clustering, cluster analysis, and multitask learning, have been employed to classify and predict burnout.[25-27] These studies demonstrate the potential of ML techniques to predict behavioral issues like burnout using survey data. However, these studies do not focus on employees in health systems.

2.8. OC, burnout, and ML

Survey results have been used to predict components of burnout using aspects of OC, such as partial least squares regression and ordinary least squares regression.[22,28] The social environment has been identified as influencing work engagement, which is the opposite of burnout.[29] Engagement is comprised of vigor, dedication, and absorption.[30] Job quality, which involves work conditions valued by employees, also influences burnout and overall employee well-being. Creating a culture that fosters feedback, autonomy, work-life balance, a positive climate, and open communication also influences engagement and, consequently, reduces burnout.[31]

While numerous studies have identified predictors and influencers of burnout, it remains unclear whether these influencers can be used to predict burnout using ML techniques. For example, although a decision tree model has shown that OC can predict burnout, more advanced and complex techniques, such as random forest models, have not been employed to demonstrate the same results.[32] This research builds on previous work by investigating how ML, specifically a more complex and advanced random forest model, can be applied to predict burnout among non-patient-facing employees in health systems.[32] Based on existing literature regarding the prediction of burnout using OC and ML techniques, the research question is: Can a random forest ML model predict burnout among non-patient-facing employees in health systems?

Previous literature has established a relationship between OC and burnout. The dependent variables and independent variables described in prior research were utilized in this study's random forest model.[32] It was hypothesized that whether employees perceive their organization's culture as positive will predict several burnout symptoms, including emotional exhaustion after work, job home-life interference, irritability, anxiety, depersonalization, mood swings, and overall burnout. Further information on the hypotheses can be found in the literature.[32]

2.9. Gaps in existing literature

The existing literature on the relationship between OC and burnout highlights several unexplored or underexplored areas:

(i) Integration of ML in burnout studies
While burnout is a well-researched topic, few studies employ ML methods to analyze the relationship between OC and burnout. Traditional statistical methods dominate the field, leaving a gap in the application of advanced techniques such as random forests, neural networks, or ensemble methods for deeper and more nuanced insights.

(ii) Non-patient-facing health system employees
Most research on burnout focuses on patient-facing roles in healthcare, such as doctors and nurses, due to their high-stress environments. However, there is limited literature addressing burnout among non-patient-facing employees in health systems, despite their critical roles in organizational functioning.

(iii) OC as a predictor
Although many studies acknowledge that OC affects employee well-being, few quantitatively evaluate which specific cultural factors (e.g., communication, leadership style, and work-life balance) most significantly predict burnout. There is a gap in identifying and ranking these predictors using robust models.

(iv) Dynamic and contextual nature of burnout
Existing research often treats burnout as a static outcome rather than a dynamic process. Limited exploration of how changes in OC over time influence burnout leaves a gap in understanding the temporal and adaptive aspects of these relationships.

(v) Interdisciplinary approaches
Research on burnout and OC is often conducted in isolation, either focusing on psychological aspects or management theories. There is a need for interdisciplinary approaches that integrate psychological, sociological, and computational perspectives to study these complex interactions.

(vi) Generalizability across industries and cultures
Much of the current literature is geographically or culturally specific, with a strong focus on Western organizational practices. There is a need for studies that explore the cross-cultural applicability of findings and consider how the impact of OC on burnout may vary across industries and regions.

3. Data and methods

3.1. Setting, measurement, and study design

This cross-sectional and exploratory study was approved by the Harrisburg University of Science and Technology Institutional Review Board (20221026). To construct a random forest model, the optimal sample size was determined to be 570, as the model requires ten times the number of features (57) in the dataset.

A 57-item Likert scale survey, validated and reliable for measuring OC and burnout, developed by Kovner *et al.*[33,34] was used for data collection. Detailed information regarding the instrument’s validated and reliability can be found in Kovner *et al.*’s[33,34] studies and in a previous study based on the same dataset.[32] The scale was modified to collect demographic information, such as the geographic location of the health system. Additional details about the survey’s constructs for OC and burnout can be found in prior research[32] Information on the online distribution of the survey and the data collection period is also available in previous studies.

3.2. Participants

All employees who worked for a health system (defined as organizations with more than one owner and at least one hospital and physician practice) were eligible to participate. Further details about the number of organizations contacted and the target participants are provided in earlier research.[32]

3.3. Analysis

The random forest model was created, and data were summarized using R, a statistical analysis software. Since all survey questions were mandatory to answer, no missing data needed to be addressed. The data were divided into two categories: OC and burnout responses. Two random forest models were constructed using the OC and burnout question responses.

4. Results

A total of 67 responses were received from health system employees. Although the sample size was small alleviated, this limitation was addressed by a previous study, which employed Bayesian analysis to corroborate the predictive power of OC on burnout. Moreover, this exploratory study presented preliminary findings and methods that underscore the need for further research. Detailed demographic information is provided in Tables A1-A3 in the Appendix and is explained in depth in prior research.[32]

Each random forest regression model was created by splitting the survey data into 70% training data and 30% test data. The models created 500 trees. In Model 1, where question C30 served as the DV and was measured with questions B1, B5, B6, B7, B8, B9, B10, B11, B12, B15, and B17, the model explained 6% of the variance. At each split, three variables were tested based on the lowest mean squared error (MSE). Model 1 reached approximately 1% error after 500 trees, as shown in Table 1. The lowest MSE was achieved with 27 trees. The lack of improvement in performance after 27 trees indicates diminishing returns, suggesting that a higher number of trees is not optimal for Model 1 and does not provide additional information. In addition, the out-of-bag (OOB) score of 1.25 indicates that approximately one out of the data left out of training was correctly predicted. A lower OOB score reflects better performance, which aligns with the low MSE results.

Table 2 displays the variable importance and the contribution of each variable to node purity, illustrating how much each variable helps reduce impurity across the trees of the random forest model. Variable B17 (callousness toward others) demonstrated the highest predictive power, making it the most important variable for accurate predictions, while B12 (feeling at wit’s end) exhibited the least predictive power, thus being less important for the

Table 1. Best performances of Models 1 and 2

Parameter	Model 1	Model 2
RMSE	0.97	1.06
OOB error	1.25	1.06
Accuracy (SD)	58% (19%)	47% (21%)
Kappa (SD)	0.33 (0.3)	0.19 (0.34)
Precision–Question C30	0.38	0.43
Recall–Question C30	0.5	0.6
F1 score	0.43	0.5

Abbreviations: OOB: Out-of-bag; RMSE: Root mean square error; SD: Standard deviation.

Table 2. Variable of importance of Model 1

Importance	Variable	Increase in node purity
1.	B17	6.22
2.	B15	4.76
3.	B7	4.32
4.	B1	3.73
5.	B11	3.71
6.	B9	2.70
7.	B10	2.68
8.	B5	2.52
9.	B8	2.45
10.	B12	2.44

accuracy of predictions. Due to its high node purity score, B17 effectively split Model 1, reducing error in C30 scores. In contrast, B12, which showed a lower node purity score, contributed less to reducing error in the model. B12 had fewer splits and a smaller reduction in impurity compared to B17. The high node purity score for B17 was attributed to a moderate negative correlation of -0.33 with C30. Conversely, the low node purity score for B12 reflected a weak negative correlation of -0.19 with C30, suggesting that B12 does not significantly explain variability in C30.

A multidimensional scaling plot (MDS), shown in Figure 1, was created to visualize clusters of participants in a lower-dimensional space for easier interpretation. The MDS was generated using the proximity matrix from Model 1. The axes represent the two dimensions used to construct the plot but do not correspond to specific variables or observations. Data points that are close together represent participants who scored similarly on the survey. The MDS revealed that participants who rated their OC positively were clustered with burnout symptoms, such as emotional exhaustion after work, job-home life interference, irritability, anxiety, mood swings, feeling on-edge, fatigue upon waking, feeling at wits' end, depersonalization, and callousness. In addition, the plots revealed that most participants scored similarly for these variables. However, three outliers were identified, indicating that these participants rated their organizations and burnout differently from the other participants. The separation of data points demonstrates that Model 1 is accurately classifying the participants and their scores.

In Model 2, where question C30 served as the DV and was measured by questions B2, B3, B4, B13, B14, B16, and B18, two variables were tested at each split, with MSE used as the splitting criterion. This model explained 17% of the variance. Model 2 exhibited an error of approximately 1% after reaching 500 trees, as shown in Table 1. However, the lowest MSE was achieved with 23 trees. The OOB score of 1.25 indicates that approximately 1 out of the 30%of the test data, which was left out, was correctly predicted. The low OOB score aligns with the low MSE.

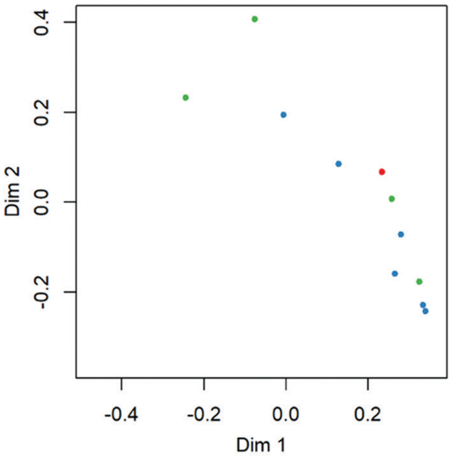


Figure 1. Multidimensional scaling plot for Model 1
Abbreviation: Dim: Dimension.

Similar to Model 1, a higher number of trees resulted in diminished performance due to the lack of additional information provided beyond 23 trees. The variable of importance is displayed in Table 3. B4 (retaking one's current job) demonstrated the most predictive power, proving to be the most important for accurate predictions and critical for Model 2's performance. In contrast, B2 (understanding patients' feelings) exhibited the least predictive power, being less important for accurate predictions and the least critical for the model's performance. Moreover, because B4 had the highest node purity score, it effectively split the model to predict scores and reduced error in the target variable C30. B2, on the other hand, had fewer splits and smaller decreases in impurity compared to B4. B4's high purity score was a result of a moderate positive correlation (0.5) between B14 and C30, whereas B2's lower purity score reflected a lower positive correlation (0.4) between B2 and C30. Although B4's high node purity score indicates its strong predictive power, it does not explain variability in C30 because it lacks a clear relationship with the target variable.

The MDS plot revealed that most participants scored differently from each other on questions related to understanding patients' and visitors' feelings, retaking one's current job, feeling stimulated when working with colleagues, effectively handling problems, feeling relaxed while managing emotional problems, and feeling exhilarated when working with and talking to patients. Several outlying participants were identified, as evidenced by data points that were not clustered. The separation of data points suggests that Model 2 is accurately classifying the data.

Table 3. Variable of importance of Model 2

Importance	Variable	Increase in node purity
1.	B4	12.81
2.	B16	9.65
3.	B13	7.13
4.	B18	6.83
5.	B14	6.65
6.	B3	4.34
7.	B2	3.43

To further assess model performance, we calculated cross-validation results, the area under the receiver operating characteristic (AUC-ROC) curves, recall, precision, and F1 scores. The models were tuned using 10-fold cross-validation, and performance was evaluated based on the number of splits in each model. Model 1's cross-validation results showed 58% accuracy and a Cohen's kappa of 34%, with two features randomly selected at each split in the decision tree. As the number of splits increased, performance worsened. At six features, accuracy and Cohen's kappa decreased to 55% and 32%, respectively. At 11 features, the accuracy dropped further to 52%, with a Cohen's kappa of 25%, as shown in Table 1. Notably, with two randomly selected features, the standard deviation of accuracy was small (19%), suggesting consistent performance across folds. The standard deviation of Cohen's kappa was also low (0.33%), indicating consistent agreement between observed and predicted scores. Despite the high accuracy at two randomly selected features, the low Cohen's kappa suggests only fair agreement between observed and predicted scores.

Similar to Model 1, Model 2 showed the best accuracy and Cohen's kappa with two randomly selected features. At two features, the model achieved 47% accuracy and a kappa of 0.19. With four features, accuracy decreased to 40%, with a kappa of 11%. At 11 features, the model showed 45% accuracy and a kappa of 17%. However, the standard deviations for both accuracy and kappa across all feature selections (2, 4, and 11) were low, indicating consistent model performance across folds and across observed and predicted values. Despite these findings, Model 1 outperformed Model 2 in terms of accuracy and kappa.

Model 1's AUC-ROC curve revealed an AUC of 0.57, indicating some predictive power, with the model correctly identifying positive and negative cases of score 5 approximately 57% of the time. However, the model demonstrated limited discriminatory power, as the AUC is relatively low, suggesting poor differentiation between classes. Given that the dataset involved a Likert scale, with many participants selecting 5 for question C30, the dataset may be imbalanced, which could explain the low AUC. The ROC curve's position above the diagonal line indicates performance slightly better than random, though the curve was not close to the top-left corner of the graph, which would indicate high sensitivity. The recall, precision, and F1 score for predicting scores of 5 were 0.5, 0.38, and 0.43, respectively, reflecting somewhat adequate performance but also indicating room for improvement in correctly identifying true positives and true negatives.

As shown in Figure 2, Model 2's AUC-ROC curve revealed an AUC of 0.6, suggesting relatively weak discriminatory power. The model correctly identified responses of 5 for the question regarding employees' perceptions of their organization's culture approximately 60% of the time, as illustrated in Figure 3. Model 2 showed weak discriminatory power, though its AUC, as shown in Figure 4, was slightly higher than that of Model 1, indicating somewhat better classification performance. The ROC curve, positioned above the diagonal line, reflects performance better than random, but it was still not close to the top-left corner, which would indicate high sensitivity and correct identification of positive perceptions of OC. The curve's position farther to the right suggests a higher false-positive rate, indicating that more scores that are not 5 are classified as 5. In addition, the recall of 0.6 for scores of 5 demonstrates that the model predicts 60% of true positives of score 5 correctly, while the precision of 0.43 means that 60% of the instances classified as positive are truly positive. The F1 score for Model 2's ability to correctly predict scores of 5 is 0.5, indicating somewhat adequate performance with room for improvement.

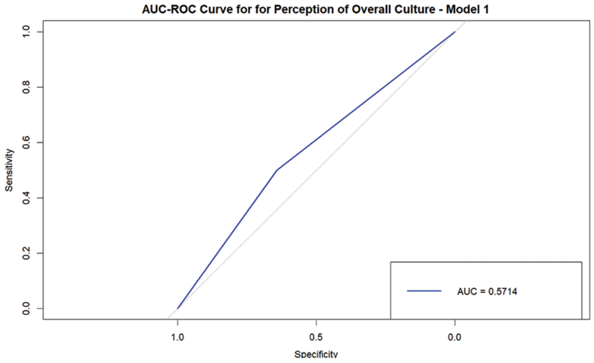


Figure 2. Area under (AUC) the receiver operating characteristic (ROC) curve for Model 1

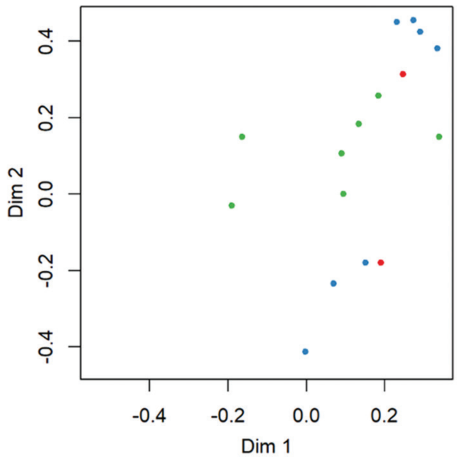


Figure 3. Multidimensional scaling plot for Model 2
Abbreviation: Dim: Dimension.

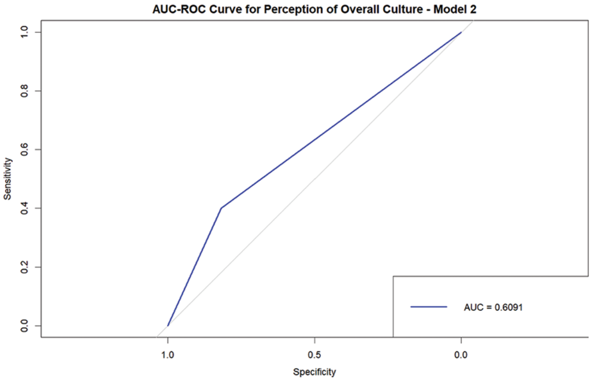


Figure 4. Area under (AUC) the receiver operating characteristic (ROC) curve for Model 2

5. Discussion

The analysis and model results demonstrate that employees' beliefs about OC predict and affect burnout symptoms. In Model 1, depersonalizing others – by feeling callous toward them and treating them as impersonal objects – was the most powerful predictor of other burnout symptoms and perceptions of overall OC. These findings align with several studies that emphasize the importance of depersonalization in predicting burnout and the impact of workplace factors on burnout.[35-37] However, this study highlights the significance of OC when predicting other burnout symptoms. One study also emphasizes the importance of physical characteristics, such as age, sex, and prior medical history (e.g., painkiller dependence), in predicting burnout.[38] Therefore, if an individual notices callousness toward others in themselves or their coworkers, it may be an indication of burnout symptoms linked to an OC that employees perceive negatively.

In addition, Model 1 reveals that feeling fatigued when waking in the morning has strong predictive power and may signal a negatively perceived OC. This finding aligns with another study showing that OC and fatigue are related and can impact turnover intentions.[39] For example, a volatile culture – where organizational elements such as rules and leadership continuously change – can lead to employee fatigue.[40] Further research is needed to identify specific aspects of OC that contribute to burnout symptoms, like fatigue.

On the other hand, feeling at the wits' end showed the least predictive power. Several studies have highlighted the importance of cultivating an OC where colleagues and supervisors support each other and where supervisors demonstrate empathy toward employees. However, feeling at the wits' end is a more individualized experience and a symptom of burnout, which may explain why it does not strongly predict perceptions of OC. Leaders should focus on a range of burnout symptoms, not just one, before determining whether their employees are burned out and respond by implementing additional safeguards, such as more flexible work schedules. This study suggests that feeling at the wits' end has a smaller impact on perceptions of OC compared to other variables. Further research is needed to quantify the impact of all variables included in the model on perceptions of OC.

Model 1's MDS plot in Figure 1 demonstrates a negative correlation between positive perceptions of OC and burnout symptoms. The more positively employees perceive their OC, the less likely they are to exhibit burnout symptoms, which aligns with findings in the literature.[37,41,42]

Model 2 shows that several engagement and personalization characteristics – such as feeling stimulated when working with colleagues and exhilarated when working with or talking to patients – are the most powerful predictors. Employees who are engaged in their work are less likely to experience burnout.[43,44] Models 1 and 2 underscore the importance of recognizing whether employees empathize with patients and coworkers, as this can predict burnout and whether OC is perceived positively.

Moreover, Model 2 demonstrates that the likelihood of an employee wanting to take their current job again holds significant predictive power. Turnover intention is important in recognizing burnout and indicating whether OC is positively perceived. Previous research has found similar results, showing that OC is related to turnover intention.[45-47] However, this study further illustrates how perceptions of OC can predict turnover intent.

Although depersonalizing others by treating them as impersonal objects and being callous toward them was identified as the most significant predictor of burnout using OC, another aspect of personalization – understanding visitors' feelings – was found to be the least important. This finding contrasts with those of previous studies. Empathizing with visitors, with whom non-patient-facing employees may have limited interact, may not be as crucial in predicting burnout as the way employees interact with coworkers and patients. Further research is needed to explore why empathizing with visitors is less important in predicting burnout than other depersonalization factors.

Model 2's MDS plot in Figure 3 reveals a positive correlation between perceptions of OC and burnout-related questions. The more positively employees perceive their OC, the more likely they are to be protected against burnout symptoms such as depersonalization. For example, employees who view their OC positively are more likely to understand visitors' feelings and feel stimulated when working with colleagues. This finding aligns with several studies showing that OC influences burnout symptoms, including depersonalization.[48-50] However, our approach is innovative in its ability to use a complex data model to predict burnout based on OC. The aforementioned studies did not employ a random forest model to predict burnout scores among non-patient-facing and patient-facing employees, highlighting the novelty of using a random forest algorithm for this purpose. Moreover, the models' strong performance, despite the small sample size, suggests that this approach can be expanded to predict burnout using OC and potentially other factors such as workload. The model's ability to predict burnout based on OC is supported by a previous study that used a decision tree model and Bayesian analysis on the same dataset.[32] The Bayesian analysis, which, like a random forest model, is suitable for small datasets, corroborates the results of this study.

The AUC-ROC curves for Models 1 and 2 demonstrate that the model classified, with some accuracy, whether employees perceived their organization's culture positively. Furthermore, the model suggests that burnout symptoms are more likely to be linked to perceptions of OC. This finding indicates to leaders that, if employees exhibit burnout symptoms, they may also perceive the OC as negative. Leaders could use the model's identification of specific burnout symptoms – such as depersonalization, as indicated by callousness toward others – to evaluate perceptions of OC. For instance, if leaders recognize that their employees are treating others harshly and exhibiting other burnout symptoms identified by the model (e.g., viewing others as impersonal objects), they could begin to foster a more positive OC. Strategies could include providing employees with the necessary equipment and resources to succeed, as well as offering flexible work schedules to promote work-life balance and reduce burnout symptoms.

While this exploratory study introduced a novel random forest method for predicting burnout using OC, it has several limitations. One notable limitation is the small sample size. Although a previous study's Bayesian analysis on the same dataset supports the results of the random forest model, other ML methods suited for small sample sizes, such as SVMs, could be explored in future studies.[32] Nonetheless, this exploratory study provides a framework and methodology for demonstrating how OC influences burnout. Moreover, this study was cross-sectional, meaning causality could not be established. It was also conducted solely in the United States. To determine the generalizability of the results, the study could be replicated in other countries.

The small dataset may have led to model overfitting, meaning it might have learned patterns specific to the training data rather than those applicable to more generalizable data. The model could have overreacted to small variations in the data, such as outliers or the responses of a few respondents who rated their OC as positive with a "1." Moreover, the small dataset may not have captured the complex, diverse perspectives of employees across different health systems, such as those who viewed their OC negatively but did not exhibit significant burnout symptoms. As a result, the model may not have fully learned all the relationships between the features, potentially performing poorly on unseen data due to noise specific to the small dataset. Future studies should incorporate larger datasets with a broader representation of various roles within health systems to improve generalizability, reduce bias, and prevent overfitting.

6. Conclusions and policy implications

This random forest model demonstrates that perceptions of OC can be used to predict specific burnout symptoms, such as engagement with others and empathy. The model also reveals that the more positively employees perceive their organization's culture, the less likely they are to exhibit burnout symptoms. Previous studies have quantified the effects of OC on burnout,[50,51] and this model further highlights that OC influences burnout more significantly than internal factors, such as resilience and self-care. If employees begin to display burnout symptoms, leaders could assess and improve the OC to mitigate these effects. Therefore, it is crucial for leaders in health systems to cultivate an OC where colleagues and employees are supported, workloads are manageable, and the resources needed to perform the job effectively and efficiently are provided.[52-55] As a result, perceptions of OC will improve, and employees will be less likely to experience burnout.

The findings also suggest that policymakers could invest in improving health systems' work environments by providing flexible work options and reducing workload through strategies such as increasing staffing. For example, given the shortage of healthcare workers, particularly nurses and physicians, in the United States, hiring international employees could help alleviate the staffing crisis. Policymakers could ease the hiring process by loosening visa requirements for international employees. In addition, health system leaders and policymakers should focus on internal factors contributing to burnout, such as mental health. Policies could be implemented to invest more in mental health resources, such as crisis hotlines. Leaders in health systems could also establish confidential peer support groups, enabling employees to discuss their mental health concerns without fear of their discussions being shared with colleagues or supervisors. Investing in OC and burnout mitigation strategies, such as peer groups, could significantly reduce burnout.

6.1. Contributions to the literature

This study makes several contributions to the existing literature. First, it is the first study to use a random forest model to predict burnout using perceptions of a positive OC as the DV, particularly in the context of the COVID-19 pandemic. A previous study only used burnout survey results to predict burnout symptoms in a machine-learning model.[23] Second, the random forest model could serve as a baseline for tuning and creating additional models to predict burnout using other aspects of OC, such as supervisor support and the availability of resources needed to perform one's job – factors not yet explored in the literature. Third, the model underscores the importance of cultivating a positively perceived OC to reduce burnout by demonstrating that OC can predict burnout scores. Fourth, these models introduce an innovative way to predict burnout and could be expanded to include employees outside of health systems. Finally, this study illustrates that advanced and complex ML techniques can effectively predict burnout.

6.2. Contributions to the healthcare industry

This study offers valuable insights with practical implications for the healthcare sector.

(i) Burnout symptom prediction and prevention
Based on the model and OC survey results, healthcare leaders, including hospital chief executive officers and human resource executives, can anticipate specific burnout symptoms, such as depersonalization among employees toward their coworkers, patients, and visitors. Depersonalization, which may negatively impact patient satisfaction, requires leaders to prioritize and improve their OC. Given that patient satisfaction is critical in healthcare, enhancing OC becomes a strategic imperative.

(ii) Burnout interconnection awareness
The findings highlight the strong predictive relationships among various burnout symptoms, such as callousness toward others and reluctance to retake one's current job. The strong association between burnout symptoms and turnover highlights the importance of addressing burnout to reduce costly staff attrition. Leaders should monitor early signs of burnout, as indicated by the model, and investigate its causes, including issues related to OC.

(iii) OC monitoring
Conducting regular employee satisfaction surveys helps leaders identify changes in OC and detect potential issues before they escalate.[56,57] Surveys should include questions specific to how employees perceive the level of support available from colleagues and supervisors.[58] When employees report feeling unsupported, leadership can take actionable steps, such as advocating for employee needs during executive discussions and ensuring that staff have the necessary resources, supplies, and equipment to perform their jobs effectively.

(iv) Actionable strategies for improvement

To enhance OC and reduce burnout, healthcare leaders can implement actionable strategies based on best practices. Some of these strategies may include:

- Flexible work schedules: Providing flexibility to employees helps them balance work and personal responsibilities, ensuring they come to work refreshed and with time for their loved ones.[59,60]
- Mental health resources: Offering access to confidential, 24/7 mental health support services can help employees manage stress, both at work and in their personal lives.[61,62]
- Promoting peer support systems: ENCOURAGING peer support groups or mentoring programs can help employees feel supported and less isolated at work.[63,64]
- Case studies for inspiration: Learning from successful case studies or applying concepts from recent literature can guide leaders in designing interventions aimed at reducing burnout and improving OC.

By actively using these strategies, healthcare organizations can create an environment that prioritizes employees' well-being, reduces burnout, and ultimately improves the quality of patient care. The model's finding that "feeling at wit's end" has low predictive power on perceptions of OC suggests that organizational leaders should not focus on an individual's feelings of frustration, as these may be temporary. Instead, leaders should focus on how employees treat visitors, coworkers, patients, and others in the health system. The high predictive power of callousness toward others indicates that leaders should focus on modeling empathy and appropriate behavior expectations for employees. Leaders can demonstrate empathy by showing consideration for different perspectives, particularly when addressing dilemmas. For example, when speaking with patients, leaders could use open-ended questions and offer choices rather than presenting authoritative statements with only one solution.[65] In addition, using clear, direct language can improve patients' understanding while demonstrating personal concern, such as asking how patients are feeling, can enhance engagement.

In addition, retaking one's current job showed strong predictive power on perceptions of OC, suggesting that leaders should focus on cultivating an OC in which employees are motivated to stay and are retained. Leaders cannot control or be responsible for employees’ individual feelings, but they can make external changes to the OC that improve employee retention. Leaders should focus on creating a positive work environment by implementing changes such as flexible schedules and hybrid work arrangements, where employees work remotely for part of the week. Monitoring retention rates before and after implementing such changes can help assess the effectiveness of these efforts to cultivate an OC where employees want to remain.

Acknowledgments

The authors would like to acknowledge Mark A. Newman for his support in advising on data collection and guiding the survey design.

Funding

None.

Conflict of interest

The authors declare that they have no competing interests.

Author contributions

Conceptualization: All authors
Formal analysis: Teray Johnson
Investigation: Teray Johnson
Methodology: All authors
Writing – original draft: Teray Johnson
Writing – review & editing: Sameh Shamroukh

Ethics approval and consent to participate

This research was approved by the Harrisburg University of Science and Technology Institutional Review Board (IRB# 20221026). Participants gave informed consent to participate in this study.

Consent for publication

Written consent was obtained at the beginning of the survey, and permission was granted by each participant to publish their data.

Availability of data

Data can be obtained by contacting the corresponding author.

References

1. Schein EH. *Organizational Culture and Leadership*. San Francisco: Jossey-Bass Publishers; 2010.
2. Hofstede G. *Culture’s Consequences: Comparing Values, Behaviors, Institutions and Organizations Across Nations*. 2nd ed. Thousand Oaks: Sage; 2001.
3. Maslach C, Leiter MP. Burnout. In: *Stress: Concepts, Cognition, Emotion, and Behavior*. Netherlands: Elsevier; 2016. p. 351-357.

   doi: 10.1016/B978-0-12-800951-2.00044-3
4. Schaufeli WB, Leiter MP, Maslach C. Burnout: 35 years of research and practice. *Career Dev Int*. 2009;14(3):204-220.

   doi: 10.1108/13620430910966406
5. Tawfik DS, Scheid A, Profit J, *et al*. Evidence relating health care provider burnout and quality of care. *Ann Intern Med*. 2019;171(8):555.

   doi: 10.7326/M19-1152
6. Cameron K, Quinn R. *Diagnosing and Changing Organizational Culture: Based on the Competing Values Framework*. 3rd ed. United States: John Wiley & Sons; 2011.
7. Schaufeli WB, Bakker AB. Job demands, job resources, and their relationship with burnout and engagement: A multi-sample study. *J Organ Behav*. 2004;25(3):293-315.

   doi: 10.1002/job.248
8. Maslach C, Leiter MP. Early predictors of job burnout and engagement. *J Appl Psychol*. 2008;93(3):498-512.

   doi: 10.1037/0021-9010.93.3.498
9. Alarcon GM. A meta-analysis of burnout with job demands, resources, and attitudes. *J Vocat Behav*. 2011;79(2):549-562.

   doi: 10.1016/j.jvb.2011.03.007
10. Chatterjee S, Goyal D, Prakash A, Sharma J. Exploring healthcare/health-product ecommerce satisfaction: A text mining and machine learning application. *J Bus Res*. 2021;131:815-825.

    doi: 10.1016/j.jbusres.2020.10.043
11. Bhardwaj S, Bhardwaj H, Bhardwaj J, Gupta P. Global prediction of COVID-19 Cases and Deaths using Machine Learning. In: *2021 Sixth International Conference on Image Information Processing (ICIIP)*. IEEE; 2021. p. 422-426.

    doi: 10.1109/ICIIP53038.2021.9702560
12. Ang CY, Lee MH, Paramesran R. A survey of machine learning techniques for behavioral data analysis on mobile devices. *Inform Fusion*. 2020;49:46-60.
13. Shatte ABR, Hutchinson DM, Teague SJ. Machine learning in mental health: A scoping review of methods and applications. *Psychol Med*. 2019;49(9):1426-1448.

    doi: 10.1017/S0033291719000151
14. Podsakoff PM, MacKenzie SB, Lee JY, Podsakoff NP. Common method biases in behavioral research: A critical review of the literature and recommended remedies. *J Appl Psychol*. 2003;88(5):879-903.

doi: 10.1037/0021-9010.88.5.879

15. DeVellis RF, Thorpe CT. *Scale Development: Theory and Applications*. 5th ed. United States: SAGE Publications; 2022.

16. Braun V, Clarke V. Using thematic analysis in psychology. *Qual Res Psychol*. 2006;3(2):77-101.

    doi: 10.1191/1478088706qp063oa

17. Bakker AB, Demerouti E. The Job Demands-Resources model: State of the art. *J Manag Psychol*. 2007;22(3):309-328.

    doi: 10.1108/02683940710733115

18. Dewa CS, Loong D, Bonato S, Thanh NX, Jacobs P. How does burnout affect physician productivity? A systematic literature review. *BMC Health Serv Res*. 2014;14(1):325.

    doi: 10.1186/1472-6963-14-325

19. Sonnentag S, Frese M. Stress in organizations. In: Borman WC, Ilgen DR, Kilmoski RJ, editors. *Comprehensive Handbook of Psychology*. Vol. 12. United States: Wiley; 2003. p. 454-481.

20. Oei MW, Pasinringi SA, Sidin AI, Noor NB, Rivai F, Healty. The influence of organizational culture and group climate on innovation readiness at Hasanuddin university hospital. *Pharmacogn J*. 2024;16(1):150-155.

    doi: 10.5530/pj.2024.16.21

21. Lee T, Yoon YS, Ji Y. Predicting new graduate nurses' retention during transition using decision tree methods: A longitudinal study. *J Nurs Manag*. 2024;2024:1-11.

    doi: 10.1155/2024/4687000

22. Louhenapessy EL, Lindawati T. Effect of emotional intelligence and organizational culture on performance with organizational commitment as mediation. *Eduvest J. Univ. Stud*. 2022;2(3):554-570.

    doi: 10.36418/edv.v2i3.384

23. Tavella G, Spoelma M, Parker G. Detecting burnout: Identifying key symptoms using standard and machine learning methods. *Int J Stress Manag*. 2023;30(4):366-375.

    doi: 10.1037/STR0000296

24. Grządzielewska M. Using machine learning in burnout prediction: A survey. *Child Adolesc Soc Work J*. 2021;38(2):175-180.

    doi: 10.1007/s10560-020-00733-w

25. Taylor S, Jaques N, Nosakhare E, Sano A, Picard R. Personalized multitask learning for predicting tomorrow's mood, stress, and health. *IEEE Trans Affect Comput*. 2020;11(2):200-213.

    doi: 10.1109/TAFFC.2017.2784832

26. Bauernhofer K, Bassa D, Canazei M, *et al.* Subtypes in clinical burnout patients enrolled in an employee rehabilitation program: Differences in burnout profiles, depression, and recovery/resources-stress balance. *BMC Psychiatry*. 2018;18(1):10.

    doi: 10.1186/s12888-018-1589-y

27. Lee YL, Chou W, Chien TW, Chou PH, Yeh YT, Lee HF. An app developed for detecting nurse burnouts using the convolutional neural networks in microsoft excel: Population-based questionnaire study. *JMIR Med Inform*. 2020;8(5):e16528.

    doi: 10.2196/16528

28. Park J, Feng Y, Jeong SP. Developing an advanced prediction model for new employee turnover intention utilizing machine learning techniques. *Sci Rep*. 2024;14(1):1221.

    doi: 10.1038/s41598-023-50593-4

29. del Pozo-Antúnez JJ, Molina-Sánchez H, Ariza-Montes A, Fernández-Navarro F. Promoting work engagement in the accounting profession: A machine learning approach. *Soc Indic Res*. 2021;157(2):653-670.

    doi: 10.1007/s11205-021-02665-z

30. Schaufeli WB, Salanova M, González-romá V, Bakker AB. The measurement of engagement and burnout: A two sample confirmatory factor analytic approach. *J Happiness Stud*. 2002;3(1):71-92.

    doi: 10.1023/A:1015630930326

31. Bailey C, Madden A, Alfes K, Fletcher L. The meaning, antecedents and outcomes of employee engagement: A narrative synthesis. *Int J Manage Rev*. 2017;19(1):31-53.

    doi: 10.1111/ijmr.12077

32. Johnson T, Shamroukh S. Predictive modeling of burnout based on organizational culture perceptions among health systems employees: A comparative study using correlation, decision tree, and Bayesian analyses. *Sci Rep*. 2024;14(1):6083.

    doi: 10.1038/s41598-024-56771-2

33. Kovner CT, Brewer CS, Fatehi F, Katigbak C. Changing trends in newly licensed RNs. *Am J Nurs*. 2014;114(2):26-34; quiz 35.

    doi: 10.1097/01.NAJ.0000443767.20011.7f

34. Kovner CT, Brewer CS, Fairchild S, Poornima S, Kim H, Djukic M. Newly licensed RNs' characteristics, work attitudes, and intentions to work. *Am J Nurs*. 2007;107(9):58-70.

    doi: 10.1097/01.NAJ.0000287512.31006.66

35. Havaei F, Ji XR, MacPhee M, Straight H. Identifying the most important workplace factors in predicting nurse mental health using machine learning techniques. *BMC Nurs*. 2021;20(1):216.

    doi: 10.1186/S12912-021-00742-9

36. Lee MCC, Ding AYL. The relationship between market culture, clan culture, benevolent leadership, work engagement, and job performance: Leader's dark triad as a

moderator. *Psychol Rep*. 2024;127(2):887-911.

doi: 10.1177/00332941221121564

37. Roth C, Berger S, Krug K, Mahler C, Wensing M. Internationally trained nurses and host nurses’ perceptions of safety culture, work-life-balance, burnout, and job demand during workplace integration: A cross-sectional study. *BMC Nurs*. 2021;20(1):77.

    doi: 10.1186/s12912-021-00581-8

38. Sun JW, Franklin JM, Rough K, *et al.* Predicting overdose among individuals prescribed opioids using routinely collected healthcare utilization data. *PLoS One*. 2020;15(10):e0241083.

    doi: 10.1371/JOURNAL.PONE.0241083

39. Lee E, Jang I. Nurses’ fatigue, job stress, organizational culture, and turnover intention: A culture-work-health model. *West J Nurs Res*. 2020;42(2):108-116.

    doi: 10.1177/0193945919839189

40. Yildizoğlu H, Cemaloğlu N. Examination of the relationship between teachers’ organizational change fatigue levels and the organizational resilience levels of secondary education institutions. *Kastamonu Eğitim Derg*. 2023:186-201.

    doi: 10.24106/KEFDERGI.1271471

41. Leiter M, Maslach C. *The Truth About Burnout: How Organizations Cause Personal Stress and What to.- Christina Maslach, Michael P. Leiter-Google Books*. 1st ed. Jossey-Bass; 1997. Available from: https://books.google.com/books?hl=en&lr=&id=uzomWhbqcCAC&oi=fnd&pg=PR7&dq=the+truth+about+burnout+how+organizations+cause+personal+stress&ots=9B9sqoSFTB&sig=L4MGTLfkwsYzLRc4Jj4TZXr-cak#v=onepage&q=the%20truth%20about%20burnout%20how%20organizations%20cause%20personal%20stress&f=false [Last accessed on 2022 Jun 28].

42. Montgomery A, Todorova I, Baban A, Panagopoulou E. Improving quality and safety in the hospital: The link between organizational culture, burnout, and quality of care. *Br J Health Psychol*. 2013;18(3):656-662.

    doi: 10.1111/bjhp.12045

43. Maslach C. Understanding burnout: Definitional issues in analyzing a complex phenomenon. In: *Job Stress and Burnout*. Sage; 1982. p. 29-40. Available from: https://scholar.google.com/scholar?hl=en&as_sdt=0%2C31&q=Understanding+burnout%3A+Definitional+issues+in+analyzing+a+complex+phenomenon.+InJob+Stress+and+Burnout&btnG= [Last accessed on 2022 Jun 28].

44. Maslach C, Leiter M. Burnout and engagement in the workplace: A contextual analysis. *Adv Motivat Achiev*. 1999;11:275-302.

45. Gardner JK, Thomas-Hawkins C, Fogg L, Latham CE. The relationships between nurses’ perceptions of the hemodialysis unit work environment and nurse turnover, patient satisfaction, and hospitalizations. *Nephrol Nurs J*. 2007;34:271-281.

46. Wright ES. The effects of organizational culture on employee turnover. *Perform Improv Q*. 2021;34(3):303-319.

    doi: 10.1002/PIQ.21372

47. Lee MS, Lee BS, Kim TK, Lee JY. The factors affecting the turnover intention of nurses at a hospital. *Indian J Public Health Res Dev*. 2018;9(3):539-544.

    doi: 10.5958/0976-5506.2018.00346.7

48. Lim S, Shin S. Moderating effect of nursing organizational culture on the relationship between verbal violence and burnout in operating room nurses. *Korean J Adult Nurs*. 2022;34(2):196.

    doi: 10.7475/kjan.2022.34.2.196

49. Lopes MCC, Oliva CCC, Bezerra NMS, Silva MT, Galvão TF. Relationship between depressive symptoms, burnout, job satisfaction and patient safety culture among workers at a university hospital in the Brazilian Amazon region: Cross-sectional study with structural equation modeling. *Sao Paulo Med J*. 2022;140(3):412-421.

    doi: 10.1590/1516-3180.2021.0614.15092021

50. de Lima Garcia C, Bezerra IMP, Ramos JLS, do Valle JETMR, de Oliveira MLB, de Abreu LC. Association between culture of patient safety and burnout in pediatric hospitals. *PLoS One*. 2019;14(6):e0218756.

    doi: 10.1371/journal.pone.0218756

51. Williams ES, Manwell LB, Konrad TR, Linzer M. The relationship of organizational culture, stress, satisfaction, and burnout with physician-reported error and suboptimal patient care: Results from the MEMO study. *Health Care Manage Rev*. 2007;32(3):203-212.

    doi: 10.1097/01.HMR.0000281626.28363.59

52. Hossny EK, Morsy SM, Ahmed AM, Saleh MSM, Alenezi A, Sorour MS. Management of the COVID-19 pandemic: Challenges, practices, and organizational support. *BMC Nurs*. 2022;21(1):196.

    doi: 10.1186/S12912-022-00972-5

53. Konlan KD, Asampong E, Dako-Gyeke P, Glozah FN. Burnout and allostatic load among health workers engaged in human resourced-constrained hospitals in Accra, Ghana. *BMC Health Serv Res*. 2022;22(1):1163.

    doi: 10.1186/S12913-022-08539-5

54. Patterson PB, Mumtaz Z, Chirwa E, Mambulasa J, Kachale F, Nyagero J. Culture’s place in quality of care in a resource-constrained health system: Comparison between three Malawi districts. *Qual Health Res*. 2021;31(13):2528-2541.

    doi: 10.1177/10497323211037636

55. Welp A, Meier LL, Manser T. Emotional exhaustion and workload predict clinician-rated and objective patient safety. *Front Psychol*. 2015;5:1573.

doi: 10.3389/fpsyg.2014.01573

56. Kepplinger A, Braun A, Fringer A, Roes M. Opportunities for nurses to address employee voice in health care providers: A scoping review. *BMC Nurs*. 2024;23(1):651.

doi: 10.1186/s12912-024-02331-y

57. Pamungkas W. Determining employee job satisfaction and work productivity: An analysis of HR competency and organizational culture at PT. XYZ. *Dinasti Int J Manage Sci*. 2024;5(5):1017-1026.

58. Gill-Bonanca K. Mentorship: A strategy for nursing retention. *Am Nurse J*. 2024;19(8):6-12.

doi: 10.51256/ANJ082406

59. Reddy S, Sai P. Impact of hybrid work model on employee engagement and productivity. *Amity Bus Rev*. 2024;25(1):42-51.

60. Lee SE, Shin G. The effect of perceived organizational and supervisory support on employee engagement during COVID-19 crises: Mediating effect of work-life balance policy. *Public Pers Manage*. 2023;52(3):401-428.

doi: 10.1177/00910260231171395

61. Melnyk BM, Davidson JE, Mayfield C, *et al*. A study protocol for the modified interactive screening program plus MINDBODYSTRONG© RCT: A mental health resiliency intervention for nurses. *PLoS One*. 2024;19(6):e0303425.

doi: 10.1371/journal.pone.0303425

62. Virani T, Ismail F, Brooke J. Beyond traditional practices: Innovating workplace mental health in the community. *Healthc Q*. 2024;27(3):28-33.

doi: 10.12927/hcq.2024.27444

63. Maloney LM, Hoffman J, Peralta E, *et al*. Informal peer support and intentional acts of kindness may attenuate the impact of work-related stressors on compassion satisfaction, secondary traumatic stress, and burnout of emergency medical services clinicians. *Air Med J*. 2024;43(4):333-339.

doi: 10.1016/j.amj.2024.03.005

64. Jensen B. South Dakota physician well-being program starts peer-support groups: Social gatherings combat burnout and improve resiliency. *S D Med*. 2023;76(3):138-139.

65. Alam F, Danieli M, Riccardi G. Annotating and modeling empathy in spoken conversations. *Comput Speech Lang*. 2018;50:40-61.

doi: 10.1016/j.csl.2017.12.003

Appendix

Table A1. Departments

Department	Number of respondents
Administration	20
Board of Directors	4
Community Health	3
Diagnostic Imaging/Radiology	1
ED	4
Inpatient Psychiatry	1
Inpatient Surgery	2
Lab Services	1
Med/surg	2
OR/PACU	1
Pharmacy	1
Quality and Patient Safety	4
Step down/Transitional	1
Analytics	4
Case Management/Social Work	1
Finance	1
Graduate Medical Education	1
Information Technology	2
Marketing/Business Development	2
Operations	1
Strategy	5
Patient Experience	1
Radiation Oncology	2
Refused to Answer	3

Abbreviations: ED: Emergency department; OR: Operating theatre; PACU: Post-Anesthesia care unit.

Table A2. Positions

Position title	Number of respondents
Administrator/Assistant Administrator/Supervisor	18
Board Member	3
Consultant	4
Case Manager	1
Clerical	1
IT Support	2
Medical Assistant	12
Medical Doctor	10
Patient Coordinator/Access Representative	3
Chaplain	1
Physician Chief/Chair	1
Project Manager	1
Registered Nurse	7
Residency Coordinator	1
Others	2

Abbreviation: Information technology.

Table A3. Hospital size, location, and type

Parameter	Number of respondents
Number of beds	
6 – 24	3
25 – 49	0
50 – 99	3
100 – 199	7
200 – 299	3
300 – 399	5
400 – 499	8
500 or more	34
Location	
Rural	22
Urban	41
Type	
Federal	2
For-profit	16
Non-profit	45

Key

C30–Overall, the culture of the hospital is positive.
B1–When I go home after work, I feel emotionally drained.
B2–I understand my patients' feelings.
B3–I understand visitors' feelings.
B4–Knowing what I know now, I would take my current job all over again.
B5–My job interferes with my home-life.
B6–My job keeps me from spending the amount of time I would like with my family.
B7–I often get irritated by little annoyances.
B8–I suffer from anxiety.
B9–My mood often goes up and down.
B10–There are days when I'm "on edge" all the time.
B11–I feel fatigued when I get up in the morning.
B12–I feel like I'm at my wits' end.
B13–I feel stimulated when I work with my colleagues.
B14–I deal very effectively with the problems of my patients and/or coworkers.
B15–I treat some of my patients and/or coworkers like they're impersonal objects.
B16–In my work, I'm very relaxed when dealing with emotional problems.
B17–I've become more callous toward people since starting my current job.
B18–I feel exhilarated after working with or talking to patients.

ACCSCIENCE PUBLISHING

Artificial Intelligence in Health

ORIGINAL RESEARCH ARTICLE

Enhancing COVID-19 severity assessment with artificial intelligence-based bone suppression technique in chest radiography

Asumi Yamazaki[1], **Masashi Seki**[2], **and Takayuki Ishida**[1]*

[1]Division of Health Sciences, Osaka University Graduate School of Medicine, Suita, Osaka, Japan

[2]Department of Radiology, Kitasato University Hospital, Sagamihara, Kanagawa, Japan

Abstract

Chest radiography (CXR) is widely used for initial respiratory assessment, but its lesion detection capability is typically inferior to that of computed tomography. Several studies have reported that artificial intelligence (AI)-based bone suppression techniques can enhance the accuracy of lesion detection and disease classification. Previously, we developed an AI-based bone suppression system based on dual-energy subtraction principles. However, the subtraction process limited its versatility and introduced significant artifacts. To overcome these challenges, we improved the system to generate bone-suppressed images directly, eliminating the need for subtraction. This study demonstrates the utility of the updated bone suppression system as a pre-processing tool for regression analyses in assessing coronavirus disease 2019 severity. Four regression models – DenseNet, ResNet18, ResNet50, and RegNetY-120 – were employed to predict the severity based on scores annotated by radiologists. Except for DenseNet, all models showed statistically significant improvements in Pearson correlation coefficients (PCCs) when using bone-suppressed images generated by the updated model. The highest PCC, 0.895, was achieved by the ResNet18 model. The direct image generation process improved the clinical practicality of the bone suppression system while reducing artifacts. Furthermore, the significant improvement in linearity suggests that AI-driven bone suppression enhances the visibility of abnormalities and improves the accuracy for pulmonary condition assessments. These advancements could expand the application of bone suppression techniques in various regression analyses, including disease severity, progression, and recurrence risk. Nonetheless, further validation using larger and more diverse datasets, as well as a broader range of prediction models, is necessary.

Keywords: Bone suppression; Artificial intelligence; COVID-19; Regression model; Chest radiography

***Corresponding author:**
Takayuki Ishida
(tishida@sahs.med.osaka-u.ac.jp)

Citation: Yamazaki A, Seki M, Ishida T. Enhancing COVID-19 severity assessment with artificial intelligence-based bone suppression technique in chest radiography. *Artif Intell Health*. 2025;2(3):95-106. doi: 10.36922/aih.5608

Received: October 28, 2024

1st revised: January 10, 2025

2nd revised: February 11, 2025

Accepted: February 21, 2025

Published online: March 5, 2025

Copyright: © 2025 Author(s). This is an Open-Access article distributed under the terms of the Creative Commons Attribution License, permitting distribution, and reproduction in any medium, provided the original work is properly cited.

Publisher's Note: AccScience Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

Coronavirus disease 2019 (COVID-19) emerged in late 2019,[1] rapidly developing into a global pandemic that overwhelmed healthcare systems due to its high contagiousness, unprecedented morbidity, and mortality rates.[2,3] Since then, with the spread of various mutant strains, over seven million deaths have been reported worldwide.[4] While severe cases of COVID-19 can lead to death, many patients remain asymptomatic or experience mild symptoms.[5] Mild cases are often self-limiting and do not require hospitalization, whereas severe cases necessitate admission to the intensive care unit; mechanical ventilation, including extracorporeal membrane oxygenation; and treatment with anti-inflammatory agents.[6] An early and accurate diagnosis, along with a reliable severity assessment, is critical for effective patient management and preventing the overburdening of healthcare facilities.

Chest radiography (CXR) is widely used for initial evaluations in medical emergencies due to its low cost, rapid examination, and widespread availability.[7] While CXR is useful for detecting COVID-19, particularly during pandemic situations, its sensitivity is lower than that of chest computed tomography (CT).[8,9] Typical imaging patterns observed in COVID-19 patients include bilateral, peripheral, and basal-predominant ground-glass opacities without pleural effusion.[8,10,11] Although CT is generally considered more reliable for identifying such subtle contrast lesions,[9] the potential of image pre-processing techniques to enhance COVID-19 detection using CXR has been extensively explored.[12-14] For instance, Sharifrazi *et al.*[12] demonstrated that Sobel filter pre-processing improves the performance of convolutional neural network (CNN) models for detecting COVID-19 using CXR. Similarly, Arias-Garzón *et al.*[13] reported that pre-processing with segmentation techniques leads to a more accurate and reliable COVID-19 classification. Avolio *et al.*[14] further suggested that pre-processing methods could benefit not only CNN-based approaches but also other machine learning techniques, such as multiple-instance learning.

Furthermore, Takaki *et al.*[15] demonstrated that bone suppression processing through temporal subtraction significantly increased radiologists' sensitivity and reduced their false-positive rates in detecting pulmonary lesions on CXR. In contrast, van der Heyden[16] suggested that dual-energy subtraction (DES) technology enhances the diagnostic accuracy of CXR by eliminating bone structures and improving soft-tissue visualization. Building on these conventional bone suppression methods, advancements in artificial intelligence (AI) have introduced innovative alternatives for enhancing bone suppression.[17-30] Unlike conventional DES systems, which rely on dual X-ray exposure and specialized detectors equipped with copper plates, AI-based DES (AI-DES) approaches improve accessibility while offering additional advantages, such as reduced motion artifacts and improved noise characteristics. A representative study employed a generative adversarial network (GAN)-based model to suppress bone structures in CXR without the need for labeled data, utilizing digitally reconstructed radiographs from CT data.[18] Recent advancements have further extended these applications to pediatric imaging.[19,20] For example, Xie *et al.*[20] utilized bone edge detection to generate both bone-enhanced and bone-suppressed images, demonstrating the potential for improved pneumonia diagnosis in children.

We also previously developed an AI-DES system that successfully generated bone-suppressed images; however, its clinical applicability was limited by the need for raw high- and low-energy images, which are often unavailable, and by labor-intensive weighted subtraction processing.[22] To overcome these limitations, we updated the system to generate bone-suppressed images directly from routine chest radiographs in this study. This improvement enhances the system's versatility and clinical utility by streamlining the entire image-processing workflow.

Several studies have demonstrated the advantages of AI-based bone suppression techniques in classification tasks related to various lung diseases, including COVID-19.[23-29] In particular, Rani *et al.*[23] proposed a model that preserves spatial features, suggesting that effectively pre-processed radiographs could enhance diagnostic performance. The authors further reported that integrating AI-based bone suppression with pre-processing techniques, such as lung segmentation and augmentation, significantly improves the classification accuracy of pneumonia, including COVID-19.[24] Lam *et al.*[25] reported that bone-suppressed images, generated using a CNN-based model proposed by Rajaraman *et al.*,[26] significantly increased the area under the receiver operating characteristic curve for COVID-19 classification tasks, compared to standard radiographs using a modified VGG16 model. Similarly, Xu *et al.*[30] developed a CNN-based rib removal model, SADXNet, which showed superior performance in lung nodule detection, lung anomaly classification, and localization tasks. These findings highlight the importance of bone suppression in enhancing the accuracy of lung disease classification.

Given these promising results, we hypothesize that bone suppression techniques may also be effective in regression tasks, such as assessing disease severity, predicting the risk of progression, and estimating patient prognosis. Regression models, which predict continuous values by identifying subtle variations, can be more complex and challenging to construct than classification models, particularly when dealing with high-dimensional or imbalanced data.[31] This study explores the utility of AI-based bone suppression in COVID-19 severity assessment using CXR, comparing the performance of regression models with and without bone suppression to validate its effectiveness.

Recent studies have proposed AI-based prediction models for COVID-19 severity assessment.[32-34] Cohen *et al.*[32,33] used DenseNet-based regression models to evaluate COVID-19 severity based on the extent of lung involvement and the degree of opacity, achieving Pearson correlation coefficients (PCCs) of 0.80 and 0.78 for these tasks, respectively. Signoroni *et al.*[34] introduced BS-Net, an end-to-end architecture, to segment, align, and quantify lung compromise based on the Brixia score.[35,36] The performance of BS-Net was evaluated not only for classification tasks but also for regression tasks using linear regression of the Brixia score, with the highest PCC reaching 0.85. These studies underscore the potential of AI approaches in evaluating COVID-19 severity.

Moreover, transparency, explainability, and interpretability are critical components of AI, especially in medical applications.[37] Understanding why and how a model derives a particular decision is essential for ensuring clinical accountability and building confidence. Gradient-weighted class activation mapping (Grad-CAM) is a widely used technique to visualize decision-making processes by highlighting image regions that contribute to the model's output.[38] By generating heatmaps based on gradients from the final convolutional layer, Grad-CAM offers explainable insights to support clinical decision-making.[39,40] Talaat *et al.*[40] integrated Grad-CAM into a breast cancer classification model, providing radiologists with valuable insights into the model's decision-making process and fostering trust in the AI system. In this study, Grad-CAM is used to validate the explainability and interpretability of COVID-19 severity prediction models.

Our work integrates AI-based bone suppression pre-processing into regression models to assess COVID-19 severity using CXR. The primary aim is to expand the applications of AI-based bone suppression techniques, verifying their utility in severity assessment. By improving the accuracy of severity predictions, this approach could enhance patient monitoring and optimize healthcare resource allocation, particularly in resource-limited settings. Our findings may also validate the applications of AI-based bone suppression in regression tasks for chest image diagnosis. Moreover, this study seeks to bridge the gap between the present limitations of CXR and the superior sensitivity of CT, ultimately contributing to more efficient and scalable diagnostic tools for COVID-19 and other pulmonary diseases.

2. Data and methods

In this section, we explain the development of the bone suppression model, followed by the method for assessing COVID-19 severity.

2.1. Bone suppression model

2.1.1. Data collection

We collected chest radiographs from 600 patients using a dual-shot DES system (Discovery XR656, GE Healthcare, Chicago, IL, USA) at Kitasato University Hospital (Sagamihara City, Japan), to develop a bone suppression model. Most of these patients had pulmonary inflammatory diseases or pulmonary mass lesions. The detector specifications are detailed in our previous work.[22] Radiography was performed with tube voltages of 130 kV for high-energy images and 60 kV for low-energy images. The system produces bone-suppressed and bone-enhanced images, along with standard chest radiographs for presentation, all with a resolution of 3524 × 4288 pixels and 13-bit contrast, from the raw data of the high- and low-energy images. For training, we utilized 480 pairs of standard and bone-suppressed radiographs, while 120 pairs were reserved for testing.

2.1.2. Data pre-processing

To prepare the dataset for model training, we first cropped the standard and bone-suppressed radiographs to extract regions of interest (ROIs) centered on the lung area. The lung regions were identified using a pre-trained U-Net[41] model, which segments chest radiographs into the lung, heart, other anatomical areas, and background, assigning pixel values of 255, 85, 170, and 0, respectively, in 8-bit contrast. The U-Net architecture employed consisted of five depths, incorporating an input layer, five encoder layers, five decoder layers, and an output layer.

For training the U-Net model, we utilized all 247 chest radiographs from the Japanese Society of Radiological Technology database, along with their corresponding segmented labels.[42] The U-Net model was trained for up to 100 epochs using the RMSprop optimizer, with a learning rate of 0.0001, a weight decay of 1×10^{-8}, and a momentum of 0.9.

After training, the U-Net model was applied to identify the lung regions in the standard radiographs collected at Kitasato University Hospital, which had been converted to 8-bit contrast in advance. These identified lung regions were then cropped from both the standard and bone-suppressed radiographs. Finally, the cropped images were resized to 1024 × 1024 pixels to standardize the input size for the subsequent training of the bone suppression model.

2.1.3. Bone suppression network architecture and training settings

We employed the pix2pix[43,44] network to generate virtually bone-suppressed images from the standard chest radiographs. Figure 1 illustrates a flowchart of the bone suppression and pre-processing steps. The network architecture follows the design proposed by Isora *et al.*,[43] as described in our previous work,[22] with modifications made to the resolution of the generator and discriminator to handle 1024 × 1024 resolution images.

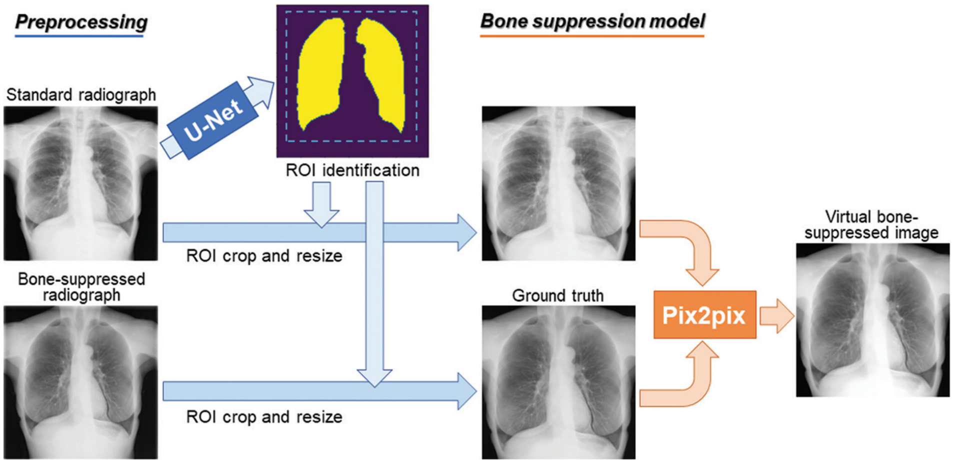


Figure 1. Flowchart of bone suppression model training
Abbreviation: ROI: Region of interest.

The input datasets in 8-bit contrast were normalized by converting pixel values to floating-point values ranging from 0 to 1. This normalization allowed the pix2pix network to process floating-point data, effectively reducing quantization errors. After processing, the network output was rescaled by multiplying the floating-point values by 255, converting them back to 8-bit images.

The network was trained for up to 4000 epochs, with a batch size of 2, using an NVIDIA TITAN RTX on an Ubuntu 20.04.6 LTS operating system. We applied the Adam optimizer with momentum parameters of $\beta 1 = 0.5$ and $\beta 2 = 0.999$. The learning rates were dynamically adjusted throughout the training: The generator started at 0.002 and decreased by 0.002 for each epoch, while the discriminator started at 0.02 and decreased by 0.02 per epoch. The implementation was carried out using Python 3.9.18 and PyTorch 1.12.0.

2.1.4. Performance evaluation

The performance of the bone suppression model was evaluated using the test dataset by measuring the similarity between the generated virtual bone-suppressed images and the ground truth images. Image similarity was assessed using two metrics: The peak signal-to-noise ratio (PSNR) and the structural similarity index (SSIM).[45,46]

The PSNR measures image similarity based on the ratio of noise to the maximum pixel values, calculated as follows:

$$PSNR = 20 \bullet \log_{10}\left(\frac{P_{max}}{MSE}\right), \qquad \text{(I)}$$

Where mean squared error *(MSE)* is the mean square error between two images, and P_{max} is the maximum value, which is 255 in this study.

The SSIM measures image similarity between two images, *x* and *y*, as follows:

$$SSIM = \frac{\left(2\mu_x\mu_y + C_1\right)\left(2\sigma_{xy} + C_2\right)}{\left(\mu_x^2 + \mu_y^2 + C_1\right)\left(\sigma_x^2 + \sigma_y^2 + C_2\right)}, \qquad \text{(II)}$$

Where μ_x and μ_y are the means; σ_x^2 and σ_y^2 are the variances; and σ_{xy} is the covariance for *x* and *y* in the ROIs, which consist of 5 × 5 pixels in this study. C_1 and C_2 are defined as $C_1 = (0.01L)^2$ and $C_2 = (0.03L)^2$, where L is 255 in this study.

2.2. COVID-19 severity assessment model

2.2.1. Data collection

We selected 192 chest radiographs from 136 COVID-19 patients in a publicly available image database provided by Cohen *et al.*[33] Although the database contains more than 700 images collected in several medical centers across 26 different countries, only 192 images were annotated using the Brixia score[33,34] by two expert radiologists: A board-certified specialist with 22 years of experience and a trainee with 2 years of experience. The scoring system was initially introduced in a radiology department in Italy during the pandemic, and it was later validated for risk stratification in a large population.[36] This score evaluates pneumonia severity by dividing the lungs into six zones and assigning an integer score ranging from 0 to 3 to each zone. Specifically, a score of 0 denotes no lung abnormalities, 1 indicates the presence of interstitial infiltrates, 2 reflects a combination of interstitial (dominant) and alveolar infiltrates, and 3 signifies interstitial and alveolar (dominant) infiltrates.

The total Brixia score, which ranges from 0 to 18, was calculated by summing the scores of all six zones. To derive the final labels for the regression models described in subsection 2.2.3., we averaged the total scores from both radiologists and normalized this value to a floating-point number between 0 and 1 by dividing by 18. These normalized scores were used as the labels for the regression models. The mean and standard deviation (SD) of the scores across 192 images were 0.380 and 0.260, respectively. Figure 2 provides an overview of the severity assessment process.

2.2.2. Data pre-processing

First, the radiographs were cropped and resized to center on the lung area, following the process described in subsection 2.1.2. Next, the images were transformed into bone-suppressed images using the AI-based bone suppression model developed in subsection 2.1. Both the standard radiographs and the bone-suppressed images were independently transformed to a resolution of 512 × 512 pixels with 8-bit contrast. Subsequently, each type of image was then randomly split into training and test data in an approximately 80:20 ratio, ensuring that all images from the same patient were grouped together in the same split. Five-fold cross-validation was applied to each dataset separately, and, to enhance robustness, this process was repeated 3 times using different random seeds.

2.2.3. Regression models and training settings

We employed several CNN architectures from different generations – DenseNet,[47] ResNet18, ResNet50,[48] and RegNetY-120[49] – all pre-trained on ImageNet.[50] To adapt these models for the regression task, we modified their final fully connected layers to have a single output that predicts a continuous value, corresponding to the normalized Brixia score. These models were trained using the MSE loss function for up to 25 epochs in each cross-validation fold. The training was conducted on an NVIDIA GeForce RTX 4070 with a Windows 11 operating system, utilizing Python 3.8.18 and PyTorch 2.2.1.

Based on our initial experiments, which indicated that the Stochastic Gradient Descent (SGD) optimizer consistently outperformed the Adam optimizer, we adopted SGD with a learning rate of 0.001 and a momentum of 0.9 for all models. In addition, a learning rate scheduler (StepLR) was applied to reduce the learning rate by a factor of 0.1 every 5 epochs.

2.3. Performance evaluation

2.3.1. Relationship between truths and predictions

We compared the performance of the severity assessment models between the standard chest radiograph dataset and the bone-suppressed image dataset by computing the mean ± SD of the mean absolute errors (MAEs) and PCCs across all folds and random seeds. The PCC quantifies the linear relationship between two variables,[51] as expressed by the following:

$$r = \frac{\sum_{i=0}^{n-1}\left(x_i-\bar{x}\right)\left(y_i-\bar{y}\right)}{\sqrt{\sum_{i=0}^{n-1}\left(x_i-\bar{x}\right)^2}\sqrt{\sum_{i=0}^{n-1}\left(y_i-\bar{y}\right)^2}}, \tag{III}$$

Where r is the PCC; x_i and y_i denote the individual sample points; and $\bar{x}$ and $\bar{y}$ are the means of x_i and y_i, respectively.

These metrics were calculated using the "mean_absolute_error" function from the Python "sklearn.metrics" library and the "pearsonr" function from the Python "scipy.stats" module.

Statistical significance tests were conducted using a two-tailed Student's *t*-test to compare the average MAEs

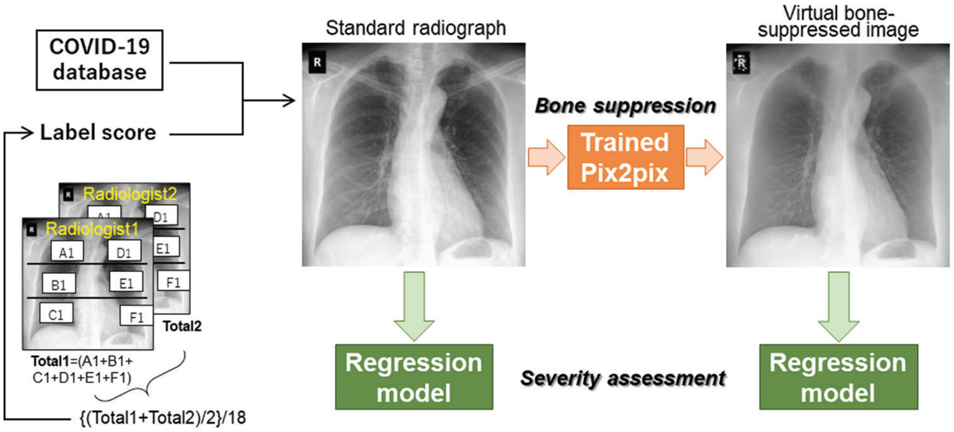


Figure 2. Flowchart of coronavirus disease 2019 severity assessment. The Brixia scoring system assigns an integer value from 0 to 3 to each of the six lung zones (A1 to F1). The total scores from two radiologists were averaged and then normalized to generate the final label scores.

and PCCs between the standard chest radiograph dataset and the bone-suppressed image dataset. The statistical tests and calculation of P-values were performed using the Python "scipy.stats" module.

2.3.2. Explainability of the severity assessment models

To validate the explainability and interpretability of the severity assessment models, we applied Grad-CAM[38] to generate heatmaps that exhibit the gradients in the final convolutional layer for the corresponding datasets tested in subsection 2.3.1. We used the "visualize_cam" function from the "gradcam.utils" module to generate these heatmaps, highlighting the regions that are most influential in predicting the severity.

3. Results

3.1. Generated bone-suppressed images

Figure 3 presents the bone-suppressed images generated by our updated bone suppression model, compared with the corresponding ground truth images for three cases from the test dataset collected at Kitasato University Hospital. The generated images closely resemble the ground truth, exhibiting a high degree of image similarity, with an average PSNR of 40.4 dB and an SSIM of 0.962 across the entire test dataset. Effective bone suppression was particularly achieved in the ribs and vertebral bones while preserving pneumonia and mass lesions.

In our previous AI-DES model, insufficient bone suppression was an avoidable issue due to enhanced quantization errors in the subtraction process.[22] In contrast, the updated model shows a significant improvement in bone suppression by directly generating bone-suppressed images, eliminating the need for the subtraction process. Furthermore, as shown in the enlarged images in Figure 3, the ground truth image of the third case exhibits motion artifacts, whereas the generated image displays a remarkable reduction in these artifacts. These findings highlight the model's ability to enhance image quality, surpassing that of the ground truth and our previous model.

Figure 4 showcases four examples of standard chest radiographs from the COVID-19 database, accompanied by the corresponding bone-suppressed images generated by our bone suppression model, and their severity score labels based on Brixia scores. This demonstrates the robust effectiveness of our bone suppression model, even when applied to an external dataset with diverse lung conditions.

3.2. Performance in COVID-19 severity assessment

Table 1 compares the performance of each trained regression model on the standard chest radiograph dataset versus the bone-suppressed image dataset, showing the averages and SDs of the MAEs and PCCs for the test data across all folds and random seeds. The table also includes the results of statistical significance tests. For cases where statistically significant differences were observed ($P < 0.05$), the better-performing averages are highlighted in bold, along with the corresponding P-values.

The ResNet18, ResNet50, and RegNetY-120 models demonstrated statistically significant improvements in the PCCs for the bone-suppressed image dataset compared to the standard chest radiograph dataset. In addition, the ResNet18 and RegNetY-120 models exhibited statistically significant lower MAEs, indicating superior predictive performance. In contrast, the DenseNet model showed similar performance on both datasets, with no statistically significant differences in either the MAEs or PCCs.

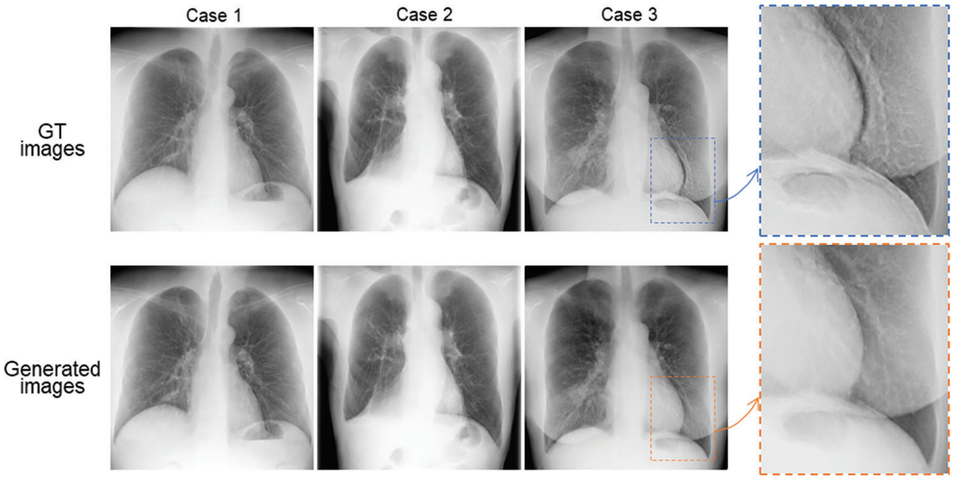


Figure 3. Comparison of virtually generated bone-suppressed images and the ground truth images. The third case also presents the enlarged images of the lower left lung field.

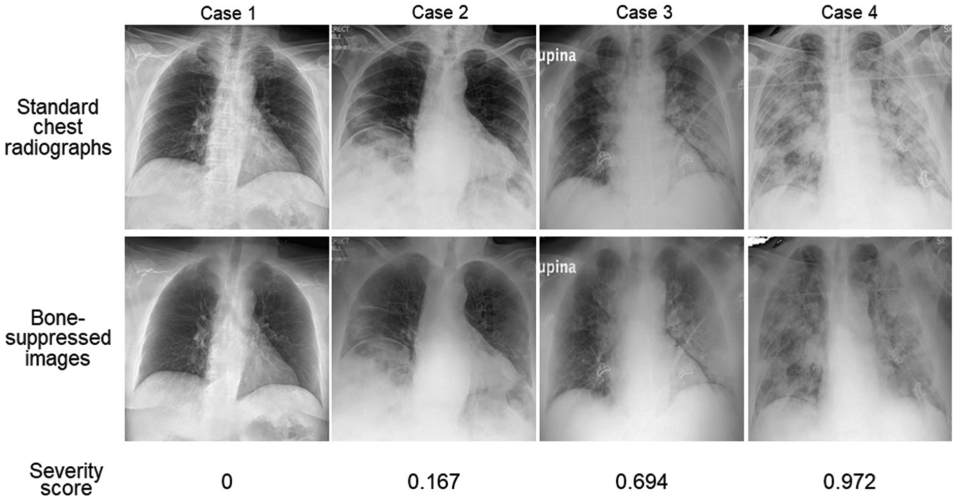


Figure 4. Examples of standard chest radiographs from a publicly available coronavirus disease 2019 database, along with the corresponding bone-suppressed images produced by our artificial intelligence model. The severity scores based on the Brixia score are displayed at the bottom of each image pair.

Table 1. Performance metrics of each model for standard chest radiographs versus bone-suppressed images in COVID-19 severity assessment

Model	MAE			PCC		
	Standard radiograph	BS image	Statistical significance	Standard radiograph	BS image	Statistical significance
DenseNet	0.0823±0.0286	0.0768±0.0451	NS	0.864±0.0953	0.873±0.164	NS
ResNet18	0.112±0.0551	**0.0722**±0.0435	0.00225*	0.797±0.216	**0.895**±0.150	0.0175*
ResNet50	0.0843±0.0261	0.0685±0.0178	NS	0.858±0.130	**0.882**±0.105	0.0230*
RegNetY-120	0.112±0.0379	**0.0993**±0.0319	0.0448*	0.811±0.143	**0.868**±0.0984	0.0181*

Note: Means±standard deviations across all folds are presented for the MAEs and PCCs. The statistical significance column lists *P* values for statistically significant cases (P<0.05*). Value in boldface indicated the better-performing average.

Abbreviations: BS: Bone-suppressed; MAE: Mean absolute error; NS: not significant; PCC: Pearson correlation coefficient, COVID-19: Coronavirus disease 2019.

Figures 5 and 6 illustrate examples of standard chest radiographs, bone-suppressed images, their corresponding Grad-CAM-generated heatmaps, and the scores predicted by the ResNet50 model. In the heatmaps, the color spectrum represents activation levels, with red indicating the highest activation areas, followed by yellow, blue, and transparency as activation decreases. In most cases in the test dataset, the high-activation areas were relatively more focused on the inner lung regions in the bone-suppressed images than in the standard radiographs, as shown in Figures 5 and 6. Furthermore, in the two cases in Figure 5, the predicted scores from the bone-suppressed images were closer to the true score labels than those from the standard radiographs. For example, in case 1, labeled with a true score of 0.556, the severity score predicted from the bone-suppressed image was 0.537, while the score from the standard radiograph was 0.342. In contrast, the two cases in Figure 6 illustrate instances where the predicted scores from the bone-suppressed images are deviated further from the true labels than those from the standard radiographs, despite the heatmaps consistently indicating high-activation areas in the lung regions. In particular, in case 2, which has a true score of 0.472, the heatmap for the bone-suppressed image indicates the highest activation in the right lung area; however, the predicted score of 0.401 was further from the true label than the score of 0.493 predicted from the standard radiograph.

4. Discussion

In this study, we developed an AI-based bone suppression model for CXR and applied it to a publicly available COVID-19 image database. The pix2pix model demonstrated a high degree of image similarity to the ground truth images, achieving PSNR and SSIM metrics comparable to those reported by existing bone suppression models for chest radiographs.[17,23,26-27] As a result, our present model effectively removes bone structures while enhancing the visibility of lung tumors and inflammation

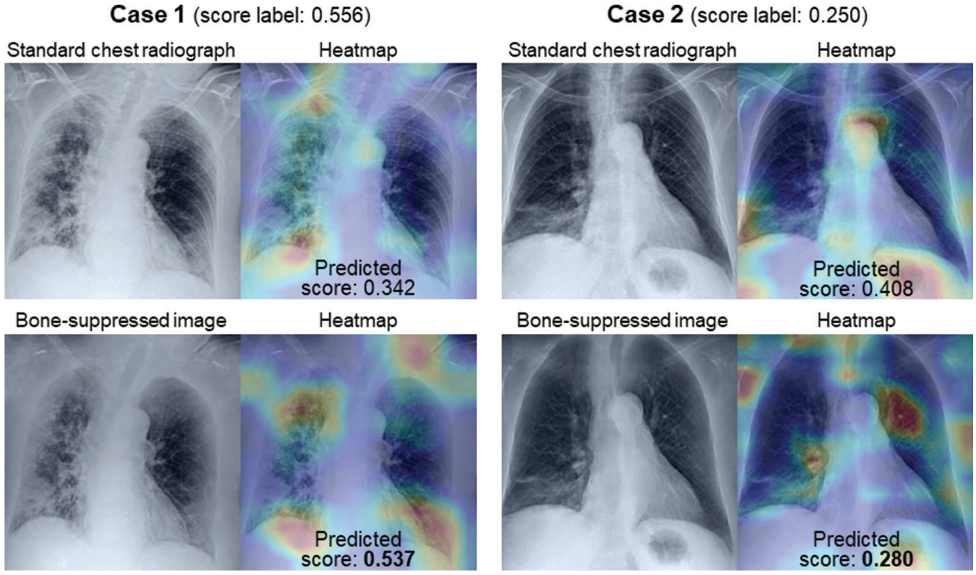


Figure 5. Standard chest radiographs, bone-suppressed images, their corresponding heatmaps, and the scores predicted by the ResNet50 model for two test cases, where the predicted scores from the bone-suppressed images are closer to the true score labels than those from the standard radiographs.

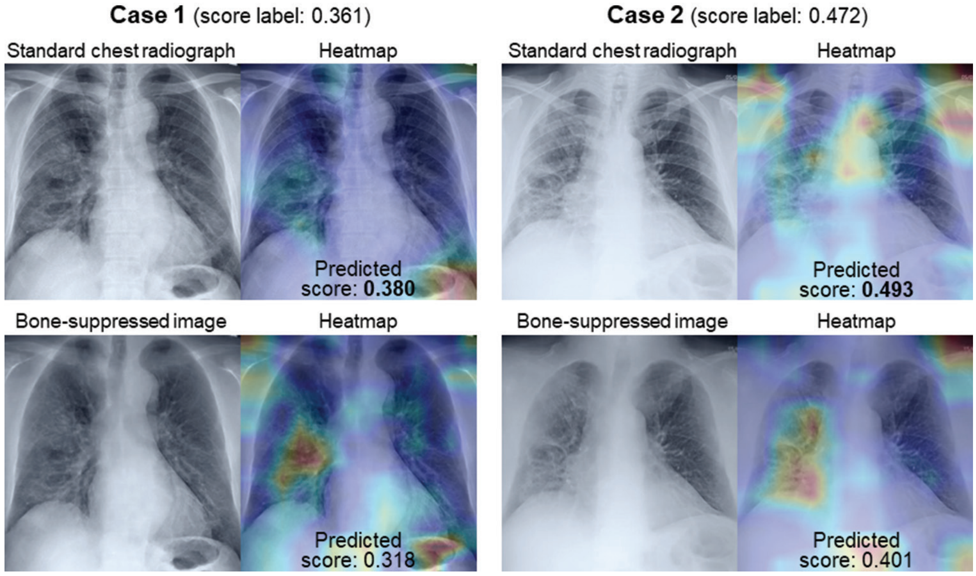


Figure 6. Standard chest radiographs, bone-suppressed images, their corresponding heatmaps, and the scores predicted by the ResNet50 model for two test cases, where the predicted scores derived from the standard radiographs are closer to the true score labels than those from the bone-suppressed images.

and reducing motion artifacts, addressing the limitations of traditional dual-shot DES systems.[52,53] Further improvements to the model architecture and parameter optimization could enhance its performance. For instance, Rani *et al.*[23] combined the pix2pix discriminator with Wasserstein GAN with gradient penalty,[54] achieving higher similarity with a PSNR of 43.588 and an SSIM of 0.989.

The enhanced image quality and practicality of our updated system, which eliminates the need for labor-intensive subtraction processing, effectively address the limitations of our previous system.[22] Moreover, this improved model has the potential to significantly enhance the diagnostic capabilities of CXR while maintaining the cost-effectiveness and time-efficiency benefits over CT scans. Its clinical applicability is further supported by its ability to generate high-quality images when applied to an external dataset, suggesting robustness across diverse scenarios, including different races, clinical conditions, and imaging systems. However, the limited sample size of the COVID-19 images used in this study underscores the need for further validation, as AI model performance is often influenced by biases in the training data. Specifically, the small dataset size may have led to imbalances in disease severity, patient age distribution, and gender ratio. In fact, the mean label score of 0.380 and SD of 0.260 suggest that most of the images are associated with lower severity scores, potentially limiting the model's performance and generalizability. Consequently, the effectiveness of bone suppression for severity assessment may not be applicable to all patient populations or clinical settings. Furthermore, artificially generated images can exhibit undesirable artifacts or implausible shadows, even when produced by cutting-edge models.[55] Therefore, the quality of the bone-suppressed images should be carefully inspected by experts, such as radiologists and respiratory medicine physicians, to identify potential artifacts and ensure that the visibility of lesions is preserved.

We applied the updated bone suppression model as a pre-processing step in COVID-19 severity assessment using regression models. By utilizing the bone-suppressed images, we observed statistically significant improvements in linear correlations between predicted scores and true labels across various regression models, including classical models such as ResNet18 and the newer model RegNetY-120. Notably, the highest PCC of 0.895 exceeded the performance reported in related studies.[32,34] However, the DenseNet model showed no significant differences in performance with and without bone suppression. We acknowledge that, again, the small dataset size remains a significant limitation, potentially affecting the generalizability of this study. Furthermore, the variability in performance across different regression models warrants further investigation. In future work, we plan to confirm the effectiveness and robustness of bone suppression techniques using larger and more diverse datasets and a broader range of prediction models.

The clinical implications of this study are promising. The proposed bone suppression system has the potential to enhance the diagnostic performance of CXR by improving the visibility of abnormalities and enabling more accurate disease assessment. Beyond its established effectiveness in classification tasks, bone suppression techniques could facilitate regression-based evaluations of disease severity, progression, and recurrence risk. Integrating this system into diagnostic image viewing software would allow radiologists and physicians to access bone-suppressed images as additional clinical information. Moreover, as demonstrated in our assessment of COVID-19 severity, these techniques could serve as a pre-processing tool for computer-aided diagnosis systems. This integration may help reduce the discrepancy in diagnostic accuracy between CXR and CT.

In addition, we employed Grad-CAM to visualize the rationale for the model's decisions. We confirmed that the bone suppression techniques effectively directed the model's focus to the inner lung field. The effect could enhance the explainability and interpretability of the models, thereby increasing confidence in their predictions. However, even when the activated areas of the models were focused within the lung fields, the predicted scores did not consistently align with the true scores. Consequently, improving prediction accuracy while fostering explainability remains a challenge for achieving broader acceptance as a reliable diagnostic support tool.

In summary, this study improved the clinical applicability of the AI-based bone suppression system by eliminating the need for subtraction processing. The updated system enhanced the visibility of lung abnormalities, leading to more accurate predictions of pneumonia severity, as demonstrated by statistically significant improvements in the linear correlation between predicted severity scores and actual labels. These findings underscore the utility of AI-driven bone suppression in CXR, particularly for regression tasks related to disease severity, progression, and recurrence risk. Furthermore, the bone suppression techniques guided the prediction model to concentrate on the inner lung fields, suggesting potential improvements in the reliability of clinical assessments. The application of bone suppression in assessing COVID-19 severity could optimize patient monitoring and healthcare resource allocation. In addition, this advancement has the potential to elevate the diagnostic accuracy of CXR, providing valuable tools to overcome existing limitations, such as inferior contrast resolution and the superimposition of anatomical structures compared to CT.

5. Conclusion

This study successfully developed and validated an AI-based bone suppression model for CXR, which effectively removes bone structures while highlighting lung abnormalities. The model not only improves image quality but also streamlines the entire image processing workflow, increasing its clinical practicality. We applied the bone suppression model as a pre-processing step to facilitate more accurate predictions of COVID-19 severity. The findings demonstrate the potential of bone suppression techniques in assessing various pulmonary conditions, particularly through regression analyses. Future research should focus on validating bone suppression techniques with larger and more diverse datasets, as well as exploring a range of prediction models. In addition, addressing potential biases in AI outputs and enhancing the model's explainability will be essential for ensuring its reliable integration into routine clinical practice.

Acknowledgments

None.

Funding

This work was supported by the Kondou Kinen Medical Foundation.

Conflict of interest

The authors declare they have no competing interests.

Author contributions

Conceptualization: Asumi Yamazaki, Takayuki Ishida
Formal analysis: Asumi Yamazaki
Investigation: Asumi Yamazaki, Masashi Seki
Methodology: Asumi Yamazaki
Writing – original draft: Asumi Yamazaki
Writing – review & editing: Masashi Seki, Takayuki Ishida

Ethics approval and consent to participate

This study involved human subjects and was approved by the Research Ethics Review Committee of Kitasato University Hospital (approval number: C22-064) and Osaka University Graduate School of Medicine (approval number: 22061-6).

Consent for publication

A waiver of informed consent was granted by the Research Ethics Review Committee of Kitasato University Hospital and Osaka University Graduate School of Medicine due to the retrospective nature of this study. In accordance with Japan's Ethical Guidelines for Medical and Health Research Involving Human Subjects, patients were provided with the opportunity to "opt out."

Availability of data

The imaging data are not available due to institutional and ethical restrictions. The publicly available dataset can be accessed through the references cited in this article.

References

1. World Health Organization. *Coronavirus Disease (COVID-19) Pandemic*. Available from: https://www.who.int/emergencies/diseases/novel-coronavirus-2019 [Last accessed on 2024 Oct 15].
2. Beeraka NM, Sukocheva OA, Lukina E, Liu J, Fan R. Development of antibody resistance in emerging mutant strains of SARS CoV-2: Impediment for COVID-19 vaccines. *Rev Med Virol*. 2022;32(5):e2346.

   doi: 10.1002/rmv.2346
3. Ramirez FE, Sanchez A, Pirskanen A.T. Hydrothermotherapy in prevention and treatment of mild to moderate cases of COVID-19. *Med Hypotheses*. 2021;146:110363.

   doi: 10.1016/j.mehy.2020.110363
4. World Health Organization. *WHO Coronavirus (COVID-19) Dashboard*. Available from: https://data.who.int/dashboards/covid19/deaths?n=c [Last accessed on 2024 Oct 15].
5. Gao Z, Xu Y, Sun C, *et al*. A systematic review of asymptomatic infections with COVID-19. *J Microbiol Immunol Infect*. 2021;54(1):12-16.

   doi: 10.1016/j.jmii.2020.05.001
6. Kim MS, An MH, Kim WJ, Hwang TH. Comparative efficacy and safety of pharmacological interventions for the treatment of COVID-19: A systematic review and network meta-analysis. *PLoS Med*. 2020;17(12):e1003501.

   doi: 10.1371/journal.pmed.1003501
7. Long B, Carius BM, Chavez S, *et al*. Clinical update on COVID-19 for the emergency clinician: Presentation and evaluation. *Am J Emerg Med*. 2022;54:46-57.

   doi: 10.1016/j.ajem.2022.01.028
8. Wong HYF, Lam HYS, Fong AH, *et al*. Frequency and distribution of chest radiographic findings in patients positive for COVID-19. *Radiology*. 2020;296(2):E72-E78.

   doi: 10.1148/radiol.2020201160
9. Sverzellati N, Ryerson CJ, Milanese G, *et al*. Chest radiography or computed tomography for COVID-19 pneumonia? Comparative study in a simulated triage setting. *Eur Respir J*. 2021;58(3):2004188.

   doi: 10.1183/13993003.04188-2020
10. Tang JS, Lai JK, McCusker MW, Irving L, Pascoe DM, Heinze SB. Chest imaging findings in COVID-19-positive patients in an Australian tertiary hospital. *J Med Imaging Radiat Oncol*. 2022;66(6):755-760.

    doi: 10.1111/1754-9485.13339
11. Landini N, Colzani G, Ciet P, *et al*. Chest radiography findings of COVID-19 pneumonia: A specific pattern for a confident differential diagnosis. *Acta Radiol*. 2022;63(12):1619-1626.

    doi: 10.1177/02841851211055163
12. Sharifrazi D, Alizadehsani R, Roshanzamir M, *et al*. Fusion of convolution neural network, support vector machine and Sobel filter for accurate detection of COVID-19 patients using X-ray images. *Biomed Signal Process Control*. 2021;68:102622.

    doi: 10.1016/j.bspc.2021.102622
13. Arias-Garzón D, Alzate-Grisales JA, Orozco-Arias S, *et al*. COVID-19 detection in X-ray images using convolutional neural networks. *Mach Learn Appl*. 2021;6:100138.

    doi: 10.1016/j.mlwa.2021.100138

14. Avolio M, Fuduli A, Vocaturo E, Zumpano E. A comparative study of linear type multiple instance learning techniques for detecting COVID-19 by chest X-ray images. *Prog Artif Intell.* 2024.

    doi: 10.1007/s13748-024-00332-1

15. Takaki T, Murakami S, Tani N, Aoki T. Evaluation of the clinical utility of temporal subtraction using bone suppression processing in digital chest radiography. *Heliyon*. 2023;9(1):e13004.

    doi: 10.1016/j.heliyon.2023

16. Van der Heyden B. The potential application of dual-energy subtraction radiography for COVID-19 pneumonia imaging. *Br J Radiol*. 2021;94(1120):20201384.

    doi: 10.1259/bjr.20201384

17. Matsubara N, Teramoto A, Saito K, Fujita H. Bone suppression for chest X-ray image using a convolutional neural filter. *Australas Phys Eng Sci Med*. 2019;43:97-108.

    doi: 10.1007/s13246-019-00822-w

18. Han L, Lyu Y, Peng C, Zhou SK. GAN-based disentanglement learning for chest X-ray rib suppression. *Med Image Anal*. 2022;77:102369.

    doi: 10.1016/j.media.2022.102369

19. Cho K, Seo J, Kyung S, Kim M, Hong GS, Kim N. Bone suppression on pediatric chest radiographs via a deep learning-based cascade model. *Comput Methods Programs Biomed*. 2022;215:106627.

    doi: 10.1016/j.cmpb.2022.106627

20. Xie W, Gan M, Tan X, Li M, Yang W, Wang W. Efficient labeling for fine-tuning chest X-ray bone-suppression networks for pediatric patients. *Med Phys*. 2024;52:978-992.

    doi: 10.1002/mp.17516

21. Ait Nasser A, Akhloufi MA. A review of recent advances in deep learning models for chest disease detection using radiography. *Diagnostics* (*Basel*). 2023;13(1):159.

    doi: 10.3390/diagnostics13010159

22. Yamazaki A, Koshida A, Tanaka T, Seki M, Ishida T. Development of artificial intelligence-based dual-energy subtraction for chest radiography. *Appl Sci*. 2023;13(12):7220.

    doi: 10.3390/app13127220

23. Rani G, Misra A, Dhaka VS, Zumpano E, Vocaturo E. Spatial feature and resolution maximization GAN for bone suppression in chest radiographs. *Comput Methods Programs Biomed*. 2022;224:107024.

    doi: 10.1016/j.cmpb.2022.107024

24. Rani G, Misra A, Dhaka VS, *et al*. A multi-modal bone suppression, lung segmentation, and classification approach for accurate COVID-19 detection using chest radiographs. *Intell Syst Appl*. 2022;16:200148.

    doi: 10.1016/j.iswa.2022.200148

25. Lam NFD, Sun H, Song L, *et al*. Development and validation of bone-suppressed deep learning classification of COVID-19 presentation in chest radiographs. *Quant Imaging Med Surg*. 2022;12(7):3917-3931.

    doi: 10.21037/qims-21-791

26. Rajaraman S, Zamzmi G, Folio L, Alderson P, Antani S. Chest X-ray bone suppression for improving classification of tuberculosis-consistent findings. *Diagnostics* (*Basel*). 2021;11(5):840.

    doi: 10.3390/diagnostics11050840

27. Rajaraman S, Cohen G, Spear L, Folio L, Antani S. DeBoNet: A deep bone suppression model ensemble to improve disease detection in chest radiographs. *PLoS One*. 2022;17(3):e0265691.

    doi: 10.1371/journal.pone.0265691

28. Kim H, Lee KH, Han K, *et al*. Development and validation of a deep learning-based synthetic bone-suppressed model for pulmonary nodule detection in chest radiographs. *JAMA Netw Open*. 2023;6(1):e2253820.

    doi: 10.1001/jamanetworkopen.2022.53820

29. Bae K, Oh DY, Yun ID, Jeon KN. Bone suppression on chest radiographs for pulmonary nodule detection: Comparison between a generative adversarial network and dual-energy subtraction. *Korean J Radiol*. 2022;23(1):139-149.

    doi: 10.3348/kjr.2021.0146

30. Xu D, Xu Q, Nhieu K, Ruan D, Sheng K. An efficient and robust method for chest X-ray rib suppression that improves pulmonary abnormality diagnosis. *Diagnostics* (*Basel*). 2023;13(9):1652.

    doi: 10.3390/diagnostics13091652

31. Thölke P, Mantilla-Ramos YJ, Abdelhedi H, *et al*. Class imbalance should not throw you off balance: Choosing the right classifiers and performance metrics for brain decoding with imbalanced data. *Neuroimage*. 2023;277:120253.

    doi: 10.1016/j.neuroimage.2023.120253

32. Cohen JP, Dao L, Roth K, *et al*. Predicting COVID-19 pneumonia severity on chest X-ray with deep learning. *Cureus*. 2020;12(7):e9448.

    doi: 10.7759/cureus.9448

33. Cohen JP, Morrison P, Dao L, Roth K, Duong TQ, Ghassemi M. COVID-19 image data collection: Prospective predictions are the future. *J Mach Learn Biomed Imaging*. 2020;2:1-38.

    doi: 10.59275/j.melba.2020-48g7

34. Signoroni A, Savardi M, Benini S, *et al*. BS-Net: Learning COVID-19 pneumonia severity on a large chest X-ray dataset. *Med Image Anal*. 2021;71:102046.

doi: 10.1016/j.media.2021.102046

35. Borghesi A, Maroldi R. COVID-19 outbreak in Italy: Experimental chest X-ray scoring system for quantifying and monitoring disease progression. *Radiol Med*. 2020;125:509-513.

    doi: 10.1007/s11547-020-01200-3

36. Borghesi A, Zigliani A, Masciullo R, *et al*. Radiographic severity index in COVID-19 pneumonia: Relationship to age and sex in 783 Italian patients. *Radiol Med*. 2020;125:461-464.

    doi: 10.1007/s11547-020-01202-1

37. Caroprese L, Vocaturo E, Zumpano E. Argumentation approaches for explanaible AI in medical informatics. *Intell Syst Appl*. 2022;16:200109.

    doi: 10.1016/j.iswa.2022.200109

38. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. In *2017 IEEE International Conference on Computer Vision* (*ICCV*). Venice, Italy: IEEE; 2017. p. 618-626.

    doi: 10.1109/ICCV.2017.74

39. Musthafa MM, Mahesh TR, Vinoth Kumar V, Guluwadi S. Enhancing brain tumor detection in MRI images through explainable AI using Grad-CAM with resnet 50. *BMC Med Imaging*. 2024;24(1):107.

    doi: 10.1186/s12880-024-01292-7

40. Talaat FM, Gamel SA, El-Balka RM, Shehata M, ZainEldin H. Grad-CAM enabled breast cancer classification with a 3D inception-ResNet V2: Empowering radiologists with explainable insights. *Cancers* (*Basel*). 2024;16(21):3668.

    doi: 10.3390/cancers16213668

41. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. In: *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015*. Cham: Springer; 2015. p. 234-241.

42. *JSRT Database*. Available from: https://db.jsrt.or.jp/eng.php [Last accessed on 2024 Oct 15].

43. Isora P, Zhu JY, Zhou T, Efros AA. Image-to-image Translation with Conditional Adversarial Networks. In: *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*). Honolulu, HI, USA: IEEE; 2017. p. 5967-5976.

    doi: 10.1109/CVPR.2017.632

44. *Junyanz/Pytorch-CycleGAN-and-Pix2pix*. Available from: https://github.com/junyanz/pytorch-cyclegan-and-pix2pix [Last accessed on 2024 Oct 15].

45. Horé A, Ziou D. Image Quality Metrics: PSNR vs. SSIM. In: *2010 20th International Conference on Pattern Recognition*. Istanbul, Turkey: IEEE; 2010. p. 2366-2369.

    doi: 10.1109/ICPR.2010.579

46. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: From error visibility to structural similarity. *IEEE Trans Image Process*. 2004;13(4):600-612.

    doi: 10.1109/tip.2003.819861

47. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely Connected Convolutional Networks. In: *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*). Honolulu, HI, USA: IEEE; 2017. p. 2261-2269.

    doi: 10.1109/CVPR.2017.243

48. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. In: *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*). Las Vegas, NV, USA: IEEE; 2016. p. 770-778.

    doi: 10.1109/CVPR.2016.90

49. Radosavovic I, Kosaraju RP, Girshick R, He K, Dollár P. Designing Network Design Spaces. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*); 2020. p. 10428-10436.

    doi: 10.48550/arXiv.2003.13678

50. Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L. ImageNet: A Large-scale Hierarchical Image Database. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*). Miami, FL, USA: IEEE; 2009. p. 248-255.

    doi: 10.1109/CVPR.2009.5206848

51. Schober P, Boer C, Schwarte LA. Correlation coefficients: Appropriate use and interpretation. *Anesth Analg*. 2018;126(5):1763-1768.

    doi: 10.1213/ANE.0000000000002864

52. Karim KS, Tilley IS. Portable single-exposure dual-energy X-ray detector for improved point-of-care diagnostic imaging. *Mil Med*. 2023;188(Suppl 6):84-91.

    doi: 10.1093/milmed/usad034

53. Shunkov YE, Kobylkin IS, Prokhorov AV, *et al.* Motion artefact reduction in dual-energy radiography. *Biomed Eng.* 2022;55:415-419.

    doi: 10.1007/s10527-022-10148-9

54. Gulrajani I, Ahmed F, Arjovsky M, Dumoulin V, Courville A. Improved Training of Wasserstein GANs. In: *Neural Information Processing Systems 30* (*NIPS 2017*). San Diego, CA, USA; 2017. p. 5769-5779.

55. Zhang L, Xu Z, Barnes C, *et al.* Perceptual Artifacts Localization for Image Synthesis Tasks. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision* (*ICCV*). Paris, France: IEEE; 2023. p. 7545-7556.

    doi: 10.1109/ICCV51070.2023.00697

ACCSCIENCE PUBLISHING

Artificial Intelligence in Health

ORIGINAL RESEARCH ARTICLE

Artificial intelligence in cardiac rhythm diagnostics and management: Challenges and opportunities

Robert Splinter[1,2,3,4]*

[1]Intelligent Bioinformatics Ltd, Thirsk, North Yorkshire, United Kingdom

[2]Intelligent Bioinformatics LLC, Mooresville, North Carolina, United States of America

[3]Departments of Physics and Optical Science, University of North Carolina at Charlotte, Charlotte, North Carolina, United States of America

[4]Departments of Mathematics and Statistics, University of North Carolina at Charlotte, Charlotte, North Carolina, United States of America

Abstract

Each day, one million people undergo electrocardiogram diagnostics. The diagnostic process is time-consuming and often yields incomplete or inconclusive results, placing significant strain on physicians. Artificial intelligence (AI)-assisted diagnosis can significantly alleviate this burden by enhancing diagnostic accuracy and efficiency, and its application is gaining traction across various fields. With the increasing number of patients and a growing backlog of diagnostic appointments, AI can offer physicians benefits such as accurate, timely, and reliable assistance in reviewing vital signs and conducting physical examinations for individual patients. As physicians face mounting pressure from insurance companies and government guidelines for consultation time, AI can help streamline the diagnostic process. In particular, with the growing global attention on cardiac health (and the overall decline thereof), the range of automated diagnostic opportunities is expanding rapidly. Additional mathematical processing tools can provide probabilistic assessments of various cardiac conditions, reducing physicians' workload while enhancing treatment options. AI has already demonstrated success in expediting the detection of pathological cardiac depolarization abnormalities and shortening diagnostic time frames. However, AI-based diagnostics requires further validation and safeguards to minimize diagnostic inaccuracies, ensuring its reliability and safety in clinical practice.

Keywords: Machine learning; Diagnostics; Medicine; Risk stratification; Screening; Signal-processing; Matched filter; Wavelet analysis

***Corresponding author:**
Robert Splinter
(rsplinter@ibi-gb.com)

Citation: Splinter R. Artificial intelligence in cardiac rhythm diagnostics and management: Challenges and opportunities. *Artif Intell Health*. 2025;2(3):107-124. doi: 10.36922/aih.8468

Received: January 9, 2025

Revised: March 3, 2025

Accepted: March 12, 2025

Published online: April 2, 2025

Copyright: © 2025 Author(s). This is an Open-Access article distributed under the terms of the Creative Commons Attribution License, permitting distribution, and reproduction in any medium, provided the original work is properly cited.

Publisher's Note: AccScience Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

Data from four large population-based registries, which contain emergency medical service data collected between 2012 and 2017 across major European Union (EU) countries, report at least 450,000 cases of sudden cardiac death (SCD) annually.[1] Similarly, heart failure poses a significant economic burden on the United States (US), with costs projected to reach US $70 billion by 2030. The cost of heart failure varies widely, from less than US $1,000 per patient in low-income countries to €5,000 – 15,000 in Europe and US $17,000 – 30,000 in the US. Accurate and timely diagnostic screening, coupled with advanced warning systems, could save approximately 1,250 lives per day in the US, 40 in the United Kingdom (UK), and 1,100 across the EU. Notably, data from Asia and Russia remain incomplete. SCD, often caused by ventricular fibrillation (VF), affects individuals of all ages, depending on factors such as congenital disorders and lifestyles. The incidence in pediatric and juvenile populations can be as high as 1 in 10,000. Women are generally disproportionately affected and often face underdiagnoses or misdiagnoses. In addition, people of African descent and Asian Indian men are more frequently affected than Caucasians. The use of AI in medical diagnostics can save time, enhance the level of detail, and improve early detection of pathological conditions. It can uncover warning signs of deviations in the image and signal patterns that may require follow-up investigations and potentially additional tests.[2-7]

The limitations of manual analysis of diagnostic results include personal preferences, bias, high dependency on professional experience, skills, and abilities, and personal health conditions, such as fatigue, eyesight problems, and personal medical history (e.g., color blindness). For example, traditional X-ray imaging alone may fail to fully assess coronary stenoses or predict the progression of life-threatening cardiovascular conditions. One important characteristic of coronary health, particularly cardiac depolarization health, is the recognition of fatal factors underlying disease progression. AI-driven computational analysis can enhance pattern recognition and identify clusters of subtle abnormalities that may be overlooked by visual inspection, such as those in peak-to-peak sequencings and groupings like image-clustering.[8-11]

One area of interest in diagnostics that may benefit from automated screening through software algorithms, often mathematically based, is electrocardiogram (ECG, also known as electrocardiogram in German) acquisition. The ECG was introduced in 1903 by Willem Einthoven from the Netherlands, who pioneered the use of a specially designed galvanometer to record the action potentials. Early recordings were obtained by immersing hands and feet in saltwater to enhance conductivity. The ECG captures the electrical depolarization and repolarization of cardiac muscle cells.[12,13] These electrical signals, known as action potentials, result from the flow of ions across the cell membrane, leading to muscle contraction. A single ECG cycle is illustrated in Figure 1A, while a representative healthy ECG is shown in Figure 1B. The ECG waveform typically consists of five distinct peaks labeled P, Q, R, S, and T, with a frequently observed sixth peak, the U wave. The P wave is the result of the depolarization of the atria, while the remaining waves are caused by the depolarization of the ventricles. The R wave represents the synchronized depolarization of the right and left ventricles over time.

Cardiac muscles contract as a direct result of cellular electrical excitation, mediated by active and passive ion transport across cell membranes. Cardiac cells are not directly stimulated by neurons but instead initiate depolarization through an intrinsic excitation process. The depolarization of one cell triggers a cascade of depolarization and contraction in neighboring cells, resulting in coordinated atria and ventricular contractions that pump blood through the heart. The electrical activation of each cardiac cell serves as an indication of its health and functional state. The summation of these electrical activities can be represented as a vector, which changes direction as depolarization propagates through the heart (Figure 2). The details of this vector rotation can be captured by placing multiple electrodes on the skin.

Modern ECG systems employ advanced signal processing techniques to enhance the signal-to-noise ratio, enabling precise analysis of arrhythmogenic tendencies. The ECG is, therefore, the cumulative result of the depolarization of individual cardiac muscle cells over time, occurring in a controlled and repetitive manner. The 12-lead ECG, the most detailed configuration, offers spatial information about depolarization patterns over time (Figure 3). In contrast, most other detection and monitoring techniques only provide temporal information. By recording the depolarization of cardiac muscle cells over extended periods, clinicians can diagnose abnormalities, identify malfunctioning regions, and determine the need for further medical intervention. Deviations from a typical ECG pattern, whether detected visually or through signal processing, can be classified as specific cardiac disorders. During VF, the depolarization vector will no longer be recognizable due to the random and chaotic nature of electrical activity in the ventricles.[9,10,14,15]

This paper introduces the pathological conditions that can potentially be identified using various mathematical techniques under AI-driven signal processing. Each condition may require unique technical approaches. The ultimate goal is to provide predictive mechanisms for early detection and prevention of cardiac disorders. The early detection and intervention of cardiac disorders require a highly regulated control system, which is currently under development and may take several years for market release. It is important to recognize that AI-based diagnosis is not a standalone solution. Medical screening requires a

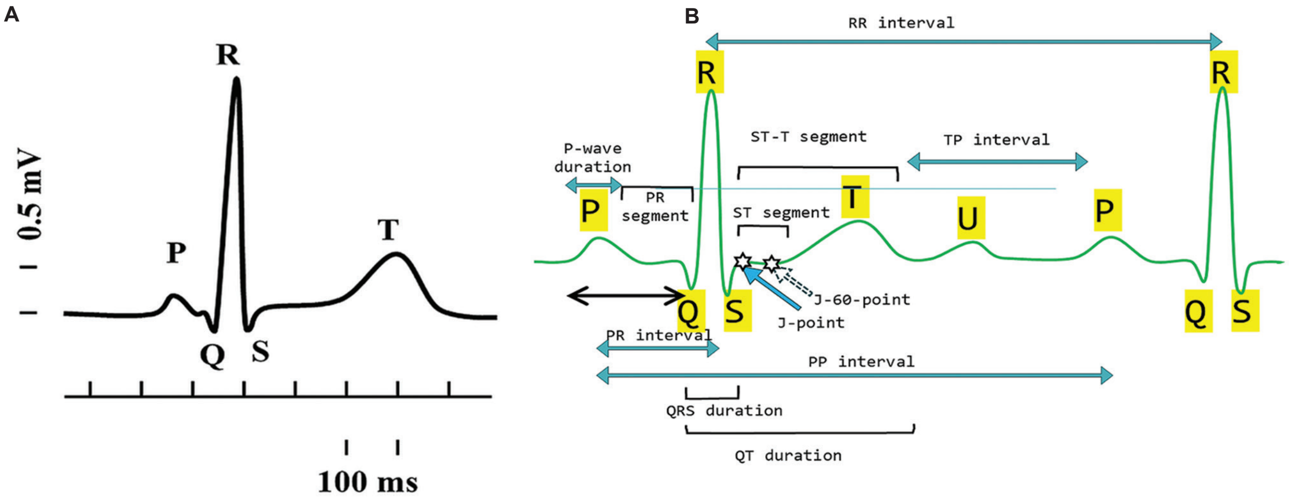


Figure 1. Synopsis of the representative and essential components in the electrocardiogram recording. (A) A standard single-period heartbeat depolarization pattern illustrates atrial depolarization (P), the ventricular depolarization sequence (QRS complex), and ventricular repolarization (T), followed by a possible U wave, representing the final phase of ventricular repolarization. Prominent U waves are characteristic of hypokalemia. (B) Heart depolarization intervals and critical markers. The PR interval typically ranges from 0.12 to 0.22 s. The QRS complex duration in a healthy person is generally <0.12 s. The J-point serves as a reference for detecting ST segment elevation or depression. The J-60 point is used to assess ST segment depression, particularly during exercise stress testing. The QT interval varies between males and females. In healthy adults, the QT interval is generally <0.45 s for males and <0.47 s for females. For calibration, the PR segment provides the most relevant baseline and isoelectric level information. The PR segment level is crucial for accurately determining ST-segment elevation or depression.

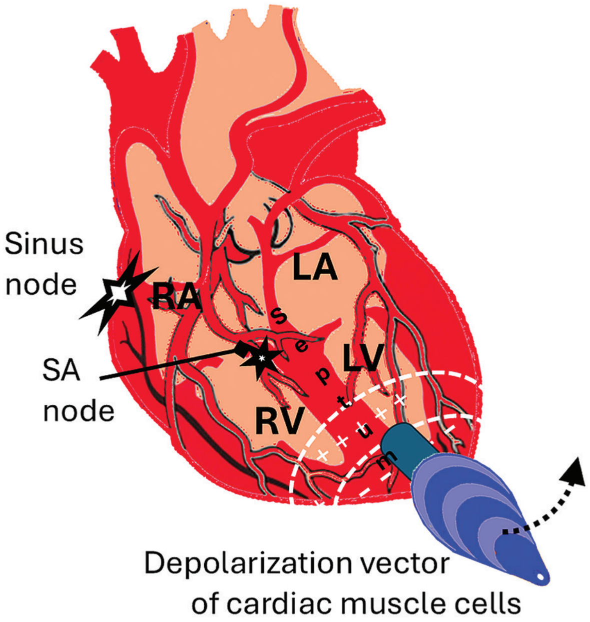


Figure 2. Representative illustration of the depolarization vector rotating through three-dimensional space during a cardiac beat (P-QRS), facilitating organized contraction.
Abbreviations: LA: Left atrium; LV: Left ventricle; RA: Right atrium; RV: Right ventricle; SA: Sinoatrial.

multidisciplinary approach that accounts for individual variability in vital signs, psychosomatic modulation (e.g., the white coat syndrome, infatuation, fear, and anxiety), and biochemical interactions. Although AI can provide recommendations for potential diagnostic conclusions and estimate the probabilities of various pathological conditions identified through mathematical models, the final diagnosis remains the responsibility of the attending physician.

2. Background

Accurate, rapid, and reliable diagnostics and patient care fundamentally depend on the integration of biochemical and physiological information obtained from various sources through multiple techniques. The acquisition and computerized processing of as much relevant information as possible, for instance, using AI, will enhance diagnostic precision and facilitate the development of effective and reliable treatment regimens.[8] Access to a patient's complete medical history is crucial in a diagnosis, as is the ability to compare the patient's data against known pathological patterns from a broad population database. Such comparisons enable early screening, identification of health conditions, and determination of the most appropriate treatment modalities or the need for additional follow-up tests targeting specific physiological, anatomical, or biochemical factors. Equally important is the ability to situate the patient within the appropriate segment of the statistical distribution relevant to their suspected pathology. Even for the same individual, physiological values can vary significantly due to factors such as activity levels, time of day

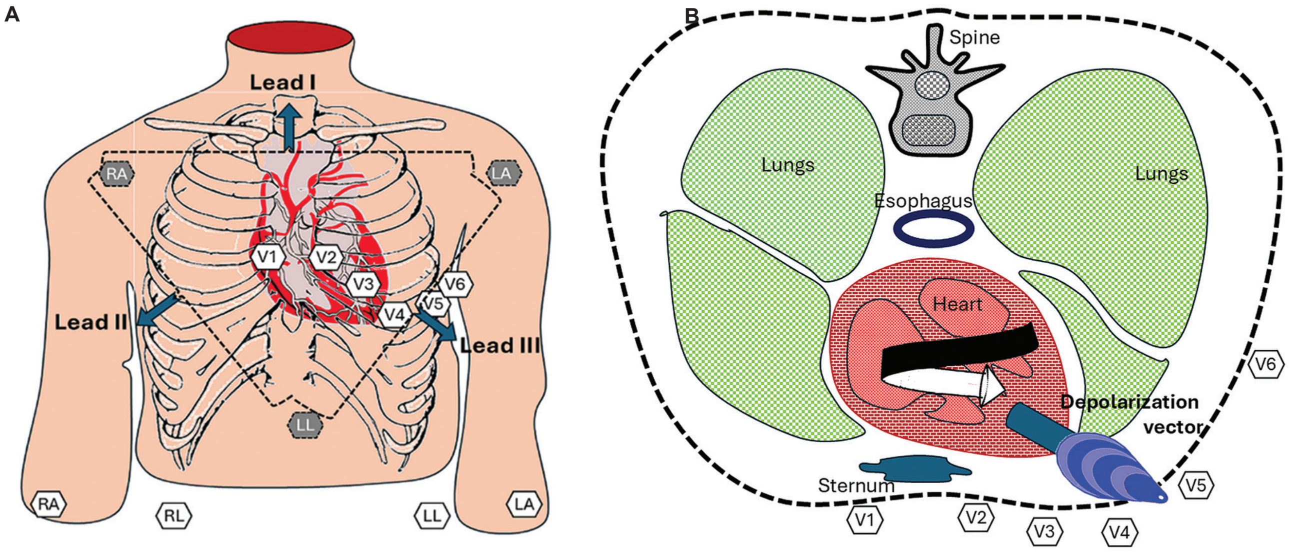


Figure 3. Typical placement of leads in a 12-lead electrocardiogram. (A) Coronal (frontal) view. (B) Axial (transverse) view
Abbreviations: LA: Left arm; LL: Left leg; RA: Right arm; RL: Right leg; V1: 4th intercostal space, right margin of the sternum; V2: 4th intercostal space, left sternal edge; V3: Midpoint between V2 and V4; V4: 5th intercostal space, midclavicular line (symmetrically opposed to V2; V5: Anterior axillary line; V6: Midaxillary line, forming a straight line with V4 and V5.

(e.g., morning vs. evening), gender (e.g., male vs. female), hormonal status, emotional state (e.g., aggression), and other boundary conditions. Anatomical deviations further complicate diagnostics. For example, while over 99% of the population has a heart located on the left side of the chest, a small minority (<1%) exhibit dextrocardia, where the heart is located on the right side.[16-18] This anatomical anomaly has profound implications for diagnostic imaging and interpretation. For instance, a chest X-ray of a patient with dextrocardia will appear markedly different from the norm. Similarly, the analysis based on a 12-lead ECG will deviate significantly from conventional patterns: the R wave amplitude in electrodes V1 through V6 will be diminished, and the P wave, QRS complex, and T wave will appear inverted in leads I and augmented vector left, as illustrated in Figure 4.

The variability in data between individuals, as well as within the same individual under different conditions, underscores the importance of verifying and validating preliminary diagnoses through multiple approaches, such as follow-up examinations, professional expertise in root-cause analysis, and consideration of personal history (including genetics) of the patient and the associated patient cohort. Another important diagnostic aspect is risk stratification, which assesses the severity of the patient's condition and determines the urgency of therapeutic intervention. Cardiovascular disease represents a financial burden, costing the EU €282 billion annually.[19] In 2022 alone, general cardiac-related healthcare and long-term medical support accounted for €155 billion. Cardiac health costs the EU 11% of its health budget. Furthermore, productivity losses add in an additional €48 billion, while the costs for out-of-hospital expenses, including in-home support, contribute an additional €79 billion each year – a figure that continues to grow.

3. Methods

In this article, examples of the application of decade-old AI in diagnostics, focusing on signal processing techniques, are described, along with more recent advancements. Some of the AI applications may not solely rely on software techniques. Here, some recent feasibility and pilot diagnostic stages of discovery are also discussed, though, due to the early stage of exploration and proprietary considerations, not all developmental stages are discussed. At this exploratory stage, no claims can be made regarding the accuracy and reliability of these AI-driven diagnostic tools for identifying certain pathological cardiac rhythm disorders. All diagnostic outcomes were compared against documented disorders in the data files and physician reviews. However, given the limited scope of conditions examined and the preliminary nature of the research, a comprehensive statistical validation is not yet feasible. It is important to note that AI-backed diagnostic systems fall under the classification of Class IIb medical devices, which mandates rigorous statistical analysis, including extensive animal and human trials. No animal or human trials have been conducted at this point. The long-term development of these medical device applications is part of a greater corporate plan.

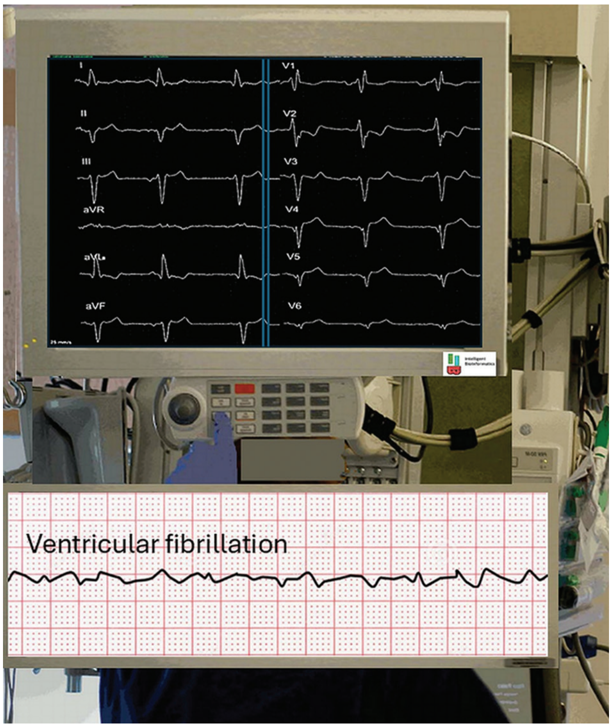


Figure 4. Representation of the acquired 12-lead electrocardiogram recordings: V1, V2, V3, V4, V5, V6, I, II, III, augmented vector right (aVR), augmented vector left (aVL), augmented vector foot (aVF). The bottom section illustrates the complete lack of signal cohesion in ventricular fibrillation. Recording locations: V1: Refer to Figure 3; V2–V4: Record from the front of the heart and are the anterior leads; V5 and V6: Record from the left side of the heart and are the lateral leads; Lead I (bipolar): Between aVL and aVR, creating a second high lateral lead; Lead II (bipolar): Between aVR and aVF, generating the second inferior lead; Lead III (bipolar): Between aVL and aVF, generating the third inferior lead; aVF: Left ankle or left lower abdomen; aVL: Left wrist or shoulder; aVR: Right wrist or shoulder.

All data processing for this research is conducted using high-end personal computers. While some of these analytical procedures can be programmed into existing approved monitoring devices, more complex approaches, such as wavelet and matched filter techniques, require specialized programming adjustments for use in standalone clinical monitoring devices. Notably, no final clinical solutions are provided at this point.

3.1. Programming and its mechanisms

Various programming languages are used to achieve diagnostic modalities and protocols. The input signal is often managed using various mathematical techniques, including noise reduction, filtering, feature extraction, and compression. Examples of the programming languages being used are Machine Language, C, C++, Java, Python, R, Structured Query Language, Swift, MatLab, and Perl. The choice of programming language is often determined by processing speed, convenience of data routing, and the implementation of higher mathematical techniques whenever necessary for complex signal processing. Examples of signal processing mechanisms are the Fourier transform, wavelet transform, Wiener filter, Kalman filter, whitening filter, Butterworth filter, Whiterock filter, whitening transformation, and matched filter processing.[19-28] While a majority of these processing tools are applied in signal denoising techniques, others have more specific tasks in pattern recognition, particularly isolated pathological deviations in an individual's ECG signal. Analog-based signal processing techniques, such as frequency cutoffs (low-pass, high-pass, and band-pass filtering), can be implemented using electronic hardware and computationally, often in conjunction with Fourier transform processing. In machine learning applications, neural networks are used to acquire signals and generate screening templates that represent various pathological conditions. However, significant work remains in defining these templates and customizing them for individual patients. With more than 50 cardiac conditions to consider, distinguishing risk factors and electrophysiological characteristics presents a significant challenge.

Diagnostic screening is further complicated by the stringent confidence levels required for Class IIb medical devices. The infrequency of certain arrhythmogenic episodes adds another layer of complexity, necessitating careful selection of start and end times for the data stream used for the creation of analytical templates. Automated selection of these time frames can be achieved using neural network processing and other signal-processing techniques, such as identifying deviations from established reference depolarization patterns by omitting peaks at specific temporal locations. These computational procedures are integral to the broader toolkit of AI-driven neural network processing. In most cases, analog-to-digital conversion is required before automated signal processing can be performed. Various off-the-shelf analog-to-digital converters are available to perform this task electronically, enabling digital signal processing. Analog signal processing, on the other hand, includes traditional techniques, such as signal amplitude control, modulation/demodulation (e.g., frequency modulation and amplitude modulation), and a plethora of electronic error correction techniques.

3.2. Signal processing: a theoretical approach to the determination of depolarization pattern

An important temporal feature in ECG analysis is the duration of the QRS complex, which is typically identified by its characteristic shape and relatively stable time constant within the waveform. Another critical feature is the time interval between the T wave and the subsequent P wave. This measure is important because it reflects the separation between two key cardiac events, that is, the excitation pulse from the sinus node, represented by the P wave, and the repolarization of the ventricles, which generates the T wave.

To successfully capture and analyze various cellular depolarization frequency patterns, multiple signal processing approaches can be employed, including analog processing, digital signal processing, or a mixed signal processing format. Other techniques include peak detection, frequency determination, the time interval between specific peaks analysis (either same function or different function [Figure 2]: e.g., QT separation, QQ separation, and P-wave repetition rate [which can indicate atrial fibrillation]), pulse widths of various components in the PQRST complex analysis, wavelet transform (however limited due to the temporal nature of pacing events), discrete wavelet transform, Butterworth filtering, discrete path transform, Fourier transform, short-term Fourier transform, Laplace signal transform, and compressed or full matched filter analysis.

Time constants, such as heart rate and various interval period durations (e.g., PR interval, QT interval, and P-wave repetition rate), can be efficiently determined using peak detection or Fourier transform techniques. These computational methods are applicable to any ECG data acquisition system. Notably, peak detection and heart rate variability tests are decade-old techniques that have been used in standard monitoring devices.

For preliminary analyses, various data banks containing healthy and pathological ECG recordings are available, such as those hosted on PhysioNet (https://physionet.org/). Collaborations with cardiology groups have also yielded highly specific data streams. All analytical signals in this study were acquired using routine clinical monitoring equipment. Some analyses were conducted as single-blinded studies, where the analytical team was provided with a well-established arrhythmogenic signal without knowing that it was abnormal. The physicians supplying the signals were fully aware of the deviations, which were often clearly defined. However, due to the early stages of development, not all hidden characteristics and atypical ECG patterns have been investigated for diagnostic purposes at this time.

3.3. Signal analysis in amplitude, temporal framework, spatial framework, and frequency spectrum

Various approaches are employed to evaluate the temporal aspect of ECG signals, such as peak-to-peak analysis, amplitude variability within specific time frames, durations and intervals between sequences of events, and correlation analysis between signals or between certain variables. The most critical temporal feature in ECG analysis is the heart depolarization cycle period, which is determined by measuring the duration from one R wave to the next. Other important features include the duration of individual waves and the time interval between them (e.g., the TP interval: time interval between the T and P waves). The TP interval is important because it reflects the separation between two important events, that is, the pulse rate of the sinus node, which is represented by the P wave, and the ventricular repolarization, which generates the T wave. A broad range of signal processing techniques, applicable in the time and frequency domains, are summarized in Table 1.[11,29-34] While it is beyond the scope of this article to discuss each of these techniques, they can be utilized for different purposes, such as signal pre-processing, noise reduction, and identifying the pathological origins of signal deviations. For instance, unique, idiosyncratic, or infrequent deviations from the normal ECG pattern may be detected using specialized approaches like modified matched filter analysis. The spatial domain is particularly significant in 12-lead ECG and multidimensional decomposition techniques, such as the matched filter approach.

Various standard ECG characteristics can be resolved using straightforward mathematical techniques. For instance, peak detection can quickly determine heart rate, while wavelet transform and matched filter approaches can define specific intervals and segments in the time domain (Figure 1). However, due to continuous advancements in signal processing and innovative approaches, it is impossible to cover all mechanisms within this article. Peak detection results related to the time domain can also be used for the recognition of certain arrhythmias. Besides, atrial and VF require a more complex mathematical approach for early and proactive detection.

An important time-domain feature in ECG is the duration of the QRS complex. In general, the QRS complex is identified by its characteristic shape and relatively stable time constant within the repetitive ECG pattern. In terms of frequency content, the ECG waveform, including the QRS complex, is primarily confined to the high-frequency region, while the P and T waves represent the lower-frequency components. The ST segment, on the other hand, is time-restricted and characterized by its low-frequency content.[8]

The frequency content of a normal ECG often differs significantly from that of a pathological ECG. For instance, a normal heart rate ranges between 60 and 100 beats/min, whereas arrhythmias or fibrillation can result in heart rates

Table 1. Overview of common temporal and frequency analysis tools

Time domain	Frequency domain
Analog-to-digital conversion	Adaptive filtering
Amplitude	Band-pass filtering
Autoregressive model	Continuous wavelet transform
Bilinear transform	Convolution
Continuous	Correlation
Convolution	Discrete cosine transform
Correlation	Discrete Fourier transform
Discrete	Empirical mode decomposition
Discriminant functions (least squares/synthetic)	Fast Fourier transform
Duration	Finite impulse response filtering
Event analysis (missing events)	Fourier transform
Finite impulse response filtering	High-pass filtering
Infinite impulse response filtering	Infinite impulse response filtering
Interval	Inverse Fourier transform
Least squares discriminant functions	Laplace transform
Least squares method (in time series)	Least-squares discriminant analysis
Matched filter technique	Least-squares wavelet analysis
Recurring sequence of events	Low-pass filtering
Sampling property of unit impulse	Matched filter techniques
Synthetic discriminant functions	Phase-shift analysis
Time-reversal technique	Sampling (e.g., Nyquist–Shannon sampling theorem), analog-to-digital conversion
Trigger point	Spectral confinement
Wavelet transform	Spectral decomposition
Z-transform	Synthetic discriminant functions
	S-transform
	Wavelet transform
	Windowing
	Z-transform

Notes: Some specific frequency filters: Wiener filter, Kalman filter, whitening filter, Butterworth filter, Whiterock filter, and whitening transformation; [11] All diagnostic modalities are represented in no particular order, alphabetically ranked.

exceeding 200 beats/min. It is worth noting that a teenager engaged in sporting activities may easily reach or exceed a heart rate of 240 beats/min without immediate health concerns. Beyond differences in heart rate, depolarization and repolarization (on- and off-ramps) can be altered under pathological conditions, requiring an extensive analytical frequency bandwidth to accurately capture these variations in the frequency domain.[8]

In the Fourier domain, a standard (healthy) ECG can be accurately described using the first eight harmonics of the heart rate. However, even minor high-frequency deviations from the conventional ECG waveform often introduce alterations, requiring a larger number of higher harmonics to characterize the frequency-domain features. The baseline ECG typically has a root frequency of approximately 1 Hz, with higher harmonics extending up to 8 Hz for root analysis. As a general guideline, frequency analysis of an ECG should at least cover a range of 0.0001–100 Hz for a normal ECG, though the fundamental frequency may be excluded from the analysis. For arrhythmogenic tendencies, spectral analysis, such as the Fourier transform, may need to extend beyond 200 Hz to capture relevant features. Further analysis is needed to determine the type of arrhythmias, requiring advanced mathematical methods to investigate the isolation and confidence levels of the specific options identified for the final diagnosis. Nonetheless, the final determination of the applicable arrhythmogenic pathology may still rest with the physician. It is important to note that extending the analysis to even higher frequencies is unlikely to yield additional diagnostic information, as the spectrum beyond this point is typically dominated by noise.

The signal processing techniques discussed here are examples of early use of AI in medical diagnostics, particularly in cardiac rhythm analysis, risk assessment, and pathological conditions screening. With advancements in knowledge, wavelet analysis has become increasingly utilized, while the matched filter approach is gaining prominence in analyzing more complex aspects of cardiac depolarization pathology.

3.4. Wavelet transform analysis

Since action potentials are mainly stochastic, wavelet analysis may not always provide the volume or quality of data required for precise diagnosis. However, when analyzing a repetitive signal as the vector summation of multiple action potentials, wavelet-domain features can effectively classify the relative contributions of higher harmonics. The wavelet features used in the analysis of an ECG signal mainly detect scaled or shifted versions of typical patterns or the aspects of depolarization wave. Wavelet decomposition using a mother wavelet that closely resembles the general shape of the QRS complex (i.e., baseline template: generic QRS complex), regardless of electrode numbering and placement, enables the quantitative identification of the QRS complex's location, amplitude, and scaling.[29] In addition, wavelet analysis can reveal features that may not be apparent when using Daubechies and Coiflet wavelets.[34-37]

One clinically significant application of wavelet analysis is the separation of a maternal ECG from the fetal ECG during pregnancy. While the waveforms and wavelet structures of the two ECGs may share similarities, a major difference is that the fetal ECG typically has a higher repetition rate and a smaller amplitude compared to the maternal ECG. The distinct differences in the wavelet signals of fetal ECG and maternal ECG arise because the two ECG signals exist at different time and frequency scales.

3.5. Matched filter analysis

To extract maximum detail from an ECG, the optimal approach is to perform an AI-based mathematical analysis of the four-dimensional vector array derived from 12-lead data acquisition, represented in spherical coordinates:

$$\vec{v} = v(\rho,\phi,\theta,t) \tag{I}$$

This vector array is deconvolved for processing using advanced tools (Figure 5), such as the matched filter approach, which introduces a sparse multidimensional vector $\vec{w}$. In this approach, a well-defined temporal segment of the signal, representing a specific rhythm condition (template) or wavelet, is used to isolate unique depolarization events within the ECG data stream.

Biological signals, such as the ECG, are often stochastic and complex. These signals may vary rapidly due to biological and electrochemical influences. The matched filter approach can be applied to signals that lack symmetry at the zero-amplitude axis, exhibit varying repetition frequencies, or have inconsistent repeatable amplitudes over time.[37-39] The erratic signal is presumably due to chemical, mental, or physical influences, but it is not necessarily a threat or deviation from life-sustaining data streams. The matched filter approach can be used to quantify the similarity between the acquired continuous signal and a configured template, expressed by a cross-correlation coefficient that defines the match. In this way, any deviations from the template can be identified and isolated for diagnostic evaluation.

It has been shown that a small number of linear measurements of the ECG contain adequate information to reconstruct a sparse or compressible signal (i.e., compressive sensing/data acquisition).[39-41] In practice, this sparse signal will be redefined in multidimensional space. Compressive sensing is a technique for efficiently acquiring data and reconstructing signals within limits by facilitating data reduction below

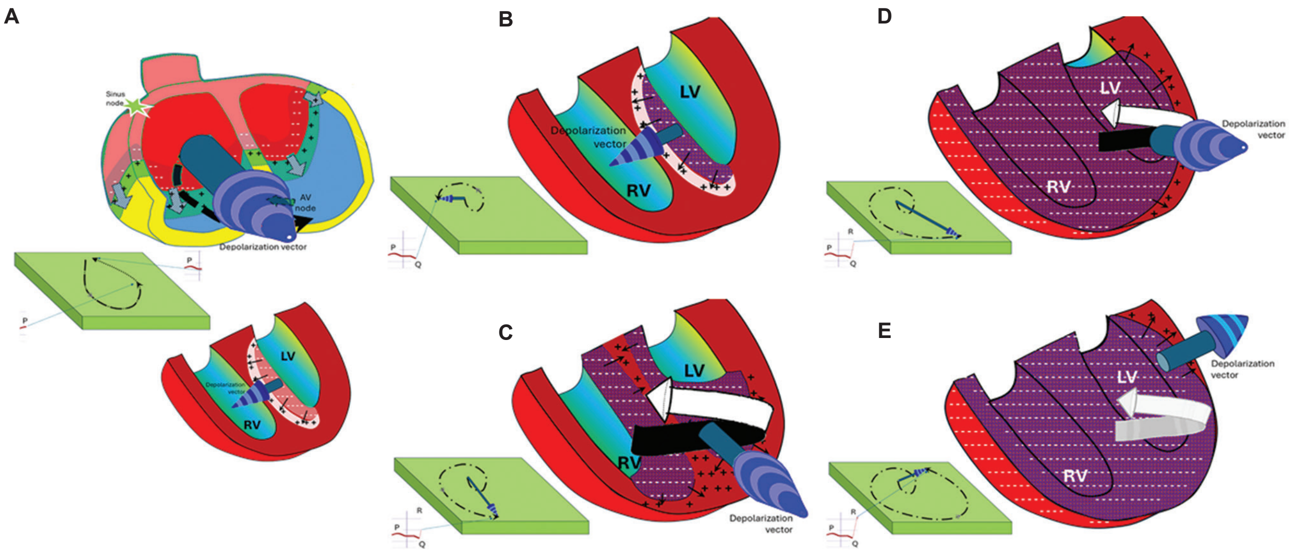


Figure 5. A three-dimensional rotating vector representing the cascading depolarization across the entire cardiac volume. (A) Initiated in the atria by a discharge from the sinus node (the natural cardiac pacemaker), represented by the P wave. (B) The onset of septal and left ventricular depolarization, originating from the passage of the excitation pulse through the Purkinje fibers to the apex of the heart, represented by the Q wave. (C) Progression into full ventricular walls illustrates the initial phases of the R wave. (D) Middle progression of the R wave. (E) Completion of the R wave depolarization front. Note that the trajectory of the rotating vector in the respective left corner is a two-dimensional projection of the full three-dimensional vector orientation as a function of time.

Abbreviations: AV: Atrioventricular; LV: left ventricle; RV: Right ventricle.

the Nyquist-Shannon sampling theorem's threshold.[10] These limits are determined through risk management by setting the accuracy, reliability, and confidence levels. For class IIb medical devices (such as important diagnostic software), these limits are set at 95% accuracy and 95% reliability. The application of compressive sensing in signal processing results in a significant reduction in computation time, offering timely updates on the patient's conditions.

Consider a J-dimensional signal vector representing the 12-lead ECG vector array (Figure 5). If this newly defined multidimensional vector has only $\varkappa < \mathrm{J}$ non-empty components (defining a $\varkappa$-sparse signal), it can be compressed using $c'\varkappa$ linear measurements (typically, $c' \equiv 3$ or 4). This J-dimensional vector now represents the inner products of the non-overlapping vectors through a set of random vectors, ultimately enabling the reconstruction of signals with high accuracy and probability. Even if the signal is not inherently sparse, it can be expressed as a sparse combination of $\varkappa$ vectors from an appropriate source.

Compressive sensing is particularly advantageous for low-complexity compression solutions, such as those involving low-bitrate signals or data-acquisition devices with low sensitivity (e.g., poorly impedance-matched skin electrodes). By computing random compressive measurements, the sensor encoder operates efficiently without requiring specific assumptions about the data stream, other than sparsity. Moreover, these measurements can be efficiently computed using analog, digital, or hybrid methods.

Given a non-uniform signal of length over time, compressive sensing can be performed on any arbitrary section of disjointed vector segments:

$$\vec{x} = \vec{v} + \vec{n} \tag{II}$$

Where $\vec{n}$ represents the noise elements in multidimensional space. For example, when

$$x_i = v_i + n_i \tag{III}$$

The arbitrary length I aims to obtain a compressed version of the acquired signal. The vector containing the acquired signal data is represented by the vector $\vec{v}$, where the length of the vector satisfies I << J, defining the compression ratio. J represents the number of signal blocks that the signal is partitioned into (signal block denoted as $\mathfrak{U}_r$), with each block having a finite length and spanning a finite time frame. The signal component v_i is scaled by a factor η_0 against a function (f_i) that is offset by a shift of ι_0, as expressed by Equation IV:

$$v_i = \eta_0 \cdot f_{i-\iota_0} \tag{IV}$$

The measured data signal vector $\vec{v}$ can be mathematically defined using N linear projections of the vector $\vec{w}$, with the new measurements represented as a linear combination of the original causative M-dimensional signal, described as:

$$\vec{v} = \Omega\vec{w} \tag{V}$$

Where Ω represents an I × J random (but identified and defined) multidimensional data matrix, typically referred to as the data acquisition sensing matrix.

Under the aforementioned conditions, with a series of J finite-length signal blocks, $\mathfrak{U}_r(\mu)$, the correlation for each segment can be computationally derived from the inner product of the signal $\mathfrak{U}_r$ with the appropriately defined vector φ_n (τ), as a function of time τ. The non-zero component of this inner product (also referred to as scalar product or dot product) represents the respective QRS complex. The matched filter approach provides an expression that represents the conditions of the R wave aspect through estimation in the compressed domain:

$$\varphi_n\left(\mathbb{R}_{\mathfrak{U}\varphi}\left(\tau\right) = \vec{v}, \varphi_\tau\right) \tag{VI}$$

The indices in Equation VI represent the respective signal amplitude extremes (primarily R, followed by P and T), which are to be eliminated and are used to determine the time instance $\theta_{\chi i}$. The chevrons denote the inner product. The time instance provides the basis for calculating the number of events per time frame, such as beats per minute in an ECG.

The ECG signal is ideally acquired through 12 electrodes placed on the chest, with $\mathfrak{U}_r(\tau)$ representing a recording acquired over a time interval T long enough to include at least one QRS complex (one single heartbeat), expressed as:

$$\mathfrak{U}_r\left(\tau\right) = \mathfrak{U}\left(\tau\right) + \mathcal{R}\left(\tau\right) = \sum_i \left\{\alpha_i \varphi\left(\tau - \theta_{\chi i}\right) + r(\tau)\right\} \tag{VII}$$

Where α_i represents the amplitude and $\theta_{\chi i}$ represents the center of a given signal kernel. The kernel encompasses the QRS complex of the individual heartbeat in the ECG signal. An additional consideration for noise contributions is represented by $r(\tau)$, as a function of time τ. This noise includes deviations from the QRS template due to transition resistance between the skin and the recording electrode(s), as well as the distorting influences of the P and T waves on the signal definition and peak detection. The heart rate is determined based on the R peak in the QRS complex and the respective duration between sequential R peaks. The heart rate is subsequently calculated by counting the number of R peaks over a duration of 60 s.

Under the compressive sensing method described above, the one-pulse ECG signal, $\mathfrak{U}_r(\tau)$, is transformed into a J-dimensional vector:

$$\overrightarrow{\mathfrak{U}_r}(\tau) = \vec{\mathfrak{U}} + \vec{\tau} \tag{VIII}$$

While incorporating several random measurements acquired during the recording, represented by:

$$\vec{y} = \Phi\vec{\mathfrak{U}} + \vec{\tau} \tag{IX}$$

Where $\vec{y}$ falls in the domain $\vec{y} \in \Re^{\mathbb{J}}$, while $\vec{\tau}$ identifies the compressed noise and random signal influences based on the measurement and skin preparation techniques. The solution involves calculating the position $\theta_{\chi i}$ of the R-wave peak in the acquired signal using the information embedded in $\vec{y}$ for the short compressive sensing of the signal.

Based on the conditions of operating with only compressive measurements, the data do not allow for pre-processing, such as removing other signal components (e.g., the P and T waves in the ECG) or artifacts like baseline drift. Furthermore, it is not a common practice to perform pre-processing within the compressive sensing sensor. Mathematical operations are designed to account for all deviations, allowing them to be processed appropriately.

The matched filter approach will effectively determine the relevant depolarization peak. With an appropriate choice of signal template, the signal-to-noise ratio can be significantly enhanced without pre-processing. However, the variability in biological data poses a significant risk of false positives or false negatives when applying a healthy or pathological template to compressive sensing signal analysis.

In the QRS complex, the matched filter approach is used in compressive sensing to locate and determine its magnitude, based on an appropriately defined filter. This is done by performing compressive matched filtering on a relatively small number of random frequency-domain samples. In this approach, the data stream measurements are considered projections under the application of a random sensing matrix. The complete signal can subsequently be reconstructed using the results obtained from the compressive sensing approach.

To determine the correlation [$R_{x\varphi}$ (τ)] between the compressed template and the compressed cardiac muscle depolarization pattern [expressed as $\mathfrak{U}_r(\tau)$], white noise is introduced, and samples are taken with a time constant $\theta_{\chi i}$, as defined in Equation V. This time constant is determined by searching for local maxima using the matched filter approach.[35,38,39]

The QRS complex can now be expressed in the compressed domain using Equation X. The solution to the signal data stream measurements, as defined by:

$$\vec{w} = \Omega\vec{v} \tag{X}$$

Can be derived using direct estimators. The associated correlation $R_{x\psi}(n)$ can be assessed by using the direct estimator ($\breve{\mathbb{R}}_{\mathfrak{U}\varphi}$) through matrix multiplication, where $\Omega\varphi_n$ is decomposed in terms of its rows, using:

$$\vartheta_n = \vec{w}, \Omega\overrightarrow{\varphi_n} \tag{XI}$$

Hence, the correlation can be obtained:

$$\breve{\mathbb{R}}_{\mathfrak{U}\varphi} = \left[\vartheta_1, \vartheta_2, \vartheta_3, \ldots, \vartheta_{n-1}, \vartheta_n\right] \tag{XII}$$

The scalar products, or dot products, in Equations XI and XII, involve vectors of length n, illustrating the degree of alignment between the respective vectors. This does not directly require the reconstruction of the measured signal. Alternatively, the use of the orthogonalized estimator ($\breve{\mathbb{R}}_{x\varphi n}$) is often more relevant, as it is calculated by the averaged matrix product, denoted by the chevrons in the following:

$$\breve{\mathbb{R}}_{x\varphi n} = \frac{\mathbb{J}}{\mathbb{I}}\vec{w}.\left(\Omega\Omega^T\right)^{-1}\Omega\varphi_n \tag{XIII}$$

The baseline drift can be compensated for by subtracting the signal mean over time from each signal block before applying the direct estimator in Equation XIII. The signal mean can be estimated using an appropriately chosen symmetric multiplication vector, defined by the transposed unified vector array:

$$\mathbb{L}_m = \left[\frac{1}{\mathbb{J}}, \frac{1}{\mathbb{J}}, \frac{1}{\mathbb{J}}, \frac{1}{\mathbb{J}}, \ldots\ldots\frac{1}{\mathbb{J}}\right]^t \tag{XIV}$$

Subsequently yielding the vector average resulting from the vector product:

$$\widehat{\zeta_{\mathfrak{U}_r}} = \vec{\mathfrak{U}}_r, \overrightarrow{\mathbb{L}_{\boldsymbol{m}}} \tag{XV}$$

This orthogonal estimator ($\breve{\zeta}_{\mathfrak{U}_r}$) can be calculated by multiplying the data acquisition sensing matrix (Ω) with its transposed (Ω^T), as expressed by:

$$\breve{\zeta}_{\mathfrak{U}_r} = \vec{w}.\left(\Omega\Omega^T\right)^{-1}\Omega\mathbb{L}_m \tag{XVI}$$

Subsequently, subtracting the mean signal from the signal data stream holds:

$$\overrightarrow{w'} = \vec{w} - \Omega\breve{\zeta}_{\mathfrak{U}_r}\left[1,1,\ldots,1\right]^t \tag{XVII}$$

Representing the alternating signal free of drift.

Applying these matrix multiplication procedures to a healthy baseline signal will generate the template for analytical computations using the matched filter method. If the healthy signal is unavailable for the patient being investigated for pathological cardiac conditions, an alternative approach is to compute an average from a broad dataset of healthy individuals within the same boundary conditions as the patient. The more closely the boundary conditions match, the more effective the analytical template will be in the matched filter approach, helping to identify deviations within the repetitive full PQRST(U) pulse trains.

In the matched filter approach, the correlation that reveals the optimal estimates for the unknowns shift, $\vec{\xi}$ with respect to ξ_0, and the unknown scaling factor, $\vec{\eta}$ with respect to η_0, is implemented by utilizing the probing sequence or template (H^{ξ}_{-k}), defined by the dot product, or scalar product, of the two respective vectors:

$$\begin{aligned}\vec{\xi}\cdot\vec{\eta} &= \arg\min_{\xi,\eta}\sum\nolimits_k\left(x_k-\eta\cdot H_{\xi-k}\right)^2 = \\ &\arg\max_{\xi,\eta}\left[2\sum\nolimits_k v_k H_{\xi-k}-\eta^2\sum\nolimits_k H_{\xi-k}{}^2\right]\end{aligned} \tag{XVIII}$$

It then follows the removal of noise (with an average value of 0), and the reversal of the sign of the equation, converting the minimum into a maximum. This results in a least-squares minimization within the domain of the data stream. By applying the Cauchy-Schwarz inequality, the two estimators reduce to, for the offset:

$$\vec{\xi} \leq \sum\nolimits_i v_i^2 = Constant \tag{XIX}$$

And for the scaling factor:[40-44]

$$\vec{\eta} = \frac{\sum_k v_k H_{\xi-k}}{\sum_k H_{\xi-k}{}^2} \tag{XX}$$

An estimation is introduced into the correlation between the compressed cardiac muscle depolarization signal and a compressed template (i.e., the average depolarization pattern complex in the compressed domain) using a matched filter template. The process employs the impulse response, which is equated to the temporal inversion of the root base signal description, $\phi(\tau)$. The computed output can be represented by Equation XXI:

$$T_{x\varphi}\left(\tau\right) = \sum\nolimits_{\mu}\mathfrak{U}_r\left(\mu\right)\varphi\left(\mu-\tau\right) \tag{XXI}$$

In this case, the amplitude of $T_{x\varphi}(\tau)$ at specific moments in relative time (τ) directly indicates the degree of cross-correlation agreement between the template and the measured signal. The amplitude maxima are located where the match is highest.

As the data compression ratio ($1-\frac{I}{J}$) increases, the estimated matches of the cross-correlation function become more susceptible to noise. When all these steps are implemented in the signal processing pipeline for the temporal data acquisition of the 12-lead vectorial ECG signal, the continuous cross-correlation process enables the identification and isolation of matches and miss-matches within subsets of the input signal over time. These subsets can then be analytically compared against matched filters representing any of the over 30 pathological conditions related to cardiac rhythm abnormalities (Table 2). Following this, a root-cause analysis is conducted to determine the underlying electrochemical and electrophysiological factors contributing to the observed ECG deviations. These factors may include cell damage, malfunction, and congenital predisposing that disrupt normal cardiac function.

In the AI and machine learning framework, the computational processing of the ECG data stream begins with a comparison of the temporal pattern against a baseline normal ECG template (H^{ξ}_{-k}). Subsequently, the signal is compared against templates corresponding to the pathological conditions outlined in Table 2. This machine learning-driven process aims to identify the highest likelihood of a pathological match, such as a diseased left ventricle, as illustrated in Figure 6. A probability distribution is computed, summarizing the degree of agreement between the acquired signal and various templates. This distribution highlights both the interval matches and specific pathological conditions, providing a comprehensive diagnostic overview for physicians.[44-47] The physician then evaluates these findings in the context of the patient's background and may consult colleagues for a second opinion. Based on this thorough evaluation, a tailored treatment plan is developed.

4. Results and discussion

Long-term ECG recordings are often analyzed manually, relying on the expertise of a physician, typically an electrophysiologist, to identify episodes of arrhythmogenic cardiac depolarization or sequences of depolarization pulses that indicate the development or progression of a pathological cardiac condition. These findings guide

Table 2. A prospective spectrum of cardiac rhythm diagnostics and screening that can be quickly and automatically derived from an electrocardiogram

Number	Content
1	1st degree atrioventricular block
2	Atrial fibrillation
3	Atrial flutter
4	Bradycardia
5	Complete the right bundle branch block
6	Hypertrophic cardiomyopathy (also known as left ventricular hypertrophy)
7	Incomplete right bundle branch block
8	J-point
9	J-60 point
10	Left-axis deviation (Purkinje fibers)
11	Left anterior fascicular block
12	Left bundle branch block
13	Left ventricular dysfunction (defined as a left ventricular ejection fraction≤35%)
14	Low QRS voltages
15	Non-specific Intraventricular conduction disorder
16	PP interval
17	PR interval
18	PR segment
19	P wave duration
20	P top amplitude (in reference to QRS)
21	Pacing rhythm (sinoatrial node functionality)
22	Premature atrial contraction
23	Premature ventricular contraction
24	Prolongation of interval (e.g., long QT interval, ST segment duration, and PQ interval)
25	Prolongation of PR interval
26	Q wave abnormalities (e.g., duration, amplitude, deletion/reduction of the follow-on R wave.)
27	RR interval (i.e., derive heart rate)
28	Right-axis deviation
29	Right bundle branch block
30	Sinus arrhythmia
31	Sinus bradycardia
32	Sinus rhythm
33	Sinus Tachycardia
34	ST segment
35	ST–T segment
36	Supraventricular premature beats
37	TP interval
38	T wave abnormalities
39	T wave inversion
40	Ventricular premature beats
41	Ventricular fibrillation

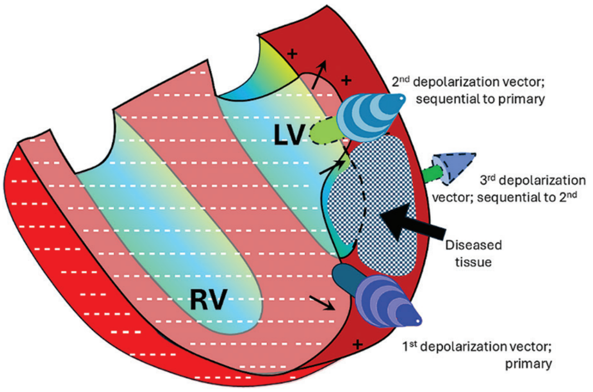


Figure 6. The infarcted volume in the left ventricular (LV) wall will cause a deformation of the R wave (as shown in Figure 7) in each recursive electrocardiogram cycle. This deformation, presenting itself as multiple conjoined peaks, a stretched peak, or other phenomena, can be detected using adaptive filter signal analysis, wavelet computational comparison, or matched filter signal processing. These methods compare the recorded signal to a healthy QRS template, as well as several other templates derived from a worldwide database of pathological and healthy electrocardiogram patterns.

the choice of treatments, which may include chemical interventions (e.g., medication), therapeutic interventions (e.g., cryoablation, alcohol ablation, and surgery), and long-term device implementation (e.g., pacemaker and implantable cardioverter defibrillator. In the worst case, the patient may require a heart transplant. These electrophysiological deviations are not age-specific. Individuals may have congenital defects that predispose them to developing pathological cardiac conditions later in life. Notably, some of these conditions can be life-threatening.

Deviations in cardiac depolarization patterns are generally classified as arrhythmias. There are numerous types of arrhythmia (as illustrated in Figure 7), each with distinct root causes and treatment options. Some of these conditions include implantation of pacemakers or implantable cardioverter defibrillators to manage or correct abnormal rhythms.

One particularly life-threatening arrhythmia is VF, which can lead to SCD if not immediately treated.[4,47-50] VF can be triggered by heart block and coronary artery disease, which is characterized by impaired perfusion and reduced oxygenation of the affected cardiac muscle cells. In cases of severe heart failure, where contractility is diminished, and intraventricular diastolic filling pressure is elevated, ventricular tachycardia can deteriorate into VF. Other life-related factors that can contribute to heart failure include:

(i) Prolonged or deep general anesthesia, for example, during extensive myocardial hypoxia or at the onset of anesthesia, especially in severely diseased patients

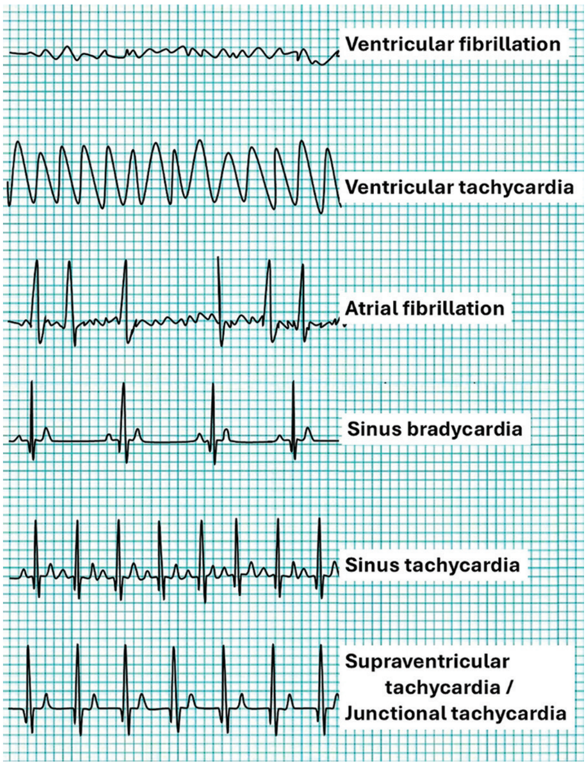


Figure 7. Representation of several characteristic electrocardiogram patterns associated with common pathological cardiac conditions

(ii) Mechanical trauma, for example, being crushed by heavy, moving participants during an American football game or other sporting events where hormonal levels are elevated
(iii) Respiratory arrest and ventilatory failure are more commonly associated with VF than primary organic cardiac disease
(iv) Many dangerous ventricular arrhythmias, which, although resembling asystole hemodynamically and electrocardiographically, are finely constructed VF.

VF is also prevalent in other mammals, including canines, felines, and equines. The most well-known dog breeds that suffer from congenital arrhythmias include the Dalmatian, Boxer, Doberman, and German Shepherd. In mammals, including humans, arrhythmias can also result from conditions like myocarditis or, more rarely, infective endocarditis (a bacterial infection of the heart valves).

A common pathological rhythm deviation observed in the frequency domain is sinus tachycardia, defined as a sinus rhythm exceeding 100 beats per minute. This condition is often due to a malfunctioning or overactive sinus node. It can also occur as a physiological response to physical exercise or psychological stress. When the sinus rhythm becomes irregular, with the longest PP or RR intervals exceeding the shortest by 0.16 seconds or more, the condition is diagnosed as sinus arrhythmia. This condition is common across all age groups but is particularly prevalent in teenagers and pre-teens, where it is sometimes not considered a pathological condition. One potential cause of sinus arrhythmia is the influence of the vagus nerve, which regulates both respiration and heart rhythm. Through the effects of the vagus nerve on the sinus node, the heart rate is increased during inspiration and decreased during expiration.

Accurate detection of fetal cardiac signals during pregnancy is critical for identifying potential congenital heart conditions. Fetal ECG recordings are obtained using leads placed in a special arrangement on the abdomen of the mother.[9] While the maternal ECG is dominant, the fetal ECG, which is superimposed on the maternal ECG, can be isolated using frequency-domain filtering. The fetal heart rate is generally higher than the maternal heart rate, enabling effective separation of the two signals using filters designed and configured in the frequency domain. Advanced frequency and wavelet analysis techniques are essential for extracting meaningful diagnostic information from these recordings.

ECG signals are stochastic and rarely perfectly reproducible. Direct analytical methods based solely on sine and cosine transformations often yield unsatisfactory results. Instead, the correlation function is typically estimated from the raw ECG data, followed by power spectrum analysis to derive frequency-domain analysis. This approach, however, is not discussed in detail in this article, as much of the verification and validation processes are proprietary.

Several frequency effects measured by the electrodes are not directly related to cardiac activity. These include noise from electrode movement, respiratory signals, and skeletal muscle activity (electromyogram). Such noise is typically filtered out in the frequency domain. A notch filter is commonly applied to eliminate frequency interference from equipment and reduce capacitive noise. Other common automated diagnostic modalities include derivation of average heart rate, peak heart rate, and low-level heart rate (e.g., recovery after exercise), as well as intervals between specific peaks in the PQRSTU complex.

More than 30 pathological cardiac rhythm deviations can now be identified using AI-driven analysis. The primary method for achieving this level of detail is 12-lead ECG monitoring over an extended period (Figure 4). The 12-lead ECG recording provides a three-dimensional view of cardiac depolarization, allowing for precise localization and characterization of abnormalities.[50-66] The approach captures the rotating depolarization vector of the heart, enabling the detection of both spatial and temporal pathological conditions. This provides additional information beyond standard cardiac rhythm identifiers, such as heart rate and various specific interval measurements.

To isolate the true depolarization sequence, noise must be eliminated using techniques such as cutoff filtering, frequency filtering, wavelet filtering, and Wiener filtering. These methods can be applied to eliminate motion artifacts, impedance mismatches, and tissue conductivity variation.

The data acquisition device will also incorporate electronic mechanisms for signal (pre-) processing, including gain adjustment, offset correction, frequency filtering (primarily low-pass), and hardware-based automated noise reduction.[65,66] In addition, frequent and sequential repetition of deviations in the ECG pattern can be analyzed using Fourier analysis, wavelet filtering, and matched filter analysis. The wavelet, matched filter, and Wiener filtering analysis rely on an accurate base template for analysis, which must be customized to match the individual's specific history. The choice of template used for screening and diagnosis is typically based on the boundary conditions of the patient group being investigated, such as athlete status, activity level, age, gender, genetics, weight, habits, and disabilities. Women are often misdiagnosed due to the use of male-centered cardiac rhythm templates. Therefore, great care must be taken to apply the appropriate screening boundary conditions and construct a respective diagnostic template based on available prior monitoring data for the individual or a similar group. By utilizing existing data streams under healthy conditions, neural network-generated templates can be used to detect deviations and reconstruct a template that matches the suspected pathological condition through machine learning analysis.

The use of AI and deep learning techniques can be enhanced through convolutional neural network processing to provide scheduled updates, as well as to verify and validate the sampling templates used in each analytical procedure for the individual patient and their corresponding patient group selected as the baseline.[46-59] The automated ECG analysis routines currently extract the following values from the data: corrected QT interval, heart rate, P height, PQR-interval, QRS width, QT interval, R height, RR interval, and ST interval.[61]

Several verified diagnoses, as well as several desired requirements for cardiac health monitoring, are listed in Table 2. Some of these diagnostic modalities are applicable to 3-lead, 5-lead, and 12-lead ECG.[64-66] In addition, numerous diagnostic techniques are currently being developed and are in various stages of preparation for commercial release under regulatory constraints. The associated pathological conditions include but are not limited to, heart attack, arrhythmias, heart failure, cardiomyopathy, and heart valve disease. Additional diagnoses of other pathological conditions, including several rare or atypical cardiac depolarization patterns, are not listed due to their infrequent occurrence and the current limitations of the ongoing research and development program. Furthermore, many arrhythmias have yet to be investigated for AI-based discovery. For quantifiable time-specific conditions, refer to Figure 1.

5. Conclusion

Artificial intelligence-supported diagnostics offer a powerful tool for rapid patient screening and identification of cardiac abnormalities. However, a thorough investigation into the root cause of these deviations is essential to develop a comprehensive, patient-specific treatment.[65,66] The growing recognition of AI's accuracy and reliability in diagnostics is gaining widespread acceptance. In particular, AI-driven screening and risk-stratification, based on comparison with specific signal patterns of a broad range of pathological cardiac conditions, has demonstrated significant value. These tools not only improve diagnostic precision but also save time in clinical settings. Time-domain and frequency-domain filtering and analyses have long been used with excellent results. However, the introduction of more advanced techniques, such as wavelet analysis and matched filtering, has enabled the identification of complex disease patterns with higher accuracy. AI-based diagnosis can statistically determine the prevalence of certain arrhythmogenic conditions by matching ECG data with predefined templates or groups of arrhythmias. Each classification may encompass various cardiac rhythm morphologies. In addition, AI can offer details about the duration of pathological events and the frequency of specific phenomena over time, which may indicate the presence of one or more arrhythmias. The feasibility study described here provides preliminary insights into the use of AI for cardiac health screening. Nonetheless, further analysis by a physician, incorporating the patient's history and likely additional tests, will lead to a patient-specific diagnosis. Alternatively, the physician may treat the arrhythmia spectrum identified by AI as a broad-based issue, addressing it with a single, multipurpose medication.[67,68] Several drugs currently available on the market can effectively manage a wide range of cardiac rhythm problems. However, certain conditions, such as VF, cannot be controlled pharmacologically and require an implantable cardioverter defibrillator. Similarly, other arrhythmias that cannot be well-controlled by pharmaceutical means require the use of a pacemaker, as determined based on follow-up examination by the physician.[67-74]

Acknowledgments

The author wants to acknowledge the discussions with and data repository sharing by the cardiologists and electrophysiologists at the following institutions: Cleveland Clinic, Cleveland, OH, USA and London, UK, Manchester University, Manchester, UK, Cambridge University, Cambridge, UK, Havard Medical School, Cambridge, MA, USA, and University of São Paulo Medical School, São Paulo, Brazil.

Funding

None.

Conflict of interest

The author declares no conflicts of interest.

Author contributions

This is a single-authored article.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data

Data cannot be shared due to confidentiality concerns and proprietary analysis procedures.

References

1. Empana JP, Lerner I, Valentin E, *et al*. Incidence of Sudden cardiac death in the European Union. *J Am Coll Cardiol*. 2022;79(18):1818-1827.

   doi: 10.1016/j.jacc.2022.02.041

2. Zwietering PJ, Knottnerus JA, Rinkens PE, Kleijne MA, Gorgels AP. Arrhythmias in general practice: Diagnostic value of patient characteristics, medical history and symptoms. *Fam Pract*. 1998;15:343-353.

   doi: 10.1093/fampra/15.4.343

3. Hoefman E, Van Weert HCPM, Reisma JB, Koster RW, Bindels PJE. Diagnostic yield of patient-activated loop recorders for detecting heart rhythm abnormalities in general practice: A randomised clinical trial. *Fam Pract*. 2005;22:478-484.

   doi: 10.1093/fampra/cmi048

4. Hoefman E, Bindels PJE, Van Weert HCPM. Efficacy of diagnostic tools for detecting cardiac arrhythmias: Systemic literature search. *Neth Heart J*. 2010;18:543-551.

   doi: 10.1007/s12471-010-0831-0

5. Auer R, Bauer DC, Marques-Vidal P, Butler J, Min LJ, Cornuz J. Association of major and minor ECG abnormalities with coronary heart disease events. *JAMA*. 2012;307:1497-1505.

   doi: 10.1001/jama.2012.434

6. Mai VT, Alattas KA, Bouteraa Y, Ghaderpour E, Mohammadzadeh A. Personalized blood pressure control by machine learning for remote patient monitoring. *IEEE Access*. 2024;12:83994-84004.

   doi: 10.1109/ACCESS.2024.3413572

7. Cheng CH, Wong KW, Chin JW, Chan TT, So RHY. Deep learning methods for remote heart rate measurement: A review and future research agenda. *Sensors*. 2021;21(18):6296.

   doi: 10.3390/s21186296

8. Androulakis E, Fielder C. *Artificial Intelligence in ECG Diagnostics-where are we now*? European Society of Cardiology; 2024. Available from: https://www.escardio.org/councils/council-for-cardiology-practice-(ccp)/cardiopractice/artificial-intelligence-in-ecg-diagnostics-where-are-we-now [Last accessed on 2025 Mar 18].

9. Sau A, Pastika L, Sieliwonczyk E, *et al*. Artificial intelligence-enabled electrocardiogram for mortality and cardiovascular risk estimation: A model development and validation study. *Lancet Digit Health*. 2024;6(11):e791-e802.

   doi: 10.1016/S2589-7500(24)00172-9

10. Siontis KC, Noseworthy PA, Attia ZI, Friedman PA. Artificial intelligence-enhanced electrocardiography in cardiovascular disease management. *Nat Rev Cardiol*. 2021;18(7):465-478.

    doi: 10.1038/s41569-020-00503-2

11. Najarian K, Splinter R. *Biomedical Signal and Image Processing*. 2nd ed. Boca Raton, FL: CRC Press; 2012.

    doi: 10.1201/b11978

12. Einthoven W. Galvanometrische registratie van het menschelijk electrocardiogram. In: *Herinneringsbundel Prof. S.S. Rosenstern*. Leiden: Eduard Ijdo; 1902. p. 101-106.

13. Einthoven W. Het tele-cardiogram. *Ned Tijdschr Geneeskd*. 1906;22:1517-1547.

14. Attia ZI, Noseworthy PA, Lopez-Jimenez F, *et al*. An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: A retrospective analysis of outcome prediction. *Lancet*. 2019;394(10201):861-867.

15. Yusuf SW, Durand JB, Lenihan DJ, Swafford J. Dextrocardia: An incidental finding. *Tex Heart Inst J*. 2009;36(4):358-359.

16. Bernasconi A, Azancot A, Simpson JM, Jones A, Sharland GK. Fetal dextrocardia: Diagnosis and outcome in two tertiary centres. *Heart*. 2005;91(12):1590-1594.

    doi: 10.1136/hrt.2004.048330

17. Garg N, Agarwal BL, Modi N, Radhakrishnan S, Sinha N. Dextrocardia: An analysis of cardiac structures in 125 patients. *Int J Cardiol*. 2003;88(2-3):143-156.

doi: 10.1016/s0167-5273(02)00539-9

18. Wiener N. *The Interpolation, Extrapolation and Smoothing of Stationary Time Series', Report of the Services 19, Research Project DIC-6037 MIT*; 1942.

19. Luengo-Fernandez R, Walli-Attaei M, Gray A, *et al*. Economic burden of cardiovascular diseases in the European Union: A population-based cost study. *Eur Heart J*. 2023;44(45):4752-4767.

doi: 10.1093/eurheartj/ehad583

20. Brown RG, Hwang PYC. *Introduction to Random Signals and Applied Kalman Filtering*. 3rd ed. New York, NY: John Wiley & Sons; 1996.

21. Welch LR. *Wiener-Hopf Theory 2006-11-25*. Available from: https://csi.usc.edu/PDF/wienerhopf.pdf [Last accessed on 2025 Mar 18].

22. Wiener N, Hopf E. *Ueber Eine Klasse Singulärer Integralgleichungen*. Berlin: Sitzungber. Akad. Wiss; 1931. p. 696-706.

23. Wiener N. *Extrapolation, Interpolation, and Smoothing of Stationary Time Series.* New York NY: John Wiley & Sons; 1949.

24. Kolmogorov AN. Stationary sequences in Hilbert space', (In Russian) Bull. *Moscow Univ*. 1941;2(6):1-40. English translation In: Kailath T, editor. *Linear Least Squares Estimation.* Stroudsburg, PA; Dowden, Hutchinson & Ross; 1977.

25. Tison GH, Zhang J, Delling FN, Deo RC. Automated and interpretable patient ECG profiles for disease detection, tracking, and discovery. *Circ Cardiovasc Qual Outcomes*. 2019;12(9):e005289.

doi: 10.1161/CIRCOUTCOMES.118.005289

26. Christopoulos G, Graff-Radford J, Lopez CL, *et al*. Artificial intelligence-electrocardiography to predict incident atrial fibrillation: A population-based study. *Circ Arrhythm Electrophysiol*. 2020;13(12):e009355.

doi: 10.1161/CIRCEP.120.009355

27. Yamashita R, Nishio M, Do RKG, Togashi K. Convolutional neural networks: An overview and application in radiology. *Insights Imaging*. 2018;9(4):611-629.

doi: 10.1007/s13244-018-0639-9

28. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat Mach Intell*. 2019;1(5):206-215.

doi: 10.1038/s42256-019-0048-x

29. Ashok A, Babburi A, Ardra T, Gayathri KS, Indu RJ, Narayanan G. Performance Comparison of Matched Filter, Wavelet Denoising and Wiener Filter Technique in Communication Receivers. In: *2018 3rd IEEE International Conference on Recent Trends in Electronics, Information and Communication Technology (RTEICT).*

doi: 10.1109/rteict42901.2018.9012

30. Proakis JG, Manolakis DG. *Digital Signal Processing*. Available from: https://uvceee.wordpress.com/wp-content/uploads/2016/09/digital_signal_processing_principles_algorithms_and_applications_third_edition.pdf [Last accessed on 2025 Mar 18].

31. Bhattacharyya SS, Deprettere EF, Leupers R, Takala J. *Handbook of Signal Processing Systems*. 3rd ed. New York, NY: Springer; 2019. Available from: https://link.springer.com/book/10.1007/978-3-319-91734-4 [Last accessed on 2025 Mar 18].

32. Lyons RG. *Understanding Digital Signal Processing*. Hoboken, NJ: Prentics Hall; 2010.

33. Casasent D, Chang WT. Correlation synthetic discriminant functions. *Appl Optics*. 1986;25(14):2343-2350.

doi: 10.1364/AO.25.002343

34. Holton T. *Digital Signal Processing, Principles and Applications*. Cambridge, UK: Cambridge University Press; 2021.

doi: 10.1017/9781108290050

35. Alizadehsani R, Roshanzamir M, Hussain S, *et al*. Handling of uncertainty in medical data using machine learning and probability theory techniques: A review of 30 years (1991-2020). *Ann Oper Res*. 2021;21:1-42.

doi: 10.1007/s10479-021-04006-2

36. Hannun AY, Rajpurkar P, Haghpanahi M, *et al*. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nat Med.* 2019;25(1):65-69.

doi: 10.1038/s41591-018-0268-3

37. Turin GL. An introduction to matched filters. *IRE Trans Inform Theory*. 1960;6(3):311-332.

doi: 10.1109/TIT.1960.1057571

38. Davenpoort MA, Boufounos P, Wakin M, Baraniuk R. Signal processing with compressive measurements. *IEEE J Select Top Signal Process*. 2010;4(2):445-460.

doi: 10.1109/JSTSP.2009.2039178

39. Wimalajeewa T, Varshney PK. Sparse signal detection with compressive measurements via partial support set estimation. *IEEE Trans Signal Inform Process Over Netw*. 2017;3(1):46-60.

doi: 10.1109/TSIPN.2016.2601025

40. Fornasier M, Rauhut H. Compressive sensing. In: Scherzer O, editor. *Handbook of Mathematical Methods in Imaging*. New York, NY: Springer; 2011.

doi: 10.1007/978-0-387-92920-0_6

41. Cauchy AL. *Extrait du Mémoire sur Quelques Séries Analogues*

À La Série de Lagrange, sur les Fonctions Symétriques, et sur la Formation Directe des Équations Que Produit L'élimination des Inconnues Entre des Équations Algébriques Données. Mémoires de l'Académie des Sciences, t. IX; 1830. p. 73-78. Available from: https://gallica.bnf.fr/ark:/12148/bpt6k901828/f79 [Last accessed on 2025 Mar 18].

42. Steele JM. *The Cauchy-Schwarz Master Class: An Introduction to the Art of Mathematical Inequalities*. The Mathematical Association of America. Cambridge, UK: Cambridge University Press; 2004.

43. Steiger WL. On a generalization of the Cauchy-Schwarz in equality. *Am Math Monthly*. 1969;76:815-816.

    doi: 10.1080/00029890.1969.12000339

44. Pishro-Nik H. *Introduction to Probability, Statistics, and Random Processes*. Sunderland MA: Kappa Research LLC; 2014. Available from: https://www.probabilitycourse.com [Last accessed on 2025 Mar 18].

45. Shamloo AS. Risk assessment in cardiac arrhythmias. *Eur Heart J*. 2020;41(47):4455-4457.

    doi: 10.1093/eurheartj/ehaa808

46. Takase B, Ikeda T, Shimizu W, *et al*. JCS/JHRS 2022 guideline on diagnosis and risk assessment of arrhythmia. *J Arrhythm*. 2024;40(4):655-752.

    doi: 10.1002/joa3.13052

47. Redfield MM, Jacobsen SJ, Burnett JC Jr., Mahoney DW, Bailey KR, Rodeheffer RJ. Burden of systolic and diastolic ventricular dysfunction in the community: Appreciating the scope of the heart failure epidemic. *JAMA*. 2003;289(2):194-202.

    doi: 10.1001/jama.289.2.194

48. McDonagh TA, Metra M, Adamo M, *et al*. 2023 Focused Update of the 2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure. *Eur Heart J*. 2023;44(37):3627-3639.

    doi: 10.1093/eurheartj/ehad195

49. Attia ZI, Kapa S, Lopez-Jimenez F, *et al*. Screening for cardiac contractile dysfunction using an artificial intelligence-enabled electrocardiogram. *Nat Med*. 2019;25(1):70-74.

    doi: 10.1038/s41591-018-0240-2

50. Ribeiro AH, Ribeiro MH, Paixão GMM, *et al*. Automatic diagnosis of the 12-lead ECG using a deep neural network. *Nat Commun*. 2020;11(1):1760.

    doi: 10.1038/s41467-020-15432-4

51. Rawi AA, Elbashir MK, Ahme AM. Classification of 27 heart abnormalities using 12-lead ECG signals with combined deep learning techniques. *Bull Electrical Eng Inform*. 2023;12(4):2220-2235.

    doi: 10.11591/eei.v12i4.4668

52. Attia ZI, Kapa S, Yao X, *et al*. Prospective validation of a deep learning electrocardiogram algorithm for the detection of left ventricular systolic dysfunction. *J Cardiovasc Electrophysiol*. 2019;30(5):668-674.

    doi: 10.1111/jce.13889

53. Kwon JM, Kim KH, Jeon KH, *et al*. Development and validation of deep-learning algorithm for electrocardiography-based heart failure identification. *Korean Circ J*. 2019;49(7):629-639.

    doi: 10.1097/mat.0000000000001218

54. Cho J, Lee B, Kwon JM, *et al*. Artificial intelligence algorithm for screening heart failure with reduced ejection fraction using electrocardiography. *ASAIO J*. 2021;67(3):314-321.

    doi: 10.1097/mat.0000000000001218

55. Pickham D, Zarafshar S, Sani D, Kumar N, Froelicher V. Comparison of three ECG criteria for athlete pre-participation screening. *J Electrocardiol*. 2014;47(6):769-774.

    doi: 10.1016/j.jelectrocard.2014.07.019

56. Sheikh N, Papadakis M, Ghani S, *et al*. Comparison of electrocardiographic criteria for the detection of cardiac abnormalities in elite black and white athletes. *Circulation*. 2014;129(16):1637-1649.

    doi: 10.1161/CIRCULATIONAHA.113.006179

57. Ko WY, Siontis KC, Attia ZI, *et al*. Detection of hypertrophic cardiomyopathy using a convolutional neural network-enabled electrocardiogram. *J Am Coll Cardiol*. 2020;75(7):722-733.

    doi: 10.1016/j.jacc.2019.12.030

58. Kwon JM, Jeon KH, Kim HM, *et al*. Comparing the performance of artificial intelligence and conventional diagnosis criteria for detecting left ventricular hypertrophy using electrocardiography. *Europace*. 2020;22(3):412-419.

    doi: 10.1093/europace/euz324

59. Han L, Askari M, Altman RB, *et al*. Atrial fibrillation burden signature and near-term prediction of stroke: A machine learning analysis. *Circ Cardiovasc Qual Outcomes*. 2019;12(10):e005595.

    doi: 10.1161/CIRCOUTCOMES.118.005595

60. Bos JM, Attia ZI, Albert DE, Noseworthy PA, Friedman PA, Ackerman MJ. Use of artificial intelligence and deep neural networks in evaluation of patients with electrocardiographically concealed long QT syndrome from the surface 12-lead electrocardiogram. *JAMA Cardiol*. 2021;6(5):532-538.

    doi: 10.1001/jamacardio.2020.7422

61. Aufiero S, Bleijendaal H, Robyns T, *et al*. A deep learning approach identifies new ECG features in congenital long QT syndrome. *BMC Med*. 2022;20(1):162.

    doi: 10.1186/s12916-022-02350-z

62. Available from: https://www.biopac.com/knowledge-base/

ecg-analysis [Last accessed on 2025 Mar 18].

63. Vahanian A, Beyersdorf F, Praz F, *et al*. 2021 ESC/EACTS guidelines for the management of valvular heart disease. *Eur Heart J.* 2022;43(7):561-632.

doi: 10.1093/eurheartj/ehab395

64. Kwon JM, Kim KH, Akkus Z, Jeon KH, Park J, Oh BH. Artificial intelligence for detecting mitral regurgitation using electrocardiography. *J Electrocardiol*. 2020;59:151-157.

doi: 10.1016/j.jelectrocard.2020.02.008

65. Elias P, Poterucha TJ, Rajaram V, *et al*. Deep learning electrocardiographic analysis for detection of left-sided valvular heart disease. *J Am Coll Cardiol.* 2022;80(6):613-626.

doi: 10.1016/j.jacc.2022.05.029

66. Splinter R, editor. *Handbook of Physics in Medicine and Biology*. 1st ed. Baco Raton, FL: CRC Press; 2010.

doi: 10.1201/9781420075250

67. *How are Arrhythmias Treated? National Heart, Lung, and Blood Institute*; 2011. Available from: https://www.nhlbi.nih.gov/health/arrhythmias/treatment [Last accessed on 2025 Mar 18].

68. Saljic A, Heijman J, Dobrev D. Recent advances in antiarrhythmic drug therapy. *Drugs*. 2023;83(13):1147-1160.

doi: 10.1007/s40265-023-01923-3

69. Tettey-Engmann FA, Parupelli SK, Desai SS. A review of biomedical devices: Classification, regulatory guidelines, human factors, software as a medical device, and cybersecurity. *Biomed Mater Dev Biomed Mater Dev*. 2024;2:316-341.

doi: 10.1007/s44174-023-00113-9

70. *Diagnosing Arrhythmias; Arrhythmia Alliance*. Available from: https://www.lhch.nhs.uk/media/.resources/64a7ecda3bd825.77246728.pdf [Last accessed on 2025 Mar 18].

71. Desai DS, Hajouli S. *Arrhythmias.* St. Petersburg, FL, USA: StatPearls Publishing; 2025. Available from: https://www.ncbi.nlm.nih.gov/books/NBK558923 [Last accessed on 2025 Mar 18].

72. Classification of cardiac arrhythmias and conduction disturbances. *Am Heart J.* 1979;98(2):263-267.

doi: 10.1016/0002-8703(79)90230-8

73. Antzelevitch C, Burashnikov A. Overview of basic mechanisms of cardiac arrhythmia. *Card Electrophysiol Clin.* 2011;3(1):23-45.

doi: 10.1016/j.ccep.2010.10.012

74. Kiladze MR, Lyakhova UA, Lyakhov PA, Nagornov NN, Vahabi M. Multimodal neural network for recognition of cardiac arrhythmias based on 12-load electrocardiogram signals. *IEEE Access.* 2023;11:133744-133754.

doi: 10.1109/ACCESS.2023.3335176

ACCSCIENCE PUBLISHING

Artificial Intelligence in Health

ORIGINAL RESEARCH ARTICLE

Screening and early detection of cervical intraepithelial neoplasia and cervicitis using a hemoglobin absorption map-derived machine learning algorithm

Phebe George[1], Rekha Upadhya[2], Rinoy Suvarnadas[1], Niranjana Sampthalia[3], and Subhash Narayanan[1]*

[1]Research and Development Division, Sascan Meditech Pvt. Ltd., TIMed, Sree Chitra Tirunal Institute for Medical Sciences and Technology, Thiruvananthapuram, Kerala, India

[2]Department of Obstetrics and Gynaecology, Kasturba Medical College, Manipal Academy of Higher Education, Manipal, Karnataka, India

[3]Department of Biomedical Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, Karnataka, India

Abstract

Early and non-invasive detection of cervical malignancy holds great clinical significance. Diffuse reflectance (DR) spectroscopy has the capability to map tissue transformation at the biochemical, morphological, and cellular levels. We have developed a non-invasive, multimodal imaging system to map changes in tissue autofluorescence using DR for the screening and early detection of cervical cancer and cervical inflammation (cervicitis). The developed multispectral imaging device consists of light-emitting diodes (LED) emitting at 375, 545, 575, and 610 nm wavelengths, along with a 5-megapixel monochrome camera for image acquisition. Camera operation and image analysis are controlled using proprietary software installed on a Windows tablet. The 375 nm LED-excited autofluorescence, and the elastically backscattered light at 545, 575, and 610 nm originating from the cervix tissue are captured by the camera and processed to assess tissue abnormalities. A machine learning (ML) algorithm based on DR image intensity ratio values was developed for tissue classification. It was observed that the R610/R545 image ratio could discriminate malignant cervical sites from normal tissues, achieving a sensitivity of 100% and specificity of 93%. In comparison, cervicitis could be discriminated from normal tissues using the R610/R575 ratio, with a sensitivity of 91.6% and specificity of 94.4%. The study demonstrates the potential of DR imaging in conjunction with ML algorithm to non-invasively screen and detect cervical intraepithelial neoplasia and cervicitis in real time. As compared to the existing practice of Pap smear and colposcopy-directed biopsy, which are subjective and require a waiting period for results, objective screening using CerviScan would help reduce patient anxiety, unnecessary biopsies, and treatment costs. With increased patient screening, the accuracy of the ML algorithm would improve. When integrated into a cloud server, the system could address the needs of multiple users in a field setting.

Keywords: Cervical intraepithelial neoplasia; Cervical inflammation; Diffuse reflectance image intensity ratio; Machine learning algorithm

***Corresponding author:**
Subhash Narayanan
(subhash@sascan.in)

Citation: George P, Upadhya R, Suvarnadas R, Sampthalia N, Narayanan S. Screening and early detection of cervical intraepithelial neoplasia and cervicitis using a hemoglobin absorption map-derived machine learning algorithm. *Artif Intell Health*. 2025;2(3):125-137. doi: 10.36922/aih.8527

Received: January 14, 2025

Revised: March 6, 2025

Accepted: April 10, 2025

Published online: May 2, 2025

Copyright: © 2025 Author(s). This is an Open-Access article distributed under the terms of the Creative Commons Attribution License, permitting distribution, and reproduction in any medium, provided the original work is properly cited.

Publisher's Note: AccScience Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

Cervical cancer is the most common type of cancer affecting women and one of the major public health problems in the world.[1] While it is manageable if detected early, cervical cancer continues to be a leading cause of cancer-related death – particularly in underdeveloped countries – where higher mortality rates are often observed among women from low socioeconomic backgrounds.[2] According to the World Cancer Report 2020, India reports one-fifth of the global burden of cervical cancer,[3] and the disease is the major cause of mortality in countries with a low Human Development Index.

Despite the high mortality rates, cervical cancer is one of the most preventable types of cancer due to its slow progression, the presence of detectable lesions using cytology, and the availability of effective treatments.[4] However, cytology-based Pap smear tests inherently have low sensitivity (<50%) and specificity (60 – 70%).[5] Furthermore, the accuracy of the diagnosis depends on the expertise and skill of the gynecologist collecting the exfoliated cervical cells. The presence of blood, the absence of squamous-columnar junction cells, or the clustering of the cells may interfere with the cytology results. These factors contribute to a high number of false negatives or undetected cases of early-stage or low-grade squamous intraepithelial neoplasia (LSIL) that have a favorable prognosis.

Positive Pap cases are typically referred for conventional visual inspection with acetic acid method of examination, and colposcopy-guided biopsies are considered the gold standard for cervical cancer detection. However, these tests are inherently subjective in nature, with the selection of the most malignant or abnormal site depending on the clinician's expertise.[6] Furthermore, there is a long waiting period before pathology results become available for patient care. Thus, there is a long-felt need for screening tests that are objective, with improved diagnostic accuracy for the detection of LSILs in real time. Factors including lack of awareness, discomfort during the screening procedure, and the need for multiple visits prevent women from periodic screening.[7] A non-invasive and low-cost modality that provides real-time screening results is essential to make women come forward for periodic check-ups. This would enable early detection of the disease, leading to a reduction in the mortality rates over time. With the introduction of vaccines against human papillomavirus (HPV) infection, cervical cancer prevention has undergone dramatic changes over the past decade. Although some countries have seen a decline in HPV-mediated cervical disease, widespread implementation of vaccination is limited by economic considerations, vaccine availability, and vaccine adoption hesitancy among the population.

Cervicitis refers to inflammation of the cervical tissue. This could be asymptomatic or symptomatic, including vaginal discharge, bleeding, and irritation, as well as intermenstrual or postcoital bleeding and lower abdominal pain. Cervicitis can be caused by either bacteria such as *Chlamydia*, *Gonorrhea*, and *Trichomoniasis*, or non-infectious exposure to chemicals, including condoms, tampons, and cervical caps. The current diagnosis of cervicitis is the same as for cervical cancer.

When light falls on tissue, various processes such as absorption, reflection, refraction, transmission, and scattering occur. The absorption and scattering properties of light depend on the tissue's internal structure, layer thickness, biochemical constitution, and the wavelength of the incident light. Therefore, spectroscopy-based techniques – such as those utilizing tissue fluorescence, scattering, and diffuse reflectance (DR) – are effective in identifying the structural, morphological, and biochemical changes in cervical tissues during cancer development.

The hemoglobin (Hb) and oxyhemoglobin (HbO_2) have distinct absorption spectral features in the visible spectrum, with HbO_2 showing two distinct peaks around 545 nm and 575 nm. However, Hb exhibits a higher molar extinction coefficient in the 600 – 650 nm spectral range compared to HbO_2.[8] This differential absorption cross-section can be utilized to discriminate between normal and abnormal cervical tissues.

Studies have been conducted to assess the potential of DR spectroscopy in discriminating cervical cancer. The chromophore Hb has absorption peaks around 280 nm, 420 nm, 540 nm, and 580 nm.[9] The DR spectrum of cervical tissues exhibits changes in the HbO_2 absorption at 542 nm and 577 nm due to a reduction in heme production during malignant transformation.[10,11] Shaikh *et al.*[12] cross-validated the DR technique with Raman spectroscopy to compare the classification accuracy for the detection of tumor sites in normal patients. In another clinical study, Prabitha *et al.*[13] reported an increase in the R545/R575 image intensity ratio with an increase in the grade of cervical cancer. Similarly, the DR image ratio R545/R575 was used in the detection of early-stage malignancies in the oral cavity with good diagnostic accuracy.[14,15]

Prasanth *et al.*[16,17] also utilized DR spectral measurements and multispectral imaging techniques to detect gingivitis and periodontitis in patients with gum inflammation. Significant changes in the recorded spectra were observed at 545 nm and 575 nm with gingival inflammation, and the DR spectral ratio R620/R575 was found to give a good correlation with different grades of gingival inflammation. A sensitivity of 91.6% and a specificity of 93% were observed in discriminating healthy and mildly inflamed gingiva using the R620/R575 image intensity ratio. In another study, the R620/R575 spectral ratio was able to discriminate between healthy gingiva from gingivitis, with 90% sensitivity and 94% specificity, whereas a sensitivity of 91% and a specificity of 100% were obtained when discriminating gingivitis from periodontitis.[18] Recently, Barik *et al.*[19] recorded the fluorescence emission of the cervix on excitation at 325 nm, and the spectral features were used to discriminate dysplasia from inflammatory changes. An increase in fluorescence emission intensity in cervicitis samples was observed at 435 nm compared to normal tissue, which was attributed to the fluorophore nicotinamide adenine dinucleotide. Zhang *et al.*[20] utilized a feature fusion method to extract information from Raman spectra and its derivatives. They classified inflamed cervical tissues as high-grade squamous intraepithelial lesions and LSIL based on the intensity of the prominent spectral peaks at 548, 640, 1,452, and 1,664 cm^{-1}. In another study, colposcopy images, in conjunction with deep learning, achieved an accuracy of 95.2% in discriminating chronic cervicitis from cervical cancer.[21]

A recent study combining fluorescence and DR imaging for the detection of oral cavity lesions reported improved diagnostic accuracy in discriminating potentially malignant oral lesions from normal tissues. This was achieved using a machine learning (ML) algorithm based on the DR image ratio R620/R545, which represents changes in the deoxy to oxygenated Hb absorption in tissue.[8] The study highlights the effectiveness of non-invasive screening modalities to provide real-time user feedback and to enhance compliance, particularly for the early detection of cervical cancers.

In the past decade, ML models have been widely used for medical diagnosis. Dong *et al.*[22] developed an ML model for cervical cancer risk stratification using full-genotyping of high-risk HPV test data. The study compared four ML models zs– XGBoost, support vector machine (SVM), random forest, and naïve Bayes – where the XGBoost model was found to be the most effective model.

Surface plasmon resonance biosensor with ML optimization was developed by Wekalao *et al.*[23] for cervical cancer detection. This support vector regression was found to enhance the sensor's predictive capability, reducing the stimulation time by 80%. Several deep learning techniques have been used for cervical cancer diagnosis using pathology slides and colposcopy images.[24]

As part of this study, we have developed a multispectral imaging system for multimodal imaging of the cervix. With the help of an ML algorithm, the DR imaging system evaluates the accuracy of the DR image ratios, R610/R545 and R610/R575, for the detection of cervical cancer and cervicitis, respectively. The study conducted using this device demonstrates its potential for real-time tissue status assessment and its ability to detect the most malignant site for biopsy. The DR image ratios were correlated with the histopathology results of guided biopsies to develop an ML algorithm. The diagnostic accuracy of the screening was determined from the scatter plot diagrams, and the results were presented accordingly.

2. Materials and methods

2.1. Methodology

The CerviScan (Figure 1) consisted of a bimodal imaging camera designed for fluorescence and DR imaging of the cervix. The device was equipped with a 5-megapixel monochrome camera (MU9PM-MH, Ximea, GmbH, Germany) to record the fluorescence emission of collagen on excitation with 370 nm LEDs (ATS2012UV365, Kingbright, United States). The HbO_2 absorption changes in cervical tissues were assessed from the DR images captured under illumination with LEDs emitting at 545 nm (L1C1-GRN1000000, LUMI LEDs, Netherlands) and 575 nm (SMP2-SGC, Bivar, United States), with both overlapping the HbO_2 absorption spectra at 542 nm and 577 nm. The DR images were also captured following LED illumination at 610 nm (SMP2-SOC, Bivar, United States), where the absorption changes in cervical tissues due to Hb were stronger compared to HbO_2. A Windows tablet (Chuwi, China) was connected to a USB camera (Ximea, GmbH, Germany), with proprietary software installed for camera control, image capture, and analytics. The monochrome images (Im545, Im610, Im575, and F370) recorded were processed by software in real time to generate ratio images (R610/R545 and R610/R575) and their corresponding pseudocolor maps (PCM). The region of interest (ROI) was then marked on these ratio

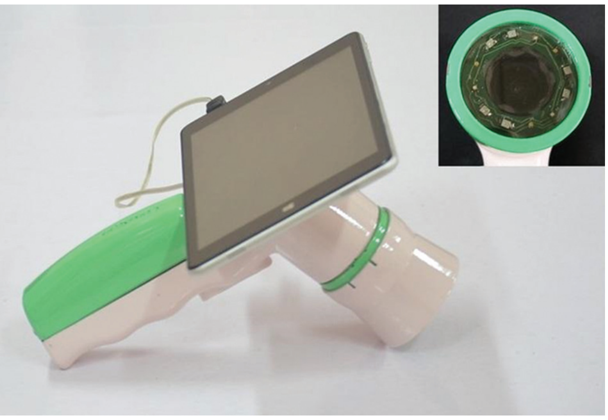


Figure 1. Hand-held CerviScan device developed for cervical cancer detection. The inset shows the front view of the device with the light-emitting diodes positioned around the camera.

images, with reference to the pseudocolor intensity maps of autofluorescence and the DR images displayed on the screen. The software detected and marked the site with the highest ratio value in the abnormal area or ROI. A biopsy was obtained from this site, and the results were correlated with the histopathology findings.

Before the initiation of the imaging procedure, the camera sensor was calibrated to the ambient light conditions using a tissue phantom. The device was positioned at a working distance of 15 cm from the tissue phantom, and the calibration icon in the software was clicked. The four sets of LEDs blinked sequentially until the calibration was successfully completed. During the calibration process, the shutter speed of the imaging camera was adjusted relative to the LED intensity to accommodate the dark room conditions to provide the same levels of image contrast, as verified in the captured image following the calibration procedure. A calibration window appeared after the calibration was successfully completed. The device needs to be recalibrated when there is a change in the room light intensity.

2.2. Study design and patient enrollment

The clinical study was conducted at the Department of Obstetrics and Gynecology, Kasturba Medical College, Manipal, India, after obtaining approval from the Institutional Ethics Committee (Ref no. 152/2020). This trial was also registered with the Clinical Trial Registry of India (REF/2020/08/036212).

Study participants were enrolled according to the study's inclusion criteria. Women who were referred for a Pap smear test and those with clinically suspicious lesions in their cervix were included in the study. Women with transmittable diseases and those who had previously received treatment for cervical cancer were excluded from the study. Signed informed consent was obtained from all participants before the initiation of any examination procedures.

2.3. Data acquisition

The patient was asked to lie down in the supine lithotomy position. After cleaning the vaginal area with a soft cotton swab, the gynecologist inserted a speculum for proper visualization of the cervix. A colored photo of the cervix was taken as a reference. The CerviScan device was held steadily and moved back and forth to obtain a focused image on the screen. The capture switch on the device was pressed, and four monochrome images of the cervix were recorded and stored in the computer following sequential illumination at 370 nm, 545 nm, 575 nm, and 610 nm wavelengths. The captured images were reviewed for clarity, and if required, the images were recaptured. The imaging of the entire cervix enabled clinicians to identify the optimal areas for further investigation.

The cervical smear was collected from the patients referred for Pap smear testing using a cervical broom and sent for cytology examination. Women with clinically suspected cervical malignancy underwent either a conventional or colposcopy-guided cervical biopsy. The site for the biopsy was decided based on the visual impression and the gynecologist's experience. The cervix was anatomically labeled using clock positions, with the anterior tip at the 12 o'clock position and the posterior tip at the 6 o'clock position. The cervical arteries and veins run parallel in 3 and 9 o'clock positions, and biopsies were generally avoided in this area.

2.4. Data processing

Figure 2 illustrates the data processing workflow of CerviScan. In the first step, the camera captured four monochrome images. In the pre-processing step, the software created pixel-by-pixel ratio images of DR-recorded images (R610/R545 and R610/R575). The PCMs of these ratio images, along with the PCM of the F375 (fluorescence) image, were displayed from blue to red, with increasing grades of malignancy. For feature extraction, the areas of the cervix with clinical significance were identified by adjusting the color bar in the PCM image of tissue autofluorescence and the DR ratio image (R610/R545), followed by marking the ROI on either the fluorescence or DR ratio images. In the post-processing, the software identified the highest ratio value in the marked ROI for abnormal tissues and an average of the values for normal tissue. The program identified the most malignant site within the ROI by locating the pixel with the highest ratio value.

3. Results

A total of 109 patients were recruited for this study. Out of the study population, 42 patients with histopathology results were included in the analysis (Table A1). The results of histopathology and cytology were correlated with the R610/R545 ratio values of the biopsy site. The patients who underwent hysterectomy following a negative Pap smear test were also included as healthy subjects. The results of histopathology or cytology were correlated with the R610/R545 ratio values of the biopsy sites. The processed DR image ratio R610/R575 was analyzed to discriminate cervicitis from normal tissues.

3.1. Case studies

The severity of the disease in the cervix was assessed based on the percentage increase in the DR ratio R610/R545 value compared to that of the normal cervix. Low DR ratio values indicate normal tissues, while higher ratio values represent higher grades of malignancy. The most malignant site was identified as a site with the highest DR image ratio R610/R545 value, and a biopsy was obtained from this site for an accurate determination of the cancer grade through histopathology.

Typical cases of patients with cervical intraepithelial neoplasia (CIN) and one case of cervicitis are presented below:

3.1.1. Case 1

Figure 3 shows a 56-year-old patient with complaints of post-menopausal bleeding and abdominal pain. The gynecologist identified a massive growth covering the anterior and posterior lips of the cervix and the upper part of the vagina. The patient was diagnosed with squamous cell carcinoma based on a Pap smear test. A conventional biopsy was obtained from the growth. Our device recorded a R610/R545 ratio value of 2.078. The histopathology result diagnosed the disease as grade 2 squamous cell carcinoma.

3.1.2. Case 2

This is the case of a 50-year-old patient presenting with foul-smelling discharge and blood stains for the past 2 years, along with passage of clots, abdominal pain, general weakness, and loss of appetite. The physical examination showed an irregular mass on the anterior side of the cervix, with the involvement of the posterior and lateral fornix, as well as the lower part of the vagina. The histopathology result following cervical biopsy confirmed malignancy. Immunohistochemistry and immunofluorescence, along with histopathology results, confirmed the diagnosis of carcinosarcoma (squamous cell carcinoma with sarcoma), with an R610/R545 image ratio value of 3.005, as illustrated in Figure 4.

3.1.3. Case 3

A 50-year-old patient with cervicitis was studied based on the DR image ratio R610/R575 value. The patient had difficulty in passing urine, and her cervix showed erosions (Figure 5). Based on the biopsy results, the patient was diagnosed with chronic cervicitis and no dysplasia. This case demonstrated a relatively higher R610/R575 ratio value of 2.433, compared to a lower R610/R545 ratio value of 1.461.

3.2. Scatter plot and algorithm development

The scatter plot diagram of the R610/R545 ratio values, correlated with the corresponding pathological results, was used to classify normal and malignant cervical tissues

Figure 2. Block diagram of data processing and analysis using CerviScan

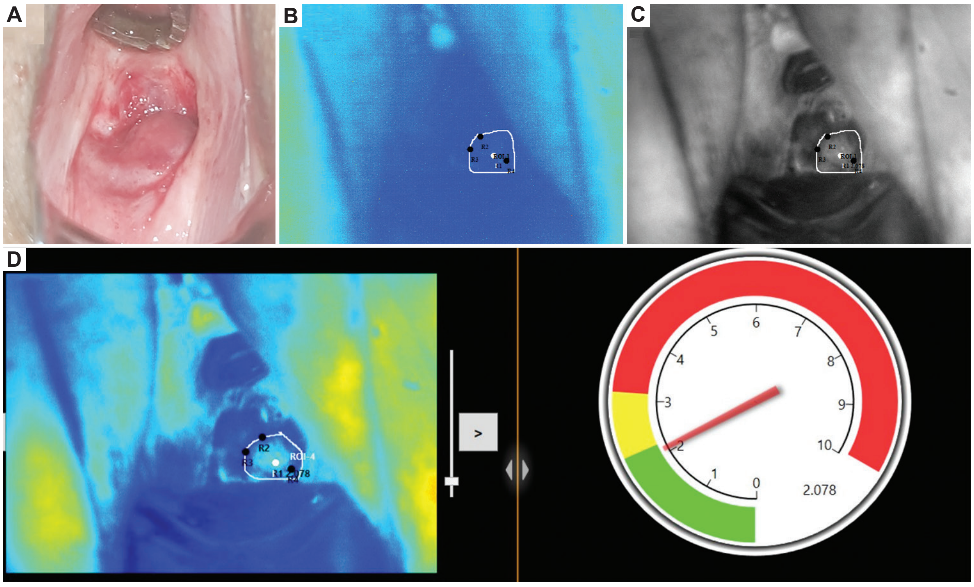


Figure 3. Images of grade 2 cervical carcinoma: (A) Clinical impression, (B) pseudocolor fluorescence, (C) monochrome diffuse reflectance image ratio R610/R545, and (D) pseudocolor diffuse reflectance image ratio R610/R545, with a dial indicating a ratio value of 2.078 in the marked region of interest

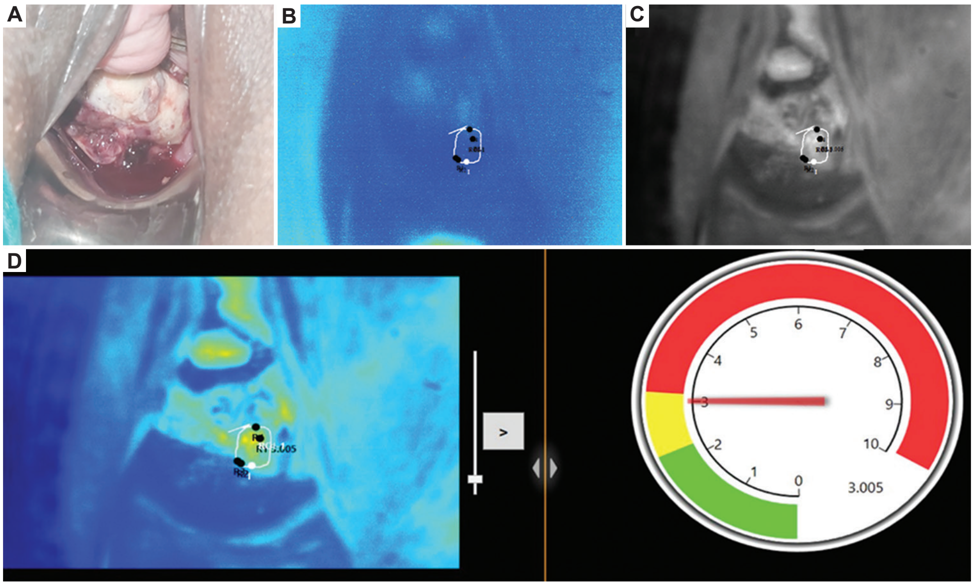


Figure 4. Images of cervical carcinosarcoma: (A) Clinical impression, (B) pseudocolor fluorescence, (C) monochrome diffuse reflectance image ratio R610/R545, and (D) pseudocolor diffuse reflectance image ratio R610/R545, with a dial indicating a ratio value of 3.005 in the marked region of interest

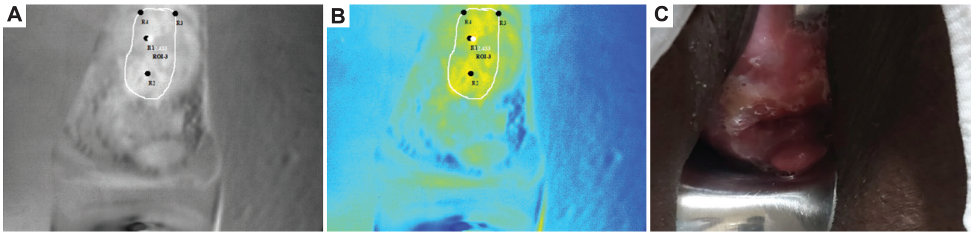


Figure 5. Patient with chronic cervicitis: (A) Monochrome diffuse reflectance image ratio R610/R575 with marked region of interest, (B) pseudocolor diffuse reflectance image ratio R610/R575, and (C) clinical image of the cervix

(Figure 6). The discrimination line was drawn at 1.470, which is the average DR ratio value of all normal and abnormal sites. Patients with no dysplasia were classified as normal. Among 27 normal cases, three cases were misclassified as abnormal, whereas all malignant cases were correctly classified, leading to a specificity of 88% and sensitivity of 100%. The mean DR ratio value for normal sites was 1.367, whereas the mean DR ratio value for all malignant cases was 1.871.

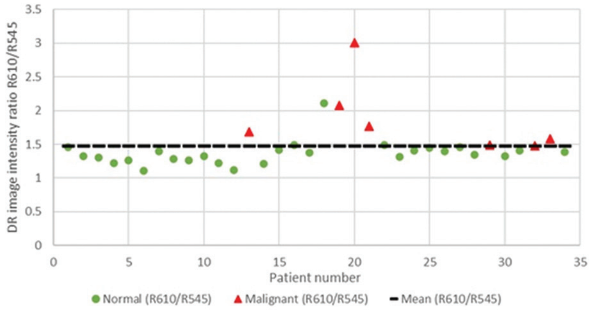


Figure 6. Scatter plot algorithm of DR image intensity ratio R610/R545 used to classify normal and malignant tissues, with a discrimination line drawn at the mean ratio value of 1.47

Abbreviation: DR: Diffuse reflectance.

Figure 7 shows the scatter plot of the R610/R575 image ratio used to discriminate cervicitis from the normal cervix. There were 18 patients in the normal group, with an average R610/R575 ratio value of 1.266, whereas the abnormal group of 17 patients with cervicitis showed an average R610/R575 ratio value of 1.663. Tissue classification was carried out by drawing the discrimination line at 1.459, representing the mean R610/R575 ratio for all normal (1.266) and abnormal cases with cervicitis (1.663). None of the malignant cases were included in this classification. Using the R610/R575 ratio, a sensitivity of 70.58% and specificity of 83.33% were achieved in discriminating cervicitis from normal tissues.

The key features of the ML algorithm were in regard to its potential for mapping changes in HbO_2 absorption, quantifying tissue abnormalities, and biopsy guidance. The key inputs to this algorithm were the ratio values derived from the images of the cervix. The processed ratio images (R610/R545 and R610/R575) were used as the primary feature for classification. The R610/R545 ratio value-based algorithm classifies cervical tissues as normal or malignant, while the R610/R575 ratio value-based algorithm classifies cervical tissues as normal or cervicitis. The model was trained using histopathology-verified datasets from 109 patients, with 35 biopsy-confirmed cases. The model provides feedback to the clinician or its user in real-time and provides biopsy guidance. It was observed that the increase in the R610/R545 ratio value correlated with the malignant status of the tissues, whereas the increase in the ratio R610/R575 correlated with cervicitis. The collagen fluorescence intensity was expected to decrease in malignant sites, and the decreasing trend was presented on screen as a PCM.

The receiver operator characteristics (ROC) curves were plotted for the R610/R545 ratio value algorithm to discriminate normal tissues from malignant tissues, and the R610/R575 ratio to discriminate normal tissues from cervicitis (Figure 8). The area under the curve was found to be 0.957 for the ROC of R610/R545 ratio and 0.803 for the ROC of R610/R575 ratio, classifying normal tissues from malignant tissues and normal tissues from cervicitis, respectively.

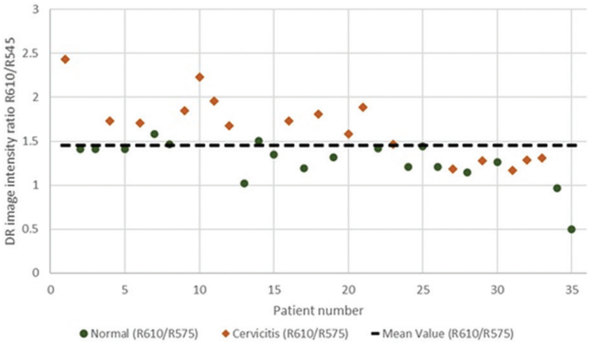


Figure 7. Scatter plot algorithm of DR image intensity ratio R610/R575 used to classify normal tissues and cervicitis, with a discriminating line drawn at the mean ratio value of 1.459
Abbreviation: DR: Diffuse reflectance.

The box plot (Figure 9) represents the distribution of R610/R545 image intensity ratio values for normal and malignant tissues. The horizontal line inside the box represents the median of both data sets. The wider box for malignant samples suggests a larger variation in intensity values compared to normal samples. The malignant samples demonstrated higher intensity values compared to the normal samples. This suggests a significant difference in the R610/R545 image intensity ratio between normal and malignant cervix arising from the pathological changes.

4. Discussion

In this study, a combination of DR and fluorescence imaging was used to identify cervical abnormalities. The higher percentage variance in R610/R545 ratio values highlights the device's ability to detect cervical precancers in a clinical setting. In histopathology, patients with cervicitis but no dysplasia were classified as unhealthy for cervicitis classification; however, these patients were considered normal for the R610/R545 image ratio classification.

In a clinical study conducted by Prabitha *et al.*,[10] it was reported that absorption at 542 nm and 574 nm in the DR spectrum due to HbO_2 becomes stronger owing to reduced heme production in malignant tissues, whereas no significant change in intensity was observed at 610 nm. Stephen *et al.*[15] pointed out that the heme production in malignant tissues is altered as a result of the reduced

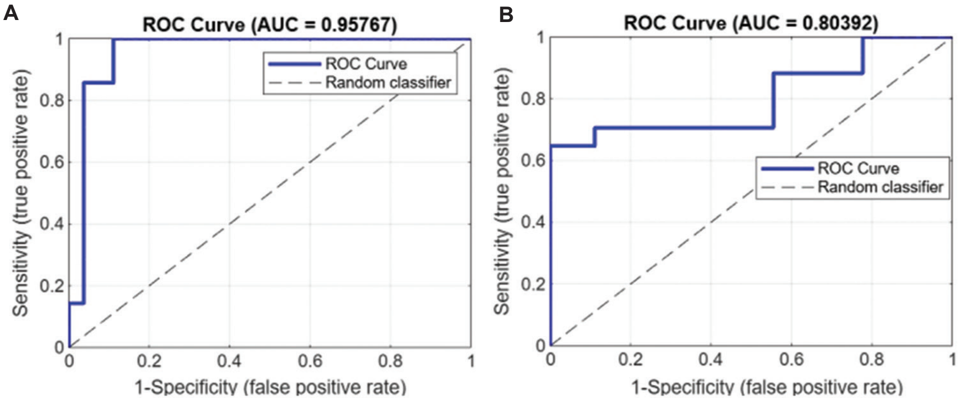


Figure 8. ROC-AUC for image intensity ratios: (A) R610/R545 ratio for normal tissues versus malignant tissues, and (B) R610/R575 ratio for normal tissues versus cervicitis
Abbreviations: AUC: Area under curve; ROC: Receiver operator characteristics.

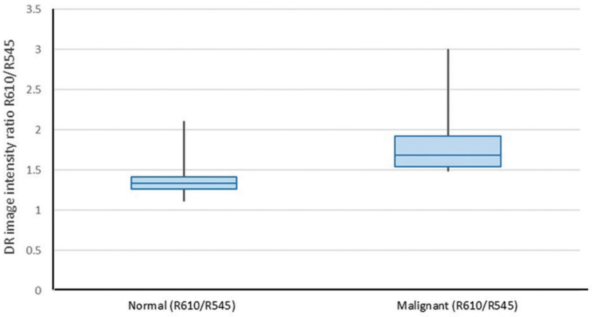


Figure 9. Box plot representing the distribution of DR image intensity ratio (R610/R545) value for normal and malignant tissues
Abbreviation: DR: Diffuse reflectance.

activity of the ferrochelatase enzyme. This leads to lower Hb production and, consequently, lower absorption at the HbO_2 absorption peak. Our study showed a linear relationship between the R610/R545 image ratio value with tissue status (Table A1), highlighting the potential of the R610/R545 image ratio to discriminate between healthy and cancerous lesions of the cervix. In line with our results, Narayanan *et al.*[8] reported a sensitivity of 82.2% and specificity of 96.63% for the R610/R545 image ratio to discriminate oral potentially malignant lesions from normal tissues.

In this analysis, squamous metaplasia occurring at the squamocolumnar junction was considered healthy. Two patients (No. 19 and 21 in Table A1) diagnosed with squamous metaplasia were classified as true negatives, with an average DR image ratio (R610/R545) value of 1.397. Endometrial adenocarcinoma was reported in one patient (No. 22 in Table A1); however, the biopsy results confirmed that the cervix was free from carcinoma. The R610/R545 image ratio value of this patient was 2.11 and was classified as false positive. Case 2 represents carcinosarcoma, a malignancy involving the epithelial and mesenchymal parts of the cervix.[25] Although this disease is reportedly challenging to diagnose, the device recorded the highest DR image ratio value of 3.005, suggesting its ability to discriminate between normal and carcinosarcoma tissues.[26]

As this was a pilot trial, only a small number of patients were recruited for this study. However, with more patients enrolled, the scatter plots developed for determining the sensitivity and specificity of detection could progress to a cloud-based ML algorithm that could classify tissue inflammation and different grades of cancer/CIN from normal tissues. Point-of-care detection of cervical diseases, made possible through a cloud-based ML algorithm, would establish CerviScan as a low-cost alternative for population-based screening. With a greater number of users sharing their screening data and pathology results with the ML algorithm, its robustness in providing feedback on tissue status would increase over time, leading to a more accurate, objective, and real-time tool for cervical cancer screening. With the availability of a larger data set, a supervised ML algorithm with SVM could be configured to classify different grades of cervical cancer, such as LSIL, HSIL, and CIN grades 1, 2, and 3, as well as carcinoma *in situ*. The SVM classifies the data by identifying the best line to separate data points, with the support vectors being the closest points to the hyperplane. These support vectors influence the position of hyperplane, leading to a more generalized cloud-based classification model. A similar cloud-based integration of an ML algorithm was utilized in the case of OralScan, an intraoral multimodal camera for oral cancer screening and biopsy guidance.[8] In this ML algorithm, the R610/R545 image intensity ratio was utilized to classify both normal and potentially malignant tissues, as well as normal versus abnormal tissues. This algorithm processed oral tissues from different anatomical sites in the oral cavity with diverse morphologies. The similarity of this study with our study is the use of R610/R545 image ratio value for cancer screening and biopsy guidance. The hypothesis is that the cervical and oral tissues are made of squamous epithelium, and hence, the optical features are expected to be similar.

In this study, the collagen fluorescence intensity was recorded in all patients using excitation at 370 nm. The recorded fluorescence images appeared dark, possibly due to the low penetration of ultraviolet light in cervical tissues. The stroma of the cervix, concealed by epithelial tissue, is the source of collagen fluorescence. Hence, the excitation light would find it difficult to reach the stroma residing below the basal layer of epithelium.[27] Furthermore, the absorption of the 370 nm light by other biochemical constituents of the cervix – which absorbs at this wavelength and emits with increased intensity during malignant transformations in the tissue – could also lead to an unpredictable fluorescence emission.[28]

In a recent study conducted on 160 patients, the sensitivity and specificity of the Pap smear test were observed to be 47.19% and 64.79%, respectively, for the detection of premalignant lesions of the cervix, and 64.72% and 52.74%, respectively, for colposcopy.[29] Although colposcopy shows a comparatively higher diagnostic accuracy, both of these modalities are subjective. In comparison, our DR imaging technique is objective, with the DR ratio value representing changes in the deoxy- and oxygenated Hb absorption. This ratio correlates with variations in backscattered light intensity due to changes at the cellular level.

The most common symptoms reported by patients enrolled in the study (Table A1) were abdominal pain and abnormal bleeding. Fewer patients had spotting, abnormal discharge, and irregular menstrual cycles. The presence of blood or moisture in the cervix could affect the DR image ratio values. These cervical conditions may cause light to reflect off the surface, resulting in specular reflection. The resulting ratio values from specular reflection can distort DR measurements. Therefore, when a patient has bleeding or experiences bleeding on touch, the DR ratio values may be misleading. This limitation can be minimized by cleaning the cervix before imaging. Another limitation of this study is the low number of malignant cases. Further multicenter studies are planned to improve the algorithm and develop a grade-specific classification.

During conventional cervical check-ups, most women who undergo a Pap smear test do not return to collect the reports, which are usually available in 2 weeks for further follow-up. The non-availability of real-time screening results is a significant drawback for screening modalities such as Pap smears or colposcopy-directed biopsies. The ability of the multimodal imaging system (CerviScan) to provide real-time results, with the help of an ML algorithm, presents a strong rationale for adopting this novel technology in screening women above 30 years of age. CerviScan can help to identify, locate, and treat CIN, while minimizing its possibility to progress to higher grades of CIN and reducing mortality. In addition, the capability of CerviScan to locate the most malignant site for biopsy enhances its value by providing a more accurate pathology. The real-time image ratio (R610/R545 and R610/R575) analytics and PCM display of autofluorescence, as well as HbO_2 absorption maps, help in the speedy assessment of malignant status and/or cervicitis.

5. Conclusion

Quantitative detection of cervical malignancies without the need for tissue incision holds great clinical significance. The algorithm developed for non-invasive multispectral widefield imaging using CerviScan provides an opportunity to objectively screen and detect cervical cancer in real time, while also assisting the gynecologist in locating the malignant site for biopsy. The cloud-based platform approach will present a more accessible and cost-effective solution for cervical cancer screening. The data from different centers, including patients with diverse age groups, menstrual phases, and gynecological histories, would refine the algorithm over time. As there are changes in the cervical anatomy during the menstrual span of women, age-specific algorithms would enhance the diagnostic accuracy. ML algorithms can outdo the interobserver variability among cytology results and could detect subtle cellular changes that might be missed by human experts. ML-powered digital diagnosis and ML-based screening would enable remote diagnosis in low-resource settings with limited access to specialists.

In this study, an overall diagnostic accuracy of 94% was achieved using the DR image ratio R610/R540 algorithm for the discrimination of malignant tissues from normal tissues. However, the data derived using collagen fluorescence could not lead to a confirmative diagnosis. Alternative markers relying on protoporphyrin IX fluorescence with excitation at 405 nm could be utilized, as in OralScan, to improve screening accuracy. Furthermore, the intensity and uniformity of illumination could be enhanced by the addition of more LEDs.

The mortality rates due to cervical cancer could be minimized by utilizing CerviScan in large-scale screening programs across the country. We believe that early detection of cervical cancer would reduce healthcare costs and minimize the number of patients lost during follow-up, as often occurs with traditional screening methods such as Pap smear and HPV deoxyribonucleic acid tests, which require a waiting period of 1 week to get the screening results. The potential of the device to screen and detect cervical cancers at the point-of-care in real time, along with its capability for biopsy guidance – without the need for visual inspection with acetic acid/visual inspection with Lugol's iodine staining – will improve patient compliance and its effectiveness as a novel modality for the early direction of cervical cancers.

Acknowledgments

The authors are grateful to acknowledge the support received from the Ethics Committee at the Kasturba Medical College (KMC), Manipal, India. The authors would like to thank Prof. Dr. Stanley Mathew, Department of Surgery, KMC, for his invaluable guidance and advice. The cooperation and support received from the enrolled patients was commendable.

Funding

The project was partially supported through a grant received under the ELEVATE 100 program of the Government of Karnataka in 2017.

Conflict of interest

The authors declare that they have no competing interests.

Author contributions

Conceptualization: Subhash Narayanan
Formal analysis: Phebe George, Rinoy Suvarnadas
Investigation: Phebe George, Rekha Upadhya

Methodology: Subhash Narayanan
Project administration: Rinoy Suvarnadas
Resources: Rekha Upadhya, Niranjana Sampthalia
Supervision: Subhash Narayanan
Writing – original draft: Phebe George
Writing – review & editing: Subhash Narayanan

Ethics approval and consent to participate

This study was approved by the Institutional Ethics Committee of Kasturba Medical College, Manipal, Karnataka, India (Ethics approval number: 152/2020). Written informed consent was obtained from all participants included in the study.

Consent for publication

All patients signed informed consent forms before enrollment in the study.

Availability of data

Data presented in this paper will be available from the corresponding author on reasonable request.

References

1. Ferlay J, Soerjomataram I, Dikshit R, *et al*. Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer*. 2015;136(5):E359-E386.

   doi: 10.1002/ijc.29210
2. Denny L. Cervical cancer: Prevention and treatment. *Discov Med*. 2012;14(75):125-131.
3. Sankaranarayanan R, Ramadas K. Socioeconomic factors and cancer prevention in India: Diverse interventions are needed. In: *World Cancer Report Cancer Research for Cancer Prevention*. Vol. 4.4. France: International Agency for Research on Cancer; 2020. p. 258-265.
4. Leyden WA, Manos MM, Geiger AM, *et al*. Cervical cancer in women with comprehensive health care access: Attributable factors in the screening process. *J Natl Cancer Inst*. 2005;97(9):675-683.

   doi: 10.1093/jnci/dji115
5. Barut MU, Kale A, Kuyumcuoğlu U, *et al*. Analysis of sensitivity, specificity, and positive and negative predictive values of smear and colposcopy in diagnosis of premalignant and malignant cervical lesions. *Med Sci Monit*. 2015;21:3860-3867.

   doi: 10.12659/MSM.895227
6. Rahmadwati G, Naghdy G, Ross M, Todd C, Norachmawati E. Classification of Cervical Cancer Using Histology Images. In: *2010 Second International Conference on Computer Engineering and Applications*. Bali, Indonesia: IEEE; 2010. p. 515-519.

   doi: 10.1109/ICCEA.2010.105
7. Bosch FX, Burchell AN, Schiffman M, *et al*. Epidemiology and natural history of human papillomavirus infections and type-specific implications in cervical neoplasia. *Vaccine*. 2008;26(Suppl 10):K1-K16.

   doi: 10.1016/j.vaccine.2008.05.064
8. Narayanan S, Anand S, Prasanna R, *et al*. Bimodal multispectral imaging system with cloud-based machine learning algorithm for real-time screening and detection of oral potentially malignant lesions and biopsy guidance. *J Biomed Opt*. 2021;26(8):086003.

   doi: 10.1117/1.JBO.26.8.086003
9. Prasad PN. *Introduction to Biophotonics*. 1st ed. United States: John Wiley and Sons; 2003.
10. Prabitha VG, Suchetha S, Jayanthi JL, *et al*. Detection of cervical lesions by multivariate analysis of diffuse reflectance spectra: A clinical study. *Lasers Med Sci*. 2016;31(1):67-75.

    doi: 10.1007/s10103-015-1829-z
11. Mirabal YN, Chang SK, Atkinson EN, Malpica A, Follen M, Richards-Kortum R. Reflectance spectroscopy for *in vivo* detection of cervical precancer. *J Biomed Opt*. 2002;7(4):587-594.

    doi: 10.1117/1.1502675
12. Shaikh R, Prabitha VG, Dora TK, *et al*. A comparative evaluation of diffuse reflectance and Raman spectroscopy in the detection of cervical cancer. *J Biophotonics*. 2016;10(2):242-252.

    doi: 10.1002/jbio.201500248
13. Prabitha VG, Suchetha S, Jayanthi JL, *et al*. Multi-spectral diffuse reflectance imaging for detection of cervical lesions: A pilot study. *Int J Eng Sci Innov Technol*. 2014;3(6):169-177.
14. Messadi DV, Younai FS, Liu HH, Guo G, Wang CY. The clinical effectiveness of reflectance optical spectroscopy for the *in vivo* diagnosis of oral lesions. *Int J Oral Sci*. 2014;6:162-167.

    doi: 10.1038/ijos.2014.39
15. Stephen MM, Jayanthi JL, Unni NG, *et al*. Diagnostic accuracy of diffuse reflectance imaging for early detection of pre-malignant and malignant changes in the oral cavity: A feasibility study. *BMC Cancer*. 2013;13:278.

    doi: 10.1186/1471-2407-13-278
16. Prasanth CS, Stephen MM, Dora TK, Maheshwari A, Subhash N. Non-invasive detection of periodontal disease using diffuse reflectance spectroscopy: A clinical study. In: *Biomedical Applications of Light Scattering VI*. Vol. 8230. Bellingham: SPIE; 2012. p. 117-127.
17. Prasanth CS, Betsy J, Jayanthi JL, Nisha UG, Prasantila J,

Subhash N. *In vivo* inflammation mapping of periodontal disease based on diffuse reflectance spectral imaging: A clinical study. *J Biomed Opt*. 2013;18(2):026019.

doi: 10.1117/1.JBO.18.2.026019

18. Chandra Sekhar P, Betsy J, Presanthila J, Subhash N. Discrimination of periodontal diseases using diffuse reflectance spectral intensity ratios. *J Biomed Opt*. 2012;17(2):027001.

    doi: 10.1117/1.JBO.17.2.027001

19. Barik AK, Pavithran MS, Nelliat M, *et al*. Laser induced fluorescence of cervical tissues: An *in vitro* study for the diagnosis of cervical cancer from the cervicitis. *J Opt*. 2022;24(5):054002.

    doi: 10.1088/2040-8986/ac59e2

20. Zhang H, Chen C, Ma C, *et al*. Feature fusion combined with Raman spectroscopy for early diagnosis of cervical cancer. *IEEE Photonics J*. 2021;13(3):3900311.

    doi: 10.1109/JPHOT.2021.3075958

21. Huang W, Sun S, Yu Z, Lu S, Feng H. Chronic cervicitis and cervical cancer detection based on deep learning of colposcopy images toward translational pharmacology. *Front Pharmacol*. 2022;13:911962.

    doi: 10.3389/fphar.2022.911962

22. Dong B, Lu Z, Yang T, *et al*. Development, validation, and clinical application of a machine learning model for risk stratification and management of cervical cancer screening based on full-genotyping hrHPV test (SMART-HPV): A modelling study. *Lancet Reg Health West Pac*. 2025;55:101480.

    doi: 10.1016/j.lanwpc.2025.101480

23. Wekalao J, Kumaresan MS, Mallan S, *et al*. Metasurface based surface plasmon resonance (SPR) biosensor for cervical cancer detection with behaviour prediction using machine learning optimization based on support vector regression. *Plasmonics*. 2024:1-24.

    doi: 10.1007/s11468-024-02623-8

24. Sarhangi HA, Beigifard D, Farmani E, Bolhasani H. Deep learning techniques for cervical cancer diagnosis based on pathology and colposcopy images. *Inform Med Unlocked*. 2024;47:101503.

    doi: 10.1016/j.imu.2024.101503

25. El-Nashar SA, Mariani A. Uterine carcinosarcoma. *Clin Obstet Gynecol*. 2011;54(2):292-304.

    doi: 10.1097/GRF.0b013e31821ac635

26. Gupta N, Dudding N, Smith JHF. Eight cases of malignant mixed Müllerian tumor (carcinosarcoma) of the uterus: Findings in SurePath™ cervical cytology. *Diagn Cytopathol*. 2014;42(2):165-169.

    doi: 10.1002/dc.22910

27. Arbyn M, Weiderpass E, Bruni L, *et al*. Estimates of incidence and mortality of cervical cancer in 2018: A worldwide analysis. *Lancet Glob Health*. 2020;8(2):e191-e203.

    doi: 10.1016/S2214-109X(19)30482-6

28. Nordstrom RJ, Burke L, Niloff JM, Myrtle JF. Identification of cervical intraepithelial neoplasia (CIN) using UV-excited fluorescence and diffuse-reflectance tissue spectroscopy. *Lasers Surg Med*. 2001;29(2):118-127.

    doi: 10.1002/lsm.1097

29. Najib FS, Hashemi M, Shiravani Z, Poordast T, Sharifi S, Askary E. Diagnostic accuracy of cervical pap smear and colposcopy in detecting premalignant and malignant lesions of cervix. *Indian J Surg Oncol*. 2020;11(3):453-458.

    doi: 10.1007/s13193-020-01118-2

Appendix

Table A1. Patient details, obstetric history, clinical symptoms, physical findings, histopathology results, and the image intensity ratio values (R610/R545 and R610/R575)

Patient no.	Age	Obstetric history	Menopausal status	Symptoms	Physical findings of examination	Histopathology results	R610/R545	R610/R575
1	50	P3L3	Pre	Difficulty in passing urine	Unhealthy cervix; cervical erosion observed	Chronic cervicitis with no dysplasia	1.461	2.433
2	64	P3L3	Post	History of foul-smelling discharge	UV prolapse	Normal	1.329	1.412
3	46	P2L2	Pre	Abdominal pain	Healthy cervix	Normal	1.307	1.408
4	53	P3L3	Pre	Abdominal pain; heavy menstrual bleeding	Endocervical polyp; bleeds on touch	Chronic cervicitis		1.73
5	45	P3L3	Pre	Abdominal distension	Healthy cervix	Normal	1.219	1.41
6	41	P2L2	Pre	Irregular spotting; breast pain during menstrual cycles	Polyp from the cervical canal; bleeds on touch	Chronic cervicitis with no dysplasia	1.263	1.708
7	43	P2L2	Pre	Abdominal pain; amenorrhea	Healthy cervix	Normal	1.112	1.583
8	49	Nulliparous	Pre	Abdominal pain; heavy menstrual bleeding	Erosions present; bleeding from the cervical os	Normal	1.401	1.463
9	47	P1L1	Pre	Heavy menstrual bleeding	Healthy cervix; copious amount of bleeding present	Chronic cervicitis with no dysplasia	1.28	1.845
10	49	P2L2	Pre	Lower abdominal pain; menstrual cramps	Indurated cervix; endometrial polyp present	Chronic cervicitis		2.227
11	47	P2L2	Pre	Burning micturition; continuous spotting; mass in descending pelvis	Healthy cervix; hypertrophied anterior lip with grade 1 descent	Chronic papillary endocervicitis		1.954
12	43	P2L2	Pre	Heavy menstrual bleeding; irregular cycles; mass per vagina	Healthy and elongated cervix; second-degree UV prolapse	Chronic papillary cervicitis		1.677
13	34	P2L2	Pre	On and off abdominal pain	Unhealthy, hypertrophied cervix with erosions; small cervical polyp; mass per vagina; third-degree UV prolapse	Normal	1.26	1.02
14	48	P2L2	Post	Abdominal pain;	WDPV; cervical erosions present;	Normal	1.328	1.507
15	29	P4 L3	Pre	itching in the perineal region; mass per vagina; second-degree of UV prolapses	elongated and hypertrophied cervix	Normal	1.226	1.348
16	53	P2L2	Pre	Itching in the vulval region; lower abdominal pain; dysuria	Cervix pulled up	Inflammatory infiltrate with no dysplasia	1.12	1.732
17	58	P3L3	Pre	Heavy menstrual bleeding		Low-grade squamous intraepithelial CIN-I	1.688	
18	45	P2L2	Pre	Heavy menstrual bleeding; frequent cycles	Healthy cervix; adenofibroma on posterior wall	Normal	1.212	1.188
19	37	P2L2	Pre	Heavy menstrual bleeding; irregular cycles	Bulky cervix	Chronic cervicitis with squamous metaplasia	1.414	1.812
20	47	P2L2	Pre	Heavy menstrual bleeding	Healthy cervix	Normal	1.49	1.32

(Cont'd...)

Table A1. ***(Continued)***

Patient no.	Age	Obstetric history	Menopausal status	Symptoms	Physical findings of examination	Histopathology results	R610/ R545	R610/ R575
21	41	P2L2	Pre	Heavy menstrual bleeding	Bulky cervix; shaped slit; mucoid discharge	Chronic cervicitis with squamous metaplasia	1.381	1.584
22	52	P2L2	Pre	Lower abdominal pain; spotting; back pain; bleeding from the cervical os	Healthy cervix	Chronic inflammation and cervix-free carcinoma	2.11	1.885
23	56	P2L2	Post	Post-menopausal bleeding; abdominal pain	Mass lesion in the lower vaginal cavity; stenosis of the vaginal cavity	Moderately differentiated SCC Grade 2	2.078	
24	50	P1L1	Pre	Foul-smelling blood stain; abdominal pain; general weakness; WDPV; loss of appetite	6 – 7 cm growth on the anterior side of the cervix; growth observed from the cervix into the vagina	Carcinosarcoma	3.005	
25	42	P4L2CO2	Pre	Lower abdominal pain; foul-smelling watery discharge	Hard growth on the lower lip of the cervix; cervical os is not visible; bleeds on touch	Moderately differentiated invasive SCC Grade 2	1.769	
26	76	P5L5	Post	Post-menopausal bleeding	Cervical wounds	Normal	1.499	1.417
27	50	P3L3	Post	Lower abdominal pain	Ovarian cyst	Chronic cervicitis		1.464
28	46	P1L1	Pre	Asymptomatic, abdominal tightness	Healthy cervix	Normal	1.196	1.211
29	75	P4L2CD2Ab4	Post	WDPV; foul-smelling discharge	Healthy cervix	Normal	1.41	1.441
30	49	P3L3	Pre	Heavy menstrual bleeding	Healthy cervix with grade 1 descent	Normal	1.448	1.209
31	42	P2L2	Pre	Abdominal pain	Mass in descending pelvis	Chronic cervicitis with no dysplasia	1.396	1.186
32	48	P2L2	Pre	Burning micturition; itching around the labia	WDPV; cervical anterior lip erosions present	Normal	1.464	1.143
33	49	P7L7Ab1	Post	Abdominal pain	Minimal erosion	Chronic cervicitis with no malignancy		1.277
34	71	P8L6CD2	Post	Abdominal pain	Healthy cervix	Normal	1.344	1.259
35	45	P2L2	Pre	Heavy menstrual bleeding	Healthy cervix; bleeding from the cervical os	Leiomyoma and chronic cervicitis		1.166
36	39	P1L1Ab2	Pre	Tiredness; minimal WDPV	Healthy cervix	High-grade squamous intraepithelial lesion CIN grade 2 – 3	1.495	
37	53	P2L2	Post	Mass per vagina	Healthy cervix; second-degree prolapses	Chronic cervicitis with no dysplasia	1.327	1.283
38	32	P2L2 LSCS	Pre	Heavy menstrual bleeding		Chronic cervicitis		1.31
39	51	P2L2	Post	Lower abdominal pain; post-menopausal bleeding	Healthy cervix	Normal	1.412	0.965
40	58	P3L3	Post	Post-menopausal bleeding	Unhealthy cervix; bleeds on touch	Keratinizing moderately differentiated SCC	1.481	
41	74	P5L4D1	Post	Abdominal pain; back pain; WDPV	2 cm hard mass lesion; bleeds on touch; clinical stage 3B	Moderately differentiated SCC	1.585	
42	55	P3L3	Post	Post-menopausal bleeding; back pain	Present of erosions on the posterior cervix	Normal	1.39	0.499

Abbreviations: SCC: Squamous cell carcinoma; UV: Uterovaginal; WDPV: White discharge per vagina.

ACCSCIENCE PUBLISHING

Artificial Intelligence in Health

ORIGINAL RESEARCH ARTICLE

Explainable solutions from artificial intelligence for health-care support systems

Valeriya Gribova[1,2] and **Elena Shalfeeva**[1,2*]

[1]Laboratory of Intelligent Systems, Institute of Automation and Control Processes, Far Eastern Branch of RAS, Vladivostok, Primorsky Krai, Russia

[2]Department of Software Engineering and Artificial Intelligence, Far Eastern Federal University, Vladivostok, Primorsky Krai, Russia

Abstract

For decades, efforts to standardize medical care have struggled to fundamentally reduce errors and unjustified variations in medical practice, largely due to the influence of the human factor. The formalization of clinical guidelines and computer-assisted interpretation makes it possible to provide decision-support tools to improve health-care quality. They can better influence clinician behavior than narrative guidelines. Medical ontologies and algorithms based on such ontologies allow the interpretation of formalized clinical documents (guidelines). To support health professionals as consultants, systems must provide reliable knowledge and rely on approaches explicitly explaining their recommendations. Integrating software engineering, knowledge engineering, and artificial intelligence advancements can provide health-care professionals with computer-interpretable clinical guidelines. These should be decision-support complexes combined under a common terminological framework capable of understanding patient health documents. The research focuses on an emerging concept of manufacturing systems working with digital clinical guidelines. The paper presents an architectural principle, a new technology for creating viable clinical decision support systems. It presents a development environment for constructing and controlling the system's improvements. The main contributions of the study include the automation of multiple physician tasks by filling a single structured "medical history," integration of formalized knowledge from clinical guidelines and other reliable sources to satisfy both the relevance of the methods used and personalization to patient, transparency of all applicable knowledge, explainability of advice based on the essence of the knowledge and linked to the source, and the integrability of decision-support complexes with neural network services, capable of inputting data from a structured medical history.

Keywords: Explanatory decision support system; Knowledge-enabled system; Interpretable clinical guideline; Knowledge ontology; Viable system

***Corresponding author:**
Elena Shalfeeva
(shalf@iacp.dvo.ru)

Citation: Gribova V, Shalfeeva E. Explainable solutions from artificial intelligence for health-care support systems. *Artif Intell Health*. 2025;2(3):138-153. doi: 10.36922/aih.5736

Received: October 31, 2024

Revised: March 1, 2025

Accepted: April 25, 2025

Published online: May 8, 2025

Copyright: © 2025 Author(s). This is an Open-Access article distributed under the terms of the Creative Commons Attribution License, permitting distribution, and reproduction in any medium, provided the original work is properly cited.

Publisher's Note: AccScience Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

One way to improve health-care quality, reduce unwarranted variation in practice, and lower health-care costs is through e-consultants. These tools are intelligent services and computer-interpreted clinical recommendations integrated into the workflow of medical care.[1,2] To support health professionals as consultants, such tools (e-consultants) must be based on solid, substantial knowledge. Machine learning algorithms, including deep learning techniques, provide consultations by being trained on massive datasets. However, the specificity of medical information means no datasets are adequate for solving the problem. An exception exists in simplified prediction or risk assessment tasks, where datasets consist of thousands of uniform tuples of predictors. For pre-diagnostic tasks, a large number of symptom checkers are provided.

This study showed that for 10 randomly selected clinical cases with a reliable diagnosis (five from PubMed publications and five from the city hospital) and the five most promising symptom checkers, the correct diagnosis was in the top five in 40% of cases, and became the most likely in only two cases. This confirmed that artificial intelligence (AI) services trained on data from one institution work well on simple cases or data from the same place. Fortunately, methods based on explicitly presented and manageable information are being developed.

In medicine, collections of text-based clinical guidelines for physicians, doctors, and clinicians are constantly being expanded and modified. Formalized clinical guidelines, which a program can interpret, make it possible to provide decision support tools with better chances of influencing clinician behavior than narrative guidelines. The methodology of creating interpretable clinical guidelines is rightly associated with knowledge-based systems,[3] a software system with a particularly important information component, the knowledge base. This additional architectural component is created with the participation of specialists and experts. The methods of automated recognition of such medical texts are still being developed and improved. Experts input these texts into electronic databases in a conscious and meaningful way. However, this process requires appropriate formalization tools.

The integration of explicitly written knowledge with machine learning allows the vast amount of human knowledge and the capabilities of machine learning to be used to achieve previously unattainable performance. Integration can increase the reliability and robustness of machine learning, facilitate interaction between humans and machine learning systems, and make system decisions understandable to humans.[4] Knowledge-enabled systems (KES) are applicable to mediate between machine learning algorithms and human users.[5]

To support health professionals as consultants, not only is high-quality knowledge important in such systems, but these systems must also rely on approaches that explicitly explain their recommendations. However, machine learning algorithms have not been able to generate an explanation of their decisions.

Time-consuming and complex systems are designed for long-term operation and use. However, the terms of use change over time, as do the user's requirements for functionality, interface, and even expertise. In medicine, clinical guidelines, which reflect clinical knowledge and practice, are subject to change. We refer to the software system for physicians that can update its knowledge as medicine evolves as a "viable" system.

Integrating advances in AI, software engineering, and knowledge engineering can offer clinicians comprehensive, medically intelligent systems that can provide credible answers, understand patients' histories and medical records, and evolve with the knowledge of specialists.

This paper aims to present an architectural principle and a new technology for creating explanatory clinical decision support systems (ExClDSS) and a tooling environment to ensure their evolution.

The ontology-oriented production method is proposed, the contribution of which can be expressed in five aspects:

(i) Independent (but coordinated) creation and support of each type of architectural component: knowledge bases, clinical dictionaries, software reasoners with explanations (ontological interpreters), and user interface. A single ontology ensures their consistency
(ii) Cognitive scientists are engaged in the ontology of knowledge, laying the foundation for forming various types of cause-and-effect, spatial, and temporal relationships of medical concepts. Therefore, the knowledge base is close to the real knowledge that doctors are taught at universities
(iii) Explanation is formed based on the applied knowledge base (in the format of the "ontology of explanation" approved by doctors)
(iv) Evaluation of the created knowledge base with various methods and tools. We evaluate the formal completeness and integrity of the knowledge base and conduct a monitoring assessment of the current state and correctness of the replenished collection of cases (precedents), ensuring a monotonous improvement of the knowledge base
(v) Software reasoners (together with a graphical user interface [GUI]) do not depend on knowledge and reference books. Therefore, they are repeatedly used to lay out many systems with knowledge bases.

The research's novelty is as follows. Unlike most existing methods of knowledge-based product constructing – that assume either the presence of a sufficient set of data to extract the necessary knowledge from them or the presence of an expert who can formulate an adequate set of knowledge – our research attempts to address the problem across the compatibility and complementarity of these paths, rather than interchangeability.

Thus, the research scope is the manufacturing technology of continuously developing trusted systems to support difficult clinical decisions.

2. Materials and methods

2.1. The concepts of explanation in a decision support system

Medical knowledge results from a person's theoretical and practical activities, reflecting the accumulation of previous experience. Knowledge is the regularities of the domain (relationships and rules) necessary to solve problems based on new data (facts). Clinicians expect recommendations or hypotheses from e-consultants with convincing explanations and systems with solid, reliable knowledge. The production of detailed explanations is an important element of decision support systems in general and computer-interpreted clinical guidelines in particular.

The generation of interpretable medical knowledge requires additional specialized mechanisms. The universal representation of knowledge in the form of rules processed by a single "inference engine" has a limitation in medicine: since the number of rules is measured in thousands, it becomes virtually impossible for experts to review, verify, refine, and correct them.

Semantic models of the medical knowledge domain are required to formulate clinical recommendations as components for ClDSS. These models must reflect the logic that physicians use when appealing to their sentences and fragments. The description of a set of concepts, relationships, and constraints used by specialists when solving problems and transferring knowledge between specialists is what we refer to as a knowledge ontology.

The hierarchies of classes of entities and their binary relations are already a step toward the declarative representation of a part of knowledge, but they cannot cover the most important clinical connections (and we do not consider them to be a full ontological language). Knowledge ontology is a part of the semantic description of the medical domain, and the other part is the semantic representation of data and documents. In domain ontology, all the concepts of specialists, relations of concepts, and restrictions on interpreting their meanings are defined. Together with them, the required types of statements are specified about factors such as necessary conditions, grouping, and cause-effect relationships. This ontological approach makes it possible to develop systems that offer detailed explanations of their recommendations for knowledge bases. Various intelligent components of the system work with a single semantic description of the source data, reading the elements they require.

Each intelligent component (algorithm) usually interprets its specific blocks of knowledge in clinical guidelines: diagnostic rules or treatment recommendations. The engineering of such interpretable knowledge as independent architectural components (knowledge about diagnosis, risks, and treatment) is done based on their ontologies.[3] Existing methods and technologies[6,7] make creating, testing, and deploying knowledge bases and knowledge-based systems, or KES, possible.

The most detailed part of the explanation is formed based on ontological knowledge (selected fragments in the analysis) and, in some cases, partially copies its cause-effect and structural connections. Forming the explanation of the results in terms that are understandable to the doctor and that correspond to his logic allows the application of a full-fledged medical ontology.

The explicit representation of all concepts and their relations makes it possible to input, shape, store data (and knowledge), and demonstrate results that can be generated automatically. Such ontological algorithms could be used in conjunction with machine learning approaches, either as a source of ground truth or as a thematic layer that could be used to promote interaction or improve explainability. When other AI modules are connected to generate advice, recommendations, or solutions, they should work with a single semantic description of the source data to ensure compatibility in explaining all results.

Using medical ontology allows for creating a GUI in terms that are understandable to the doctor and also, as a rule, forms a dialog script that corresponds to the logic of explaining the results. To support the doctor at all stages, it is important to combine AI modules as components of medical knowledge, obtained in different ways, into a united ExClDSS, all based on a single terminological framework.

Existing explainable AI methodologies use large language models, requiring large capacities and rechecking to mitigate potential hallucinations.[8,9] The path used in this study is an improvement (based on ontology) of classical explanation generation: during the decision process, the reasoner records its arguments in the knowledge base.

2.2. Decision support system maintenance and viability concepts

The developer of a clinical system has the task of designing and implementing the mechanisms that will provide further maintenance to meet the changing knowledge and service conditions in the created software product. The same mechanisms help solve problems in implementing "continuous delivery," a process in which software is always kept relevant.

The application of typical architectural solutions, declarative representation of components, and separation of competencies between developers of components of different types are all used to create maintainable decision support systems. Information technology managers struggle to scale AI projects because they lack the tools to create and manage a "production-grade AI pipeline."[10]

With the advent of complex software systems, the problem of their long-term maintenance has become more and more critical. Maintenance is the possibility of adaptation (to hardware and system software, to new types of human-machine interfaces and users) and extensibility (at the request of users). Modification of software systems is due to changes in operating conditions, user requirements, and subject area. In the operation of applied systems to support professional activity, there is a need to add new user-defined functions (and adapt to new devices or user interface changes). The average maintenance cost of the software system life cycle is about 50%,[11] but according to some reports, it can reach 80 – 90%.[12]

In addition to maintenance, viability has become a modern, useful property of software systems. It is defined as sustainability in a changing environment (maintaining usefulness and operability),[13] and the ability to evolve, as the ability to adapt with the least possible cost to requirements' variability, maintaining architectural integrity.[14] We will specify "the viability" as software system resilience to some functioning environment changes (the maintenance of working capacity) and the ability to develop over the "life" (evolvability).

In the case of applied decision support systems in intelligent tasks such as diagnosis, planning, and forecasting, the situation is different. Here, knowledge variability and the emergence of new solutions, such as creating new diagnostic methods and identifying new influencing factors, are expected, rather than just the extension of user functions. Therefore, the approach to maintaining ClDSS should not be similar to maintaining application software systems.

Many well-known tasks in diagnosis, treatment, and prognosis in general and medicine, in particular, are quite stable. Algorithms for solving them are described and can be qualitatively programmed once for long-term use. However, this is not the case for medical knowledge and clinical guidelines. They cannot be "sewn" into programs because they are regularly updated in this dynamic field.

Clinical decision support systems should have one part (knowledge) that is constantly evolving and another part that can read and understand it, i.e., be an interpreter. Medical knowledge, such as clinical guidelines, is an evolving part of ExClDSS. ExClDSS are expected to remain useful and effective in an environment of changing knowledge. Under conditions of variability in clinical knowledge, the viability of the medical system is manifested in its ability to adapt and update in response to new information and evolving practices.

In medical knowledge, the influence of factors and events on the patient's state, their change over time, individual characteristics, and some of their processes on others is important. The development (evolution) of such complex knowledge bases is the main "challenge" of modern "conditions" with (Ex)ClDSSs. "The ability to adapt under a change in the set of facts and knowledge" is one of the aspects of intelligence.[15(p.5)] For medicine related to solving intellectual problems, this implies the evolution of knowledge. The ontological approach to knowledge and programming for working with them was sufficient.

3. Clinical decision support systems as software systems that apply understandable knowledge

To ensure that doctors trust ClDSSs, their developers need to demonstrate correctness (sometimes accuracy) on subsets of precedents (cases from practice), implement the ability to explain the proposed solution or hypothesis (the explanation must be understandable, consistent with formalized knowledge), have a mechanism for permanent improvement of the knowledge base that does not worsen its correctness, apply procedures for regular evaluation of stored clinical knowledge, and provide the opportunity for specialists to read and evaluate the included knowledge.

Knowledge must be formed considering standardized clinical guidelines and under domain experts' control. One method is to use trained text recognizers. Knowledge can sometimes be created by experts themselves (possibly with knowledge engineers and cognitive scientists). In this case, experts fully participate in the development and maintenance process with programmers and designers. This requires knowledge bases to be presented in a form understandable to medical domain experts. When knowledge is isolated and framed in independent architectural components and knowledge bases, the system using them becomes a KES.

Several AI, mathematical modeling, and machine learning methods for solving practical problems provide medical services based on hidden knowledge.[16] In medicine, these are most often the tasks of risk assessment (for example, risk of hypokalemia in patients with arterial hypertension) and predicting complications.[17,18] Almost all services based on machine learning provide versions of a preliminary diagnosis without considering the dynamics of the patient's observations.[19,20] For such tasks (preliminary diagnostics, risk assessment, or forecasting), as a rule, an intelligent service becomes inaccurate if it was "trained" on the data of one institution and it tries to operate in other circumstances.

There are tasks for which no one has yet accumulated adequate training material. In medicine, these are corrections to disease treatment and differential diagnosis. For this, intelligent services consultants trained on text corpora (and GPT helpers) within the idea of hybrid services to support the doctor's work may be used.

3.1. The influence of the ontological model on the properties of system components

For ClDSS to correctly formulate advice or results for solving medical problems (risk, diagnostics, treatment, and prognosis), it needs to operate with concepts that specialists use. For example, for knowledge of the task of monitoring the recovery process, one of the most common types of sentences (statements) is:

<process type$_k$, set-of {(period$_{ik}$ + interval$_{ki}$), set-of {characteristic$_j$, characteristic values range$_{ijk}$}}>.

For the diagnosis of diseases, statements about the relationships are needed:

<diagnosis$_k$, process' existence necessary condition$_k$, set-of {factor$_{km}$}>;

<diagnosis$_j$, set-of {symptom complex$_{jk}$ | variant$_{jk}$ of disease course}, [necessary condition$_j$]>;

<symptom complex$_k$, set-of {feature$_j$, range$_{kj}$ of values of feature}>;

<symptom complex$_k$, set-of {sign$_{jk}$, {period$_{ik}$, duration of period$_i$, range$_{ijk}$ of values of sign$_j$ in period$_i$}}>;

<variant$_n$ of disease course, set-of {(period$_{in}$ + interval$_{in}$), set-of {observation$_k$, set-of {observation element$_{jk}$, range of values$_{ijkn}$>}}>;

<necessary condition$_k$, set-of {factor$_{km}$}>.

All such concepts and relationships are explicitly written and "available" to algorithms when they are developed based on ontologies and access to explicit knowledge resources. The medical ontology was created by experts and knowledge engineers. The factors without which a disease does not begin can be the events or properties of an organism; they can also determine one of the options for the development of the disease. For the treatment of diseases, the statements about the method to eliminate the cause of the disease:

<diagnosis$_k$, event$_u$, set-of {(observation$_{jk}$, new-range$_{iu}$ of observation values)}, delay$_{ku}$>;

The statements about the impact on an organism for recovery start:

<process variant$_{kn}$, (period$_{in}$, interval$_{in}$), treatment event$_u$, set-of {observation$_{jkn}$, value range$_{jkn}$ (period$_{(i+1)n}$, interval$_{(i+1)n}$)>}}>;

The statements about acting on a symptom to alleviate it:

<process variant$_{kn}$, (period$_{in}$, interval$_{in}$), observation$_{jk}$, values$_{jkn}$ range, (event$_u$, delay$_{kj}$), values new range$_{iu}$>.

The domain ontology represents all types of statements as a structural language (template) for introducing or describing knowledge. The knowledge base explicitly contains sets of statements of the corresponding type sufficient for this profession.

The traditional architecture of KES is the knowledge base + fact base + intellectual problem solver + intellectual GUI.[15Finn2004] Knowledge bases are generated manually or inductively, including training samples from archives and databases; this process involves inductive generalizations in machine learning.[16] Bayesian classifiers, clustering algorithms, and reinforcement learning[21,22] are sometimes involved in this process.

The approach to creating systems with transparent knowledge bases is based on an architecture expanded by a new component: ontology + ontological knowledge base + ontological fact base + ontological interpretator (problem solver) + intelligent graphical interface. It can include databases with reference information, operational information, and work with files.

Often in one domain, a set of interrelated tasks is solved; examples of related tasks are diagnosis, treatment, and prognosis. To solve all tasks, one formal domain ontology can be created, but it is more convenient to map a separate ontological resource to separate tasks solved in the domain. A set of formal ontologies for related tasks may be required when designing applied systems with knowledge bases, which is followed by the creation of a set of ontological knowledge bases (for each task).

3.2. The influence of ontology on the integrability of various components within a single architecture

To combine various achievements of knowledge engineering and AI in complex medical intelligent trustworthy systems, it is reasonable to choose the document "patient's medical record" as an integration point. The document “medical history of any patient” contains a structured set of facts observed or objectively measured in the considered situation (medical case) regarding which the problem is solved.

All results and decisions are recorded in the same document (medical history), regardless of the method in which they were collected. The place in the document structure must be strictly connected with the essence of the result (diagnosis in one place and prognosis in another). Such a document structure is part of the domain ontology. Thus, it can be asserted that ontology ensures the integrability of various components.

Often, it is necessary to add a pre-existing software service with hidden knowledge (trained model) to the system to solve a specific problem. Typically, this task falls into one of three categories: risk assessment, prediction, or recognition of a class of pathologies. A structured description of the service is sought, which includes the following elements: (i) name and author of the method, (ii) essence of the result, (iii) vector of initial data, (iv) conditions of applicability (entering values in limited ranges), (v) manner of launching the service, and (vi) if the expected response of the service is numeric, then the description should also include the interpretation of the result.

For the mutual exchange of data and results with software services (with hidden knowledge), a single semantic template is used: <name of method, author, essence of the result, vector of initial data, [conditions of applicability], description of result interpretation, launch method or full address of microservice>. For example, for a software service for assessing the risk of developing a disease, the description of the interpretation of the result = a set of pairs <threat level value, range of calculated values>. Adding such a semantic template (with a description of the interpretation of the result) to the medicine ontology ensures the explainability of connected, intelligent services. The vector of initial data (from the semantic template) should be formed only with the help of the terms of the “medicine” thesaurus. The “medicine” thesaurus (a dictionary of terms for observing a patient and studying the patient’s body) is traditionally considered part of the ontology of this domain area. Hence, the ontology is a structural basis of both tools for experts and users (editing tools) and for software components of KESs.

4. The viability model of clinical knowledge-enabled decision support systems

An ontology, as a structural basis for viewing and editing knowledge bases, provides a basis with a declarative property. However, for the knowledge base to be adaptable, the user interface of the knowledge editing tool must meet the requirements and expectations of domain experts.

The main challenge in medical systems manufacturing is ensuring that the knowledge reflects current knowledge (e.g., clinical guidelines) and continuous improvement. Continuous improvement of the knowledge base allows it to become a reliable source (repository) of expert knowledge, hoping to create a “reference” knowledge base. Its quality will determine the success of the use of this knowledge.

The relevance of a knowledge base is achieved through three main ways:[23-25] interactive change of the knowledge base, usage of machine learning methods (tools of inductively generating knowledge from selected precedents and tools of knowledge discovery from “big data”), or a combination thereof. The “success” of adaptability depends on several conditions and principles.

4.1. Architectural properties of clinical systems enabled with declarative knowledge

This intelligent software system class, which explains decisions, requires specialized development and maintenance tools. The key principle is the special role of the knowledge ontology (as a model of professional concept relations). Its formalized representation, separated from the professional knowledge itself, allows for the independent development of each ontological component, relying on its integrability. The medical ontology makes it possible to create a GUI that is understandable to the doctor and, as a rule, to create a dialogue script corresponding to the logic of explaining the results.

The interpretation of knowledge consists of choosing each hypothesis and transitioning from it (along the chain of connections) to the expected values areas of observations for subsequent comparison with facts, as well as constructing an explanation with the collected arguments. The “structural” complexity of the ontological interpretation algorithm is determined by the number and the length of the chains of cause-and-effect relationships in the statements, the degree of fuzziness prescribed in cause-and-effect relationships and statements, and the structure of observation, description, or conditions for a decision.

To develop interpreters (task solvers), coding tools for new software units or their new versions, tools for cataloging units for reuse, and tools for integrating reusable units and new ones into new solvers or their new versions are needed. Solvers built according to a given ontology (for problems of diagnosis, prognosis, etc.) must be reusable reasoning engines of medical services. A version of clinical knowledge is their input parameter. Therefore, regular updating of the knowledge base does not require changes in other components of the KES.

A connector module is used to connect an external microservice to the medical service, which solves an additional task for the same data (documents about the patient). It consists of standard tasks such as reading the list of names of the required data in the declaration (specification) of this external service, finding the values of this data in the input document, composing a "PUT" request with the specified uniform resource locator, and sending it. If the microservice is not interactive, it is necessary to wait for a response, select fragments (specified in the declaration) for explanation, and add them to the provided substructure of the final explanation.

4.2. Ontological approach to support and develop KESs

The dependency of all KES components on a domain ontology supports the viability properties, which include the replaceability of components for their improved versions, admissibility of the improvement of the decision method, permissibility of changing or adding functions (for example, the formation of additional results), adaptability of the user interface due to changes in the input data, and permissibility of expanding the ontology (adding concepts and relationships).

As a result, the structure of the KES and its components does not require changes due to current maintenance and sustainable development.[26] As mentioned above, this class of software systems (ClDSSs) requires specialized maintenance tools because clinical guidelines (and other medical knowledge) are constantly evolving. Due to the importance of evolving knowledge bases, only ClDSSs integrated with a knowledge base management system should be considered.

A toolkit for building application systems with declarative (interpretable) knowledge is based on a domain and problem ontology. If a separate formal ontology is created for each task (diagnosis, treatment, prognosis, etc.), it is easier to develop tools to ensure the quality of homogeneous, localized knowledge. The tools for developing and verifying knowledge bases are desirable to be integrated into the architecture of the decision support system (to be a part of the integrated architecture of the decision support system). Thus, a maintenance environment has to provide knowledge base editors, tools for assessing knowledge bases by archives of etalons (solved problems), tools for checking and evaluating the quality of knowledge bases, and tools for the inductive formation of knowledge base fragments.

If the pre-existing solver is in accordance with the problem statement and "building up" additional functionality is not supposed to be used, then coding tools for program units may be unnecessary.

4.3. Testing the quality of the system with a knowledge base

When testing the quality of ExClDSS work, "control sets of clinical cases" should be used. This is a carefully selected set of documented medical histories containing the correct solution and facts sufficient to develop the correct solution. Based on the "control set of clinical cases," a "control set of test cases" (CSTS) should be prepared by clearing out information that is not important for the target task. This, in particular, depends on the task being tested (treatment, diagnostics, prognosis, or prevention).

We believe it is important to use metrics to assess the quality of the ExClDSS components (such as sensitivity, specificity, positive predictive value, and precision) and metrics to evaluate the quality of ExClDSS performance in supporting the solution of specific clinical problems. These metrics are defined to the CSTS. For example, for the task of forming hypotheses about a diagnosis, we use: CorrectnessEstimation (number of tests-with-a-finding/card of CSTS) and AccuracyEstimation (number of tests-with-a-hit/card of CSTS), where a test-with-a-finding is a test clinical case of a certain disease, for which the ExClDSS generated many hypotheses during testing, among which was this disease. The test-with-a-hit is a test clinical case for which the ExClDSS generated its disease as the only hypothesis. The card of CSTS is the power of the set of all prepared test clinical cases.

These quantitative metrics are associated with clinician satisfaction. For almost every clinical task (except for differential diagnostics), the metrics for assessing the accuracy and correctness should be determined separately for each of the diseases, knowledge of which is "embedded" in the ExClDSS. Each of these tasks has its own requirement for the set of patient data in the test clinical cases used, and they are not the same for acute, slowly progressive, chronic, and hereditary diseases.

5. Manufacturing environment for viable clinical decision support systems

As noted above, this class of clinical software systems in the construction of <ontology + set of ontological knowledge bases, set of ontological interpreters, set of user interface components, sets of facts> requires specialized development and maintenance tools where coding tools for new software units or their new versions may turn out to be unnecessary. The authors are aware of ontological portals but not of development environments focused on quality assurance and the development of ontological components for such systems.

5.1. Implementing KESs in manufacturing environments

An example of an environment for producing clinical KESs is thematic medical ontological portals, and an example of a toolkit for modern KES production and maintenance environments is the Intelligent Adaptive Clinical Platform as a Service (IACPaaS) cloud platform. The creation of an ontological portal starts with the creation of an ontology as a semantically structured basis for the creation and processing of ontological information resources (Figure 1).

An ontology (as a template or "meta-information") defines a semantic model, structure, rules for the formation of information resources, limitations of its interpretation, or processing rules.[27] Usually, cognitive scientists (users of the IACPaaS platform) form such semantic structures and rules for the community of experts and specialists (users of the IACPaaS portals) (Figure 1). The tool for ontology formation is the IACPaaS meta-information-editor "ontology editor."

Creating an ontology is a creative process requiring extensive analytical work and a systematic domain analysis to identify common patterns in forming knowledge, structure, and integrity constraints. The ontology is generally not changeable throughout the life cycle of the KES. The separation of an ontology from a knowledge base (and a set of facts) leads to the ability to interpret them with a specialized ontology-based algorithm. The algorithm (ontological reasoner) searches for or refutes hypotheses by traversing the (declarative) knowledge base. It "sorts" the knowledge base statements of each type related to the hypothesis, comparing these sentences of input information (patient's document).

The medicine ontology was carried out by cognitive scientists (knowledge engineers) together with experts. It "covers" several classes of tasks (for example, diagnosis of diseases regardless of their etiology[28]). The knowledge base is an ever-changing component and should be formed based on the created ontology. KES is a special case of applied software services on the IACPaaS platform (IACPaaS services). They need the information resources (of the portal), such as knowledge bases. The set of IACPaaS-portal tools for the formation of KES's information components are the IACPaaS-editor of knowledge base, generated in terms of ontology with a self-adaptive user interface (Figure 2), and the IACPaaS data editor (with self-adaptive user interface) (Figure 3).

The regular generator of information editors makes knowledge base editors available and constantly operating in the production environment (Figure 2).

In addition, the IACPaaS toolkit to form software solvers (in addition to coding tools for software units) contains (Figure 4) (i) an IACPaaS meta-information editor to explain the resulting structure, (ii) a "master" of the formation of declarative parts of software units and their blocks-reasoners, which conduct reasoning on ontological information, (iii) a generator of code blanks (according to declarative parts) for new IACPaaS software units, (iv) a solver constructor from GUI (or "root"-unit or IACPaaS-agent), a software unit (being represented by its declarative parts), including connector modules, and (v) tools for testing IACPaaS agents and preparing them for reuse.

5.2. Creating clinical KESs using Intelligent Adaptive Clinical Platform as a Service environment tools

The KES design technology in the proposed environment provides for a sequence of activities.

(i) Find pre-existing knowledge ontologies whose concepts and relationships are sufficient for the tasks and data ontology. If the IACPaaS platform does not already have an ontology for the problem under

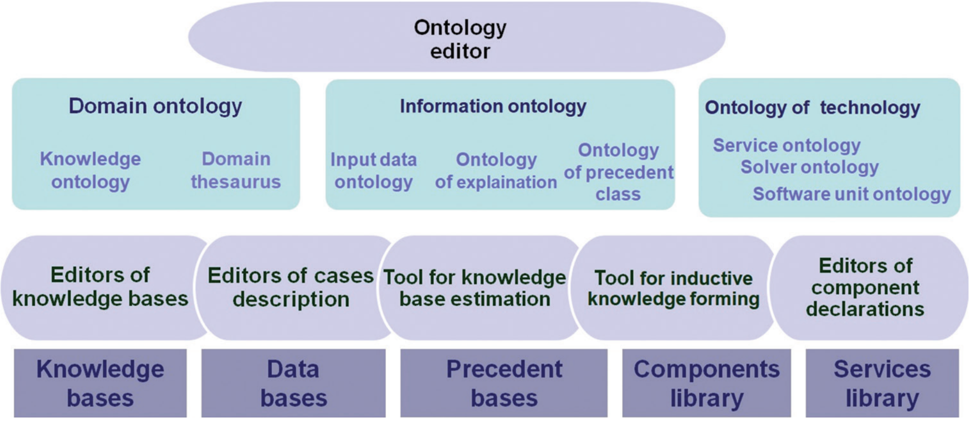
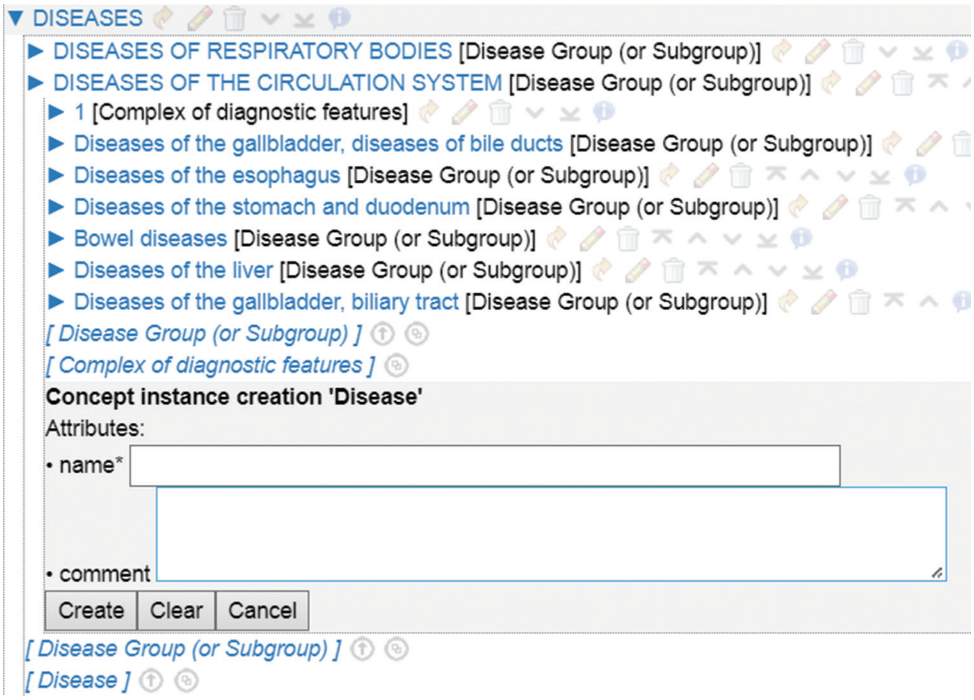


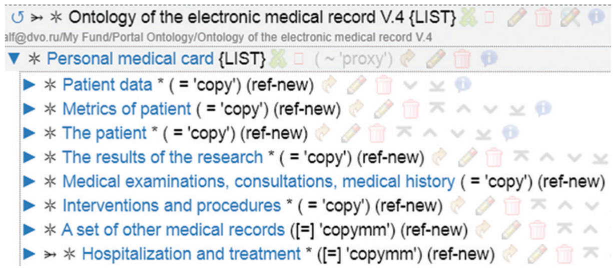
Figure 1. Basic components of the environment for developing basic components of a knowledge-enabled system



▼ DISEASES
▶ DISEASES OF RESPIRATORY BODIES [Disease Group (or Subgroup)]
▶ DISEASES OF THE CIRCULATION SYSTEM [Disease Group (or Subgroup)]
  ▶ 1 [Complex of diagnostic features]
  ▶ Diseases of the gallbladder, diseases of bile ducts [Disease Group (or Subgroup)]
  ▶ Diseases of the esophagus [Disease Group (or Subgroup)]
  ▶ Diseases of the stomach and duodenum [Disease Group (or Subgroup)]
  ▶ Bowel diseases [Disease Group (or Subgroup)]
  ▶ Diseases of the liver [Disease Group (or Subgroup)]
  ▶ Diseases of the gallbladder, biliary tract [Disease Group (or Subgroup)]
  [Disease Group (or Subgroup)]
  [Complex of diagnostic features]
  Concept instance creation 'Disease'
  Attributes:
  • name*
  • comment

[Disease Group (or Subgroup)]
[Disease]

Figure 2. Development process by the Intelligent Adaptive Clinical Platform as a Service editor of the knowledge base



↺ ⤚ ✻ Ontology of the electronic medical record V.4 {LIST}
alf@dvo.ru/My Fund/Portal Ontology/Ontology of the electronic medical record V.4
▼ ✻ Personal medical card {LIST} (~ 'proxy')
  ▶ ✻ Patient data * (= 'copy') (ref-new)
  ▶ ✻ Metrics of patient (= 'copy') (ref-new)
  ▶ ✻ The patient * (= 'copy') (ref-new)
  ▶ ✻ The results of the research * (= 'copy') (ref-new)
  ▶ ✻ Medical examinations, consultations, medical history (= 'copy') (ref-new)
  ▶ ✻ Interventions and procedures * (= 'copy') (ref-new)
  ▶ ✻ A set of other medical records ([=] 'copymm') (ref-new)
  ▶ ⤚ ✻ Hospitalization and treatment * ([=] 'copymm') (ref-new)

Figure 3. The example of the Intelligent Adaptive Clinical Platform as a Service medical document ontology

consideration, e.g., for risk assessment or prevention, the knowledge engineer uses the ontology editor to describe a set of relevant concepts, relationships, and constraints (creating a knowledge ontology for the problem or task). Often, the engineer will work with the expert to establish a set of decision rules (ontological agreements). A new task may require an extension of the thesaurus

(ii) Formation of the knowledge base by medical experts. The IACPaaS knowledge base editor (for example, based on the extended knowledge ontology) may be required. It is possible to formulate clinical guidelines without obstacles. Using the tools for translating text documents into a given structure is advisable. As a rule, creating the knowledge base is a collective process

(iii) If the creation of a solver is required, then the designer uses the ontology editor to create an ontology to explain the results. Using software engineering techniques, designers and programmers then declare and implement solver components (agents). Notably, creating a knowledge base and creating a problem solver (particularly assembling program blocks from the library) are parallel processes. The ontological structure ensures the compatibility of the components within the portals created.

5.3. Ensuring clinical KESs' viability

The properties of KESs related to viability are provided by the development environments implemented on the IACPaaS platform. Consider the example of the knowledge updatability property and support for knowledge updating. When an update to the knowledge base is required in connection with the acquisition of new knowledge (statements), it is possible to modify it manually. This development (production) environment has knowledge base editors.

Then, it is important to evaluate the consistency with the available facts. It must check the non-decrease of a set of correctly solved tasks when replacing the version of the knowledge base (according to the importance of monotonous improvement of knowledge bases[29]). The procedure for checking the non-deterioration of knowledge is as follows: from each reference task (precedent), enter input conditions into the solver integrated with the new version of the knowledge base, and obtain the explanation

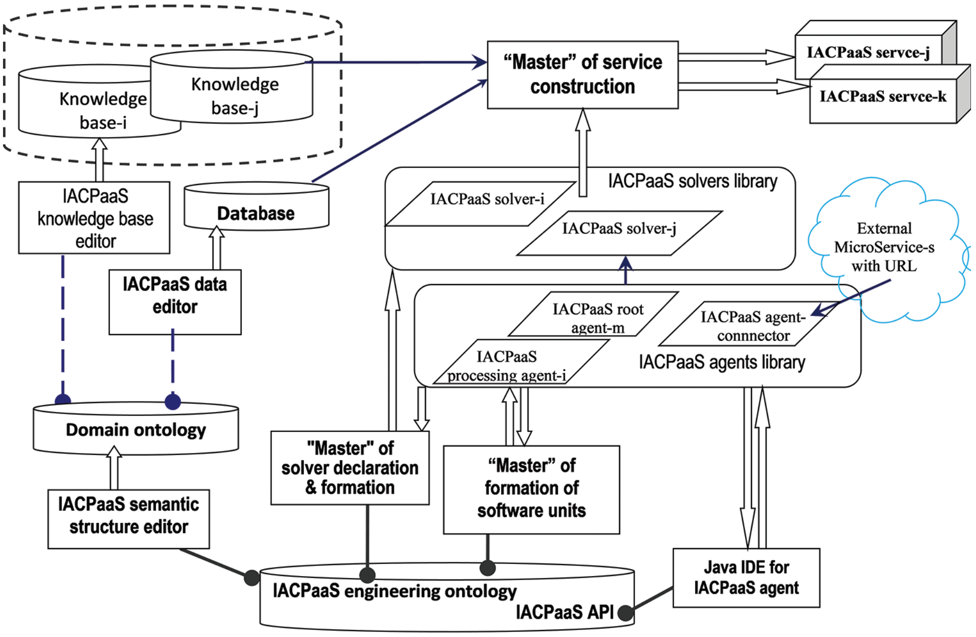


Figure 4. The model of the KESs manufacturing environment
Abbreviations: API: Application programming interface; IACPaaS: Intelligent Adaptive Clinical Platform as a Service; URL: Uniform resource locator.

to compare it with the result, fixed in this precedent. It is advisable to develop a tool for checking the quality of the knowledge base together with a solver since they have many common software blocks.

If it is necessary to update the knowledge base in connection with obtaining precedents that do not correspond to the knowledge (if the precedent contains the correct result of solving the problem, which does not correspond to the result obtained with the help of KES), then it is effective to form a new version of the knowledge base automatically, based on the methods of inductive formation of fragments of the knowledge base.[29] An example of such an update of knowledge "from practice" occurs when, after a certain period, the correct result of diagnosis or treatment becomes available from a medical institution. This outcome is then compared with the result from KES: are they contradictory? In such cases, evaluating the consistency of the updated KES's work result with existing precedents is preferable.

To implement the process of monotonous improvement of knowledge bases, the following means are required: (i) inductive knowledge creation tools for each intellectual problem solved (diagnosis, planning, forecasting, etc.), (ii) tools to support the choice of precedents (correctly solved problems in a statement), and (iii) tools for verifying the correctness (quality) of the new version of the knowledge base (the same as described above). The "knowledge updatability" property depends on the availability and performance of the above tools; other tools are added to the development environment (as a framework*)* one after another.

We compared the process of building a complex of interconnected evolving knowledge bases to make medical systems using the Protégé[6] and IACPaaS[30] tools. We created classes of diseases, symptoms, and drugs on the web, in Protégé. Next, we had to associate diseases with symptoms using the object property mechanism (some acute diseases required a dynamic description of the clinical picture). However, these mechanisms did not describe knowledge in a way that doctors would need and understand, and explaining the dynamics of disease development was particularly difficult. It should be noted that the Protege tools are incomprehensible and difficult for doctors; they are intended for knowledge engineers (although describing the dynamics proved difficult for them). Similar difficulties were encountered when trying to describe treatment protocols, taking into account the specifics of drug use and patient characteristics. Protégé's mechanisms did not allow the patient's history to be formed as a single document. In contrast, the advantages of IACPaaS in addressing this limitation have been demonstrated.

5.4. Implementing a clinical knowledge-based decision support system

The formation of a Medical Portal, Med-IACPaaS (https://iacpaas.dvo.ru/), began with the development of medical ontologies and editing tools. Previously formed by experts and knowledge engineers, the medical ontology has been formalized as a hierarchical semantic network. It includes a glossary of terms (more than 25,000) for describing the patient's anamnesis, current state, complaints, objective results, laboratory, and instrumental research. The glossary contains commonly used and specific terms, such as the basis of symptoms of cardiological pathologies and neurological terminology. Additionally, the ontology also includes an ontology of medical diagnostics (about 70 concepts and 100 relations between concepts) as one of the knowledge ontologies, the ontology of knowledge about the nomenclature and effects of medicines on the human body with various impaired functioning (about 80 concepts), the ontology of treatment regimens of diseases (about 80 concepts), and a knowledge base (for several nosology groups) formed based on the ontology, the ontology of medical case records history (more than 120 concepts) nodes of the intended for describing information about characteristics of an organism, facts, events, and observations in patients, and the formats of structured reports with analysis and explaining hypotheses on decision on the base of knowledge.

In terms of such networks, experts then began to create knowledge bases without the participation of knowledge engineers. To date, experts and doctors have created and maintained clinical information guides as knowledge bases on a wide range of nosologies (Figure 5). The diagnosis knowledge base, formed in terms of the ontology of medical diagnostics, currently includes more than 250,000 concepts describing the diagnosis of 35 diseases from 10 groups. The knowledge base formed in terms of the treatment ontology currently contains more than 90,000 concepts.

The IACPaaS clinical services were created on various nosologies: viral, diseases of the oral cavity, salivary glands, and jaws, diseases of the gallbladder, biliary tract, and pancreas, bowel diseases, hemorrhagic fevers, coronary

- ▼ ✻ DISEASES OF THE RESPIRATORY SYSTEM [A group of diseases (or Subgroup)]
  - ▼ ➳ ✻ Viral [A group of diseases (or Subgroup)]
    - ► ✻ MERS-CoV * [Disease]
    - ► ✻ Parainfluenza [Disease]
    - ▼ ✻ Flu [Disease]
      - ▼ ✻ The version of the manifestations of the "National guide 2013" [Complex of symptom
        - ▼ Complex complaints and physical examination
          - ► ✻ ↕ Body temperature [Sign]
          - ► ✻ ↕ Chills [Sign]
          - ▼ ✻ Sore throat [Sign]
            - ▼ A simple sign
              - ▼ 1 [Period dynamics]
                - ◊ 0.5 [the lower bound of the duration (sort: real)]
                - ◊ 2.0 [the upper bound of the duration (sort: real)]
                - ► 1 [Option values]
                - ◊ ✻ day [unit (sort: string)]
                - ▼ modality
                  - ◊ specificity (type: string)
              - ► 2 [Period dynamics]
          - ► ✻ Malaise [Sign]
          - ► ↕ Dizziness [Sign]
          - ► ✻ Arthralgia [Sign]
          - ► ✻ Myalgia [Sign]
          - ► ✻ General weakness [Sign]
          - ► ✻ ↕ Vomiting [Sign]
          - ► ✻ ↕ Headache [Sign]
          - ► ✻ ↕ Cough [Sign]
          - ► ✻ Cyanosis of the lips [Sign]
      - ► The causes of the disease
      - ► <no translation for 'https://klinica.com.ua/diagnoz'> [Dinamical Complex of Symptoms]
      - ► Details of the diagnosis

Figure 5. The fragment of the description of some diseases in the diagnosis knowledge base of the Medicine Intelligent Adaptive Clinical Platform

heart diseases, diseases characterized by increased blood pressure, and chronic rheumatic heart disease. Usually, experts form a clinical picture of diseases with dozens of dynamic symptoms. The knowledge about the diagnostic signs of biliary tract disease was built inductively, based on the data set of their surgical department. In addition, from the knowledge built by the linguistic processor Ontosminer (http://ontosminer.opkrt.ru/) based on the analysis of millions of documents from the free resource PubMed, some clinical guidelines were selected for which it was possible to carry out validation based on real case histories. Specialists in mucopolysaccharidoses manually created them, tested them on real examples of patients from different countries, and refined the knowledge.

Developing knowledge bases and intelligent software components for medical services was carried out in parallel. Such services are intended to help medical teams and institutions support the solution of the problems of their intellectual activity, providing a "third opinion" through cloud means. Their hypothesis explanations rely on formalized knowledge (Figure 6).

The labor costs are fully justified because each solver-interpreter works for more than one profile, and each nosological base is used to solve several intellectual problems (diagnosis, treatment, and prognosis). The "cloud" implementation of ClDSS allows the monitoring of relevance and evolving a single clinical guidelines base for several profiles and classes of tasks. The general base accumulates the experience of several professional communities and teams in addition to universal knowledge.

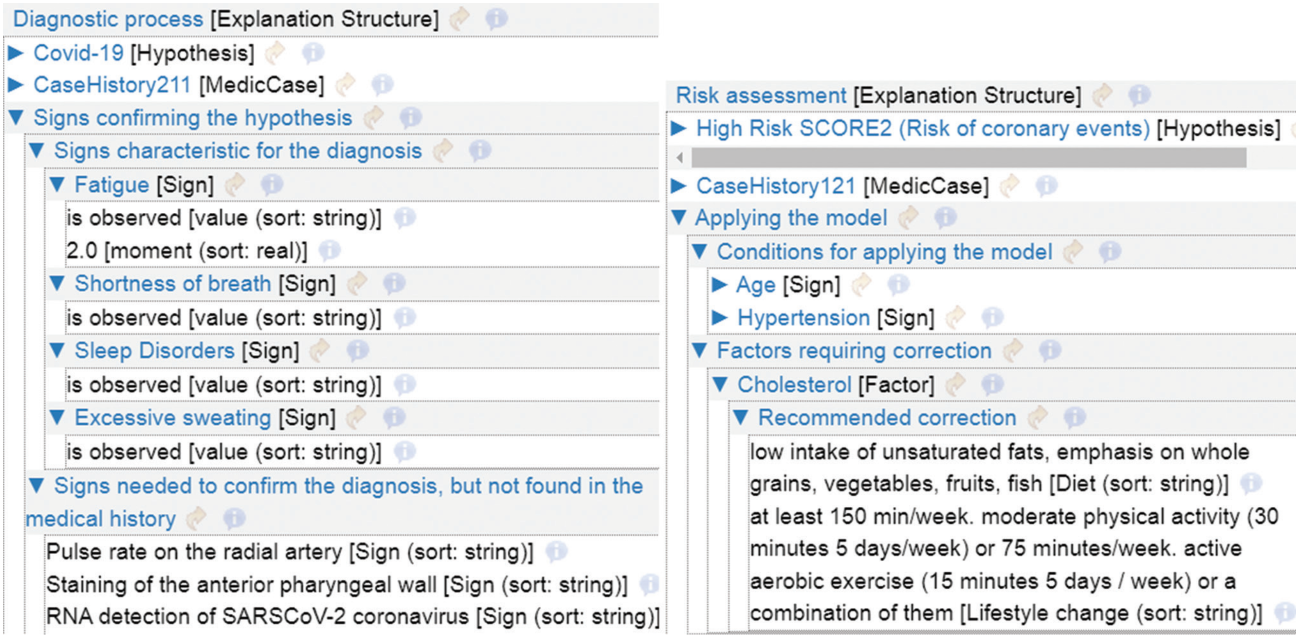
6. Comparing different approaches to the automation of medical activities

Today, specialists from different medical teams are using medical IACPaaS services to test the capabilities of AI in solving various problems for formalized health case histories. For general practitioners, differential diagnosis services are being created, and for gastroenterologists, a complex for diagnosis, treatment, and prognosis of recovery. For cardiologists, the complexes are made so that the risk assessment is carried out in two ways: based on ML and formalized knowledge. In comparison, advice on diagnostics is given in two other ways: based on knowledge and formalized precedents.

Medical software assistants (solving different tasks) are built using IACPaaS tools and components of the IACPaaS medical portal. All assembled systems work based on a single terminological base of symptoms and facts (more than 20 thousand) with synonyms.

The overall knowledge base is large and maintained by multiple experts. Updates are carried out according to procedures that keep the cloud services running. For example, knowledge on some digestive system diseases has been expanded, information on the regional manifestations of fever caused by rodents has been clarified, and new, unique knowledge on metabolic disorders of glycosaminoglycans, mucolipids, and gangliosides has been added.

Knowledge base editing and verification tools are used to update current diagnostic and therapeutic knowledge.



Diagnostic process [Explanation Structure]
▶ Covid-19 [Hypothesis]
▶ CaseHistory211 [MedicCase]
▼ Signs confirming the hypothesis
  ▼ Signs characteristic for the diagnosis
    ▼ Fatigue [Sign]
      is observed [value (sort: string)]
      2.0 [moment (sort: real)]
    ▼ Shortness of breath [Sign]
      is observed [value (sort: string)]
    ▼ Sleep Disorders [Sign]
      is observed [value (sort: string)]
    ▼ Excessive sweating [Sign]
      is observed [value (sort: string)]
  ▼ Signs needed to confirm the diagnosis, but not found in the medical history
    Pulse rate on the radial artery [Sign (sort: string)]
    Staining of the anterior pharyngeal wall [Sign (sort: string)]
    RNA detection of SARSCoV-2 coronavirus [Sign (sort: string)]

Risk assessment [Explanation Structure]
▶ High Risk SCORE2 (Risk of coronary events) [Hypothesis]
▶ CaseHistory121 [MedicCase]
▼ Applying the model
  ▼ Conditions for applying the model
    ▶ Age [Sign]
    ▶ Hypertension [Sign]
  ▼ Factors requiring correction
    ▼ Cholesterol [Factor]
      ▼ Recommended correction
        low intake of unsaturated fats, emphasis on whole grains, vegetables, fruits, fish [Diet (sort: string)]
        at least 150 min/week. moderate physical activity (30 minutes 5 days/week) or 75 minutes/week. active aerobic exercise (15 minutes 5 days / week) or a combination of them [Lifestyle change (sort: string)]

Figure 6. The fragment of hypotheses explanatory in diagnosing the Intelligent Adaptive Clinical Platform service

Inductive knowledge generation tools based on tabular data sets are used for two types of tasks. To test the knowledge, tools are used to compare solutions in the reference case history sets with the generated results of each AI component integrated into the medical IACPaaS services. Two comparative analyses were carried out: (i) opportunities for generating advice: IACPaaS-ExClDSS (including integrated ones) versus other available services, and (ii) labor costs (as the number of employees) to maintain the relevance of the knowledge base. In the cloud implementation of knowledge-enabled services, there is a division of competencies versus traditional KESs development.

The comparative analysis involved five medical internet services (https://symptoms.webmd.com, https://www.everydayhealth.com/symptom-checker, https://www.mayoclinic.org/symptom-checker, https://kiberis.ru, and https://symptomchecker.isabelhealthcare.com) and five of the most recent IACPaaS “assemblies.”

Five clinical cases were randomly selected from the accumulated set of clinical cases from practice, and five clinical cases with the same diagnoses were taken from publications on PubMed. As for internet services, for 10 reference cases, the true diagnosis was rarely in the top three (Table 1): 2 – 4 times (out of 10 stars), and more frequently in the top 10: 6 – 8 times. In contrast, with IACPaaS, the true diagnosis appeared in the top three 7 – 8 times, and always within the top 10.

The results of treatment and prognosis are difficult to compare, because only one internet service was ready to issue a cure, and two of our selected IACPaaS assemblies had a treatment module (prescription was issued for eight + six cases, 6 times it coincided with prescription from the reference case).

The prognosis is always implemented separately from other tasks. It is aimed at a specific disease, while IACPaaS services can predict disease course at the initial stages of diagnosis and predict recovery course (with the prescribed treatment).

Before creating a cloud platform and developing an ontological approach, we developed several diagnostic consultants with a knowledge base based on the rules. The development of the first version of the diagnostic service for a group of diseases (therapy, ophthalmology, and gastroenterology), including about 15 diseases, took an average of 20 months: for testing with debugging (8 months), to expand with another disease (a month), and subsequent testing with debugging (2 months). With cloud implementation of KES (with separation of competencies) for a similar group of diseases, it takes on average 10 months (due to a more understandable form of writing): Ontological Diagnostic Reasoner (5 months), testing with debugging (3 months), to extend with another disease (1 month), and new testing with debugging is a month.

Internet services provide savings by scaling their use. Theoretically, to offer many specialists and teams of the same profile, additional efforts are required only to integrate the presentations of patient data. However, in practice, services have a limited set of concepts that do not allow them to accept all the information about the patient. In the (new) cloud technology KES, with a separate declarative knowledge base and a powerful glossary of terms, there are several savings due to a single center of knowledge update and knowledge control (by accumulated case histories). Similarly, the knowledge base about treatment is improving (but there is no way to compare because the team did not have much experience in the past).

The situation to test the system’s viability was requested in early 2020 to expand the service for the diagnosis, differentiation, and treatment of COVID-19. In comparison with other service providers who presented updated versions a few months after the appearance of diagnostic guidelines (Infermedica, klinica.com.ua, and medicase.pro), in our technology, the addition of an existing knowledge base to describe several known variants of manifestation, course, and diagnosing methods of the new disease took several days.

For this extension of the KES, medical experts used two knowledge base editors, adding several dozen statements of diagnosis and treatment. The accuracy of the updated knowledge base was evaluated using the first 15 case histories available. A week later, a new cloud service was launched to search for hypotheses about a patient’s possible viral disease and differential diagnosis. This cloud service is an example of explanatory AI (Figure 6). It provides a rationale for the proposed solutions and recommendations (unlike the services of klinica.com.ua or medicase.pro). The service indicates which signs of the disease are/are not included in the clinical picture of the disease and whether additional information is needed to confirm or refute it. At

Table 1. Comparative analysis of diagnostic hypotheses of two types of Internet services

Type of service	The reference diagnosis was in the top three	The reference diagnosis was in the top 10
Internet	2 – 4	6 – 9
Intelligent adaptive clinical platform	6 – 8	10

the same time, the service asks which values of which signs need to be obtained additionally.

The new cloud service and a declarative method for accessing it (based on the existing solver) demonstrate the feasibility of the technique and approach for evolving KESs and the adequacy of the proposed infrastructure for the development and ongoing evolution of KESs.

7. Conclusion

The application of the proposed approach ensured the construction of scalable medical software services to support specialists of different profiles at different stages of work. It has been demonstrated that the proposed method for producing medical software assistants brings them to the level of explainable AI, which is the consequence of the interpretability of clinical guidelines and knowledge about the course of diseases and their management.

The proposed methodology and production environment for viable systems proved easy to learn and convenient for teamwork. For a medical diagnostic system, each significant knowledge extension (more than 20 such acts were performed in total) required from 5 h to 2 working days for an expert, 5 – 8 h for quality control, 20 min for an architect, and without a programmer, which would be unattainable in another production environment. After each update, the product characteristics analysis showed that the results were consistent with the case samples received from real practice.

Further research should focus on integrating the developed tools with textual facts, knowledge parsers, and third-party diagnostic and predictive tools. A detailed study is required to demonstrate whether the components working with structured information, verbal text, images, and digital arrays can be combined into a single complex. This approach would save valuable time for users in critical areas of activity.

Work is currently underway to expand the capabilities of the approach further. Today, the bottleneck for us is an adaptable user interface. The technology allows you to generate three user interfaces based on the explanation ontology, but these features are insufficient. We are currently working on creating tools for automatically generating an interface based on the user model, considering the usability requirements.

The main contributions of the study are: (i) the automation of multiple physician tasks by filling a single structured "medical history" (integrated with full electronic medical record), (ii) the integration of formalized knowledge from clinical guidelines and other reliable sources to satisfy both the relevance of the methods used and the personalization to the patient, (iii) the transparency of all applicable knowledge, (iv) the explainability of advice based on the essence of the knowledge and with a link to the source, and (v) the integrability of ExClDSS with neural network services, capable of inputting data from a structured document, such as the medical history. Our participation in piloting the (Ex)ClDSS in some medical clinical institutions aligns with the rhetoric of conferences emphasizing the importance of AI for healthcare.

Some of the limitations of the study include the lack of pre-existing converters of formalized knowledge (e.g., in the Protégé paradigm) into our development and support environment (this would provide an opportunity for both the integration of high-quality knowledge and the quality control of the accumulated archives of precedents), a high "entry threshold" to the IACPaaS platform for Python-savvy programmers, and insufficient attention to colorful visualization tools and flexibility of data input.

Today, we are helping to bridge the gap between AI innovation and real-world applications. The experience of moving to trial operations in 2024 has shown that doctors welcome such important general characteristics of these systems. These include the ability to explain hypotheses (results), a mechanism for adding specific knowledge (e.g., new in the clinical information guidelines), and specific properties of specific software systems (for risk assessment, diagnosis, and prognosis). Doctors particularly appreciate systems that facilitate a dialogue to increase the result's accuracy. For treatment-related software systems, the ability to apply knowledge from modern, regularly updated clinical research is also crucial. Developers of ExClDSS, using our technology, emphasize the importance of features such as procedures for regular evaluation of the knowledge base by subsets of precedents (from archived sets and cases from the practice of specialists), as well as reading and directly evaluating the knowledge contained in a specific system.

As technology developers, we consider it important to have a procedure to verify the accuracy or correctness of hypotheses based on any subsets of precedents provided by clients or potential users. Therefore, we believe there is potential for this ontological technology to bridge the gap between AI innovations and their real-world applications.

Acknowledgments

None.

Funding

The work was supported by the Ministry of Education and Science of Russia (No.: FWFW-2021-0004).

Conflict of interest

The authors declare they have no competing interests.

Author contributions

Conceptualization: Valeriya Gribova
Investigation: All authors
Methodology: All authors
Software: Elena Shalfeeva
Writing – original draft: All authors
Writing – review & editing: Elena Shalfeeva

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data

Some software and information components for the study can be obtained by contacting the authors at shalf@iacp.dvo.ru.

Further disclosure

Some of the findings have been presented in the preprint (https://doi.org/10.21203/rs.3.rs-814383/v1) deposited in the preprint server "Research Square."

References

1. Peleg M. Computer-interpretable clinical guidelines: A methodological review. *J Biomed Inform*. 2013;46(4):744-763.

   doi: 10.1016/j.jbi.2013.06.009

2. Young O, Shahar Y, Liel Y, *et al*. Runtime application of Hybrid-Asbru clinical guidelines. *J Biomed Inform*. 2007;40(5):507-526.

   doi: 10.1016/j.jbi.2006.12.004

3. Novais P, Oliveira T, Satoh K, Neves J. The role of ontologies and decision frameworks in computer-interpretable guideline execution. In: Nalepa G, Baumeister J, editors. *Synergies between Knowledge Engineering and Software Engineering. Advances in Intelligent Systems and Computing*. Vol. 626. Cham: Springer; 2018. p. 197-216.

4. Deng C, Ji X, Rainey C, Zhang J, Lu W. Integrating machine learning with human knowledge. *iScience*. 2020;23(11):101656.

   doi: 10.1016/j.isci.2020.101656

5. Müller L, Gangadharaiah R, Klein SC, *et al*. An open access medical knowledge base for community driven diagnostic decision support system development. *BMC Med Inform Decis Mak*. 2019;19:93.

   doi: 10.1186/s12911-019-0804-1

6. Musen M. The protégé project: A look back and a look forward. *AI Matters*. 2015;1(4):4-12.

   doi: 10.1145/2757001.2757003

7. Mortensen J, Minty E, Januszyk M, *et al*. Using the wisdom of the crowds to find critical errors in biomedical ontologies: A study of SNOMED CT. *J Am Med Inform Assoc*. 2015;22(3):640-648.

   doi: 10.1136/amiajnl-2014-002901

8. Loh HW, Ooi CP, Seoni S, Barua PD, Molinari F, Acharya UR. Application of explainable artificial intelligence for healthcare: A systematic review of the last decade (2011-2022). *Comput Methods Programs Biomed*. 2022;226:107161.

   doi: 10.1016/j.cmpb.2022.107161

9. Yang G, Ye Q, Xia J. Unbox the black-box for the medical explainable AI via multi-modal and multi-centre data fusion: A mini-review, two showcases and beyond. *Inf Fusion*. 2022;77:29-52.

   doi: 10.48550/arXiv.2102.01998

10. *Gartner Identifies the Top Strategic Technology Trends for 202*1. STAMFORD, Conn; 2020. Available from: https://www.gartner.com/en/newsroom/press-releases/2020-10-19-gartner-identifies-the-top-strategic-technology-trends-for-2021

11. Pressman RS. *Architectural Design, Software Engineering: Practitioner's Approach*. 7th ed. New York, NY, USA: McGraw-Hill; 2010. p. 242-275.

12. Islam M, Katiyar V. Development of a software maintenance cost estimation model: 4th GL perspective. *Int J Techn Res Appl*. 2014;2(6):65-68.

13. Izurieta C, Bieman JM. A multiple case study of design pattern decay, grime, and rot in evolving software systems. *Software Qual J*. 2013;21(2):289-323.

    doi: 10.1007/s11219-012-9175-x

14. Breivold HP, Crnkovic I, *Eriksson PJ. Analyzing Software Evolvability. COMPSAC 2008: 32nd Annual IEEE International Computer Software and Applications Conference*. Turku, Finland; 2008. p. 327-330.

15. Finn VK. *About Data Mining. Artificial Intelligence News*. (In Russ.); 2004. p. 3-18. Available from: https://masters.donntu.ru/2006/kita/balabanov/library/articles/art010.pdf

16. Weissler EH, Naumann T, Andersson T, *et al*. The role of machine learning in clinical research: Transforming the future of evidence generation. *Trials*. 2021;22:537.

    doi: 10.1186/s13063-021-05489-x

17. Lin Z, Cheng YT, Cheung BMY. Machine learning algorithms identify hypokalaemia risk in people with hypertension in the United States National Health and Nutrition Examination Survey 1999-2018. *Ann Med*. 2023;55(1):2209336.

doi: 10.1080/07853890.2023.2209336

18. Kang SH, Baek H, Cho J, *et al*. Management of cardiovascular disease using an mHealth tool: A randomized clinical trial. *NPJ Digit Med*. 2021;4:165.

    doi: 10.1038/s41746-021-00535-z

19. Zhang Y, Hu N, Li Z, Ji X, *et al. Lumbar Spine Localisation Method Based on Feature Fusion. In: CAAI Transactions on Intelligence Technology*; 2022.

    doi: 10.1049/cit2.12137

20. Wang Y, Li S, Zhao B, *et al*. A ResNet-based approach for accurate radiographic diagnosis of knee osteoarthritis. *CAAI Trans Intell Technol*. 2022;7(3):512-521.

    doi: 10.1049/cit2.12079

21. Golenkov VV, Gulyakina NA, Grakova NV, Nikulenka VY, Eremeev AP, Tarasov VB. From training intelligent systems to training their development means. *Open Semantic Technol Intell Syst*. 2018;2:81-98.

22. Nikolenko SI, Tulupiev AL. *Self-learning Systems*. Moscow, Russia: MCNMO; 2009. p. 288.(In Russ).

23. Dodonov AG, Landje DV. *Viability of Information Systems*. Kiev, Ukraine: Naukova Dumka; 2011. p. 256.(In Russ)

24. Kryazhych OO. Ensuring the viability of the information in time under its processing in the DSS. *Math Mach Syst*. 2015;2:170-176. (In Russ)

25. Gribova VV, Kleshchev AS. Paradigm for controlled intelligent systems. *Manag Syst Inform Technol*. 2016;3(65):32-38. (In Russ)

26. Chhabra JK. Improving modular structure of software system using structural and lexical dependency. *Inform Softw Technol*. 2017;82:96-120.

    doi: 10.1016/j.infsof.2016.09.011

27. Gribova VV, Kleschev AS, Moskalenko FM, Timchenko VA, Shalfeeva EA. Extensible toolkit for the development of viable systems with knowledge bases. *Softw Eng*. 2018;9(8):339-348.

    doi: 10.17587/prin.9.339-348

28. Gribova V, Shalfeeva E, Petryaeva M. Cloud infrastructure for creation of interpretable diagnostic knowledge bases of diseases regardless their etiology. *Atlantis Highlights Comput Sci*. 2019;3:79-82.

    doi: 10.2991/csit-19.2019.13

29. Kleschev AS, Chernyakhovskaya MY, Shalfeeva EA. The paradigm of an intellectual professional activity automation. Part 1. The features of an intellectual professional activity. *Ontol Design*. 2013;3(9):53-69. (In Russ).

30. Gribova V, Moskalenko P, Petryaeva M, Okun D. Cloud environment for development and use of software systems for clinical medicine and education. *Adv Intell Syst Res*. 2019;166:225-229.

    doi: 10.2991/itids-19.2019.40

ACCSCIENCE PUBLISHING

Artificial Intelligence in Health

MINI-REVIEW

Emotional intelligence and artificial intelligence as catalysts for professional development and lifelong learning among healthcare professionals: A literature review

Nadia Hachoumi[1,2]* and **Mohamed Eddabbah**[2,3]

[1]Department of Rehabilitation, Higher Institute of Nursing Professions and Health Techniques, Ministry of Health and Social Protection, Marrakesh, Marrakesh–Safi, Morocco

[2]Research Laboratory for Intelligent and Sustainable Technologies, National School of Applied Sciences, CADI AYYAD University, Marrakesh, Marrakesh–Safi, Morocco

[3]Department of computer sciences, Higher School of Technology, Cadi Ayyad University, Essaouira, Marrakesh–Safi, Morocco

Abstract

The current healthcare professionals require interdisciplinary training that integrates emotional intelligence (EI) and artificial intelligence (AI). EI equips healthcare professionals with the ability to communicate empathetically, provide patient care, and regulate emotions, while AI serves as a knowledge-based decision support system that improves decision-making and clinical efficiency. This study explores the integration of EI and AI in healthcare and examines their combined impact on both instructional methods and clinical practice. In addition, we evaluated the role of EI in fostering patient interaction, strengthening teamwork, and combating burnout, alongside the role of AI in advancing learning, improving diagnostic accuracy, and enabling personalized care. Moreover, existing literature on EI and AI is discussed in this study to highlight their complementary roles in enhancing healthcare practices. A combined EI and AI training approach can offer a holistic training model for preparing healthcare professionals. While EI enhances its ability to handle emotional challenges, AI provides data-driven information that can sharpen clinical thinking and improve efficiency. Together, EI and AI play a crucial role in enhancing patient care, decision-making, and teamwork. An integrated approach that combines both AI and EI, aimed at enhancing clinical skills and professional development, represents a promising advancement in healthcare practice. Integrating EI with AI tools optimizes both human and technological capabilities, fostering a more competent, compassionate, and productive healthcare workforce.

Keywords: Artificial intelligence; Emotional intelligence; Healthcare; Patient-centered care; Professional development

***Corresponding author:**
Nadia Hachoumi
(nadia.hachoumi@ced.uca.ma)

Citation: Hachoumi N, Eddabbah M. Emotional intelligence and artificial intelligence as catalysts for professional development and lifelong learning among healthcare professionals: A literature review. *Artif Intell Health*. 2025;2(3):154-160. doi: 10.36922/AIH025080011

Received: February 18, 2025

Revised: March 24, 2025

Accepted: April 28, 2025

Published online: May 15, 2025

Copyright: © 2025 Author(s). This is an Open-Access article distributed under the terms of the Creative Commons Attribution License, permitting distribution, and reproduction in any medium, provided the original work is properly cited.

Publisher's Note: AccScience Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

Healthcare professionals require more than just technical skills to effectively manage their responsibilities in the healthcare sector. It is imperative for them to be adaptable to rapidly changing circumstances, to make swift decisions under pressure, and to provide patient-centered, compassionate care.[1] Therefore, a varied skill set is essential. In general, a broad range of skills are really necessary, especially the combination of EI and AI as they are increasingly being recognized as key elements underpinning contemporary healthcare practice.[2]

Despite their differences, both EI and AI have made significant contributions to healthcare, impacting both clinical practice and professional development (PD) across healthcare systems. This study extends existing literature on EI and AI by exploring their complementary roles in PD and offering new insights into their combined potential to address persistent challenges in healthcare. AI provides revolutionary tools that enhance healthcare professionals' capabilities, broaden their knowledge, and improve their ability to meet patients' needs. Meanwhile, EI equips healthcare professionals with the ability to perceive, analyze, and control both their own emotions and those of others.[3]

In recent decades, EI has gained widespread recognition as a key driver of effective healthcare. The 1995 Goleman's theory highlights the importance of understanding and managing emotions in the workplace. In the healthcare sector, EI enables healthcare professionals to communicate more effectively, build stronger connections with patients, and foster collaborative environments with colleagues.[4] The importance of these skills cannot be overstated, especially in emotionally charged situations where empathy and emotional connection can significantly influence a healthcare professional's performance. Healthcare professionals with higher EI are better equipped to meet patient needs, demonstrate empathy, and respond with patience – all of which are essential aspects of comprehensive, human-centered care.[5] Moreover, EI plays a vital role in supporting healthcare professionals' well-being by helping to reduce stress, prevent burnout, and manage emotional challenges common in the profession.[6] While EI has been widely discussed, this study explores the integration of EI and AI to further support and scale emotional competencies in healthcare training.

Furthermore, AI is transforming the ways in which healthcare professionals learn, practice, and access information. AI capabilities provide healthcare professionals with unique opportunities to engage with educational content relevant to their skills and needs, enabling the development of highly personalized learning experiences.[7] This study highlights the potential of AI in creating adaptive learning pathways when combined with emotional attunement, presenting a promising advancement in health education.

AI – ranging from diagnostic robots that enhance clinical judgment to adaptive learning platforms that tailor content to the user's abilities – enables healthcare professionals to engage in lifelong learning. Using these tools, they can keep their knowledge and skills up-to-date, stay informed about healthcare advances, and continuously develop practical skills. AI-based platforms revolutionize the speed and quality of skills development by providing interactive simulations, personalized examinations, and immediate feedback.[8]

The integration of EI and AI in healthcare has created new opportunities for PD, requiring a more integrated approach to training and lifelong learning.[9] This convergence holds the potential to reshape conventional methods of healthcare education and PD. This study aims to address the gap in the existing literature by exploring the synergy between AI and EI in shaping the training of healthcare professionals who are both empathetic and technologically skilled. In such an approach, doctors and nurses are trained not only to make accurate diagnoses with the assistance of AI but also to communicate those diagnoses with empathy and EI.[10] Together, EI and AI can help foster a healthcare workforce that is not only technically skilled but also highly sensitive to the human aspects of care.[11]

This study explores the integration of EI and AI in the PD and lifelong learning of healthcare professionals. It focuses on the synergy of EI and AI in shaping more adaptable, resilient, and compassionate healthcare professionals capable of meeting the new demands of modern medicine. By exploring the promising potential of this integrated approach, this study provides a new perspective on how it may reshape healthcare training and ultimately enhance patient care.

2. Methodology

This study explores the roles of EI and AI in supporting the PD of healthcare professionals. Both the ability to engage emotionally with patients and proficiency in current technological advancements are essential in modern healthcare. This integrated approach equips healthcare professionals to better manage their responsibilities, improve patient care, and foster empathy in a rapidly changing healthcare environment. By balancing technical proficiency and patient-centered care, this approach has the potential to enhance both PD and clinical outcomes.

Therefore, this study investigates the combined impact of EI and AI in enhancing the skills of healthcare professionals and the quality of patient care. This methodological section outlines the research process, including the databases used, the keywords employed, and the inclusion and exclusion criteria applied in selecting relevant literature for analysis.

2.1. Databases

An extensive search was conducted using PubMed, Scopus, Connected Papers, EBSCO, and Google Scholar to identify relevant published studies.

2.2. Inclusion criteria

The selection criteria prioritized original and empirical research published in English between 2010 and 2024, focusing on studies that examined the impact of EI training or AI applications within the healthcare context.

2.3. Exclusion criteria

Studies were excluded if they were opinion-based, not peer-reviewed, or not focused on healthcare-related applications.

2.4. Search strategy

The search strategy incorporated keywords such as "emotional intelligence," "artificial intelligence," "healthcare professional," "lifelong learning," "clinical practice," and "skill development." Boolean operations (AND/OR) were applied to refine and narrow the search results.

2.5. Selection of articles

Articles were initially screened based on their titles and abstracts to determine their relevance. Full-text reviews were then conducted for articles that met the inclusion criteria to verify their suitability for the study.

2.6. Data extraction

Key information was extracted from the selected articles, including the authors' names, publication dates, research methodologies, demographic details of the sample group, and main findings. A predefined template was used to ensure consistency in the data extraction process.

3. Results

The literature review offers valuable insights into the relationship between EI and AI in the context of healthcare professionals. A total of 15 studies met the inclusion criteria, indicating a strong consensus on the relevance of these constructs in PD. EI consistently emerged as a key factor in enhancing the effectiveness of healthcare professionals.[5] Those with higher EI demonstrated greater abilities in recognizing and regulating both their own emotions and those of their patients.[12] This skill fosters stronger therapeutic relationships, enabling healthcare professionals to meet their patients' needs with empathy, contributing to improved patient satisfaction and health outcomes.[13]

One study found that EI plays a crucial role in mitigating burnout among caregiver – a major challenge in high-pressure healthcare settings.[14] This aligns with other findings indicating that healthcare professionals with higher EI experience lower levels of stress and burnout, suggesting that EI may act as a protective factor against the emotional strain of a career in healthcare.[14] In addition, EI has been linked to improved teamwork among healthcare professionals. By promoting effective communication and conflict management, EI contributes to a more cooperative work environment, and this sense of teamwork is especially valuable in multidisciplinary teams, where diverse expertise and perspectives are essential for delivering optimal patient care.[12]

3.1. The role of AI

AI is transforming the education and training of healthcare professionals. Adaptive learning systems are particularly promising for personalizing educational content based on individual learning styles and needs, alongside other AI technologies.[10] For example, AI-powered platforms can analyze unique learning profiles and tailor content to better suit healthcare professionals.[9] This personalized approach to learning enhances engagement and retention, which is crucial for ensuring healthcare professionals stay up-to-date with advancements in health sciences.[15] Moreover, AI tools – such as chatbots and virtual assistants – facilitate patient interactions.[16] These tools not only provide patients with immediate responses to their queries but also help healthcare professionals manage the emotional challenges associated with patient care.[3] By supporting patients' needs, AI systems allow healthcare professionals to provide the highest quality care without becoming overwhelmed by the emotional demands of their occupation.[17]

3.2. Synergistic effects of EI and AI

This review highlights the complementary benefits of integrating EI and AI in healthcare. For instance, AI tools can provide healthcare professionals with real-time feedback on patients' emotional state by analyzing cues such as the tone of their voice and body language.[18] Shaik *et al.*[19] observed that AI tools help healthcare professionals to personalize their approach, enabling more responsive care to patients' emotional needs. Incorporating EI training into the design of AI systems helps ensure that healthcare professionals use these technologies effectively and appropriately. Combining EI and AI training not only enhances healthcare professionals' technical skills but also builds the emotional skills needed to successfully manage complex interactions with patients.[20] The integration of EI and AI presents a promising opportunity to advance healthcare practice, ensuring that technological innovation complements the human aspects of care.[20]

Evidence from the existing literature shows that both EI and AI contribute to the PD of healthcare professionals. By fostering emotional competence alongside technical expertise, healthcare systems can improve the quality of patient care and support healthcare professionals' well-being. This dual focus is essential for addressing the evolving challenges within the healthcare sector.

This study identified several key studies that met the inclusion criteria, highlighting the significance of EI and AI in the PD of healthcare professionals. High EI was associated with improved communication, enhanced patient satisfaction, and reduced burnout. On the other hand, AI contributed to personalized learning tools, improved diagnostic accuracy, and supported decision-making. A summary of the studies is provided in Table 1.

Figure 1 presents a keyword cloud – generated using https://github.com/eddabbah/keywordsMapGenerator/blob/main/main.py – that visualizes the key terms and concepts emerging from the reviewed studies. This visualization highlights the frequency and significance of the central themes, offering a concise overview of the critical ideas surrounding EI and AI in the context of healthcare professionals. It provides a clear illustration of the dominant themes and their relationships as identified in the research literature.

4. Discussion

4.1. Implications of EI and AI on the PD of healthcare professionals

The integration of EI and AI in the PD of healthcare professionals presents a valuable avenue for PD. Integrating EI into training programs helps healthcare professionals to enhance their communication skills and emotional regulation – both critical for effective interactions with

Table 1. Summary of key studies on the role of emotional intelligence and artificial intelligence in healthcare professionals

Authors	Title	Year	Sample size	Methodology	Research question	Key findings
Johnson[1]	The shifting landscape of health care: toward a model of health care empowerment	2011	150	Quantitative	The impact of EI on patient care	Improved communication and reduced burnout
Bohr and Memarzadeh[10]	The rise of artificial intelligence in healthcare applications	2020	250	Systematic review	AI's role in personalized learning	Boosted engagement and personalized learning pathways
Karimi *et al.*[6]	Emotional intelligence: predictor of employees' well-being, quality of patient care, and psychological empowerment	2021	300	Quantitative	EI and caregiver burnout	Reduced burnout and improved well-being
Cao *et al.*[14]	The influence of emotional intelligence on job burnout of healthcare workers and mediating role of workplace violence: a cross-sectional study	2022	220	Cross-sectional	The effect of EI on job burnout and workplace violence	Lower burnout rates and better emotional regulation
McNulty and Politis[5]	Empathy, emotional intelligence, and interprofessional skills in healthcare education	2023	120	Qualitative	EI in team dynamics	Enhanced teamwork and patient satisfaction
Coronado-Maldonado and Benítez-Márquez[12]	Emotional intelligence, leadership, and work teams: a hybrid literature review	2023	100	Qualitative	EI and leadership in healthcare teams	Strengthened leadership skills and teamwork
Shaik *et al.*[19]	Remote patient monitoring using artificial intelligence: current state, applications, and challenges	2023	180	Longitudinal	Synergy of EI and AI	Enhanced clinical decision-making and emotional regulation
Mishra[7]	The impact of AI on improving the efficiency and accuracy of managerial decisions	2024	200	Mixed-methods	AI in diagnostics	Increased accuracy and decision efficiency
Narimisaei *et al.*[11]	Exploring emotional intelligence in artificial intelligence systems: a comprehensive analysis of emotion recognition and response mechanisms	2024	130	Mixed-methods	AI-driven emotional recognition in clinical settings	Enhanced patient-clinician emotional communication

Abbreviations: AI: Artificial intelligence; EI: Emotional intelligence.

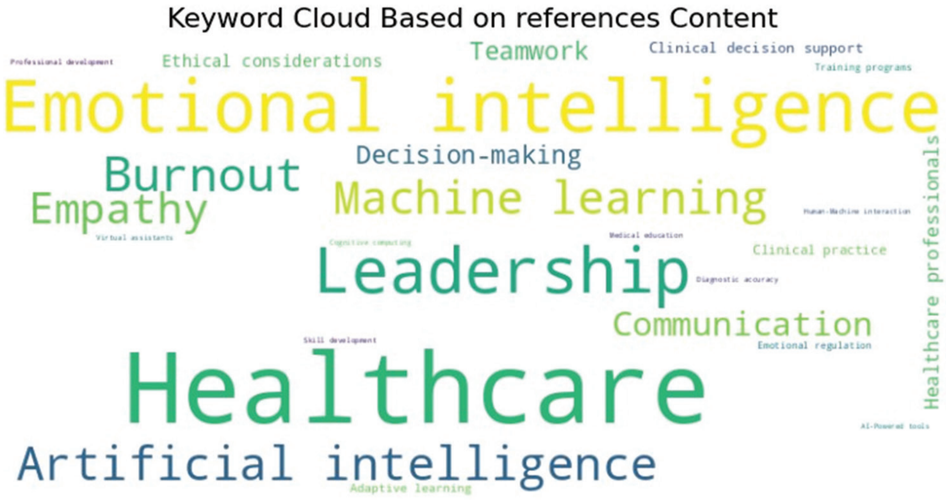


Figure 1. Keyword cloud generated from the content of the reviewed articles

patients. EI-focused modules in these programs have shown strong potential to improve healthcare professionals' ability to navigate the emotional complexities of patient care.[21]

Studies have shown that healthcare professionals with higher EI are more likely to engage in lifelong learning and effectively apply acquired knowledge in real-world settings.[9] Therefore, integrating EI with AI-driven adaptive learning technologies not only supports the PD of healthcare professionals but also aligns with institutional goals of developing a more competent and compassionate workforce. This shift in educational priorities has the potential to improve the healthcare system's ability to deliver high-quality care.[22]

4.2. The creation of collaborative learning environments

Collaborative learning plays a crucial role in healthcare education, as emotional skills are most effectively developed in group settings.[23] EI encourages open communication and conflict management, both of which are essential for delivering high-quality patient care.[5] Integrating EI with AI improves these dynamics, creating richer and more adaptive learning experiences. AI platforms can facilitate group discussions and clinical case simulations, creating opportunities for collaborative learning. These environments foster mutual learning, allowing participants to share their unique skills and perspectives, thereby enhancing the overall team performance. Such collaboration is essential for building trust and shared accountability, both of which are vital to improving patient outcomes.[24]

4.3. The impact of EI and AI on the motivation and commitment of healthcare professionals

EI has been associated with greater motivation for continuing education among healthcare professionals. Research shows that those with strong emotional regulation skills are more likely to participate in ongoing PD.[25] The use of AI in training enhances this process by providing immediate feedback, which increases engagement and sustains learning interest. AI systems can recognize when a learner reaches certain milestones and reward them for achieving them. This real-time responsiveness is particularly valuable in the fast-paced world of healthcare, where time is indeed money. By fostering a culture of lifelong learning and PD, healthcare organizations can motivate healthcare professionals to invest in their personal growth, leading to improved patient care and greater job satisfaction.[26]

5. Limitations

The integration of EI and AI in healthcare training offers promising potential; however, several limitations must be addressed. One limitation is ensuring that training programs are comprehensive enough to educate healthcare professionals on the use of AI tools and to develop their EI skills. In addition, to implement AI systems effectively, institutions must provide financial resources for their establishment and activate the necessary infrastructure. However, financial accessibility may not be guaranteed in all institutions. Furthermore, over-reliance on AI could also weaken the human touch, which is essential in patient care. Ethical concerns, such as data privacy and algorithmic biases, may further complicate the adoption of AI technologies. In addition, the subjectivity and context-sensitivity of EI make it challenging to measure the effectiveness of AI-based training. To address these limitations, further research and the development of protocols that facilitate the balanced integration of AI and EI across diverse healthcare settings are essential.

6. Future directions and research

Further investigation is needed to fully understand the clinical potential of integrating EI with AI, particularly in relation to patient outcomes and the well-being of healthcare professionals. Longitudinal studies could provide valuable insights into how this integration evolves over time and its impact on the performance of healthcare professionals and the quality of patient care.

The integration of EI and AI also holds valuable potential for guiding other aspects of healthcare, such as team organization and ethical decision-making. A deeper understanding of EI and AI in the context of leadership development and clinical ethics will be essential for the development of competent, well-rounded healthcare professionals.

Finally, it is essential to establish training programs for educators or trainers to ensure they are well-equipped to teach EI and effectively utilize AI tools. By preparing educators or trainers to educate these fundamental skills, healthcare institutions can foster a more competent and emotionally intelligent workforce capable of addressing the demands of modern healthcare.

7. Conclusion

Integrating EI with AI in healthcare education supports the PD of healthcare professionals and enhances patient care by fostering stronger emotional connections, which in turn improve empathy and patient satisfaction. At present, AI contributes to personalized learning by enabling data-driven decision-making, ultimately enhancing knowledge acquisition and clinical skills. The integration of EI and AI holds the potential to create a more technologically competent and emotionally intelligent workforce, thereby fostering stronger inter-professional collaborations and improving patient outcomes. Longitudinal studies are needed to thoroughly assess the overall impact of this innovation on PD and patient care. In turn, this new paradigm in healthcare education would shape healthcare professionals who are more emotionally and psychologically resilient.

Acknowledgments

None.

Funding

None.

Conflict of interest

The authors declare no conflict of interest.

Author contributions

Conceptualization: All authors
Writing – original draft: Nadia Hachoumi
Writing – review & editing: All authors

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data

Not applicable.

References

1. Johnson MO. The shifting landscape of health care: Toward a model of health care empowerment. *Am J Public Health*. 2011;101(2):265-270.

   doi: 10.2105/AJPH.2009.189829

2. Erol BA, Majumdar A, Benavidez P, Rad P, Choo KKR, Jamshidi M. Toward artificial emotional intelligence for cooperative social human-machine interaction. *IEEE Trans Comput Soc Syst*. 2020;7(1):234-246.

   doi: 10.1109/TCSS.2019.2922593

3. Sharma V, Kumar H. Emotional intelligence in the era of artificial intelligence for medical professionals. *J Int Med Grad*. 2023;2(2):112.

   doi: 10.56570/jimgs.v2i2.112

4. Goleman D. *Leadership: The Power of Emotional Intelligence*. United States: More Than Sound; 1995.

5. McNulty JP, Politis Y. Empathy, emotional intelligence and interprofessional skills in healthcare education. *J Med Imaging Radiat Sci*. 2023;54(2):238-246.

   doi: 10.1016/j.jmir.2023.02.014

6. Karimi L, Leggat SG, Bartram T, Afshari L, Sarkeshik S, Verulava T. Emotional intelligence: Predictor of employees wellbeing, quality of patient care, and psychological empowerment. *BMC Psychol*. 2021;9(1):93.

   doi: 10.1186/s40359-021-00593-8

7. Mishra S. *The Impact of AI on Improving the Efficiency and Accuracy of Managerial Decisions; 2024*. Available from: https://www.researchgate.net/publication/382850204-the-

impact-of-ai-on-improving-the-efficiency-and-accuracy-of-managerial-decisions [Last accessed on Mar 23 2025].

8. Dai CP, Ke F. Educational applications of artificial intelligence in simulation-based learning: A systematic mapping review. *Comput Educ Artif Intell*. 2022;3:100087.

   doi: 10.1016/j.caeai.2022.100087

9. Hachoumi N, Eddabbah M, El Adib AR. Health sciences lifelong learning and professional development in the era of artificial intelligence. *Int J Med Inform*. 2023;178:105171.

   doi: 10.1016/j.ijmedinf.2023.105171

10. Bohr A, Memarzadeh K. The rise of artificial intelligence in healthcare applications. *Artif Intell Healthc*. 2020;26:25-60.

    doi: 10.1016/B978-0-12-818438-7.00002-2

11. Narimisaei J, Naeim M, Imannezhad S, Samian P, Sobhani M. Exploring emotional intelligence in artificial intelligence systems: A comprehensive analysis of emotion recognition and response mechanisms. *Ann Med Surg (Lond)*. 2024;86(8):4657-4663.

    doi: 10.1097/MS9.0000000000002315

12. Coronado-Maldonado I, Benítez-Márquez MD. *Emotional Intelligence, Leadership, and Work Teams: A Hybrid Literature Review Science Direct; 2023*. Available from: https://www.sciencedirect.com/science/article/pii/s2405844023075643 [Last accessed on 2024 Nov 6].

13. Moudatsou M, Stavropoulou A, Philalithis A, Koukouli S. The role of empathy in health and social care professionals. *Healthcare*. 2020;8(1):26.

    doi: 10.3390/healthcare8010026

14. Cao Y, Gao L, Fan L, Jiao M, Li Y, Ma Y. The influence of emotional intelligence on job burnout of healthcare workers and mediating role of workplace violence: A cross sectional study. *Front Public Health*. 2022;10:892421.

    doi: 10.3389/fpubh.2022.892421

15. Aggarwal K, Mijwil MM, Sonia, G, *et al*. Has the future started? The current growth of artificial intelligence, machine learning, and deep learning. *Iraqi J Comput Sci Math*. 2022;3(1):115-123.

    doi: 10.52866/ijcsm.2022.01.01.013

16. Bassner P, Frankford E. *Iris: An AI-Driven Virtual Tutor for Computer Science Education; 2024*. Available from: https://arxiv.org/html/2405.08008v1 [Last accessed on 2024 May 21].

17. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. *Future Healthc J*. 2019;6:94-98.

    doi: 10.7861/futurehosp.6-2-94

18. Abdulghafor R, Abdelmohsen A, Turaev S, Ali MAH, Wani S. An analysis of body language of patients using artificial intelligence. *Healthcare*. 2022;10(12):2504.

    doi: 10.3390/healthcare10122504

19. Shaik T, Tao X, Higgins N, *et al*. Remote patient monitoring using artificial intelligence: Current state, applications, and challenges. *WIREs Data Min Knowl Discov*. 2023;13(2):e1485.

    doi: 10.1002/widm.1485

20. Mattingly V, Kraiger K. Can emotional intelligence be trained? A meta-analytical investigation. *Hum Resour Manag Rev*. 2019;29(2):140-155.

    doi: 10.1016/j.hrmr.2018.03.002

21. Ibrahim MMA. Exploring emotional intelligence in Jordan's artificial intelligence (AI) healthcare adoption: A UTAUT framework. *J Electr Syst*. 2024;20(10s):502-541.

    doi: 10.52783/jes.5143

22. Abulibdeh A, Zaidan E, Abulibdeh R. *Navigating the Confluence of Artificial Intelligence and Education for Sustainable Development in the Era of Industry 4.0: Challenges, Opportunities, and Ethical Dimensions - Science Direct; 2024*. Available from: https://www.sciencedirect.com/science/article/abs/pii/s0959652623046851 [Last accessed on 2024 Apr 27].

23. Docherty M. *Collaborative Learning: The Group is Greater than the Sum of its Parts*. Cham: Springer; 2020. p. 26-33.

24. Sethi SS, Jain K. AI technologies for social emotional learning: Recent research and future directions. *J Res Innov Teach Learn*. 2024;17(2):213-225.

    doi: 10.1108/JRIT-03-2024-0073

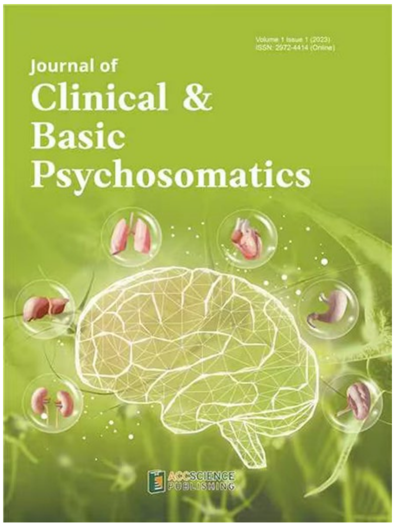
25. Cherry MG, Fletcher I, O'Sullivan H, Dornan T. Emotional intelligence in medical education: A critical review. *Med Educ*. 2014;48(5):468-478.

    doi: 10.1111/medu.12406

26. Conradty C, Bogner FX. *Learning How to Flow: Paving the Way for Lifelong Learning IntechOpen; 2024*. Available from: https://www.intechopen.com/chapters/89014 [Last accessed on 2024 Nov 11].

ACCSCIENCE
PUBLISHING

OUR JOURNALS



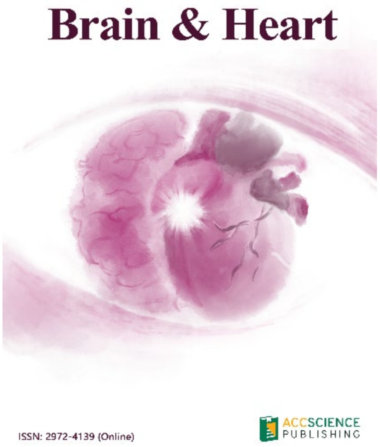
Journal of Clinical and Basic Psychosomatics (JCBP) is a quarterly journal focusing on clinical and basic research on symptoms, assessment, treatment, management, and the mechanism of psychosomatic disorders. *Journal of Clinical and Basic Psychosomatics* covers subject areas, including but not limited to the following:

- Conceptualization and classification of psychosomatic medicine
- Mechanism, biological markers, brain images, and treatment studies
- Psychosomatic reactions, syndromes, disorders, and diseases
- Psychosomatic disorders treated in general hospitals, including endocrinology, neurology, gastroenterology, dermatology, pain management, oncology, rheumatology, and other departments
- Psychological evaluation, management, rehabilitation, resilience training, and psychotherapy for general and specific populations during the pandemic
- Physiological disorders related to psychological factors (eating disorders, sleeping disorders, and sexual dysfunction)
- Somatic symptoms and related disorders and mental disorders due to somatic disease

Brain & Heart focuses on neurocardiology, a neurology and cardiology-based interdisciplinary subject that studies the circulatory mechanism of the human body, as well as the mechanisms of the interplay between the cardiovascular system and the nervous system. The journal’s scope includes:

Clinical and basic research on diseases related to the circulatory and nervous systems, such as: orthostatic dizziness, orthostatic hypotension, autonomic dysfunction, and the relationship between the autonomic nervous system and the circulatory function in cerebral degeneration;

Heart-brain research on patients with syncope, autonomic dysfunction, cryptogenic stroke, and stroke with atrial fibrillation; research on the relationship between structural heart diseases and nervous system diseases, the correlation between cardiac electrophysiology and abnormal organizational structures and the pathogenesis of stroke, as well as new ways of diagnosis, treatment and prevention of unexplained stroke.



Start a new journal

Write to us via email if you are interested to start a new journal with AccScience Publishing. Please attach your CV, professional profile page and a brief pitch proposal in your email. We shall inform you of our decision whether we are interested to collaborate in starting a new journal.

Contact: info@accscience.com

https://www.accscience.com/journal/AIH

ACCSCIENCE
PUBLISHING

Contact
www.accscience.com
9 Raffles Place, Republic Plaza 1 #06-00 Singapore 048619
Email: editorial@accscience.com
Phone: +65 8182 1586