ACCSCIENCE PUBLISHING

Artificial Intelligence in Health

ORIGINAL RESEARCH ARTICLE

Diagnosis of COVID-19 from computed tomography slices using flower pollination algorithm, k-nearest neighbor, and support vector machine classifiers

Betshrine Rachel Jibinsingh[1], Khanna Nehemiah Harichandran[1]*, Kabilasri Jayakannan[2], Rebecca Mercy Victoria Manoharan[3], and Anisha Isaac[1]

[1]Ramanujan Computing Centre, College of Engineering Guindy, Anna University, Chennai, Tamil Nadu, India

[2]Department of Information Science and Technology, College of Engineering Guindy, Anna University, Chennai, Tamil Nadu, India

[3]Department of Computer Science and Engineering, College of Engineering Guindy, Anna University, Chennai, Tamil Nadu, India

***Corresponding author:**
Khanna Nehemiah Harichandran (nehemiah@annauniv.edu)

Citation: Jibinsingh BR, Harichandran KN, Jayakannan K, Manoharan RMV, Isaac A. Diagnosis of COVID-19 from computed tomography slices using flower pollination algorithm, k-nearest neighbor, and support vector machine classifiers. *Artif Intell Health*. 2025;2(1):14-28. doi: 10.36922/aih.3349

Received: April 3, 2024

Revised: June 17, 2024

Accepted: June 24, 2024

Published Online: October 23, 2024

Abstract

Coronavirus disease 19 (COVID-19), caused by the severe acute respiratory syndrome-coronavirus-2 virus, is commonly diagnosed through imaging techniques such as computed tomography (CT) scans, which reveal characteristic lung lesions. In this study, we propose a computer-aided diagnosis (CAD) system to assist in the early detection of COVID-19 from CT lung slices, leveraging advanced machine-learning algorithms for precise and efficient analysis. To achieve this, we developed a CAD system that diagnoses COVID-19 from CT lung slices. An adaptive Wiener filter was applied to remove noise from the CT images. The chest tissues were then segmented using an optimal thresholding method to extract regions of interest, which represent the COVID-19 lesions under investigation. The feature vectors were divided into training and testing with an 80/20 ratio. A wrapper-based flower pollination algorithm was employed alongside the k-nearest neighbor classifier to select the optimal feature set. These selected features were subsequently used to train a support vector machine (SVM) classifier. With feature selection, the SVM achieved an accuracy of 91.30% on a real-time dataset, outperforming seven other machine learning classifiers (radial basis function-SVM, k nearest neighbor, linear discriminant analysis, random forest, naïve Bayes, AdaBoost, extreme gradient boosting) and four deep learning classifiers (convolutional neural network, recurrent neural network, long short term memory, Bidirectional long short term memory). For the publicly available COVID-19 CT dataset, an accuracy of 88.18% was achieved. In conclusion, our COVID-19 CAD system improves diagnostic accuracy, with future work aimed at enhancing efficiency and expanding to covariant detection and severity assessment.

Keywords: Support vector machine; Flower pollination algorithm; k-nearest neighbor; Coronavirus disease 19; Coronavirus disease 19 computed tomography dataset

1. Introduction

The lungs are a pair of spongy, air-filled organs located on either side of the chest. Each lung is roughly cone-shaped, with its base resting on the diaphragm.[1] The lung has two parts: the right lung, which is larger and has three lobes (superior, middle, and inferior), and the left lung, which is smaller and divided into superior and inferior lobes.[2] Lung diseases are conditions that obstruct normal lung function.[3] These include a variety of conditions such as chronic obstructive pulmonary disease, pneumonia, asthma, acute bronchitis, Coronavirus disease 19 (COVID-19), pulmonary edema, idiopathic pulmonary fibrosis, sarcoidosis, pleural effusion, pleurisy, bronchiectasis, cystic fibrosis, lymphangioleiomyomatosis, interstitial lung diseases, lung cancer, tuberculosis, acute respiratory distress syndrome (ARDS), and coccidioidomycosis, and so on.[4] In this research, early detection of COVID-19 is the key focus.

COVID-19 is an infectious disease caused by severe acute respiratory syndrome-coronavirus-2 (SARS-CoV-2), which transmits between humans through physical contact, respiratory droplets, and aerosols. The disease is identified by lung lesions detected through imaging techniques, such as X-rays and computed tomography (CT) scans. CT scans make it easier to assess the presence and severity of COVID-19 nodules. Moreover, considering the structural or anatomical details of the lung that are essential for the detection analysis, CT imaging outperforms X-ray radiography in providing knowledge on these.[5] The typical signs of lung lesions, such as ground glass opacity (GGO) in the early stages and consolidation in the later stages, could be observed from CT slices.[6] Studies have reported that radiological imaging, such as CT and X-rays, may be helpful in supporting the early screening of COVID-19.[7,8] Although real-time polymerase chain reaction (RT-PCR) is considered the gold standard for diagnosing COVID-19, recent advancements in medical imaging have significantly improved the diagnosis and quantification of various diseases. Using RT-PCR results as a reference, a study of 1,014 patients in Wuhan, China, achieved an accuracy of 0.68, a sensitivity of 0.97, and a specificity of 0.25 for CT slices indicating COVID-19 infection.[9]

Segmentation is the process of partitioning lung tissues with accurate boundaries from CT slices by eliminating surrounding anatomical structures, such as bones and fat tissues.[10] The objective of segmentation is to extract regions of interest (ROIs) within the lung region to differentiate abnormality from anatomical background. There are 10 different segmentation techniques for lung imaging,[11] including: thresholding,[12-15] region growing method,[16-18] watershed algorithm,[19-21] active contour model,[22,23] clustering techniques,[24-26] level set techniques,[27,28] graph cut techniques,[29,30] genetic algorithms,[31,32] artificial intelligence-based segmentation,[33,34] and hybrid algorithm.[35] In our work, an optimal thresholding approach has been used to locate a value acceptable for segmenting the lung CT slice.

Computer-aided diagnosis (CAD) systems play an important role in assisting physicians in the process of clinical decision-making.[36] In the domain of diagnostic radiology, the CAD system is designed to diagnose abnormalities in images created by imaging modalities. The imaging modalities are X-rays, CT, high-resolution CT, positron emission tomography, single-photon emission CT, and magnetic resonance imaging. A CAD system helps medical professionals by simplifying the process of interpreting numerous images created by different types of imaging, where manual involvement is time-consuming.[37] In the domain of diagnosing pulmonary disorders, the CAD system takes the input image obtained from the imaging modalities, employs computational techniques to locate suspected abnormalities present in the image, and leads to a precise diagnosis. Techniques such as machine learning (ML), image processing, pattern recognition, and deep learning (DL) are commonly employed to enhance abnormality detection in medical images.[38]

In this research, we developed a CAD system to detect the presence or absence of COVID-19. First, an adaptive Wiener filter was used to eliminate the additive noises. Then, optimal thresholding was used to segment the lungs, and relevant features were extracted. To select the optimal feature set, a bio-inspired wrapper-based flower pollination technique was employed, using the accuracy of the k-NN classifier as the fitness function. The support vector machine (SVM) classifier was then trained using the selected optimal subset of features.

This framework can be generalized for applications in biomedical lung imaging diagnosis. This manuscript is structured as follows: **S**ection 2 discusses the relevant literature; Section 3 outlines the system's methodology; Section 4 summarizes the dataset, compares classifiers, evaluates other state-of-the-art approaches, and presents the experimental findings; Section 5 offers conclusions and recommendations for future work.

2. Related works

2.1. Segmentation techniques for CAD to detect COVID-19

Segmentation is an essential step in image processing and analysis for the assessment and quantification of COVID-19. It delineates the ROIs, namely, lung, lobes, lesions or infected regions, and bronchopulmonary

segments, in the chest X-ray or CT slices. Segmented regions could be further used to extract features for diagnosis and other applications. This subsection summarizes the related segmentation works in COVID-19.

Khin *et al.*[39] have proposed a segmentation algorithm to detect COVID-19 in chest CT slices using Deeplab v3+. The dataset used was the COVID-19 radiography database, which contains a total of 15,153 images, including 10,192 normal images, 3,616 COVID-19 images, and 1,345 pneumonia images. Since the dataset was highly imbalanced, five different approaches were employed. The ensemble of convolutional neural network (CNN) with image augmentation achieved an accuracy of 99.23%.

Venkatesan *et al.*[40] have introduced an automated image processing scheme to extract the COVID-19 lesions from lung CT scan images. In their work, the firefly algorithm and Shannon Entropy-based multi-threshold were used to enhance the pneumonia lesions, followed by Markov-Random-Field segmentation to extract the lesions with better accuracy. The dataset was obtained from the COVID-19 database, which includes 100 images for training and 45 images for testing. The proposed scheme was tested and validated using a class of COVID-19 CT images, achieving a mean accuracy >92% in lesion segmentation.

Chandra[41] has demonstrated a segmentation approach using the Cuckoo search algorithm with Otsu's image thresholding for the extraction of COVID-19 pneumonia infection. The proposed approach used Otsu's/Kapur to enhance the value with a threshold of three and employed Level Set techniques to extract ROIs. The dataset included COVID-19 images from 20 patients, and the approach achieved a segmentation accuracy of 97.62.

Mohammed *et al.*[42] have proposed a CAD system for the diagnosis of COVID-19 disease from chest X-ray images. This system can be used to differentiate COVID-19 from other viral pneumonia-like Middle East respiratory syndrome, SARS, and ARDS. Segmentation was performed using Li's[42] method, followed by the application of Law's[42] masks to enhance secondary details in the segmented chest images. Texture features were then extracted using the gray-level co-occurrence matrix (GLCM). The obtained feature vectors were used to build SVM ensemble models. Then, the choices of ensemble classifiers were put together using a weighted voting method. The proposed CAD system achieved an accuracy of 98.04%.

Bhargava *et al.*[43] have introduced an automatic detection system for the diagnosis of COVID-19 from CXR and chest CT slices. Segmentation was done using the FCM algorithm. Four types of features, namely, histograms of gradients, textural, statistical, and discrete wavelet transforms, were extracted using the method of principal component analysis. In the classification, k-NN, sparse representation classifier (SRC), artificial neural network (ANN), and SVM classifiers were used for normal, pneumonia, and COVID-19 classifications. Nine different datasets collected from various sources were examined. The accuracies achieved were 91.70%, 94.40%, 96.16%, and 99.14% by k-NN, SRC, ANN, and SVM, respectively, for COVID-19 diagnosis.

Shankar *et al.*[44] have suggested a CAD system for diagnosing COVID-19 using chest X-ray images. Initially, the Wiener filter was used to pre-process images. The fusion-based feature extraction method was subsequently carried out using GLCM, gray level run length matrix, and local binary patterns. The ideal feature subset was then determined using the Salp swarm algorithm. The images were divided as infected or healthy using an ANN. The obtained outcomes outperformed state-of-the-art techniques. The proposed CAD model's experimental results showed 95.1% and 95.65% accuracy for binary and multiple classes, respectively.

Kadry *et al.*[45] have proposed a classification technique using a machine learning system (MLS) to classify the CT slices as healthy or affected by COVID-19. The MLS includes five steps, namely, tri-level thresholding, segmentation of the image, feature extraction, feature ranking, implementation of serial fusion, and classifier implementation and validation. This proposed system was tested with 500 images, which includes 250 normal and 250 COVID-19-affected images obtained from benchmark datasets (Table 1). The proposed MLS achieved an accuracy of 89.80%.

2.2. CAD system to detect COVID-19 using supervised and un-supervised techniques

Wu *et al.*[46] have proposed a classification system using a random forest (RF) classifier for the diagnosis of COVID-19 disease. The dataset description has been given in Table 1. In the proposed system, 11 key features were selected from 49 features. The model was trained with 11 key features and achieved an accuracy of 96.95%.

Banerjee *et al.*[47] have suggested a binary classification model utilizing ANN, Logistic regression (LR), and LASSO Elastic Net Regularized Generalized Linear Models. The dataset comprised 598 full blood count results obtained from COVID-19 patients. The model with LR achieved an accuracy of 87% for the diagnosis of COVID-19 disease.

Moutaz *et al.*[48] have demonstrated an artificial intelligence technique based on deep CNN to detect COVID-19 disease. The dataset was obtained from the Kaggle dataset, which has 128 images, including 28 healthy

Table 1. Comparison of computer-aided diagnosis systems for diagnosing COVID-19 and the dataset used

References	Contribution	Dataset used	Number of images	Balanced/Unbalanced	Techniques used
Khin *et al.*[39]	Deeplab v3+for diagnosing COVID-19 achieved an accuracy of 99.23%	COVID-19 radiography database	15,153 images, including 10,192 normal, 3,616 COVID-19, and 1,345 pneumonia	Highly unbalanced	Weighted loss, image augmentation, undersampling, oversampling, and hybrid resampling
Kadry *et al.*[45]	Machine learning system using SVM with an accuracy of 89.80%	LIDC-IRDI dataset, RIDER-TCIA dataset, and COVID-19 images from the Radiopedia database	500 images, including 250 normal and 250 COVID-19	Balanced	Balanced dataset from benchmark datasets
Wu *et al.*[46]	Random forest classifier with 11 key features achieved an accuracy of 96.95%	Real-time dataset	253 samples	Balanced	-
Banerjee *et al.*[47]	LR achieved an accuracy of 87%	COVID-19 Data sharing/BR initiative	5644 images, in which 598 samples are considered	Unbalanced	Tested separated for specificity and sensitivity
Moutaz *et al.*[48]	VGG16 with an accuracy of 94.80%	Kaggle dataset	128 images, including 28 healthy and 70 COVID-19 images	Balanced	Data augmentation
Najjar *et al.*[50]	Feature extraction using GLCM and classification using k-NN and SVM classifier. k-NN classifier achieved 99.96%	COVID-19 radiography database	2,399 chest X-ray images, which include 1,577 normal and 822 COVID-19 images	Unbalanced	Using the performance metrics
Maryam *et al.*[51]	Ensemble learning model	COVID-19 Data Sharing/BR initiative	5644 images	Unbalanced	Ensemble model using performance metrics
Atta *et al.*[52]	CSDC-SVM model with an accuracy of 98%	Real-time	547 samples that are classified through the SVM K-fold cross-validation method	Unbalanced	The area under the receiver operating characteristics curve, G-mean, and the F1-score
Tongxue *et al.*[54]	U-Net-based segmentation network using attention mechanism achieved a specificity of 99.3%	Italian Society of Medical and Interventional Radiology: COVID-19 CT segmentation dataset	Dataset 1: 100 axial CT slices from 60 patients with COVID-19 with pleural effusion Dataset 2: 373 slices of COVID-19 with consolidation	Unbalanced	Because of the small data in both datasets, they combine the two datasets as the final training dataset
Mobiny *et al.*[55]	Detail-oriented capsule network architecture with 83.2% accuracy	COVID-19 CT dataset	746 images, which includes 349 COVID-19 and 397 non-COVID-19 images	Unbalanced	Image-to-Image (pi×2pix) conditional GAN architecture augmentation
Hasoon *et al.*[56]	LBP-k-NN, HOG-k-NN, Haralick-k-NN, LBP-SVM, HOG-SVM, and Haralick-SVM. Achieved an accuracy of 89.2% and 98.66%	Github repository	5,000 normal and pneumonia COVID-19 images	Unbalanced	Feature-based balancing

Abbreviations: CSDC-SVM: Cloud-based smart detection algorithm using support vector machine; CT: Computed tomography; GLCM: Gray level co-occurrence matrix; HOG-KNN: Histogram of gradients k nearest neighbor; KNN: K-nearest neighbor; LBP-KNN: Local binary pattern k nearest neighbor; RF: Random forest; SVM: Support vector machine; LR: Logistic regression.

and 70 COVID-19 images. The forecasting methods, namely, the prophet algorithm, auto-regressive algorithm, integrating moving average model, and long short-term memory (LSTM), were used to predict the number of

COVID-19 confirmations. The proposed system achieved an accuracy of 94.80%.

Feng *et al.*[49] have proposed a predictive model using four classifiers, namely LR with LASSO, LR with ridge regularization, decision tree, and adaptive boosting (AB) algorithms, for the early detection of COVID-19 disease. The strength of this proposed model lies in the 46-feature selection. Based on the results, the LR with the LASSO classifier selected only 18 features and achieved an accuracy of 93.80%.

Najjar *et al.*[50] have presented a cutting-edge solution for classifying COVID-19 from chest radiography slices using the SVM and k-NN classifiers. The dataset was obtained from the COVID-19 radiography database, which included 1577 normal and 822 COVID-19 images. The proposed work produced five matrices, namely, GLCM1, GLCM2, GLCM3, GLCM4, and GLCMA, and achieved an accuracy of (95.83 – 97.07%), (95.21 – 97.03%), (95.52 – 96.87%), (95.57 – 97.24%), and (95.94 – 96.87%) with SVM and k-NN classifiers, respectively.

Maryam *et al.*[51] have proposed an ensemble learning model for the diagnosis of COVID-19 from a blood routine test. This proposed model was trained and evaluated using a publicly available dataset in Brazil, which includes 5644 images. This proposed model achieved an accuracy of 99.88% in diagnosing COVID-19 disease.

Atta *et al.*[52] have demonstrated a supervised approach named the cloud-based smart detection algorithm using SVM (CSDC-SVM), tested with 5, 10, 15, and 20 cross-fold validation. The dataset included 547 samples, which were classified using the SVM K-fold cross-validation method. The proposed CSDC-SVM model classifies COVID-19 into four categories, namely, negative, mild, moderate, and severe. The virus can be classified as negative, mild, moderate, or severe, indicating its presence at various levels. The proposed system with CSDC-SVM achieved an accuracy of 98.4% with a 15-fold cross-validation strategy.

The results presented in Table 1 show that to identify the COVID-19 infection more accurately, image-aided diagnosis is important. In addition, by providing the necessary details about the patient who had been admitted with a COVID-19 infection, this system could significantly reduce the pulmonologist's diagnostic burden. The infection rate may be precisely identified when there is an image processing system that is properly developed and implemented.

The aforementioned results were obtained by reviewing this pertinent literature. To begin, ROIs are along lung boundaries; segmenting the lung tissues is essential. Second, training the CAD system with the best ROI features promotes classification performance. Third, a wrapper-based feature selection strategy that uses bio-inspired algorithms is more robust and performs better in a variety of optimization challenges when compared to conventional approaches to feature selection.

3. Methods

The proposed CAD system illustrated in Figure 1 consists of five main steps: (i) segmentation with image enhancement, optimal thresholding, cavity filling, and background removal process; (ii) ROI extraction; (iii) GLCM feature extraction; (iv) selection of features; and (v) classification by building a set of SVM models to classify the chest image into either positive (COVID-19) or negative type (non-COVID-19).

3.1. Segmentation

The objective of segmentation is to partition lung tissues from each lung CT slice. To eliminate additive noise and improve edge sharpness, a Laplacian filter is applied. Next, lung parenchyma is partitioned using an optimal

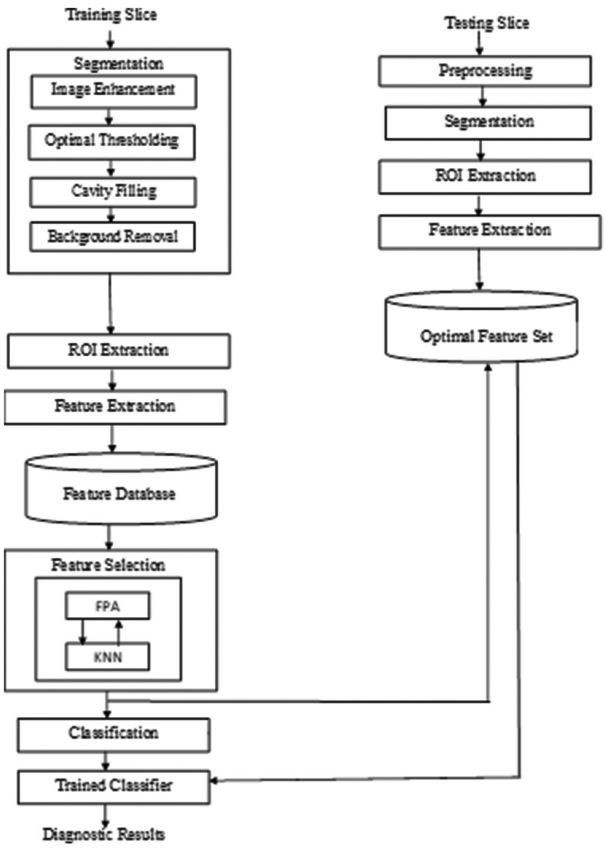

Figure 1. The proposed COVID-19 CAD system. Image created using MS Word application.
Abbreviations: CT: Computed tomography; FPA: Flower pollination algorithm; k-NN: k-nearest neighbor; ROI: Region of interest.

thresholding technique. Morphological techniques are then used to eliminate the background and holes of the lung. An adaptive Wiener filter is used to remove noises from the input CT slices. After removing the noises, optimal thresholding is applied to segment the left and right lung tissues. Optimal thresholding is a method that divides the histogram into two parts to minimize variance within the same class while maximizing separation between different classes.[53] There are two distinct types of pixels that can be seen in a CT slice of the lung, namely, high- and low-intensity pixels. Since their intensity distributions differ, an optimum thresholding approach is used to locate a value acceptable for segregating the lung slice.[57] In the cavity-filling process, the appearance of airways, small holes, or cavities in the binary slice, which represent pathogenic regions, is addressed. Morphological techniques are used to fill these cavities with intensity levels similar to those of neighboring pixels. Pixels with lower intensity values outside the chest cavity are classified as background pixels.[58] In addition, morphological operations are employed to eliminate all connected components smaller than 1000 pixels in the area from the slice.

3.2. Region of interest extraction

The ROIs considered for the COVID-19 CAD model are crazy paving, interlobular septal-thickening, patchy GGO, bilateral GGO, traction bronchiectasis, sub-pleural GGO, peripheral GGO, consolidation, bronchovascular thickening in the lesion, and GGO with consolidation. The ROIs with the pixel intensity score on the scale from 125 to 255 were extracted. The pixel intensity scores <125 are not considered. Each ROI was annotated and labeled by an expert radiologist. Then, Class Label 1 was given to the ROI diagnosed with the presence of COVID-19, and Class Label 2 was given to the ROI diagnosed with the absence of COVID-19.

3.3. Feature extraction using gray level co-occurence matrix

The GLCM-based features were extracted to differentiate between the CT slices with positive and negative cases of COVID-19. The GLCM matrix uses pixel pairs of a joint probability distribution (JPD). The JPD between pixel pairs is calculated by using angle "θ" and the distance "d." The value will be the $(i,j)^{th}$ entry in the GLCM matrix.[42] The features that are extracted from each ROI, as well as the class label that is associated with each ROI, are saved as a feature vector in a database that stores features. From the class labeled ROI, geometrical and textural features were extracted. In our work, 12 geometrical features and 17 textural features, along with four orientations (0°, 45°, 90°, and 135°) were extracted. Then, the feature vector pertaining to each ROI from the 80 extracted features (12 geometric features and 68 texture features) along with the class label were stored in the feature database.[59-61] The features that were extracted from each ROI are outlined in Table 2.

3.4. Feature selection

The goal of this step is to select the optimal feature subset from the extracted features to improve the classifier's predicted performance. The subset of features has been chosen using the Wrapper technique, which combines the Flower pollination algorithm (FPA) and the accuracy of the k-NN classifier as the fitness function.

Table 2. Outline of features extracted from each region of interest

Geometric features
1. Euler number
2. Major axis length
3. Eccentricity
4. Orientation
5. Convex area
6. Filled area
7. Solidity
8. Extent
9. Perimeter
10. Equivalent diameter
11. Minor axis length
12. Area
Texture features (0°, 45°, 90°, and 135°)
1. Sum of squares variance
2. Autocorrelation
3. Cluster prominence
4. Cluster shade
5. Information measure of correlation
6. Energy
7. Correlation
8. Difference variance
9. Dissimilarity
10. Difference entropy
11. Entropy
12. Homogeneity
13. Contrast
14. Inverse difference
15. Maximum probability
16. Sum average
17. Sum entropy

3.4.1. FPA or flower algorithm

Yang[64] introduced the FPA in 2012, inspired by the way blossoming plants attract pollinators.[62-65]

A flower is an angiosperm's bisexual reproductive shoot, with reproductive organs encircled by whorls of sterility organs. Angiosperms are distinguished by a number of characteristics, of which the flower is only one. Sepals, stamens, petals, and carpels are the four whorls that make up the flower. The sepals, which morphologically resemble a whorl of leaves, are the first whorl of the flower. The sepals, which are usually green in color, are formed as lateral extensions from the floral meristem. The petals, which are morphologically identical to leaves, make up the second whorl. The third outer whorl is the sexual organ named the stamens. Stamens and leaves share traits in common in the presence of chlorophyll and their growth form, which is elongation in a single plane (with little or no laminar growth). Female sexual parts, known as carpels, are found in the fourth and innermost whorl of the flower.

Flowers are labeled as bisexual or unisexual depending on the presence of male (gametes) as well as female (gametophyte) reproductive organs. Bisexual or hermaphrodite flowers have both male and female sexual parts. Unisexual flowers have either male sexual parts or female sexual parts. The main function of the sexual organ is to prepare seeds and fruits. The first and foremost step is achieving seeds and fruits, which is possible through pollination. Pollination is divided into two types: self-pollination and cross-pollination. Self-pollination occurs when pollen from one flower's anther is transferred to the stigma of the same bloom (autogamy) or to the stigma of another bloom on the same plant (geitonogamy). Cross-pollination occurs when pollen travels from the anthers in one individual's flower to the stigma in another. Because plants are immobile, pollen movement from plant to plant requires the use of a pollen vector, which can be abiotic or biotic.

Abiotic pollen vectors are primarily caused by water and wind. In wind pollination, the stamen filaments of wind-pollinated flowers are typically long, exposing the locules to the wind and causing an aeroelastic release of pollen as an energy that is transferred from the wind to the stamen through the long filament. Water pollination, also known as hydrophily, is a rather unusual method of gamete transfer used by a few grass and waterweed species. Most hydrophilous species release pollen below the water's surface, where it is passively conveyed by currents to female reproductive structures. Many maritime plants use this process of water pollination.

Biotic pollen vectors, on the other hand, comprise a wide range of species, particularly insects, but also birds, bats, and a small number of other vertebrates that observe flowers as a source of food. Insects are the most common biotic pollen transporter. They receive a complimentary sample of nectar, a sugar solution containing varying amounts of various different sugars as well as other nutrients, and pollen, which is high in amino acids. When animals and insects collect this food, they unintentionally touch the flower's reproductive organs, transferring pollen from stamens to their bodies and from their bodies to stigmatic surfaces. Many flowers have structural elements that promote this unintended interaction.

Pollen grains settle on the stigma's surface and germinate, forming pollen tubes. One of the pollen tubes continues to develop downward. This tube transports male gametes to the ovary. After reaching the ovule, the male gametes get released from the pollen tube and mate with the egg cell. The process of merging the male and female gametes is called fertilization. After fertilization, the ovary becomes larger and develops into a fruit. The fertilized ovule, which results from the fusion of gametes, matures into a seed. The fertilized gamete is referred to as a fertilized ovule. Other aspects of the flower, such as the sepals and the petals, will detach themselves after fertilization has taken place in the bloom. The developed ovary of the flower serves as the primary component of the fruit. The FPA parameters and their respective values are outlined in Table 3.

Output: Feature vectors.

The FPA algorithm has been outlined as follows:

Input: Feature vectors.

Process:

Step 1: Generate a random initial population that is evaluated to determine the current optimal solution.

Initialize the size of the population n, MaxGeneration and p.

Step 2: Initialize the population of n pollen gametes $x = (x_1, x_2, \ldots x_n)$ with random solutions. By calculating the fitness value of each pollen gamete in the population using the k-NN classifier, where the k-NN classifier's accuracy is regarded as the fitness function, the best solution g_* in the initial population is found.

Table 3. Parameters outlined in the flower pollination algorithm

Parameter	Value	Definition
n	10	Initial population
MaxGeneration	100	Maximum no. of iterations
p	(0, 1)	Switch probability
λ	1.5	Control parameter
γ	0.01	Scaling factor

Step 3: Determine the type of pollination based on a predetermined probability p. Generate a random number $r \in [0,1]$, and if $r < p$, where p is the switching probability, then global pollination and flower constancy take place, as described by Equation I:

$$x_i^{t+1} = x_i^t + \gamma L\left(g^* - x_i^t\right) \quad \text{(I)}$$

Where x_i^t denotes the solution of i at iteration t, γ is a scaling factor, and g_* is the current optimal solution at iteration . The parameter L is the pollination strength, in which essentially a step size is drawn from Levy flight, which is given by Equation II:

$$L \sim \frac{\lambda * \Gamma(\lambda) * \sin(\frac{\pi\lambda}{2})}{\pi} * \frac{1}{S^{1+\lambda}}, (s \gg 0) \quad \text{(II)}$$

Where, $\Gamma(\lambda)$ denotes the standard gamma function, and this distribution is for $S > 0$. λ is the tail amplitude of the distribution's control parameter. Commonly, it is recommended to use $\lambda = 1.5$, which is followed in all simulations.

Step 4: Otherwise, if $r > p$, then the local pollination and the flower constancy are performed, as described by Equation III:

$$x_i^{t+1} = x_i^t + \varepsilon\left(x_j^t - x_k^t\right) \quad \text{(III)}$$

Where, x_j^t and x_k^t are pollens from other flowers of the same plant species, with j and k chosen at random from all the solutions. $\varepsilon \in [0,1]$ is a random number.

Step 5: Evaluate each new solution x_i^{t+1} in the population and update the population according to their fitness value.

Step 6: Calculate the current best solution g^* by ranking the solution.

Step 7: Repeat Steps 3 through 6 until MaxGeneration is reached or until convergence is achieved.

In this feature selection Step 24 features have been selected namely Area, Minor Axis Length, Convex Area, Eccentricity, Cluster Prominence 2, Cluster Prominence 3, Contrast 1, Contrast 3, Correlation 2, Correlation 3, Difference Variance 4, Dissimilarity 2, Dissimilarity 4, Energy 1, Entropy 1, Entropy 2, Entropy 4, Homogeneity 4, Information Measure of Correlation 1, Information Measure of Correlation 4, Inverse Difference 1, Sum Average 1, Sum entropy 1, Sum of Squares Variance 4.

3.5. Classification

The SVM algorithm was employed to train the optimal feature subset. A SVM is a supervised learning algorithm that uses hyperplanes to separate different classes. The distance of a feature vector from these hyperplanes indicates how likely it is to belong to a specific class.[66,67] In ML, SVM is a model that classifies and predicts outcomes based on training data. The main goal of SVM is to identify the best hyperplane that divides two classes within a feature set. In other words, the SVM training method builds a model that assigns new examples into one of the two classes based on a set of training examples for binary classification. A SVM assigns training samples in a spatial arrangement that maximizes the separation between the two classes. When new samples are introduced, they are similarly positioned in this space, and their class is forecasted based on which side of the hyperplane they fall. The SVM classifier was trained using the set of FPA-selected features. Then, the performance of the trained SVM classifier was validated using the test dataset.[67]

4. Results

This section includes a description of the real-time dataset and public dataset used in this research, as well as performance evaluation, comparison results of ML and DL classifiers, and experimental results.

4.1. Dataset outline

The research utilized two datasets: a real-time dataset collected from Bharat Scan Centre, Chennai, India, and a COVID-19 CT dataset obtained from the GitHub repository. In the real-time dataset, CT slices were labeled by an expert radiologist as either "normal" or "COVID-19." This dataset includes images from 41 individuals, comprising 26 with COVID-19 and 15 with healthy lungs. Among the COVID-19 patients, 19 exhibited mild severity while seven had moderate severity; the cohort included 17 females and nine males. The ages of the COVID-19 patients ranged from 23 to 49 years, with an average of 36 years. The images in the dataset have a pixel size of 512 × 512 and are in jpg format. The nodule size ranges from 3 to 30 mm, with lesions primarily located in the sub-pleural and posterior respiratory zones. The ROIs were patchy GGO, bilateral GGO, subpleural GGO, peripheral GGO, broncho-vascular thickening, traction bronchiectasis, consolidations, and GGO with consolidations.

The datasets have been divided into training and testing datasets, with training datasets comprising 80% of the total and testing datasets 20%. To preserve privacy, we have masked all personal information from CT slices. Each ROI has been differentiated based on the opinion of an experienced radiologist. In addition, the radiologist manually identified and described each ROI. Table 4 gives an overview of the real-time dataset.

Table 4. Overview of the experimental dataset

Patient cases	Total no. of patients	Total COVID-19 CT slices considered	ROIs	Training set ROIs	Testing set ROIs
COVID-19	26	342	343	242	101
Normal	15	446	452	394	58
Total	41	788	795	636	159

Abbreviations: CT: Computed tomography; ROI: Region of interest.

For the COVID-19 CT database, a publicly available dataset was utilized to train and test the proposed model. It contains a total of 349 COVID-19 CT images from 216 patients and 463 non-COVID-19 CTs, which have been divided into two classes, namely, COVID-19 and non-COVID-19. The COVID-19 CT dataset was divided into training and testing datasets, with training datasets comprising 80% of the total and testing datasets 20%. A pre-processed version of the dataset is available at https://github.com/UCSD-AI4H/COVID-CT.

4.2. Performance evaluation

The aim of this work is to decrease the false negative and false positive values, that is, to increase the sensitivity and specificity, respectively. However, there is often a tradeoff between sensitivity and specificity; as one increases the other decreases. In the proposed research, we obtained inferences from the radiologist. He reviewed the model and provided feedback, suggesting that although it works well, more CT slices should be included so that it may be used to diagnose different lung diseases. Figures 2 and 3 display the effectiveness of the CAD system's implementation for patients with and without COVID-19. The algorithm's optimization performance was compared in terms of accuracy, precision, recall or sensitivity, and specificity, with results obtained using Equations IV–VII:

$$\text{Accuracy} = \frac{a+d}{a+b+c+d} \quad \text{(IV)}$$

$$\text{Precision} = \frac{a}{a+b} \quad \text{(V)}$$

$$\text{Sensitivity} = \frac{a}{a+c} \quad \text{(VI)}$$

$$\text{Specificity} = \frac{d}{b+d} \quad \text{(VII)}$$

Where a, b, c, and d denote actual positives, predicted positives, predicted negatives, and predicted positives, respectively. The confusion matrix obtained for FPA is shown below in Table 5.

The extraction of COVID-19 lesions from a chest CT slice demonstrating the presence of the COVID-19 disease is depicted in Figure 4A-D. Figure 4A displays the reference chest CT slice. Figure 4B and C illustrate the segmentation and feature extraction processes necessary for effectively isolating the nodules. Figure 4D displays the peripheral GGO lesion that was excised, indicating the presence of COVID-19.

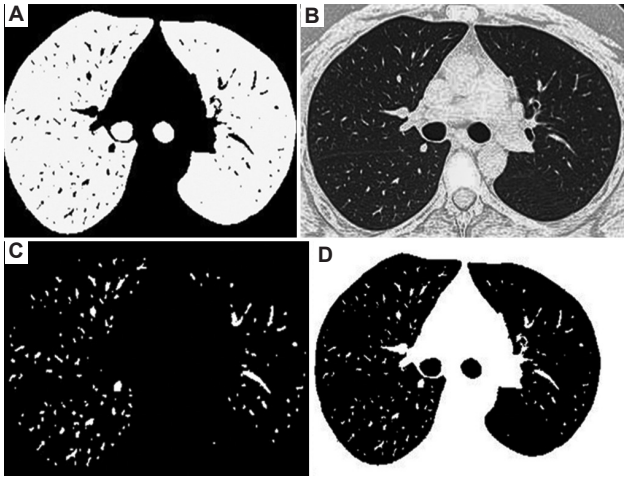

Figure 2. Experimental images of a normal lung CT slice. (A) Non-COVID-19 input CT slice. (B) Segmented image. (C) Extracted ROI. (D) Non-COVID-19 nodules. These images were generated using Python
Abbreviations: CT: Computed tomography; ROI: Region of interest.

Figures 2A-D depict the steps involved in the extraction of ROIs that indicate the absence of COVID-19 disease. The input CT slice of the lung is displayed in Figure 2A. The output image of various steps involved in extracting the nodules is shown in Figures 2B and C. The nodules extracted are shown in Figure 2D.

The CAD system that utilizes FPA for feature selection with 100 iterations produced a greater accuracy of 91.30% for the real-time dataset and 88.18% for the COVID-19 CT dataset. The performance comparison using the real-time and COVID-19 CT datasets is outlined in Table 6.

4.3. Comparison with machine learning and DL classifiers

The proposed CAD system was compared against seven traditional ML classifiers and four DL classifiers. The ML classifiers included radial basis function SVM, k-NN,

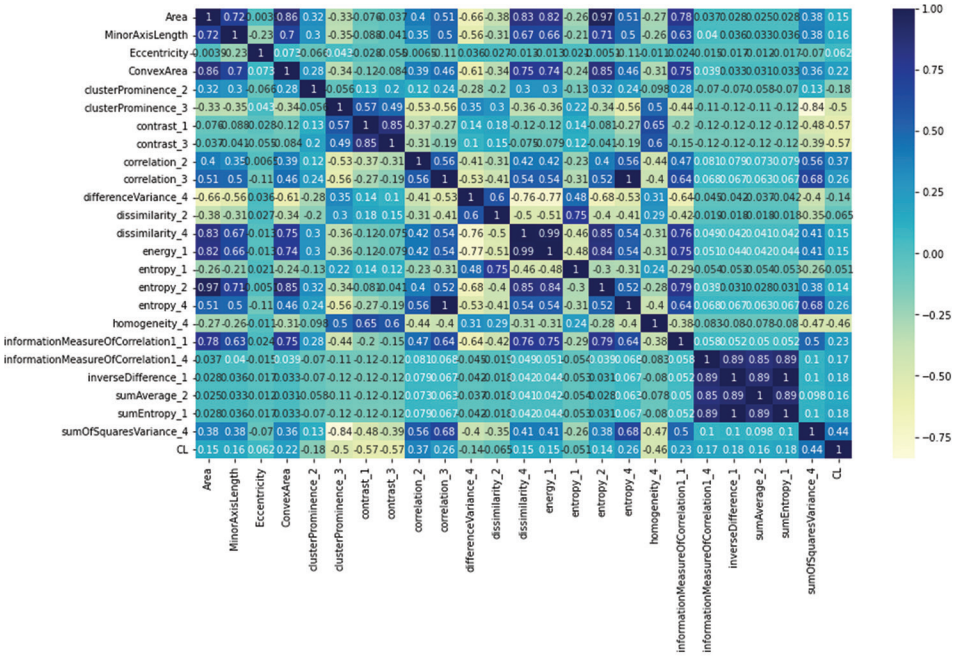

Figure 3. Kendal's rank correlation map. Output generated using the Python application

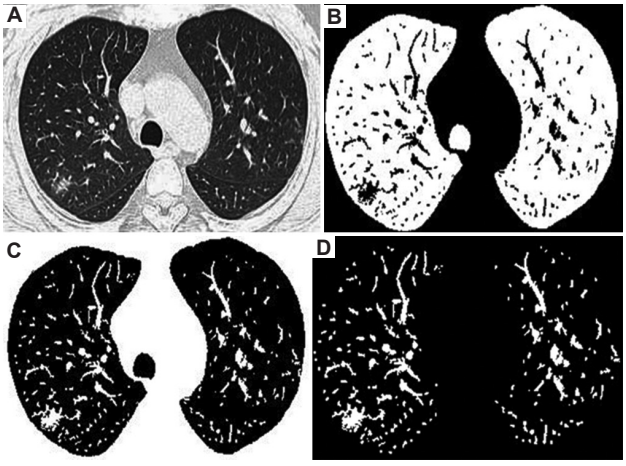

Figure 4. Experimental images obtained for COVID-19 CT slices. (A) COVID-19 input CT slice. (B) Segmented image. (C) Extracted ROI. (D) COVID-19 nodules. These images were generated using Python
Abbreviations: CT: Computed tomography; ROI: Region of interest.

linear discriminant analysis, RF, naïve bias, extreme gradient boosting, and AB. The four DL classifiers used for comparison were CNN, recurrent neural network, LSTM, and bidirectional LSTM, respectively. Our system outperformed these ML classifiers with an accuracy of 91.30%. For each model, average (± standard deviation) performance was reported over 30 iterations. The comparison of ML and DL classifiers in terms of accuracy, precision, recall, and specificity, along with mean and standard deviation values, is presented in Tables 7 and 8.

Table 5. Generated confusion matrix

Actual/predicted	Predicted positive	Predicted negative
Actual positive	94	9
Actual negative	5	51

Table 6. Performance comparison using real-time and COVID-19 dataset

Performance metrics average	Real-time dataset	COVID-19 CT dataset
Accuracy (*M*±*SD*)	0.9130±0.0177	0.8818±0.0180
Precision (*M*±*SD*)	0.8989±0.0324	0.9192±0.0280
Recall (*M*±*SD*)	0.8003±0.0340	0.8956±0.0305
Specificity (*M*±*SD*)	0.9374±0.0218	0.8574±0.0538
F1 score (*M*±*SD*)	0.9302±0.0217	0.9065±0.0140
Selected features	24	22

Abbreviation: CT: Computed tomography

4.4. Comparison with other state-of-the-art approaches using the COVID-19 CT dataset

Our proposed CAD system metrics using the COVID-19 CT dataset obtained from the GitHub repository were compared with other state-of-the-art approaches[55,69-72] for diagnosing COVID-19 disease (Table 9). A maximum accuracy of 89.36% in this comparison was achieved by Ali and Assadi[71], whereas our CAD system using the COVID-19 CT dataset produced an accuracy of 88.18%.

Table 7. Machine learning classifier comparison

Classifier/ performance metrics	RBF-SVM	k-NN	LDA	RF	NB	EB	AB	Our proposed system using real-time dataset
Accuracy (*M*±*SD*)	0.6329±0.0387	0.8572±0.0243	0.8706±0.0210	0.8996±0.0180	0.7541±0.0403	0.9044±0.0232	0.8753±0.0220	0.9130±0.0177
Precision (*M*±*SD*)	0.9189±0.0660	0.8779±0.0481	0.8861±0.0341	0.9135±0.0337	0.9093±0.0551	0.9113±0.0388	0.8697±0.0404	0.8989±0.0324
Recall (*M*±*SD*)	0.1815±0.0515	0.7855±0.0495	0.8095±0.0435	0.8524±0.0432	0.4883±0.0722	0.8673±0.0403	0.8434±0.0381	0.8003±0.0340
Specificity (*M*±*SD*)	0.9867±0.0111	0.9149±0.0320	0.8095±0.0435	0.9373±0.0235	0.9625±0.0230	0.9342±0.0281	0.9009±0.0321	0.9302±0.0217

Abbreviations: AB: AdaBoost; EB: Extreme boosting; k-NN: k-nearest neighbor; LDA: Linear discriminant analysis; NB: Naïve bias; RBF-SVM: Radial basis function-support vector machine; RF: Random forest.

Table 8. Deep learning classifier comparison

Classifiers/ performance metrics	CNN (%)	RNN (%)	LSTM (%)	BLSTM (%)
Training accuracy	89.15	84.74	80.66	83.64
Testing accuracy	89.31	85.53	83.01	83.67
Training precision	88.54	81.29	80.57	81.27
Testing precision	84.81	84.50	83.58	81.94
Training recall	85.61	83.39	71.95	80.07
Testing recall	93.05	82.33	77.77	81.94
Training specificity	87.05	82.33	76.02	80.67
Testing specificity	88.74	83.91	80.57	81.94

Abbreviations: CNN: Convolutional neural network; BLSTM: Bidirectional LSTM; LSTM: Long short-term memory; RNN: Recurrent neural network.

Table 9. Comparison of the proposed CAD system with state-of-the-art approaches for the COVID-19 CT dataset

State-of-the-art approaches	Accuracy (%)	Precision (%)	Recall (%)	Specificity (%)	F1-score (%)
Mobiny *et al.* using Inception V3.[55]	85.3	84.4	74	85.3	78.1
Mobiny *et al.* using DenseNet 121.[55]	82.5	81.5	79.4	83.9	80.1
Xingyi *et al.* using DenseNet-169.[69]	79.5	-	-	-	76
Polsinelli *et al.*[70]	85.03	85.01	81.44	88.23	83.98
Xingyi *et al.* using ResNet-50[69]	77.4%	-	-	-	74.6
Ali and Assadi[71]	89.26	-	-	-	89.18
Pedro *et al.*[72]	87.6	-	-	-	86.19
Our proposed CAD system	88.18	91.92	89.56	85.74	90.65

Abbreviations: CAD: Computer-aided diagnosis; CT: Computed tomography.

Our proposed system achieved higher precision, recall, and F1 score values, as shown in Table 9.

4.5. Statistical test

The Mann–Whitney U test revealed significant differences between the variables and the dependent variable ($P < 0.001$). The difference is statistically significant ($P < 0.001$). The $P = 0.001$, which is less than the minimum value of 0.05 for significance. Kendal's rank correlation coefficient map examines sample correlation. Kendal's correlation map for the selected attributes in the dataset is given in Figure 3.

5. Conclusion

Our proposed COVID-19 CAD system achieved an accuracy of 91.30% on a real-time dataset and 88.18% accuracy on the COVID-19 CT Public Dataset. Notably, our system demonstrated significant superiority over seven state-of-the-art ML classifiers and four DL classifiers. This shows that our COVID-19 model excels in generating robust and highly discriminative features. The primary goal of our research is to improve classification accuracy and aid physicians in clinical decision-making. Hence, time and space complexity are not the primary interests of this research work. The suggested CAD system exhibited improved accuracy when employing FPA with k-NN and SVM classifiers because it increased the test accuracy and time efficiency. Since the FPA algorithm is larger than some algorithms, more memory is needed. In addition, since this is a classification system, it does not provide information on disease severity.

In the future, this work can be extended to identify the covariants of COVID-19 and the assessment of COVID-19's severity. Optimizing the system's architecture and integrating other feature selection methods are two excellent methods to improve the rapidity of the COVID-19 CAD system. Importantly, for the COVID-19 CAD system to be clinically validated, it should be implemented in real-world settings, such as by training it on a hospital's private

database. This would allow for a thorough evaluation and improvement of its clinical validity.

Acknowledgments

None.

Funding

None.

Conflict of interest

The authors declare no conflicts of interest.

Author contributions

Conceptualization: Betshrine Rachel Jibinsingh, Khanna Nehemiah Harichandran
Formal Analysis: Betshrine Rachel Jibinsingh
Investigation: Betshrine Rachel Jibinsingh, Kabilasri Jayakannan, Rebecca Mercy Victoria Manoharan
Methodology: Betshrine Rachel Jibinsingh, Khanna Nehemiah Harichandran, Anisha Isaac
Writing – original draft: Betshrine Rachel Jibinsingh
Writing – review & editing: Betshrine Rachel Jibinsingh, Khanna Nehemiah Harichandran

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data

Data used in this work are available from the corresponding author on reasonable request.

References

1. Fuller HW. Lea HC, editor. *On Diseases of the Lungs and Air-Passages: Their Pathology, Physical Diagnosis, Symptoms, and Treatment*. Kissingert: Publishing; 1867.
2. Tomashefski JF, Farver CF. Anatomy and histology of the lung. In: *Dail and Hammar's Pulmonary Pathology*. Germany: Springer; 2008. p. 20-48.
3. Demedts M, Wells AU, Anto JM, *et al*. Interstitial lung diseases: An epidemiological overview. *Eur Respir J*. 2001;18(32):2S-16S.
4. Schwarz T, Johnson V. Lungs and bronchi. In: *Veterinary Computed Tomography*. United States: John Wiley and Sons; 2011. p. 261-262.
5. Fan DP, Zhou T, Ji GP, *et al*. Inf-net: Automatic COVID-19 lung infection segmentation from CT images. *IEEE Trans Med Imaging*. 2020;39(8):2626-2637.
6. Carotti M, Salaffi F, Puttini PS, *et al*. Chest CT features of coronavirus disease 2019 (COVID-19) pneumonia: Key points for radiologists. *J Natl Public Health Emerg Collect*. 2020;125:636-646.

   doi: 10.1007/s11547-020-01237-4
7. Fang Y, Zhang H, Xie J, *et al*. Sensitivity of chest CT for COVID-19: Comparison to RT-PCR. *Radiology*. 2020;296(2):E115-E117.

   doi: 10.1148/radiol.2020200432
8. Ng MY, Lee EY, Yang J, *et al*. Imaging profile of the COVID-19 infection: Radiologic findings and literature review. *Radiol Cardiothorac Imaging*. 2020;2(1):e200034.

   doi: 10.1148/ryct.2020200034
9. Ai T, Yang Z, Hou H, *et al*. Correlation of chest CT and RT-PCR testing in COVID-19 in China: A report of 1014 cases. *Thorac Imaging Radiol*. 2020;296(2):E32-E40.

   doi: 10.1148/radiol.2020200642
10. Pal NR, Pal SK. A review on image segmentation techniques. *Pattern Recognit*. 1993;26(9):1277-1294.

    doi: 10.1016/0031-3203(93)90135-J
11. Norouzi A, Rahim MS, Altameem A, *et al*. Medical image segmentation methods, algorithms, and applications. *IETE Tech Rev*. 2014;31(3):199-213.

    doi: 10.1080/02564602.2014.906861
12. Sitanggang S, Sonang S, Yuhandri Y, Setiawan A. Image transformation with lung image thresholding and segmentation method. *J RESTI* (*Rekayasa Sist Teknol Inform*). 2023;7(2):278-285.

    doi: 10.29207/resti.v7i2.4321
13. Yu T, Huang L. An Adaptive Thresholding Method for Automatic Lung Segmentation in CT Images. In: *IEEE AFRICON Conference*; 2009. p. 1-5.
14. Sweetlin JD, Nehemiah HK, Kannan A. Feature selection using ant colony optimization with tandem-run recruitment to diagnose bronchitis from CT scan images. *Comput Methods Programs Biomed*. 2017;145:115-125.

    doi: 10.1016/j.cmpb.2017.04.009
15. Sweetlin JD, Nehemiah HK, Kannan A. Computer aided diagnosis of drug sensitive pulmonary tuberculosis with cavities, consolidations and nodular manifestations on lung CT images. *Int J Bio Inspired Comput*. 2019;13(2):71-85.

    doi: 10.1504/IJBIC.2019.098405
16. Dehmeshki J, Amin H, Valdivieso M, Ye X. Segmentation of pulmonary nodules in thoracic CT scans: A region growing approach. *IEEE Trans Med Imaging*. 2008;27(4):467-480.

    doi: 10.1109/TMI.2007.907555
17. Nabipour S, Khorshidi A, Noorian B. Lung tumor segmentation using improved region growing algorithm.

Nucl Eng Technol. 2020;52(10):2313-2319.

doi: 10.1016/j.net.2020.03.011

18. Prabin A, Veerappan J. Automatic segmentation of lung CT images by CC based region growing. *J Theor Appl Inform Technol*. 2014;68(1):63-69.

19. Avinash S, Manjunath K, Kumar SS. An Improved Image Processing Analysis for the Detection of Lung Cancer Using Gabor Filters and Watershed Segmentation Technique. In: *IEEE International Conference on Inventive Computation Technologies*; 2016.

20. Kumar SL, Swathy M, Sathish S, Sivaraman J, Rajasekar M. Identification of lung cancer cell using watershed segmentation on CT images. *Indian J Sci Technol*. 2016;9:1-4.

doi: 10.17485/ijst/2016/v9i1/85765

21. Shojaii R, Alirezaie J, Babyn P. Automatic Lung Segmentation in CT Images Using Watershed Transform. In: *IEEE International Conference on Image Processing*; 2005.

22. Nithila EE, Kumar SS. Segmentation of lung from CT using various active contour models. *Biomed Signal Process Control*. 2019;47:57-62.

23. Kasinathan G, Jayakumar S, Gandomi AH, *et al*. Automated 3-D lung tumor detection and classification by an active contour model and CNN classifier. *Expert Syst Appl*. 2019;15(134):112-119.

doi: 10.1016/j.eswa.2019.05.041

24. Sangamithraa PB, Govindaraju S. Lung Tumor Detection and Classification Using EK-Mean Clustering. In: *IEEE International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*; 2016. p. 2201-2206.

25. Joon P, Bajaj SB, Jatain A. Segmentation and detection of lung cancer using image processing and clustering techniques. In: *Advanced Computing and Intelligent Engineering*. Germany: Springer Nature; 2019. p. 13-23.

26. Xu M, Qi S, Yue Y, *et al*. Segmentation of lung parenchyma in CT images using CNN trained with the clustering algorithm generated dataset. *Biomed Eng Online*. 2019;18:2.

doi: 10.1186/s12938-018-0619-9

27. Farag AA, Munim HE, Graham JH, Farag AA. A novel approach for lung nodules segmentation in chest CT using level sets. *IEEE Trans Image Process*. 2013;22:5202-5213.

doi: 10.1109/TIP.2013.2282899

28. Swierczynski P, Papież BW, Schnabel JA, Macdonald C. A level-set approach to joint image segmentation and registration with application to CT lung imaging. *Comput Med Imaging Graph*. 2018;65:58-68.

doi: 10.1016/j.compmedimag.2017.06.003

29. Wei J, Deihui X, Zhang B, Wang L, Kopriva I, Chen X. Random walk and graph cut for co-segmentation of lung tumor on PET-CT images. *IEEE Trans Image Process*. 2015;24(12):5854-5867.

doi: 10.1109/TIP.2015.2488902

30. Ali AM, Farag AA. Automatic Lung Segmentation of Volumetric Low-dose CT Scans Using Graph Cuts. In: *International Symposium on Visual Computing*; 2008. p. 258-267.

31. Bhuvaneswari P, Therese BA. Detection of cancer in lung with k-NN classification using genetic algorithm. *Procedia Mater Sci*. 2015;10:433-440.

doi: 10.1016/j.mspro.2015.06.077

32. Filho DC, Silva AO, Paiva AC, Nunes RA, Gattass M. Computer-aided diagnosis system for lung nodules based on computed tomography using shape analysis, a genetic algorithm, and SVM. *Med Biol Eng Comput*. 2017;55:1129-1146.

doi: 10.1007/s11517-016-1577-7

33. Herrmann P, Busana M, Cressoni M, *et al*. Using artificial intelligence for automatic segmentation of CT lung images in acute respiratory distress syndrome. *Front Physiol*. 2021;12:76118.

doi: 10.3389/fphys.2021.676118

34. Shi F, Wang J, Shi J, *et al*. Review of artificial intelligence techniques in imaging data acquisition, segmentation, and diagnosis for COVID-19. *IEEE Rev Biomed Eng*. 2021;14:4-15.

doi: 10.1109/RBME.2020.2987975

35. Van EM, Hoop D, Viergever MA, Prokop M, Ginneken BV. Automatic lung segmentation from thoracic computed tomography scans using a hybrid approach with error detection. *Med Phys*. 2009;36:2934-2947.

doi: 10.1118/1.3147146

36. Doi K. Computer-aided diagnosis in medical imaging: Historical review, current status, and future potential. *Comput Med Imaging Graph*. 2007;31(5):198-211.

doi: 10.1016/j.compmedimag.2007.02.002

37. Choi YJ, Baek JH, Park HS, *et al*. A computer-aided diagnosis system using artificial intelligence for the diagnosis and characterization of thyroid nodules on ultrasound: Initial clinical assessment. *Thyroid*. 2017;27(4):546-552.

doi: 10.1089/thy.2016.0372

38. Isaac A, Nehemiah HK, Dunston DS, Christo VRE, Kannan A. Feature selection using competitive coevolution of bio-inspired algorithms for the diagnosis of pulmonary emphysema. *Biomed Signal Process Control*. 2022;72:103340.

doi: 10.1016/j.bspc.2021.103340

39. Khin Y, Maneerat N, Sreng S, Hamamoto K. Ensemble deep learning for the detection of COVID-19 in unbalanced chest X-ray dataset. *Appl Sci*. 2021;11(22):10528.

doi: 10.3390/app112210528

40. Venkatesan R, Kadry R, Thanaraj KP, Kamalanand K, Seo S.

Firefly-Algorithm Supported Scheme to Detect COVID-19 Lesion in Lung CT Scan Images using Shannon Entropy and Markov Random Field. [arXiv Preprint].

41. Chandra SC. Segmentation and evaluation of COVID-19 lesion from CT scan slices-a study with Kapur/Otsu function and Cuckoo Search Algorithm. 2020.

    doi: 10.21203/rs.3.rs-40148/v1

42. Mohammed SN, Alkinani FS, Hassan YA. Automatic computer-aided diagnostic for COVID-19 based on chest X-ray image and particle swarm intelligence. *Int J Intell Eng Syst*. 2020;13(5):63-73.

43. Bhargava A, Bansal A, Goyal V. Machine learning-based automatic detection of novel coronavirus (COVID-19) disease. *Multimed Tools Appl*. 2022;81(10):13731-13750.

    doi: 10.1007/s11042-022-12508-9

44. Shankar K, Mohanty SN, Yadav K, Gopalakrishnan T, Elmisery AM. Automated COVID-19 diagnosis and classification using convolutional neural network with fusion based feature extraction model. *Cogn Neurodyn*. 2021;10:1-4.

    doi: 10.1007/s11571-021-09712-y

45. Kadry S, Rajinikanth V, Rho S, *et al*. *Development of a Machine-learning System to Classify Lung CT Scan Images into Normal/COVID-19 Class*. arXiv [Preprint]

46. Wu G, Zhou S, Wang Y, *et al*. A prediction model of outcome of SARS-CoV-2 pneumonia based on laboratory findings. *Sci Rep*. 2020;10(1):14042.

    doi: 10.1038/s41598-020-71114-7

47. Banerjee A, Ray S, Vorselaars B, *et al*. Use of machine learning and artificial intelligence to predict SARS-CoV-2 infection from full blood counts in a population. *Int Immunopharmacol*. 2020;86:106705.

    doi: 10.1016/j.intimp.2020.106705

48. Moutaz A, Awajan A, Mesleh A, Alhyari S. COVID-19 prediction and detection using deep learning. *Int J Comput Inf Syst Ind Manag Appl*. 2020;12:11-14.

49. Feng C, Wang L, Chen X, *et al*. *A Novel Triage Tool of Artificial Intelligence-assisted Diagnosis Aid System for Suspected COVID-19 Pneumonia in Fever Clinics*. MedRxiv; 2020.

50. Najjar FH, Kadhim KA, Kareem MH, *et al*. Classification of COVID-19 from X-ray images using GLCM features and machine learning. *Malays J Fundam Appl Sci*. 2023;19(6):389-398.

    doi: 10.11113/mjfas.v19n3.2911

51. Maryam A, Ahmad I, Imtiaz A, Mohammed A. Ensemble learning model for diagnosing COVID-19 from routine blood tests. *Inform Med Unlocked*. 2020;21:100449.

    doi: 10.1016/j.imu.2020.100449

52. Atta A, Sultan K, Naseer I, *et al*. Supervised machine learning-based prediction of COVID-19. *Comput Mater Contin*. 2021;69(1):21-34.

    doi: 10.32604/cmc.2021.013453

53. Rajinikanth V, Dey N, Raj AN, *et al*. *Harmony Search and Otsu based System for Coronavirus Disease (COVID-19) Detection Using Lung CT Scan images*. arXiv [Preprint].

54. Tongxue Z, Canu S, Ruan S. *An Automatic COVID-19 CT Segmentation Network Using Spatial and Channel Attention Mechanism*. [arXiv Preprint].

55. Mobiny A, Cicalese PA, Zare S, *et al*. *Radiologist-level COVID-19 Detection Using CT Scans with Detail-oriented Capsule Networks*. [arXiv Preprint].

56. Hasoon JN, Fadel AH, Hameed RS, *et al*. COVID-19 anomaly detection and classification method based on supervised machine learning of chest X-ray image. *Results Phys*. 2021;31:105045.

57. Mahdy LN, Ezzat KA, Elmousalami HH, *et al*. *Automatic X-ray COVID-19 Lung Image Classification System Based on Multi-level Thresholding and Support Vector Machine*. MedRxiv; 2020. p. 2020-2023.

58. Elizabeth DS, Raj CS, Nehemiah HK, Kannan A. A novel segmentation approach for improving diagnostic accuracy of CAD systems for detecting lung cancer from chest computed tomography images. *J Data Inf Qual*. 2012;3:1-16.

59. Rachel RB, Nehemiah HK, Marishanjunath CS, Manoharan RM. Diagnosis of pulmonary edema and COVID-19 from CT slices using squirrel search algorithm, support vector machine and back propagation neural network. *J Intell Fuzzy Syst*. 2023;44:1-4.

    doi: 10.3233/JIFS-222564

60. Rachel RB, Nehemiah HK, Singh VK, Manoharan RM. Diagnosis of COVID-19 from CT slices using whale optimization algorithm, support vector machine and multi-layer perceptron. *J Xray Sci Technol*. 2023;32:253-269.

    doi: 10.3233/XST-230196

61. Anisha I, Nehemiah HK, Anubha I, Kannan A. Computer-Aided Diagnosis system for diagnosis of pulmonary emphysema using bio-inspired algorithms. *Comput Biol Med*. 2020;124:103940.

    doi: 10.1016/j.compbiomed.2020.103940

62. Glover B. *Understanding Flowers and Flowering: An Integrated Approach*. Oxford: Oxford University Press; 2007.

63. Kalra S, Arora S. Firefly Algorithm Hybridized with Flower Pollination Algorithm for Multimodal Functions. In: *Proceedings of the International Congress on Information and Communication Technology*. Germany: Springer; 2016. p. 207-219.

64. Yang XS. Flower Pollination Algorithm for Global Optimization. In: *International Conference on*

Unconventional Computing and Natural Computation. Germany: Springer; 2012. p. 240-249.

65. Pavlyukevich I. Levy flights, non-local search and simulated annealing. *J Comput Phys*. 2007;226(2):1830-1844.

66. Fred AL, Daniel A, Carol JJ. SFCM for efficient brain tumour segmentation. *Int J Adv Eng Technol*. 2019.

67. He B, Zhao W, Pi JY, *et al*. A biomarker basing on radiomics for the prediction of overall survival in non-small cell lung cancer patients. *Respir Res*. 2018;19:199.

doi: 10.1186/s12931-018-0887-8

68. Isaac A, Nehemiah HK, Kannan A. Computer-aided diagnosis system for diagnosis of cavitary and miliary tuberculosis using improved artificial bee colony optimization. *IETE J Res*. 2021;69:1-20.

doi: 10.1080/03772063.2021.1946440

69. Xingyi Y, Xuehai H, Jinyu Z, *et al*. *COVID-CT Dataset: A CT Image Dataset about COVID-19*. [arXiv Preprint].

70. Polsinelli M, Cinque L, Placidi G. A light CNN for detecting COVID-19 from CT scans of the chest. *Pattern Recognit Lett.* 2020;140:95-100.

doi: 10.1016/j.patrec.2020.10.001

71. Ali AE, Assadi TA. GLCMs based multi-inputs 1D CNN deep learning neural network for COVID-19 texture feature extraction and classification. *Karbala Int J Mod Sci*. 2022;8(1):28-39.

doi: 10.33640/2405-609X.3201

72. Pedro S, Luz E, Silva G, *et al*. COVID-19 detection in CT images with deep learning: A voting-based scheme and cross-datasets analysis. *Inform Med Unlocked*. 2020;20:100427.

doi: 10.1016/j.imu.2020.100427