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

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

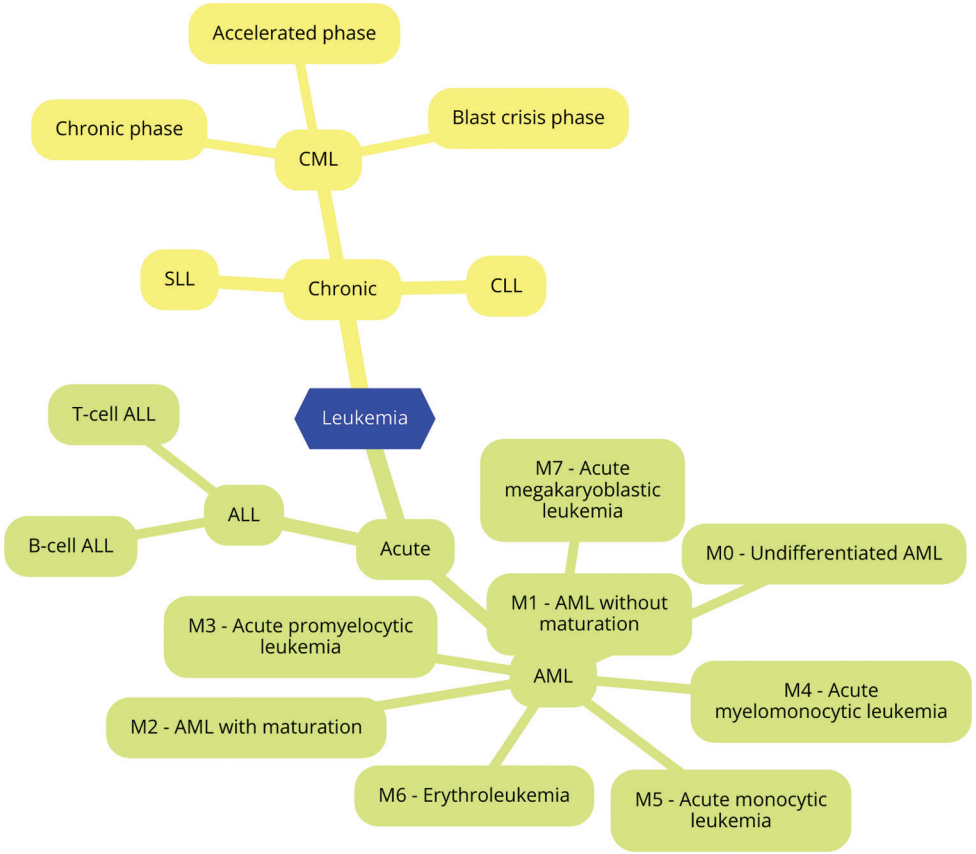

Figure 1. Categorization of different types of leukemia: (A) acute leukemia and its subtypes, including acute myeloid leukemia and acute lymphoblastic leukemia, with detailed divisions; (B) chronic leukemia and its subtypes, including chronic myeloid leukemia and chronic lymphocytic leukemia, along with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

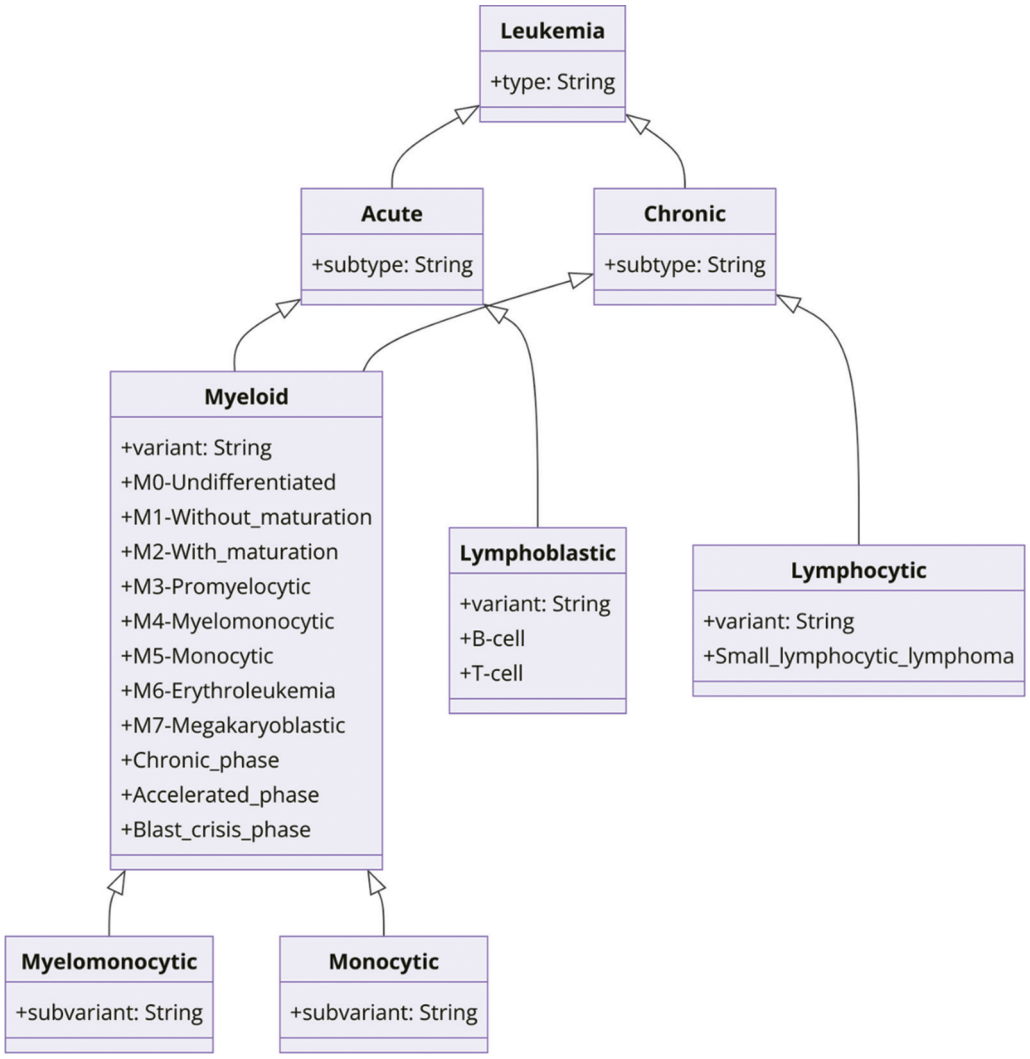

Figure 2. A UML diagram of detailed classification of leukemia depicting the object-oriented representations of acute leukemia subtypes and chronic leukemia subtypes.

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

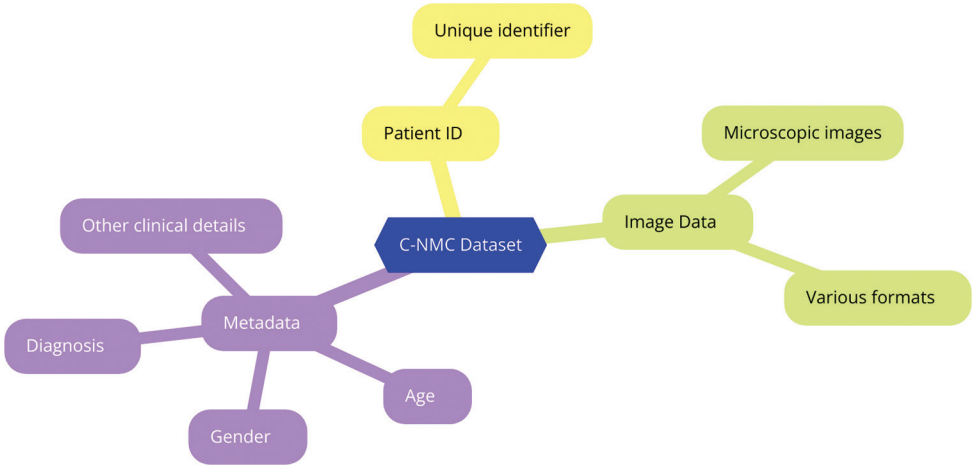

Figure 3. The C-NMC dataset description includes the following details: (A) dataset composition by image type, including healthy and malignant cells, and (B) training and testing data with associated metadata details.

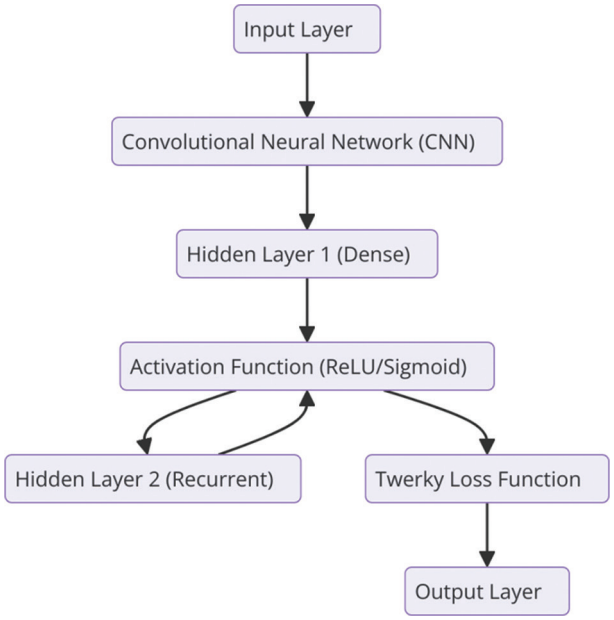

Figure 4. The proposed deep learning model architecture includes the following components: (A) convolutional layers for feature extraction, (B) dense layers for classification tasks, and (C) a Tversky loss function optimization block.

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.

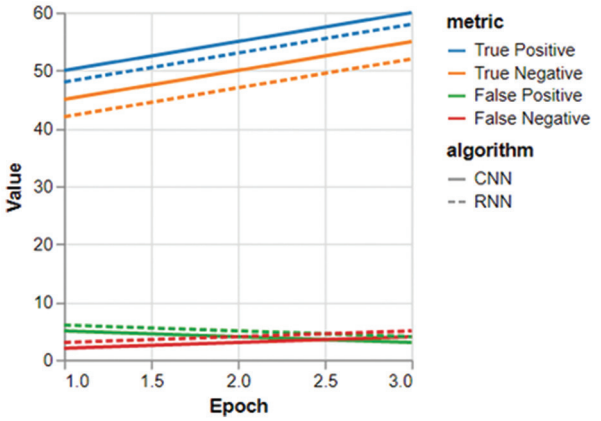

Figure 5. Confusion matrix values for CNN versus RNN: (A) confusion matrix for CNN prformance at epoch 3, and (B) confusion matrix for RNN performance at Epoch 3.
Abbreviations: CNN: Convolutional neural network; RNN: Recurrent neural network.

- *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 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

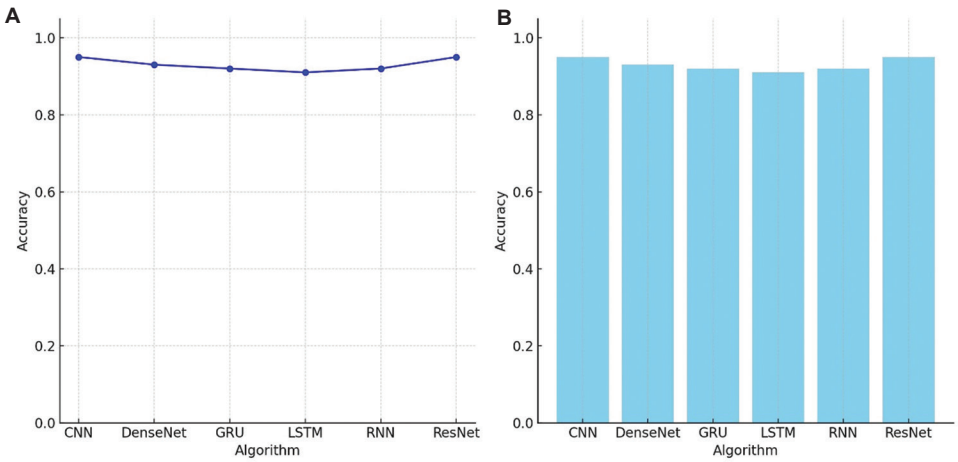

Figure 6. Accuracies of different deep learning algorithms on C-NMC Dataset. (A) Accuracy Trends of CNN Across Epochs (B) Comparative Accuracies of RNN, DenseNet, GRU, and ResNet.
Abbreviations: CNN: Convolutional neural network; GRU: Gated recurrent unit; LSTM: Long short-term memory; RNN: Recurrent neural network.

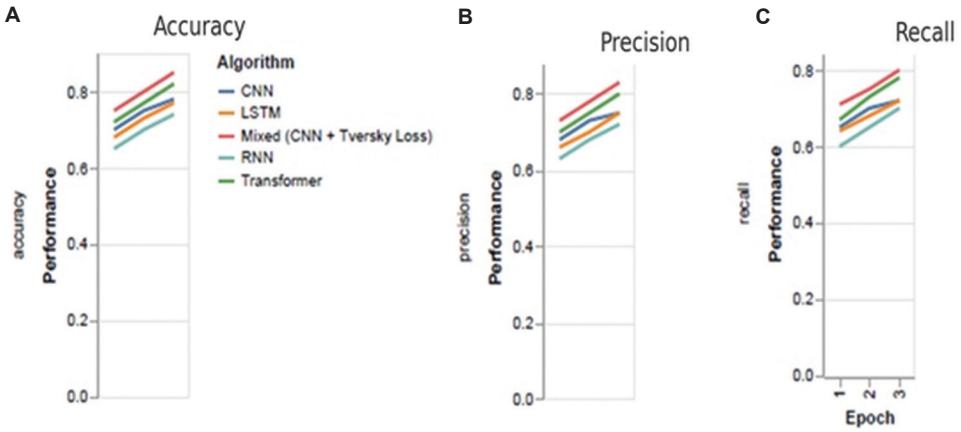

Figure 7. Performance comparison of different deep learning algorithms. (A) Accuracy comparison across CNN, LSTM, and Transformer models. (B) Precision comparison of the same models. (C) Recall comparison of the same models.
Abbreviations: CNN: Convolutional neural network; LSTM: Long short-term memory;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

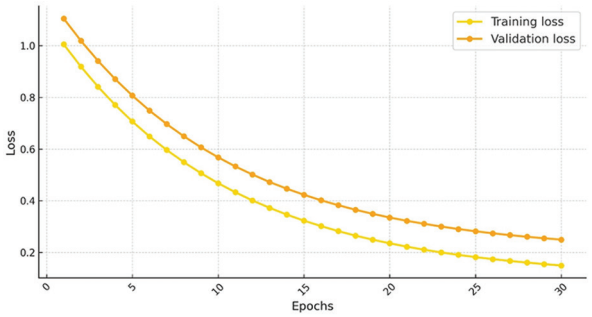

Figure 8. The trends of training and validation losses of CNN + Tversky loss on C-NMC dataset over epochs.

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

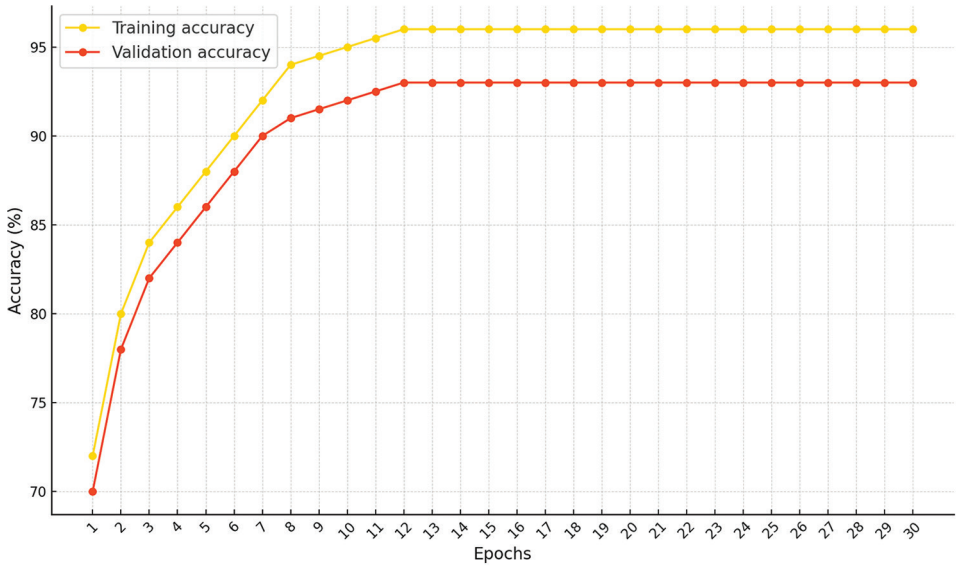

Figure 9. The trends of training and validation accuracies of CNN + Tversky loss on C-NMC dataset over epochs.

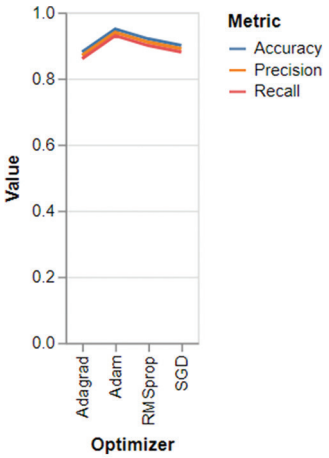

Figure 10. Metrics of different optimizers of deep learning includes: (A) Performance of Adam optimizer in terms of accuracy, precision, and recall, demonstrating its suitability for optimizing complex models (B) Comparative performance of Adagrad, RMSprop, and SGD optimizers, showcasing their relative 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 YC, Zheng WS, Lai JH, Yuen PC. An Asymmetric Distance Model for Cross-View Feature Mapping in Person Reidentification. *IEEE Trans Circuits Syst Video Technol*. 2017;27(8):1661-1675.

   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(3):567-585.

   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 AB, Ale Hassan S, Gull S. Automatic Detection of Leukemia through Convolutional Neural Network. *In: 2022 International Conference on Frontiers of Information Technology (FIT)*. IEEE; 2022:195-200.

   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)*. IEEE; 2024:173-177.

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, Ma WY, Zhao T. Automatic image dataset construction from click-through logs using deep neural network. *In: Proceedings of the 23rd ACM International Conference on Multimedia*. Association for Computing Machinery; 2015:441-450.

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