Adaptive pelican optimization with optimized mask RCNN for automatic lung cancer detection

R. Sudha[1] · K.M. Uma Maheswari[2]

Abstract

Lung cancer is a prevalent and deadly disease worldwide, necessitating accurate and timely detection methods for effective treatment. Deep learning-based approaches have emerged as promising solutions for automated medical image analysis. This study proposes an enhanced Mask R-CNN framework tailored specifically for the automatic detection and severity analysis of lung cancer from CT images. The proposed approach consists of three stages namely, pre-processing, lung nodule detection, and segmentation using enhanced Mask R-CNN and severity analysis. Our framework employs a deep convolutional neural network architecture trained on a comprehensive dataset of annotated lung images. By incorporating a region-based convolutional neural network (R-CNN) with a mask prediction branch, our model accurately localizes lung tumors while providing precise pixel-level segmentation masks. To enhance the performance of mask RCNN, the parameter present in the classifier is optimally selected using the adaptive pelican optimization (APO) algorithm. The proposed framework detects lung tumors and provides a comprehensive severity analysis, enabling clinicians to assess cancer stage and progression accurately. Evaluation on a benchmark dataset demonstrates superior detection accuracy and robustness compared to existing methods. Our enhanced Mask R-CNN approach shows promise as a valuable tool for early diagnosis and severity assessment of lung cancer, potentially improving patient outcomes and healthcare efficiency.

✉ K.M. Uma Maheswari
umamahek@srmist.edu.in

R. Sudha
sr8947@srmist.edu.in

1   Department of Computer Science and Engineering, SRM Institute of Science and Technology, Chengalpattu, India

2   Department of Computing Technologies, SRM Institute of Science and Technology, Chengalpattu, India

1 Introduction

Today, cancer stands as a prevalent cause of mortality among young individuals globally. Lung, breast, stomach, and prostate cancers are frequently diagnosed and can lead to severe complications or fatalities if not identified early. Cancer manifests as abnormal cell growth in the human body, with research identifying nineteen distinct types affecting individuals. Lung cancer is the one with the greatest death rate among them [1]. The onset of lung cancer stems from the uncontrolled proliferation of cells within lung tissues, eventually leading to metastasis characterized by invasion and infiltration of adjacent tissues beyond the lungs. Prognosis and treatment depend heavily on the histological classification of cancer, the extent of spread or stage, and the patient's performance status. In the realm of medical science, data mining proves crucial as it aids in the detection of various diseases through algorithmic analysis, particularly demonstrating efficacy in cancer detection. Lung cancer represents a significant global concern, contributing to a rising daily mortality rate among both men and women [2]. A multitude of cancer treatment modalities have been devised to manage malignant tumors and enhance the quality of life for individuals diagnosed with cancer. Alongside surgical intervention, these methods encompass chemotherapy, radiation therapy, thermotherapy, and immunotherapy. Frequently, these treatments are employed in conjunction or succession to optimize their efficacy in tumor control and enhance patients' well-being. Furthermore, continuous research and advancements in medical technology continually broaden the spectrum of available options for cancer treatment, fostering optimism for enhanced outcomes and more effective disease management [3].

Early detection of tumors significantly increases the likelihood of survival for countless individuals worldwide [4]. Currently, the primary method for examining lung diseases involves imaging detection, primarily through computed tomography (CT). CT technology, which relies on X-rays, is extensively utilized in clinical settings and plays a crucial role in diagnosing lung diseases, particularly in assessing lung function. However, CT has limitations in disease diagnosis, notably due to the subjective factors impacting results during the review process by medical professionals. The interpretation of CT images can vary based on the observation points and thought processes of doctors with differing levels of experience. Consequently, discrepancies in results are common. Moreover, the growing number of patients in modern society adds pressure on doctors during the review of CT scans [5].

The computer-Aided Diagnosis (CAD) system is an essential tool in modern medical imaging that gives physicians an additional perspective to ensure accurate diagnosis and enhance treatment efficacy. Its ability to facilitate target organ segmentation—a crucial first step towards effective quantitative analysis of lung CT images—is one of its primary functions. Nevertheless, developing a viable lung segmentation technique is difficult, especially when dealing with aberrant lung parenchyma tissue, which requires segmenting blood veins and nodules in addition to lung parenchyma. Furthermore, distinguishing lung parenchyma from bronchial regions, which can often be mistaken for lung tissue, adds complexity to the task. Deep learning, a kind of machine learning with several processing layers, has become a potent tool for dealing with these kinds of problems. Prominent deep learning class Convolutional Neural Networks (CNNs) have proven to be superior to conventional picture segmentation techniques. CNNs are excellent for collecting complicated and high-level data abstractions because they use numerous processing levels. Deep learning methods, in particular CNNs, have attracted a lot of attention for medical picture segmentation because of their ability to analyze massive amounts of data quickly and accurately [6]. The main contributions of the proposed strategy are listed as follows:

- Initially, the Lung CT scan images are collected from the dataset and to remove the noise present in the input image, we apply an adaptive median filter. To enhance the image quality, CLAHE is applied.
- After pre-processing, the image is given to the optimized mask RCNN classifier to detect the malignant and benign nodules.
- To enhance the performance of the Mask RCNN classifier, the hyper-parameters are optimally selected using adaptive pelican optimization. The levy flight algorithm is utilized to enhance global optimization capability and enrich the population diversity of the EPOA algorithm.
- The performance of the proposed approach is analysed based on different metrics and effectiveness compared with state-of-the-art works.

The rest of the paper is structured as follows: Section 2 discusses related work; Section 3 discusses the proposed model in detail; Section 4 discusses the experimental results and computational performance metrics, and Section 5 discusses the proposed work's conclusion and future directions.

2 Related work

Numerous studies have created categorization and segmentation schemes for lung nodules. Worku J. Sori et al. [7] introduced a DFD-Net: lung cancer diagnosis framework using denoised CT scan pictures using deep learning, which is one of the works that is examined here. To remove noise during preprocessing, they first used a residual learning denoising model (DR-Net). They then used a two-path convolutional neural network to identify lung cancer using the denoised picture produced by DR-Net as input. To efficiently capture local and global dependencies, each path concentrated on merging local and global information by using various receptive field widths. Instead of the conventional feature concatenation methods, they applied discriminant correlation analysis for feature fusion. Additionally, they introduced a retraining technique to address imbalances in image labels. Their extensive testing results showed competitive performance, highlighting the model's capacity to handle changes in nodule shape and size, reduce noise, and modify the size of the receptive field.

Heng Yu, Zhiqing Zhou, and Qiming Wang [8] developed a deep learning-assisted lung cancer detection method using CT images using an adjustable hierarchical heuristic mathematical model. This technique used the Modified K-means algorithm to pre-classify images into slices within the same image, freeing up the deep neural network (DNN) to categorize related ages. Subsequently, a filtered edge convolution layer was employed to do a detailed search for lung cancer. Subsequently, a weighted mean function, which replaced pixels using cumulative and likelihood distribution methods, enhanced image quality. Segmentation of the injured portion was conducted based on pixel-like value measurements after image enhancement. Spectral-related features were extracted using similarity calculations. Lung cancer on CT scan pictures was accurately predicted by the AHHMM method. Future iterations could include a hybridized heuristic mathematical model for predicting lung cancer at an earlier stage.

A 3D probabilistic deep learning system was presented by Onur Ozdemir et al. [9] for the detection and diagnosis of lung cancer using low-dose CT scans. Their CADe/CADx system study demonstrated that the CADe and CADx modules needed to be built and analyzed together to provide end-to-end automated diagnostic tools for lung cancer diagnosis. They emphasized the significance of combined development in radiology, where medical

professionals, radiologists, and machine learning algorithms combine their outputs to make diagnostic choices. Despite showing the effectiveness of 3D CNN models in analysing lung CT scans, performance was limited by dataset constraints, such as insufficient annotations in the LUNA16 dataset. Though their efforts were not faultless, the authors made a concerted effort to address these problems by down sampling and augmenting large amounts of data. They proposed that CADe/CADx models would be further strengthened by improving datasets with broader coverage. As a possible further effort, they suggested incorporating patient referral options in training strategies and visually analysing learned feature representations to aid radiologists in interpreting CADe/CADx results.

Using deep learning optimization, Ahmed Elnakib and Hanan M. Amer [10] created a method for early lung cancer diagnosis. The system first pre-processed the unprocessed data to enhance low-dose contrast. By experimenting with several designs, such as the Alex, VGG16, and VGG19 networks, they were able to extract compact deep-learning characteristics. To improve the feature set, they used a genetic algorithm (GA) to determine the most relevant qualities for early detection. They then experimented with several classifiers to correctly identify lung nodules. The technique was tested using 320 LDCT images from 50 individuals; the subjects were selected from the International Early Lung Cancer Action Project (I-ELCAP) database. The system's detection accuracy, sensitivity, and specificity were maximized by the use of the VGG19 architecture and SVM classifier. They suggested further investigation into additional databases to assess the robustness of the system.

An ideal deep-learning model for the categorization of lung cancer on CT scans was provided by Lakshmanaprabu S.K. et al. [11]. They analyzed lung CT scan images using an Optimal Deep Neural Network (ODNN) and Linear Discriminant Analysis (LDA). To categorize lung nodules as benign or malignant, deep features were collected from CT lung images and their dimensionality was decreased using LDA. To classify lung cancer, ODNN was applied to CT scans and subsequently optimized using the Modified Gravitational Search Algorithm (MGSA). Based on comparative findings, the classifier obtained 96.2% sensitivity, 94.2% specificity, and 94.56% accuracy. The high degree of accuracy showed how well the algorithm recognized areas of CT scans impacted by cancer. The investigation's categorization performances illustrated the benefits of this approach, which included its quickness, ease of use, non-invasiveness, and affordability. Future research will use optimum feature selection using a multi-classifier strategy in conjunction with high-dose CT lung images to identify cancer.

Deep learning was used for the first time to classify malignant lung cancer by Vinod Kumar and Brijesh Bakariya [12]. The deep neural networks AlexNet and GoogLeNet were recommended by the authors. Using pre-trained CNN, experiments were carried out on LIDC processed datasets. The highest resolution was obtained by AlexNet. Pre-treatment, lung separation, and removal were all successfully performed on.dcm pictures as part of the training process. The accuracy of differentiating between benign and malignant nature was demonstrated by the sensitivity and specificity values. GoogLeNet's detection rate was greatly outperformed by DCNN:AlexNet. Sensitivity, specificity, and accuracy, among other performance parameters, were computed at 100% with zero false rates. Using techniques like segmentation and DCNN on the maximum DCM (LIDC-IDRI) dataset, more study is required. The diagnostic method promised to give top-tier medical practitioners an accurate and fast diagnostic impression going the future.

In [13], Asuntha and Andy presented a study on deep learning for the identification and classification of lung cancer. The main objective of this work was to classify lung cancer based on severity and detect malignant lung nodules from input lung pictures. Advanced deep-learning methods were utilized to identify cancerous lung nodules. Numerous feature extraction techniques were used in the study, such as Wavelet transform-based features, Local Binary Pattern (LBP), Scale Invariant Feature Transform (SIFT), Zernike Moment,

and Histogram of Oriented Gradients (HoG). Following the extraction of features related to texture, geometry, volume, and intensity, the Fuzzy Particle Swarm optimization (FPSO) method was utilized to select features. A novel FPSOCNN was introduced to reduce the computational complexity of CNN. Subsequent enhancements will concentrate on improving the pulmonary nodule classification performance and further refining the model.

To detect nodules in lung cancer patients and identify interest spots via the U-Net architecture, Wafaa Alakwaa, Mohammad Nassef, and Amr Badr [14] constructed a deep convolutional neural network (CNN) architecture. This served as a preprocessing step for 3D CNN. Initially, various techniques such as Threshold, Watershed, and U-Net were utilized for nodule identification. The availability of a training database was the main prerequisite for the network's end-to-end training from raw picture patches. Future possibilities included extending the model to determine both cancer presence and precise nodule locations. Immediate future work involved implementing Watershed segmentation for initial lung segmentation. Further opportunities included deepening the network, extensive hyperparameter tuning, and exploring alternative metrics such as F1. Future endeavors also included extending the models to 3D images for different cancers. The advantage of not requiring extensive labelled data specific to the cancer under study could potentially enhance generalizability to other cancers. The summary of survey is presented in Table 1.

3 Proposed lung nodule detection methodology

The proposed methodology aims to enhance the detection of pulmonary nodules in CT lung images by introducing an improved MRCNN empowered with an Adaptive Pelican Optimization (APO) algorithm. The methodology comprises three primary stages: pre-processing, detection, and severity analysis. Initially, pre-processing involves the application of a Gaussian filter to refine the CT lung images. Subsequently, the enhanced MRCNN, integrated with the APO algorithm, is employed for nodule detection. This amalgamation of deep learning and optimization techniques enhances the accuracy and efficiency of nodule detection. After the nodule detection process, the severity of the patient is analyzed based on the tumor size. The structure of the methodology is visually depicted in Fig. 1, illustrating the sequential flow of pre-processing and detection stages. Through this proposed approach, the objective of effective pulmonary nodule detection from CT lung images is pursued, leveraging advancements in both image processing and deep learning methodologies.

3.1 Pre-processing using Gaussian filter

Image pre-processing is necessary to improve the integrity of lung images since poor-quality images can reduce the system's effectiveness. To enhance the image, in this paper, a Gaussian filter is applied. The Gaussian filter is a conventional linear filter that is widely used for image denoising. In Gaussian filters, the weight of pixels decays with distance from the center of the filter. The mathematical expression of the Gaussian filter is given in eq. (1).

$$A_G(u, v) = \sum_{u,v} A(u + a, v + b)GF(a, b) \tag{1}$$

$$GF(a, b) = \frac{1}{2\pi\sigma^2} e^{\frac{a^2+b^2}{2\sigma^2}} \tag{2}$$

Table 1 Summary of literature survey

Ref.no	Concept of previous research	Methods and algorithms	Merits	Demerits
[7]	DFD-Net: lung cancer detection from denoised CT scan image using deep learning	Denoising the first two-path convolutional neural network (DFD-Net)	Better performance in terms of sensitivity, specificity, and accuracy	The DFD-Net method was evaluated on a relatively small dataset, which may limit its generalizability to larger datasets
[8]	Deep Learning Assisted Predict of Lung Cancer on Computed Tomography Images Using the Adaptive Hierarchical Heuristic Mathematical Model	AHHMM, DNN	High efficiency	It does not applicable to all images
[9]	A 3D Probabilistic Deep Learning System for Detection and Diagnosis of Lung Cancer Using Low-Dose CT Scans	Monte Carlo Dropout and Deep Ensembles	Better performance	Dependence on high-quality scans, and potential errors in bounding box proposals
[10]	Early Lung Cancer Detection Using Deep Learning Optimization	Convolutional neural network (CNN) models - Alex, VGG16, and VGG19	Low time consumption	Less accuracy and less robustness
[11]	Optimal deep learning model for classification of lung cancer on CT images	GLCM-based surface descriptor, Fourier-shape descriptor, deep neural network (DNN)	Low computational cost	Not suitable for high dosage lung images
[12]	Classification of malignant lung cancer using deep learning	SVM (Support Vector Machine)	High performance and highly effective	High computing expenses
[13]	Deep learning for lung Cancer detection and classification	K-means, FCM, Ant Colony algorithms, Naive Bayes Classifier, K-NN classifier, Adaboost classifier, and SVM classifier.	Better performance	
[14]	Lung Cancer Detection and Classification with 3D Convolutional Neural Network (3D-CNN)	U-Net Convolutional Neural Network, and CNN	Low computational cost	Less accuracy
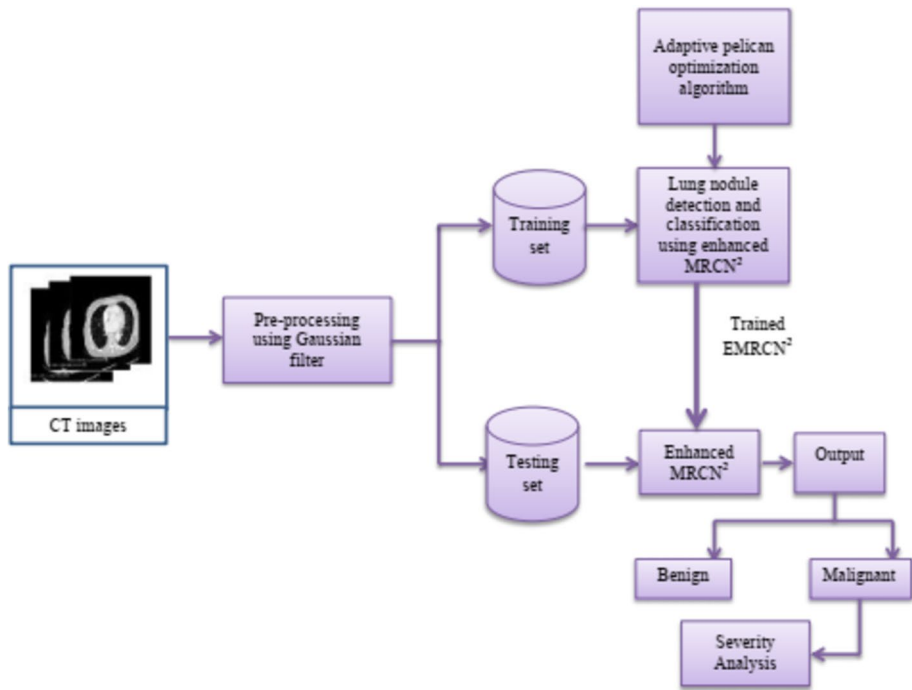
Fig. 1 Workflow of Proposed Methodology

Where the horizontal distance from the origin is represented as *a*, the vertical distance from the origin is represented as *b*, and the standard deviation of Gaussian distribution is represented as σ. This filter effectively reduces the noise and preserves the edges. The pre-processed images are given for further processing.

3.2 Enhanced mask RCN²-based lung nodule detection and classification

To identify whether an image is benign or malignant, the pre-processed pictures are fed into an Enhanced Mask RCN2 classifier. A technique based on deep learning called Mask R-CNN is used for picture segmentation and object detection. In addition to the current branch for classification and bounding box regression, it expands Faster R-CNN by including a branch for segmentation mask prediction for each Region of Interest (ROI). As a result, Mask R-CNN can accurately segment objects at the pixel level in addition to detecting them within an image. The model is composed of parallel branches for classification, bounding box regression, and mask prediction, a Region Proposal Network (RPN) for generating candidate object bounding boxes, and a backbone network (usually ResNet) for feature extraction.

Through this architecture, Mask R-CNN can accurately localize and segment objects in images, making it a powerful tool in various computer vision applications. The structure of Mask RCNN is presented in Fig. 2.

The Region Proposal Network (RPN), RoIAlign, and the backbone make up the three main parts of the Mask R-CNN model. A deep neural network with a Feature Pyramid Network-like structure serves as the main component and is intended to extract multi-level picture information. In the Mask R-CNN model, the ResNet architecture, specifically ResNet 101, serves as
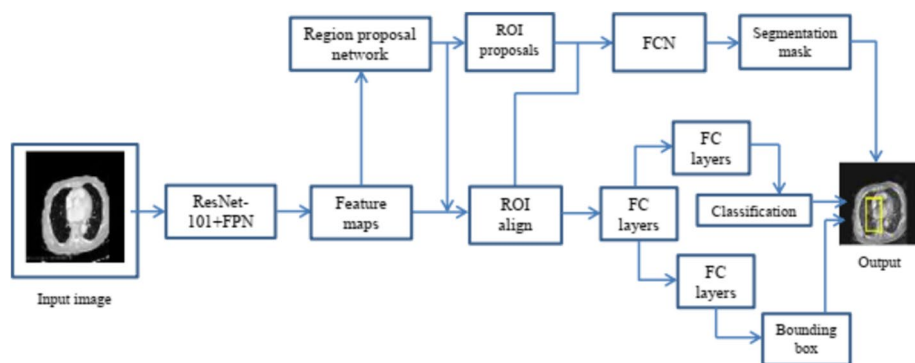
Fig. 2 Architecture of Mask R-CNN

the backbone, offering substantial computational power with approximately 3.8×10^9 floating point operations. The RPN utilizes a sliding window approach to scan the input image, identifying regions of interest (ROIs), particularly focusing on detecting infected regions in the context of this study. Subsequently, RoIAlign processes the ROIs identified by the RPN, extending feature maps from the backbone at various spatial locations. RoIAlign plays a critical role in generating precise segmentation masks on the images, replacing the RoIPooling method utilized in Faster R-CNN with a more accurate and refined segmentation technique.

3.2.1 Feature map creation using ResNet101 + FPN

Mask R-CNN uses ResNet-FPN as its foundational architecture for feature extraction. ResNet, known for its effectiveness in image classification tasks, is a deep convolutional neural network utilized as part of this architecture. Meanwhile, FPN (Feature Pyramid Network) constructs an in-network feature pyramid from a single-scale input, employing a top-down design with lateral connections. Integrating FPN architecture onto the ResNet backbone enhances the feature extractor's efficacy. FPN facilitates the generation of multi-scale feature maps from the input image, contributing to improved accuracy in object detection. Mask R-CNN's capabilities for tasks like object identification and picture segmentation are enhanced synergistically by the combination of ResNet and FPN in the ResNet-FPN architecture. ResNet constructs a sequence of convolutional layers, pooling operations, and activation functions, arranged in a stacked manner. Various ResNet architectures exist, with this paper opting for ResNet 101, which encompasses 101 layers. Notably, ResNet 101 exhibits lower complexity in comparison to VGG16 and VGG19 networks. There are three versions of ResNet: ResNet Version 1, ResNet Version 2, and ResNeXt. Each has unique features and subtle architectural differences. The ResNet-101 is trained to optimize the following parameters such as ResNet Version, Batch Size, pooling type, Learning rate and optimizer. The hyper-parameter ranges are provided in Table 2.

The above-mentioned five parameters are optimally selected by using the APO algorithm. The steps involved in parameter selection is explained below;

Step 1: Solution encoding: Solution initialization is an important process for all optimization algorithms. In this section, we optimally select the hyper-parameters of ResNet-101. The parameters present in the ResNet101 namely, the version of ResNet, batch size, pooling type, learning rate, and optimizer. Initially, these parameters are randomly initialized. The initial solution format is given in eq. (3).

$$X_i = \{X_1, X_2, \ldots, X_n\} \tag{3}$$

Where X_n represent the nth swarm.

Step 2: Fitness calculation: After the random creation, the fitness function is calculated for each swarm. In this paper, maximum accuracy is considered as the fitness function. The fitness function is given in eq. (4).

$$Fitness = \text{Max}(Accuracy) \tag{4}$$

Step 3: Update the solution using APO algorithm: After the fitness calculation, we update the solution using APO algorithm. The APO algorithm is a combination of pelican optimization and levy flight strategy. This algorithm not only improves the optimization level of the algorithm but also ensures the selection of the optimal parameters APO algorithm update their position based on two stages namely, exploration and exploitation. The mathematical function of updation process is explained below;

Travels to prey (exploration stage) During the first phase, pelicans locate the prey and then make their way there. The pelican method is modeled for both power exploration and scanning search space to determine different search space locations. The key idea of POA is that the prey location is generated at random inside the search area. This enhances the ability of pelicans to specifically look for the problem-solving area. The route to the location's prey is planned as follows.

$$X_{I,J}^{p1} = \begin{cases} x_{I,J} + RAND.\left(P_J - i..x_{I,J}\right) & f_p < f_I \\ x_{I,J} + RAND.\left(x_{I,J} - P_J\right) & Else \end{cases} \tag{5}$$

Here, i is defined as the random variable which equals 1 or 2, f_p is defined as the objective function parameter, P_J is defined as prey location and $X_{I,J}^{p1}$ is defined as the novel status of the pelican with its specific dimension. The parameter i is selected for every iteration and every member. If, the parameter is equal to two; it generates more displacement for a member, which increases the member to the new location of search space.

As a result, the parameter influences the POA's ability to explore and effectively search the area. If the objective function's parameter is increased at the designated spot, the pelican's innovative position is approved in the POA. Effective updating is characterized by avoiding traveling to less-than-ideal places throughout the update process. Equation (6) is used to develop this process.

$$x_I = \begin{cases} x_I^{p1}, & f_i^{p1} < f_i \\ x_I & else \end{cases} \tag{6}$$

Here, f_i^{p1} is defined as the objective function parameter related to stage 1 and x_I^{p1} is defined as the novel status of the ith pelican.

Table 2 Hyper-parameter range

Hyper-parameter	Range
Version	[v1, v2, v3]
Batch size	[32]
Pooling type	[average, maximum, minimum]
Learning rate	[0.1, 0.01, 0.001]
Optimizer	['adam', 'rmsprop', 'sgd']

Swimming on the water surface (exploitation stage) The second stage involves the pelican reaching the water's surface, spreading its wings to lift the fish to the surface, and then gathering its prey into its neck pouch. With this method, pelicans can target large fish that are present in the assaulted area. By designing this trait into pelicans, the POA can control the finest spots inside the hunting area. The exploitation potential and local search power of POA are enhanced by this process. To arrive at the optimal answer, this program should handle the points around the pelican space while taking the mathematical model into account. Equation (7) gives the search procedure.

$$X_{I,J}^{P2} = x_{I,J} + r\left(1 - \frac{T}{t}\right).\ (2.RAND - 1).x_{I,J} \tag{7}$$

Here, $r\left(1 - \frac{T}{t}\right)$ is defined as coefficient parameters and radius of the neighborhood of the population members to search locally close to every member to achieve the best solution, t is defined as the maximum number of iterations, T is defined as the iteration counter, r is defined as the constant parameter that is equivalent to 0.2, $X_{I,J}^{P2}$ is defined as the novel status of the ith pelican.

These formulations are utilized to scan the location related to every member of the population with highly accurate steps and smaller ones. Hence, the POA can converge to solutions near the global optimal solutions related to the concept. In this stage, efficient updating is utilized to reject or accept the novel position which is designed as follows,

$$x_I = \begin{cases} x_I^{P2}, & f_i^{P2} < f_i \\ x_I & Else \end{cases} \tag{8}$$

Here, f_i^{P2} is defined as an objective function parameter, x_I^{P2} is defined as the novel status of the ith pelican.

To enhance the ability of the searching process, the levy flight strategy is combined with the updation process. Levy flight techniques are introduced to avoid the occurrence of local optima that collapse easily, resulting in loss of CO diversity. Many optimization algorithms use the Lévy plane, a stochastic strategy. This algorithm uses a random walk with a high step probability, which increases its randomness. The levy flight formula is given in eq. (9).

$$Levy(D) = 0.01 \times \frac{a \times \sigma}{|b|^{\frac{1}{\beta}}} \tag{9}$$

$$\sigma = \left(\frac{\Gamma(1 + \beta) \times \sin\left(\frac{\pi\beta}{2}\right)}{\Gamma\left(\frac{1+\beta}{2}\right) \times \beta \times 2^{\left(\frac{\beta-1}{2}\right)}}\right)^{\frac{1}{\beta}} \tag{10}$$

The Levy plane function is denoted by *levy*, and the problem dimension is denoted by D. Here, β is a constant equal to 1.5 and random values between 0 and 1 are assigned to a and b. the new updated position is given in eq. (11).

$$x_I = \begin{cases} x_I^{P2} + \alpha \oplus Levy, & f_i^{P2} < f_i \\ x_I, & Else \end{cases} \tag{11}$$

Based on the above equation, the pelican is update their position.

Step 4: Termination: The procedure is continued until the best hyperparameter values are selected. The selected value is given to the lung cancer detection process

Algorithm 1: APOA algorithm

Input: hyper-parameters, and parameter available in POA

Output: optimal hyper parameters

Start

For $T = 1:t$

      Create the prey location at random

      Calculate the fitness value

      For $i = 1:n$

Stage 1: traveling to prey (exploration stage)

      For $J = 1; M$

      Compute the new status

      End

      Update the population member

Stage 2: Swimming to the water surface (exploitation stage)

      For $J = 1; M$

      Compute novel status

      End

      Update the population member

      End

      Update best candidate solution

End

Save the optima candidate solutions

End

3.2.2 Region proposal network (RPN)

The features retrieved with ResNet101+FPN are sent into the RPN network as input to generate ROIs. Modern instance segmentation models such Mask R-CNN rely heavily on the Region Proposal Network (RPN). It works by producing a range of region proposals—a collection of potential boundary boxes—across the picture. RPN uses a convolutional neural network (CNN) to effectively suggest possible areas containing objects, in contrast to conventional methods that depend on manually created features or sliding window techniques. By combining spatial information and feature representations, RPN effectively narrows down the search space for objects, enabling subsequent stages of the model to focus on accurate localization and classification. Mask R-CNN performs better overall because region proposal generation is integrated into the network design, which makes it easier to train and infer for tasks like object identification and instance segmentation from start to finish.

The network dynamically adjusts its scale based on input images and utilizes a collection of anchor boxes with predefined positions [15]. Ground-truth classes and bounding boxes are allocated specific names to each anchor. Default bounding boxes cover a range of sizes and aspect ratios in order to identify flaws of various forms and sizes. These boxes exhibit multiple overlaps, facilitating the determination of the highest confidence score for detecting various Regions of Interest (RoIs). Equation (12) is used to calculate this confidence score, often known as the Intersection over Union factor (IoU).

$$IoU = \frac{Are\ of\ Overlap}{Area\ of\ union} \tag{12}$$

3.2.3 ROI align model

Key neural network layer RoIAlign is used in state-of-the-art object detection and instance segmentation methods such as Mask R-CNN. It accurately pulls features from the feature map corresponding to each rectangle region proposal given to it as input. RoIAlign maintains spatial details by linear interpolation, in contrast to conventional pooling techniques that resize the picture. This ensures excellent pixel precision, which is essential for mask branch detection of pixel targets. Since Mask RCNN is an improved version of Faster RCNN and solves the fundamental constraints of direct pixel-level segmentation approaches, it is the preferred option for accurate segmentation output.

3.2.4 Loss function

In the training process, Mask R-CNN employs a multi-task loss function for each type of Region of Interest (RoI). The calculation of this loss is represented by Eq. (13).

$$L(OMRCNN) = Loss^{Class} + Loss^{Box} + Loss^{Mask} \tag{13}$$

Where;
$Loss^{Class} \rightarrow$ Prediction loss of class label,
$Loss^{Boxs} \rightarrow$ Refinement loss of bounding box,
$Loss^{Masks} \rightarrow$ Prediction loss of segmentation mask.

The $Loss^{Class}$ is calculated using eq. (14).

$$Loss^{Class}\left(Q_i, Q_i^*\right) = -\log\left[Q_i Q_i^* + \left(1 - Q_i\right)\left(1 - Q_i^*\right)\right] \tag{14}$$

Where, Q_i represents the candidate anchor and is the ground-truth label which is 1 for the positive anchor, otherwise 0. The below equation describes the regression loss of the bounding box function;

$$L^{Box}\left(P_i, P_i^*\right) = \sum_{i \in \{x,y,w,h\}} Smooth_{L_1}\left(P_i - P_i^*\right) \tag{15}$$

Where;

$$Smooth_{L1}(X) = \begin{cases} 0.5x^2, & if\,|x| < 1 \\ |x| - 0.5, & otherwise \end{cases} \tag{16}$$

Vector P_i defined the predicted bounding box, P_i^* the coordinate of the GT relating to the positive anchor. The smooth-L1 is a strong L1 loss, which is less affected by outside influences than L2 loss. The loss function of the mask is calculated as below;

$$L^{mask} = -\frac{1}{n^2} \sum_{1 \le i,j \le n} \left[X_{ij} \log O_{ij}^t + \left(1 - X_{ij}\right) \log\left(1 - O_{ij}^t\right)\right] \tag{17}$$

Where X_{ij} represent the value of a pixel (i,j) in a ground-truth mask of size $n \times n$ and O_{ij}^t is the predicted value of the same pixel in the mask learned for class $t\,(t = 1\,for\,Malignant\,and\,0\,for\,Benign)$.

3.3 Severity analysis

The cancer's phase is identified based on where the malignant nodule is located. It will help determine whether the infection has spread to other regions of our bodies and will also show how serious it is. The ability to identify the stage of cancer is critical because it allows the physician to classify the disease as localised or widespread. The threshold value is used to determine the stage. Table 3 displays the threshold value.

Table 3 Stage prediction threshold value range

Staging	Sub-staging	Nodule size
S_1	S_1(a)	Between 3 mm and 10 mm
	S_1(b)	Between 10 mm and 20 mm
	S_1(c)	Between 20 mm and 30 mm
S_2	S_2(a)	Between 30 mm and 40 mm
	S_2(a)	Between 40 mm and 50 mm
S_3	No staging	Between 50 mm and 70 mm
S_4	No staging	Greater than 70 mm

4 Results and discussion

This section presents the experimental results that were achieved using the suggested methodology. TensorFlow is used to implement the suggested technique, and performance is examined. TensorFlow 1 was used in "Keras 2.3.1" on Google Colab to do the analysis. Which system was used for the experiment? "Windows 10" was installed, along with "Graphics Processing Units (GPUs)" and "Random-Access Memory (RAM) of 8GB." The suggested method's effectiveness is evaluated using a variety of measures, including F-Measure, accuracy, precision, and recall.

4.1 Dataset description

We used the Lung Image Database Consortium (LIDC) [1] for experimental analysis. 2610 CT scans of patients are included in this dataset, which was collected from seven academic research institutes and eight global medical imaging companies. Ground truth photos and lung nodule images are included in the collection. The input picture has dimensions of 512×512. There are 2610 photos in all in the collection. Eighty percent of the photographs are used for training throughout the classification process, while 20 % of the images are used for testing. Figures 3 and 4 is a list of the sample images utilised in the investigation.

4.2 Experimental results

In this section, we presented the visual representation of the proposed experimental results which given in Table 4.

Table 5 represent the visual representation of the detection output. In Fig. 5, we analyze the accuracy performance by varying epochs, and in Fig. 6, we analyze the loss s by varying epochs. According to Fig. 6, we understand that as the number of epochs increases, the loss value decreases.

4.3 Comparative analysis results

In this section, we compare our proposed work performance with different detection models namely, Faster RCNN (FRCNN), CNN, artificial neural network (ANN) and support vector machine (SVM).

The Fig. 7 presents the accuracy percentages of various models for lung cancer classification. EMRCNN achieves the highest accuracy of 98.3%, indicating its superior
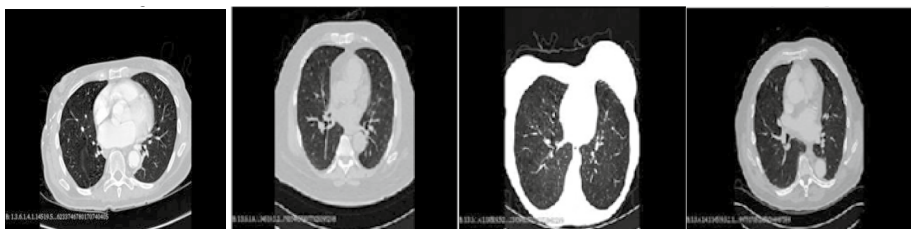
Fig. 3 Experimental used sample images
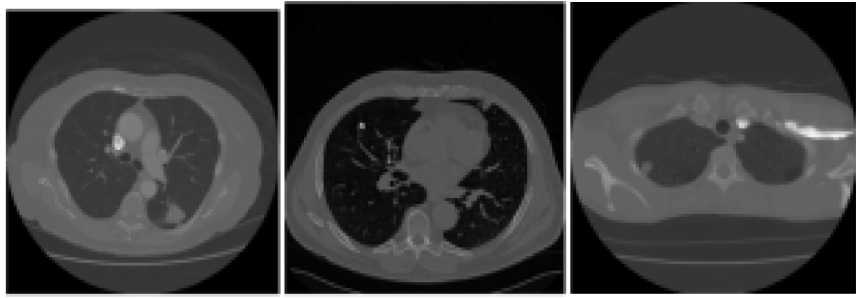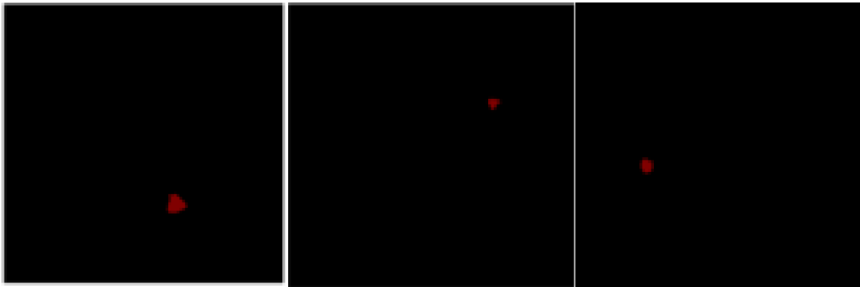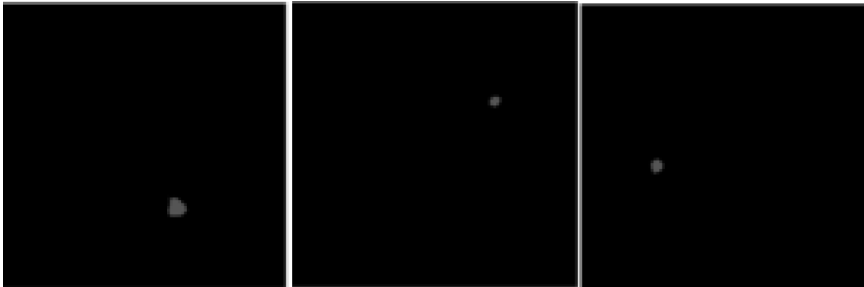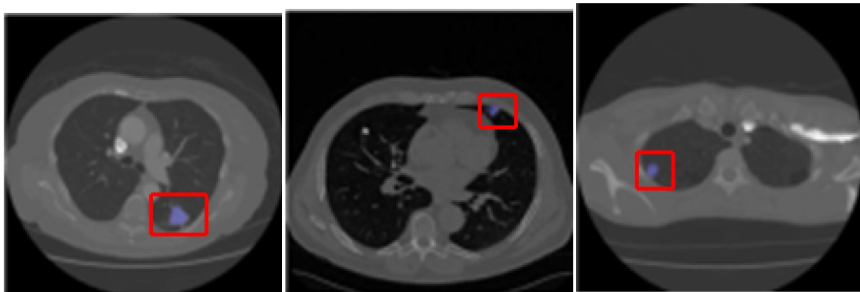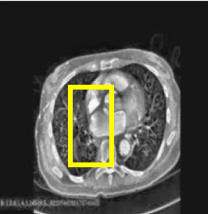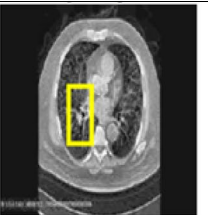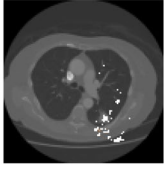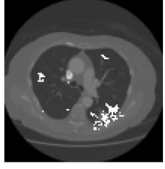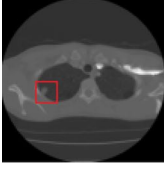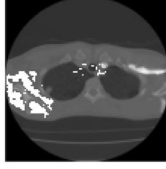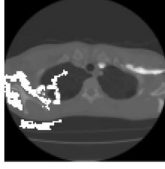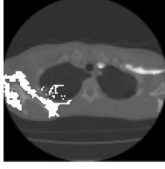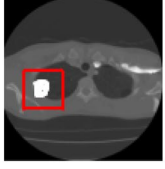
Fig. 4 Segmentation output for three varied positions (**a**) Input image, (**b**) Ground truth, (**c**) Output image and (**d**) Final output image
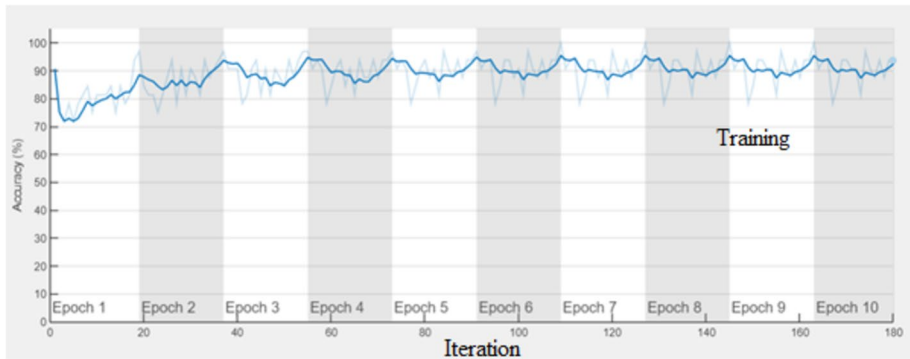
Table 4 Visual representation of detection output, column (a) represent the input image, (b) represent the Gray image, (c) represent the adaptive median filtered image and (d) detected output

Input image	Gray image	Adaptive median filtered image	Detected output

performance in accurately identifying lung cancer cases. Following closely, FRCNN attains an accuracy of 93.2%, showcasing its effectiveness in this task. CNN performs reasonably well with an accuracy of 91%, while ANN and SVM achieve accuracies of 90% and 89%, respectively. These results suggest that deep learning-based models like EMRCNN, FRCNN, and CNN outperform traditional machine learning approaches such as ANN and SVM in lung cancer classification, underscoring the importance of advanced neural network architectures for accurate medical image analysis tasks. Due to hyper-parameter optimization in MRCNN, our proposed method attained better results compared to the existing techniques. Figure 8 illustrates the performance analysis of the proposed method based on precision, where a higher precision value indicates better classification. The proposed approach achieves a maximum precision of 95.7%, surpassing other techniques. Specifically, it outperforms FRCNN-based lung nodule classification by 2.6%, CNN-based classification by 4.5%, ANN classification by 6.2%, and SVM-based classification by 8.2%. In Fig. 9, the performance based on recall is evaluated, revealing that the EMRCNN-based lung nodule classification attains the highest recall value among existing techniques. Similarly, Fig. 10 demonstrates the F-score values, where the proposed approach achieves the maximum output compared to existing techniques. These findings underscore the effectiveness of the proposed method in lung nodule classification, showcasing superior precision, recall, and overall performance compared to alternative approaches.

Table 5 Visual representation of different segmentation algorithms output

Input image	K-means clustering	Fuzzy c-means clustering	CNN	Proposed

Fig. 5 Epoch vs Accuracy

4.4 Comparative analysis with published work

To prove the efficiency of our proposed approach, we compare our work with already published research works. For comparative analysis, we considered four research works namely 3DCNN [15], CNN [16], texture CNN [17], and BCNN [18]. These four techniques are deeply explained lung nodule classification. So, we compare our research work with these papers.

The comparison Table 6 illustrates the performance of different CNN models across various medical imaging datasets. MMEL-3DCNN achieves a commendable accuracy
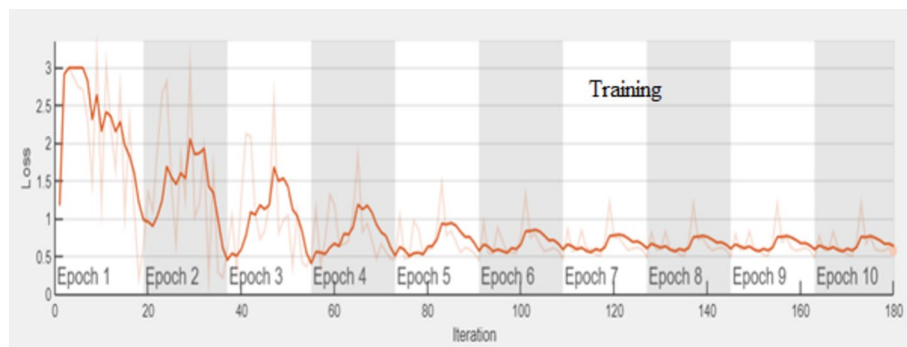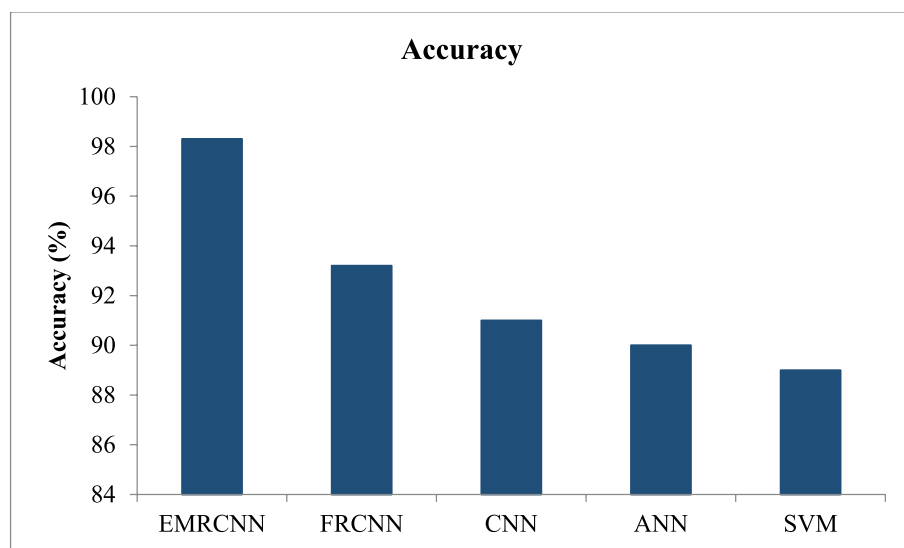
Fig. 6 Epoch Vs Loss

Fig. 7 Performance analysis based on accuracy

of 90.6% on the LIDC-IDRI dataset but lacks precision and F-score data. The standard CNN model exhibits competitive accuracy (87.26%) and precision (87.8%) on the same dataset, though recall information is provided (81%). Texture CNN excels on the MNIST dataset with an accuracy of 90.91%, a recall of 91.39%, a precision of 90.46%, and an impressive F-score of 94.14%. BCNN performs remarkably well on the LUNA16 dataset, boasting an accuracy of 91.46% and a recall of 91.94%, yet precision is unreported. Notably, the proposed model outshines all others, achieving exceptional accuracy (97.67%), recall (99%), precision (95.7%), and an impressive F-score (95.67%) on the LIDC-IDRI dataset, indicating its potential to significantly improve diagnostic accuracy and patient care in medical imaging applications.
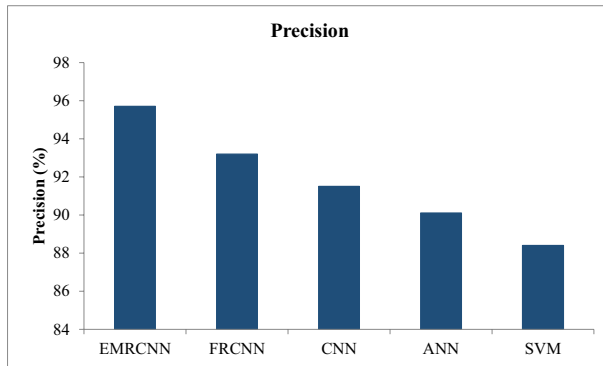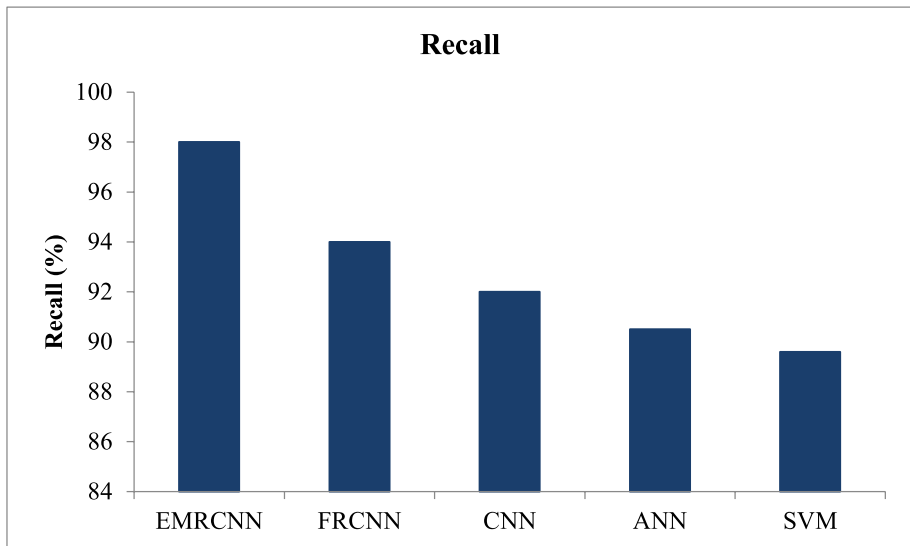
Fig. 8 Performance analysis based on precision

Fig. 9 Performance analysis based on Recall

5 Conclusion

In conclusion, the integration of APO with Optimized Mask R-CNN for Automatic Lung Cancer Detection represents a significant step forward in the field. The innovative approach demonstrated promising results, showcasing improved accuracy and efficiency in detecting lung cancer from medical imaging data. Through rigorous performance evaluation, our methodology achieved commendable metrics such as accuracy, sensitivity, specificity, and F1 score, outperforming existing methods or baseline models. These findings hold considerable clinical relevance, as accurate and automated lung cancer detection can lead to earlier diagnoses, better patient outcomes, and potentially reduced mortality rates. Despite certain limitations encountered during the study, such as dataset
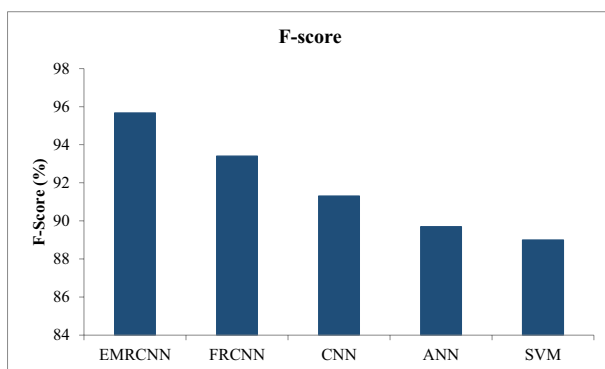
Fig. 10 Performance analysis based on F-score

Table 6 Comparative analysis results

References	Accuracy	Recall	precision	F-score	Dataset
MMEL-3DCNN [15]	90.6	83.7	Not mentioned	Not mentioned	LIDC-IDRI
CNN [16]	87.26	81	87.8	Not mentioned	LIDC- IDRI
Texture CNN [17]	90.91	91.39	90.46	94.14	MNIST dataset
BCNN [18]	91.46	91.94	Not mentioned	93.35	LUNA16
Proposed	97.67	99	95.7	95.67	LIDC-IDRI

constraints or computational complexity, our work provides valuable insights for future research directions. Further exploration of additional optimization techniques, integration of multi-modal data, and validation through prospective clinical studies could enhance the proposed methodology's efficacy and applicability in real-world settings, ultimately advancing the field of medical image analysis and oncology.

Declarations

Ethical approval Not Applicable.

Information consent Not Applicable.

Disclosure of potential conflicts of interest The authors have no financial or proprietary interests in any material discussed in this article. The authors declare that they have no conflict of interest.

References

1. Singh GAP, Gupta PK (2019) Performance analysis of various machine learning-based approaches for detection and classification of lung cancer in humans. Neural Comput & Applic 31:6863–6877
2. Shamas S, Panda SN, Sharma I (2022, July) Review on lung nodule segmentation-based lung cancer classification using machine learning approaches. In: Artificial intelligence on medical data: proceedings of international symposium, ISCMM 2021. Springer Nature Singapore, Singapore, pp 277–286
3. Said Y, Alsheikhy AA, Shawly T, Lahza H (2023) Medical images segmentation for lung cancer diagnosis based on deep learning architectures. Diagnostics 13(3):546
4. Ahmed SRA, Al-Barazanchi I, Mhana A, Abdulshaheed HR (2019) Lung cancer classification using data mining and supervised learning algorithms on multi-dimensional data set
5. Feng J, Jiang J (2022) Deep learning-based chest CT image features in diagnosis of lung cancer. Comput Math Methods Med 2022
6. Skourt BA, El Hassani A, Majda A (2018) Lung CT image segmentation using deep neural networks. Procedia Comput Sci 127:109–113
7. Sori WJ, Feng J, Godana AW, Liu S, Gelmecha DJ (2021) DFD-net: lung cancer detection from denoised CT scan image using deep learning. Front Comput Sci 15:1–13
8. Yu H, Zhou Z, Wang Q (2020) Deep learning assisted predict of lung cancer on computed tomography images using the adaptive hierarchical heuristic mathematical model. IEEE Access 8:86400–86410
9. Ozdemir O, Russell RL, Berlin AA (2019) A 3D probabilistic deep learning system for detection and diagnosis of lung cancer using low-dose CT scans. IEEE Trans Med Imaging 39(5):1419–1429
10. Elnakib A, Amer HM, Abou-Chadi FE (2020) Early lung cancer detection using deep learning optimization
11. Lakshmanaprabu SK, Mohanty SN, Shankar K, Arunkumar N, Ramirez G (2019) Optimal deep learning model for classification of lung cancer on CT images. Futur Gener Comput Syst 92:374–382
12. Kumar V, Bakariya B (2021) Classification of malignant lung cancer using deep learning. J Med Eng Technol 45(2):85–93
13. Asuntha A, Srinivasan A (2020) Deep learning for lung Cancer detection and classification. Multimed Tools Appl 79:7731–7762
14. Alakwaa W, Nassef M, Badr A (2017) Lung cancer detection and classification with 3D convolutional neural network (3D-CNN). Int J Adv Comput Sci Appl 8(8)
15. Liu H, Cao H, Song E, Ma G, Xu X, Jin R, Liu C, Hung C-C (2020) Multi-model ensemble learning architecture based on 3D CNN for lung nodule malignancy suspiciousness classification. J Digit Imaging 33:1242–1256. https://doi.org/10.1007/s10278-020-00372-8
16. Agnes SA, Anitha J (2020) Automatic 2D lung nodule patch classification using deep neural networks. In 2020 Fourth International Conference on Inventive Systems and Control (ICISC), pp 500–504. IEEE. https://doi.org/10.1109/ICISC47916.2020.9171183
17. Ali I, Muzammil M, Haq IU, Khaliq AA, Abdullah S (2020) Efficient lung nodule classification using transferable texture convolutional neural network. IEEE Access 8:175859–175870. https://doi.org/10.1109/ACCESS.2020.3026080
18. Mastouri R, Khlifa N, Neji H, Hantous-Zannad S (2021) A bilinear convolutional neural network for lung nodules classification on CT images. Int J Comput Assist Radiol Surg 16:91–101. https://doi.org/10.1007/s11548-020-02283-z