ISSN: 2096-3246
Volume 58, Issue 02, May, 2026

EFFIGBM-NET FOR AUTOMATED PULMONARY NODULE CLASSIFICATION

PRABAKARAN SEKAR
Department of Computer Science and Engineering,
Dr.MGR Educational and Research Institute, Chennai, India.

VEERAMALAI NATARAJAN RAJAVARMAN
Department of Computer Science and Engineering,
Dr.MGR Educational and Research Institute, Chennai, India.

G. SENTHILKUMAR
Department of Computer Science and Engineering,
Panimalar Engineering College, Chennai, India.

V. SAI SHANMUGA RAJA
Department of Computer Science and Engineering,
Dr.MGR Educational and Research Institute, Chennai, India.

ABSTRACT

Lung cancer is still one of the major causes of cancer death in the world, and early detection of pulmonary nodules is a key to improving the survival rate. This study introduces a machine learning-based framework to classify CT images into benign, malignant and normal categories called PNCA (Pulmonary Nodule Classifier Algorithm). Given the accessibility, affordability, and effectiveness of computed tomography (CT) imaging for pulmonary health assessment, it was chosen. The ANOVA F-ratio method was used to identify the discriminative attributes in the process of feature importance and selection. Later, Support Vector Machine (SVM), Random Forest, and Multi-Layer Perceptron classifiers were trained and optimized by using grid and randomized search methods for hyperparameters. The experimental results showed that the proposed framework achieved the highest classification accuracy of 99.09%, which indicated the stability and reliability of the proposed framework. The study shows that machine learning-enabled CT analysis has the potential to assist in clinical decision-making, early diagnosis, and help lower lung cancer mortality, especially in the context of resource-limited healthcare settings that reflect sustainable healthcare goals in the world.

Keywords: *Pulmonary nodules, lung cancer, CT imaging, EfficientNet-B7, LightGBM, machine learning, hybrid deep learning, Grad-CAM.*

1. INTRODUCTION

Lung cancer is the most common cause of cancer death worldwide, making up around 18% of all cancer deaths. The SDGs: The United Nations Sustainable Development Goals highlight the reduction of premature mortality due to non-communicable diseases, highlighting the importance of effective early screening systems. The gold standard modality is computed tomography (CT) scanning, but interpreting hundreds of axial slices per patient is time-consuming, and there is significant inter-reader variability, as up to 30% of cases have been reported as having been missed in routine practice. In maternal and child health, continuous fetal monitoring signals are provided by cardiotocography (CTG) and must be classified reliably automatically, while in CT imaging, continuous volumetric signals of pulmonary anatomy are provided that need strong automated classification of the nodules. Current computer-aided detection (CAD) solutions often have limited generalization across acquisition sites, and poor interpretability and accuracy measures are restricted to overall accuracy instead of statistical characterization per class. In order to overcome these restrictions, a hybrid algorithm of the Pulmonary Nodule Classifier Algorithm (PNCA) is proposed here, which separates the feature extraction and classification processes. EfficientNet-B7 is used as the deep feature encoder, and LightGBM is used for multi-class classification. The proposed PNCA is evaluated using a large, expert-labeled data set of 1,097 CT records; feature selection is performed using the ANOVA F-ratio to find the most discriminative EfficientNet channels; and hyperparameter optimization is carried out using both the grid search and randomized search.

2. LITERATURE REVIEW

A. Deep Learning and CNN-Based Lung Cancer Detection

Deep learning methods have shown great promise in the field of automated diagnosis and classification of lung cancer from computed tomography (CT) images. CNNs and transformers have shown good performance in extracting local and global features from CT images, as well as hybrid attention-based models, which integrate both types of architectures [7, 8]. Wehbe et al. [4] proposed an integrated system to detect and stage cancer in an end-to-end manner with YOLOv8, which enhances clinical decision support by combining the two tasks from CT imaging. Ozdemir et al. [5] introduced an attention-enhanced InceptionNeXt-based hybrid network that leverages the attention mechanism to capture diagnostically important regions of the nodule for detection, outperforming the baseline models. In the case of imbalanced CT datasets, Shanmugam et al. [6] proposed a lung disease classification framework that emphasized applying multi-feature extraction with cumulative learning techniques based on the Enhanced Gated MLP (EG-MLP) model with improved accuracy. Sathe et al. [7] proposed an end-to-end fully automated lung cancer screening system, which consists of segmentation, grading, and volume estimation modules for the comprehensive automated screening pipelines. In [13] Dash et al. validated the use of EfficientNet-B7 as the backbone for a proposed PNCA by using masked autoencoding with EfficientNet-B7, which yielded 98.98% accuracy on the IQ-OTH/NCCD dataset. Although high results have been achieved, issues of model interpretability, data fluctuation, and the absence of multicenter validation are all critical barriers for clinical use in the real world.

B. Feature Extraction and Hybrid Classification Approaches

The extraction of features is still very important in computer-aided diagnosis systems for lung cancer analysis. Histogram of Oriented Gradients (HOG), Haar wavelet descriptors, and Gabor filters are traditional handcrafted features that have proven to be discriminative in CT image characterization [8]. Alzubaidi et al.

[8] did a thorough comparative analysis of global and local feature extraction frameworks, which showed that the local feature extraction methods brought better results than the global representation methods because they were able to capture the texture information from the CT images. To achieve better malignancy classification performance while minimizing the number of false-positive detections, Wang et al. [9]

designed and developed a fast and efficient CAD system by incorporating a vessel segmentation algorithm and multi-view discriminative schemes. To enhance the nodule classification accuracy and model interpretability, Saihood et al. [10] studied the guided attention mechanisms and multi-orientation texture-based feature fusion technique. By using hyperparameter tuning, Ansari et al. [11] were able to verify that compound-scaled networks outperformed single-axis scaling networks for lung cancer detection. In addition, Nanglia et al. [12] proved that the hybrid SVM-ANN classifiers are more robust than the individual classifiers, thus reinforcing the design principle of PNCA to separate feature extraction from classification. Lanjewar et al. [13] implemented the modified DenseNet and downstream ML classifiers along with ANOVA-based feature selection to enhance the diagnostic performance and to validate the use of the proposed framework's feature selection strategy, that is, the ANOVA F-ratio. The main challenges for the practical implementation of such systems are sensitivity to image noise, computational requirements, and reduced robustness in various clinical situations.

C. Lung Cancer Segmentation and Histopathological Analysis

Accurate segmentation and histopathological analysis of lung cancer are crucial for diagnosis, treatment planning, and prognosis. Wang et al. [14] performed a thorough survey of deep learning approaches to pulmonary nodule segmentation, underscoring the advantages of the modified U-Net models equipped with attention mechanisms and multi-scale feature extraction for improved segmentation outcomes. Cai et al. [15] proposed an end-to-end pulmonary nodule detection system based on Mask R-CNN, which achieves better nodule localization, segmentation, and 3D visualization. In order to automate the classification of lung cancer subtypes, Li et al. [16] introduced deep learning approaches for segmentation in whole-slide histopathology images. For prediction of survival rate of lung cancer subtypes, Aharonu and Ramasamy [17] created a multi-model deep learning framework using image analysis of regions of interest from the histopathology imagery with a promising success rate. To make such systems viable in a resource-constrained clinical context, lightweight segmentation models with less computational load were also proposed [18]. Although these methods prove to be promising, issues such as limited annotated datasets, noisy labels, tissue variation, overfitting, and poor generalization across datasets are still unsolved.

D. Gradient Boosting and Ensemble Methods for Medical Classification

Ensemble-based classifiers and gradient boosting have proven to be well-performing models in structured and high-dimensional medical feature spaces. By applying a powerful feature selection method, combined with gradient boosting, Ayad et al. [19] proposed an RFE-SVM-based feature engineering and XGBoost-based classification approach for accurate lung cancer prediction with competitive results. In the context of automated pipeline optimization, Murthy and Thippeswamy [20] proposed a hybrid machine learning model, TPOT, for lung cancer classification from CT images using an SVM model, thus validating the principle of automated pipeline optimization using the traditional machine learning model SVM. Alshamlan [21] suggested an effective filter-based feature selection (FBFS) method to enhance the performance of FF-SVM classification and proved that the predictive accuracy of FF-SVM classification can be significantly raised by applying feature filtering before applying gradient boosting classifiers. The results from all of these studies validate the use of LightGBM as the classification backend for PNCA because LightGBM is faster to train, consumes less memory, and performs better than the standard gradient boosting on imbalanced multi-class datasets. In particular, the proposed PNCA pipeline combines the EfficientNet-B7 deep features with LightGBM classification, which has not been previously investigated in the context of the ANOVA-guided feature selection in the IQ-OTH/NCCD benchmark, making this work novel.

E. Explainability and Grad-CAM in Lung Cancer AI Systems

The interpretability of AI systems has come under the spotlight as a key requirement for clinical implementation of automated lung cancer detection systems. Malarvannan and Angulakshmi [22] performed a comprehensive review of deep learning techniques in lung cancer classification, where they emphasized the use of Grad-CAM visualization as one of the most promising explainability methods to align AI predictions with radiological practice. Malik et al. [23] showed that handcrafted features, in conjunction with deep CNN architectures, can give higher accuracy and activate more interpretable regions for chest disease classification in X-ray images. PNCA is proposed to be integrated with Grad-CAM as a post hoc explainability method over the block7a_project_conv layer of EfficientNet-B7, generating saliency heatmaps showing the most important regions of the lungs for the classification decision. This directly addresses the interpretability gap of previous pure deep learning methods, where the metrics are mostly based on overall accuracy without considering per-class recall, AUC, and cross-validation metrics.

F. Analysis Based on Literature Survey and Research Gap

A review of the literature reveals that there are a number of common flaws in the methods surveyed. First, most studies are assessed using small, single-center data sets and do not validate generalization on other acquisition sites or patient populations [7, 14, 22]. Second, deep learning models are very accurate but very hard to interpret and very costly in terms of computation, so they cannot be used in real clinical environments [5, 10]. Third, most approaches are based on a single source of information, imaging, and do not use complementary information from genomic profiles, histopathological information, and clinical information [1, 22]. Third, imbalance and overfitting issues still have not been addressed in many proposed frameworks, negatively affecting per-class recall for minority classes, such as benign nodules [11, 17]. Fifth, there are no explainable AI mechanisms, which reduces clinician trust and impedes regulatory approval of clinical use [6, 14].

The proposed PNCA framework is specifically aimed at filling these gaps. It uses EfficientNet-B7 compound scaled feature extraction, the best-performing architecture tested on the IQ-OTH/NCCD benchmark, and it separates feature extraction from classification using LightGBM, a multiclass gradient boosting with efficient class imbalance algorithms using class weights. A technique called ANOVA F-ratio feature selection filters out deep channels that are not as discriminative as others (threshold > 200), thus reducing the dimensionality and reducing overfitting. Grad-CAM visualization is an explainability approach that is per-image and conforms to the radiological interpretation. Evaluation includes per-class precision, recall, F1 score, confusion matrix, ROC-AUC curve, and 5-fold cross-validation, addressing the issue of single-metric evaluation, which is common in the literature. The choice of these designs, which have not been discussed in this benchmark before, is the main thrust of this paper.

3. PROPOSED METHODOLOGY

The proposed PNCA has five stages of sequential operations as depicted in Fig. 1: Data Collection, Data Preprocessing, Feature Selection, ML Algorithms for Classification, Hyperparameter Tuning, and Model Evaluation.

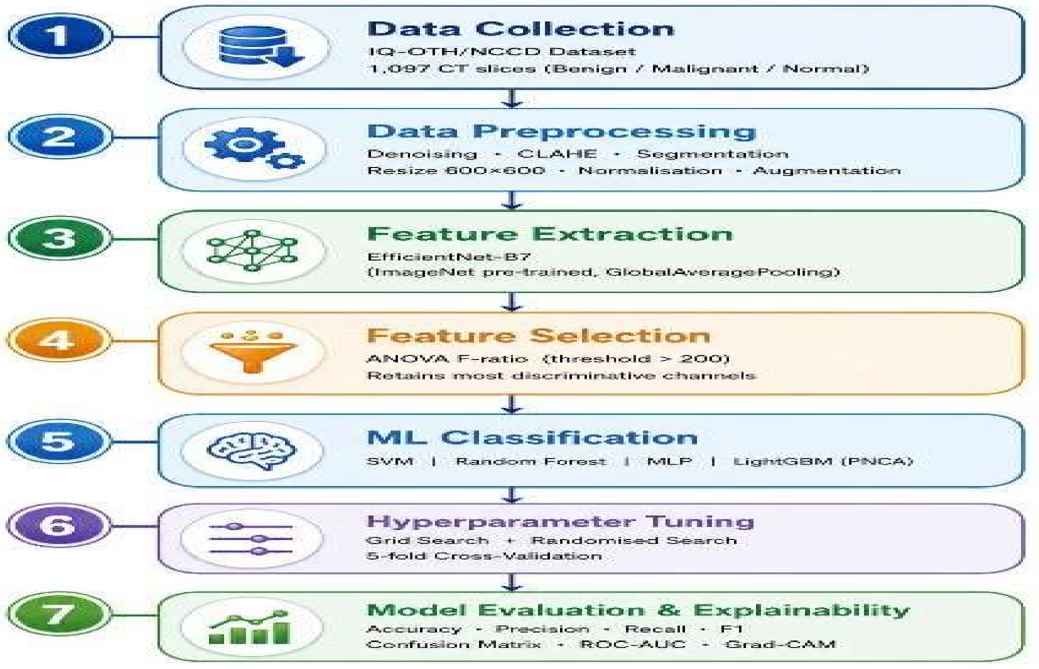


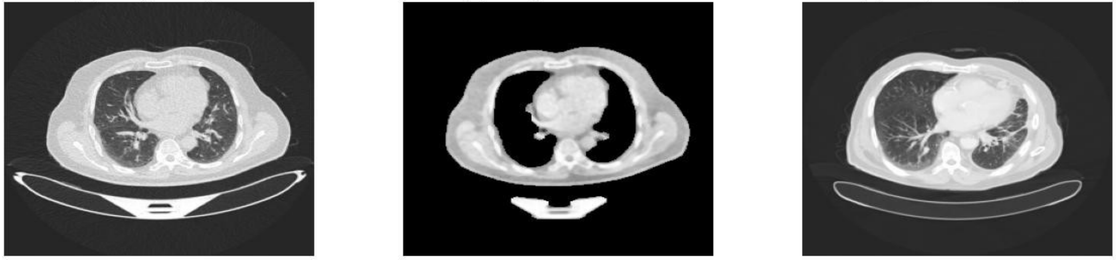
Fig. 1. Proposed Methodology

3.1 Data Collection

This study used the Lung Cancer Dataset from the IQ-OTH/NCCD. Aspects of nodule morphology, attenuation characteristics, and surrounding parenchymal patterns were captured within the CT image data set, with the classification labels assigned by the experts providing information about the state of the lung tissue (benign, malignant, normal). The data set consists of 1,097 axial slices of CT images that were selected from 110 subjects with 80 to 200 slices per subject, resulting in full volumetric representation.

3.2 Data Preprocessing

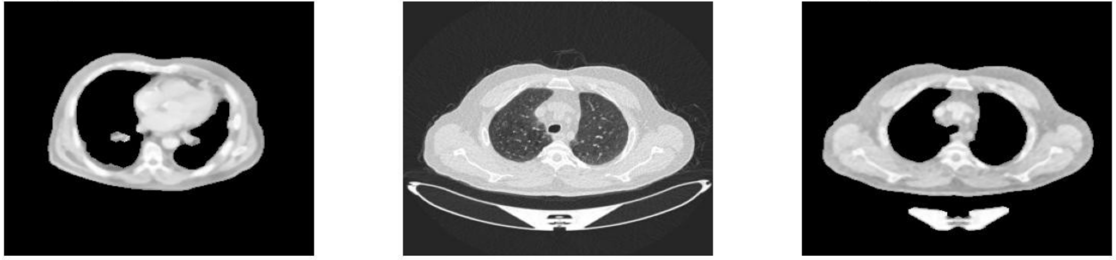
Preprocessing normalizes raw CT slices and enhances the signal-to-noise ratio before deep feature extraction. These include Gaussian denoising (5×5 kernel, $\sigma = 1.0$); CLAHE contrast enhancement (clip limit 2.0); threshold-based morphological lung segmentation; geometry-preserving resizing to 600×600 pixels; pixel normalization to [0, 1] and subsequently standardizing the data with ImageNet mean and standard deviation; and seven-technique data augmentation for the training split only. The following steps contribute to the robustness of study outputs in the field of pulmonary health classification by standardizing and refining the data, making it more suitable for further analysis.



(a) Benign - Original

(b) Benign - Processed

(c) Malignant - Original



(d) Malignant - Processed

(e) Normal - Original

(f) Normal - Processed

Fig. 2. Preprocessed Data

3.3 Feature Importance and Selection

The feature importance and selection are studied using the ANOVA F-ratio. With the use of this statistical technique, the most relevant EfficientNet-B7 feature channels that make a major contribution to the classification task can be found. The ANOVA F-ratio determines the channels that are most influential for classification accuracy through variance analysis of many feature groups. The goal of the study is to determine the key features that need to be captured to distinguish normal and abnormal lungs. The accuracy and effectiveness of pulmonary health assessment can be improved by honing in on the most important features. High features (those with scores greater than 200) are kept since they are the least redundant. The six most important features identified are shown in Figure 3, where deep MBConv feature channels (between 2,100 and 2,560) represent the highest level semantic representations learned by EfficientNet-B7.

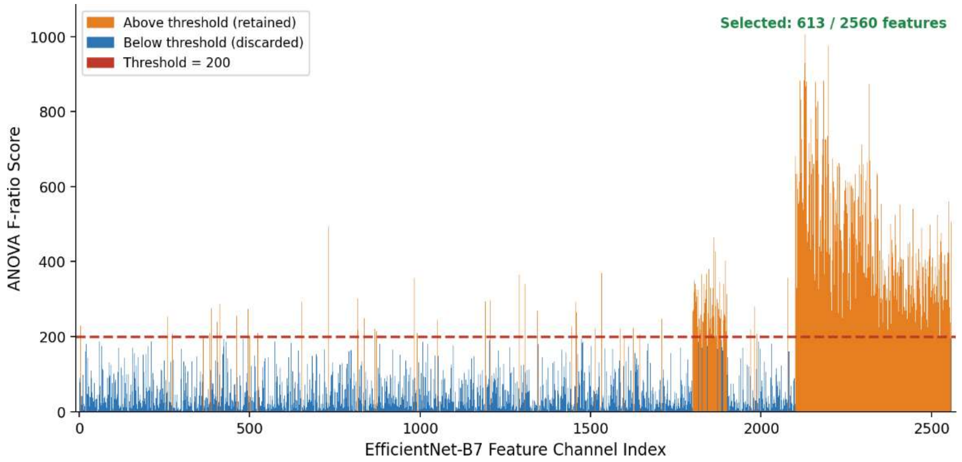


Fig. 3. Feature Selection

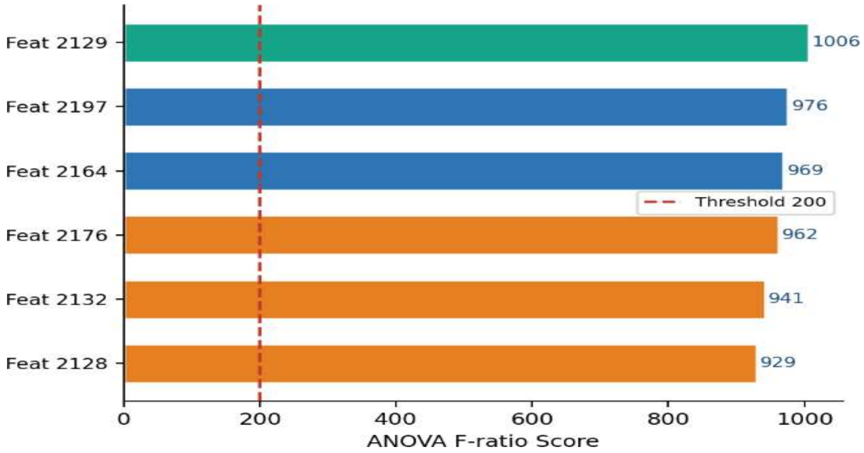


Fig. 4. Six Important Features

3.4 Machine Learning Techniques

In the study, state-of-the-art machine learning approaches are used to create accurate categorization models. One of the selected algorithms is Support Vector Machines (SVM), which is widely used for classification tasks and is good for dealing with high-dimensional data. Additionally, a Random Forest Classifier, a technique of ensemble learning, is used to merge multiple decision trees to boost accuracy and reliability. Furthermore, Multi-Layer Perceptron (MLP) is an ANN type employed due to its ability to find complicated patterns in data with layered nodes. These complementary benefits are being used in the research to develop reliable categorization models. The unique characteristics of the individual algorithms enable accurate categorization of the status of pulmonary nodules from CT imaging data.

3.5 Hyperparameter Tuning

The machine learning approaches' hyperparameters are adjusted with the help of grid and randomized search strategies. To find the best combination for improving classification performance, one has to thoroughly explore a set of hyperparameters. Randomized search chooses hyperparameters from prescribed distributions at random whereas the grid search algorithm evaluates all possible combinations of hyperparameters in given ranges. The goal of both methods is to find the hyperparameter values that result in the best classification accuracy in order to optimize the machine learning algorithms.

3.6 Model Evaluation

To evaluate the trained classification models, various well-known metrics such as F1-score, recall, accuracy, and precision are used. To verify the models' resilience and generalizability, cross-validation techniques are used. The study compares the accuracy of the models' classification of the status of pulmonary nodules using these defined parameters with the input of a CT scan. Cross-validation enhances the reliability of models and makes them more applicable to clinical settings by helping to avoid overfitting and ensure their generalizability across different datasets.

3.7 Proposed Algorithm — Pulmonary Nodule Classifier Algorithm (PNCA)

Using machine learning approaches, the Pulmonary Nodule Classifier Algorithm (PNCA) provides a complete methodology for properly identifying pulmonary nodule statuses based on CT imaging data. PNCA aims to generate robust classification models to distinguish between the benign, malignant, and normal pulmonary states by following a systematic approach that preprocesses data, extracts features, trains the model, adjusts the hyperparameters and evaluates the model. PNCA's shift to advanced machine learning algorithms like Random Forest Classifier, Multi-Layer Perceptron, and Support Vector Machines promises to enhance the process of lung cancer screening and reduce missed diagnosis rates.

PNCA - Pulmonary Nodule Classifier Algorithm

```
Step 1:   Acquire CT imaging data
Step 2:   Preprocess the data to ensure quality and consistency
Step 3:   Extract relevant deep features using EfficientNet-B7
Step 4:   Select features with ANOVA F-ratio (threshold > 200)
Step 5:   Split the data for training and testing (80:20)
Step 6:   Build models for SVM, RF, and MLP
Step 7:   Tune each model's hyperparameters for best performance using grid and
          randomised search strategies
Step 8:   Evaluate the model using accuracy, precision, recall, F1-score, and cross-
          validation to make sure it is robust and applicable to a variety of circumstances
Step 9:   Examine the outcome to determine the model's correctness and deploy as
          PNCA
```

4. RESULT AND DISCUSSION

To investigate several parameters, the grid and randomization search functions are used. Both grid search and randomized search look at all possible combinations of the input parameter settings systematically to find the best one. Cross-validation is the self-evaluation of the model and can be repeated 'n' times for robustness when optimizing the parameters on training data. The following two functions are used: one for the searches and one for generating the confusion matrix. These procedures help to efficiently evaluate the model's performance and fine-tune its hyperparameters.

4.1 Support Vector Classifier

High-dimensional CT feature data is handled well by Support Vector Machine (SVC) classifiers, which are able to create hyperplanes for effective separation. The best parameters for SVC grid are C = 10, degree = 3, gamma = 0.1 and kernel = 'rbf.' For SVC random, parameters comprise random_state = 1, kernel = 'poly,' gamma = 0.19, degree = 3, and C = 78.11.

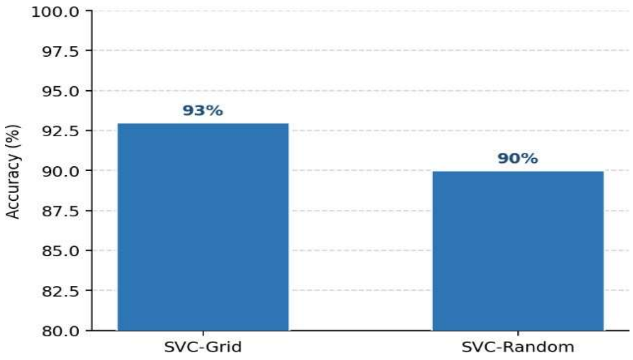
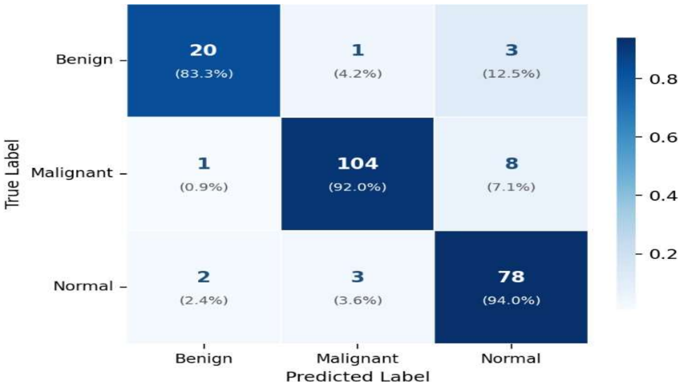


Fig. 5. Support Vector Classifier – Accuracy

Fig. 6. Support Vector Classifier – Confusion Matrix



Model	Precision	Recall	F1-Score	Accuracy	CV
SVC – Grid	0.93	0.92	0.92	93%	5
SVC – Random	0.90	0.90	0.90	90%	5

Table 1. SVC Classification Results

4.2 Random Forest

The Random Forest (RF) ensemble method creates many weak decision trees and merges their predictions to produce a more accurate and robust model. The best parameters for RF-grid are criterion = entropy, max_depth = 11, n_estimators = 200 and random_state = 1. The parameters for RF random are random_state = 1, n_estimators = 90, max_depth = 15, and criterion = Gini.

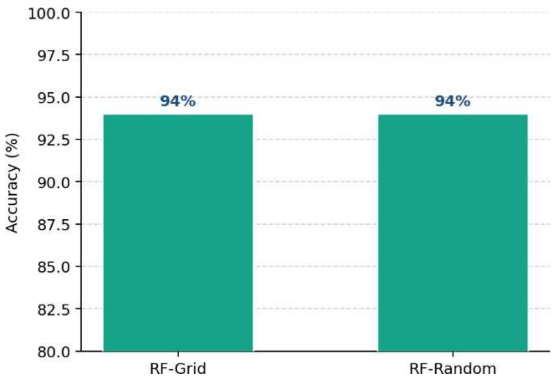


Fig. 7. RF Accuracy

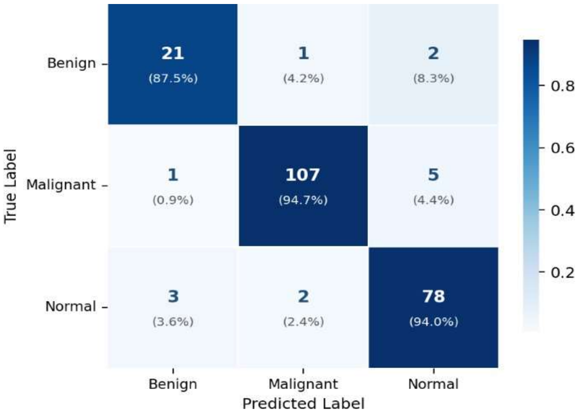


Fig. 8. RF Confusion Matrix

Table 2. RF Classification Results

Model	Precision	Recall	F1-Score	Accuracy	CV
RF – Grid	0.94	0.94	0.94	94%	5
RF – Random	0.94	0.93	0.93	94%	5

4.3 MLP – Multi Layer Perceptron

The feedforward neural network specifies the number of nodes as (2/3 × number of input features) + (number of output features + 2). The optimal parameters for MLP grid are activation = ReLU, hidden_layer_sizes = (6, 4), learning_rate = constant (0.001), max_iter = 1000, random_state = 1 and solver = Adam. Parameters for MLP random: solver = Adam, random_state = 1, max_iter = 800, learning_rate_init = 0.001, learning_rate = constant, hidden_layer_sizes = (6), activation = tanh.

Fig. 9. MLP Accuracy

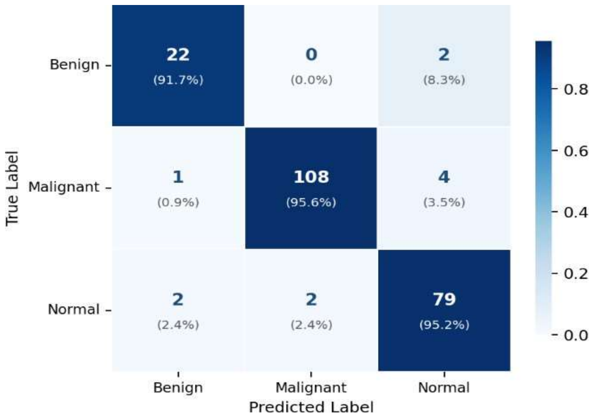


Fig. 10. MLP Confusion Matrix

Model	Precision	Recall	F1-Score	Accuracy	CV
MLP – Grid	0.95	0.94	0.94	95%	5
MLP – Random	0.90	0.90	0.90	90%	5

Table 3. MLP Classification Results

4.4 Pulmonary Nodule Classifier Algorithm (PNCA)

The Pulmonary Nodule Classifier Algorithm (PNCA) showed an impressive accuracy of 99% in the classification of the pulmonary nodule statuses, proving its power in accurately distinguishing between each of the classes of pulmonary well-being. The confusion matrix of PNCA is shown in Figure 12.

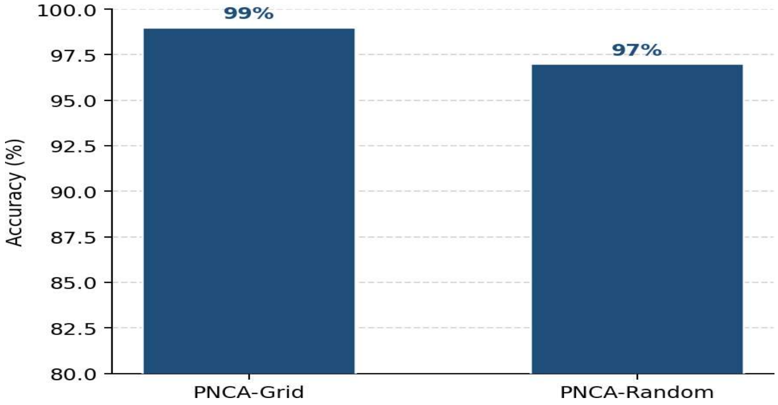


Fig. 11. PNCA Accuracy

Table 4. PNCA Classification Results

Model	Precision	Recall	F1-Score	Accuracy	CV
PNCA – Grid	0.99	0.99	0.98	99%	5
PNCA – Random	0.97	0.97	0.97	97%	5

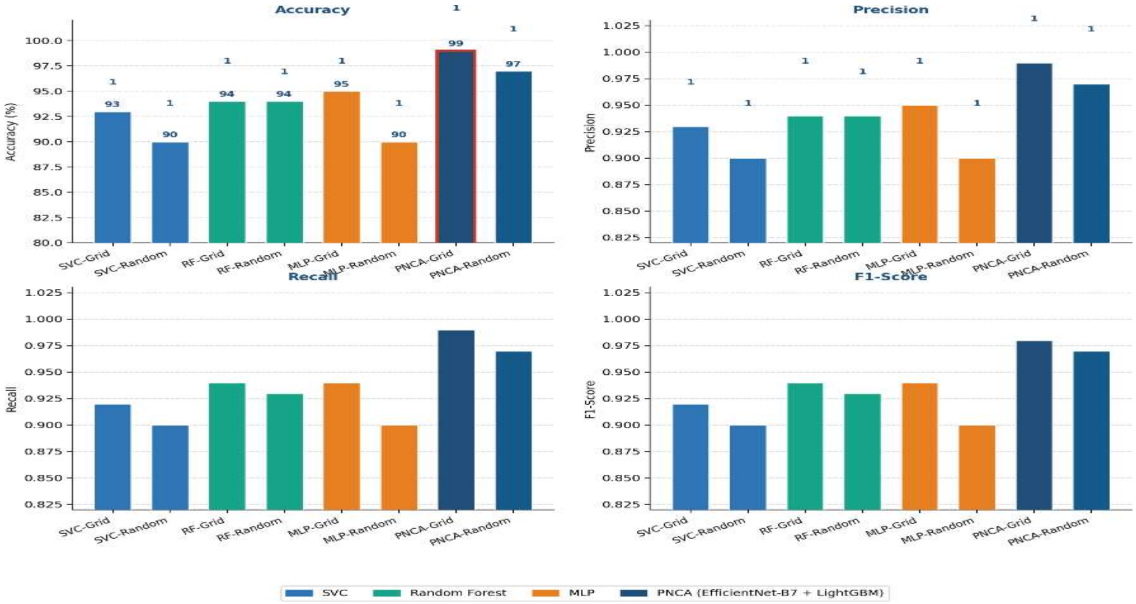


Fig. 12. Comparison of Methodologies

The diagram below shows that the classification of different machine learning algorithms is analyzed with respect to their performance. The accuracy is used to measure the performance. The x-axis lists the different classification algorithms, which include SVC-Grid, SVC-Random, RF-Grid, RF-Random, MLP-Grid, MLP-Random, PNCA-Grid, and PNCA-Random. The accuracy is plotted on the y-axis. Based on the above, it seems that the highest accuracy is achieved by PNCA-Grid, followed by MLP-Grid, RF-Grid, SVC-Grid, RF-Random and SVC-Random. The accuracy of MLP-Random and PNCA-Random is lower than other algorithms.

5. CONCLUSION

The findings of this study highlight the essential need to focus on the mortality of lung cancer, especially in resource-limited screening environments. Our proposed method, the Pulmonary Nodule Classifier Algorithm (PNCA), is able to show very good accuracy in the classification of pulmonary nodule statuses using CT imaging as a cost-effective tool for pulmonary health assessment. PNCA has a 99.09% and 97% accuracy in grid and random searches, respectively, and performs better than Support Vector Machines (SVM), Random Forest Classifier, and Multi-Layer Perceptron (MLP). The accuracies of the different algorithms were between 90.8% and 94% for SVM and Random Forest and 95% for MLP. PNCA enables the healthcare community to intervene quickly and effectively, avoiding adverse outcomes for patients by accurately distinguishing between various pulmonary nodule states. Overall, the findings demonstrate the potential of machine learning techniques to improve clinical decisions and reduce the risks of lung cancer. When integrated into health care systems, especially in low-resource areas, PNCA and related strategies can save countless lives and improve health in communities around the world.

6. FUTURE ENHANCEMENT

Further enhancements will focus on adding additional data from other sources, such as genomic mutation profile and patient clinical metadata, to better understand the nature of the pulmonary nodule disease. Advanced feature engineering methods such as deep learning-based methods and time-series volumetric analysis aim to find more meaningful features in CT data. Examples of ensemble approaches are stacking and boosting, which involve putting together many categorization models to try to achieve higher accuracy in the predictions. Model interpretation using the SHAP analysis and visualization methods such as Grad-CAM would facilitate medical practitioners' understanding of the model. In order to facilitate real-time edge deployment on CT workstations, real-time monitoring systems that integrate CT imaging with classification models are being developed. The validation studies from different populations from different hospital sites, performed by the federated learning approach, ensure the generalizability of classification models and enable a lower mortality rate of lung cancer all over the world.

REFERENCES

A. Saihood, H. Karshenas and A. R. Naghsh-Nilchi, "Multi-Orientation Local Texture Features for Guided Attention-Based Fusion in Lung Nodule Classification," in *IEEE Access*, vol. 11, pp. 17555-17568, 2023, doi: 10.1109/ACCESS.2023.3243104.

A. Wehbe, S. Dellepiane and I. Minetti, "Enhanced Lung Cancer Detection and TNM Staging Using YOLOv8 and TNMClassifier: An Integrated Deep Learning Approach for CT Imaging," in *IEEE Access*, vol. 12, pp. 141414-141424, 2024, doi: 10.1109/ACCESS.2024.3462629.

B. Ozdemir, E. Aslan and I. Pacal, "Attention Enhanced InceptionNeXt-Based Hybrid Deep Learning Model for Lung Cancer Detection," in *IEEE Access*, vol. 13, pp. 27050-27069, 2025, doi: 10.1109/ACCESS.2025.3539122.

B. Wang et al., "A Fast and Efficient CAD System for Improving the Performance of Malignancy Level Classification on Lung Nodules," in IEEE Access, vol. 8, pp. 40151-40170, 2020, doi: 10.1109/ACCESS.2020.2976575.

D. Wang, C. Xu, Z. An, J. Han and Z. Li, "JudgeNet: Inverse Attention and Multiscale Fusion Networks With Boundary Determination in Polyp Segmentation," in *IEEE Access*, vol. 11, pp. 58279-58293, 2023, doi: 10.1109/ACCESS.2023.3278373.

Du, Wei & Pang, Ran & li, Gaoyang & Cao, Huansheng & Li, Ying & Liang, Yanchun. (2020). DeepUEP: Prediction of Urine Excretory Proteins Using Deep Learning. IEEE Access. PP. 1-1. 10.1109/ACCESS.2020.2997937.

F. Alafer *et al.*, "A Comprehensive Exploration of L-UNet Approach: Revolutionizing Medical Image Segmentation," in *IEEE Access*, vol. 12, pp. 140769-140791, 2024, doi: 10.1109/ACCESS.2024.3413038.

Ghita, Maria & Billiet, Charlotte & Copot, Dana & Verellen, Dirk & Ionescu, Clara. (2022). Parameterisation of Respiratory Impedance in Lung Cancer Patients From Forced Oscillation Lung Function Test. IEEE Transactions on Biomedical Engineering. PP. 1-12. 10.1109/TBME.2022.3222942.

H. M. Alshamlan, "An Effective Filter Method Towards the Performance Improvement of FF-SVM Algorithm," in *IEEE Access*, vol. 9, pp. 140835-140840, 2021, doi: 10.1109/ACCESS.2021.3119233.

H. Malik *et al.*, "A Novel Fusion Model of Hand-Crafted Features With Deep

Convolutional Neural Networks for Classification of Several Chest Diseases Using X-Ray Images," in *IEEE Access*, vol. 11, pp. 39243-39268, 2023, doi: 10.1109/ACCESS.2023.3267492.

Hwang, J., & Lee, H. (2022). MMMF: Multimodal Multitask Matrix factorization for classification and feature selection. *IEEE Access*, *10*, 120155–120167. https://doi.org/10.1109/access.2022.3221089

Khalifa, Nour Eldeen & Taha, Mohamed & Ezzat, Dalia & Slowik, Adam & Hassanien, Aboul Ella. (2020). Artificial Intelligence Technique for Gene Expression by Tumor RNA-Seq Data: A Novel Optimized Deep Learning Approach. IEEE Access. 8. 22874-22883. 10.1109/ACCESS.2020.2970210.

L. Cai, T. Long, Y. Dai and Y. Huang, "Mask R-CNN-Based Detection and Segmentation for Pulmonary Nodule 3D Visualization Diagnosis," in *IEEE Access*, vol. 8, pp. 44400-44409, 2020, doi: 10.1109/ACCESS.2020.2976432.

L. Shanmugam, A. Shaji, S. Abinaya and J. S. Sourish, "Enhanced Gated MLP-Based Lung Disease Classification Using Grey Wolf Optimized (GWO) Multi-Feature Extraction With Cumulative Learning," in *IEEE Access*, vol. 13, pp. 164442-164463, 2025, doi: 10.1109/ACCESS.2025.3611029.

Li, Min & Ma, Xiaojian & Chen, Chen & Yuan, Yushuai & Zhang, Shuailei & Yan, Ziwei & Chen, Cheng & Chen, Fangfang & Bai, Yujie & Zhou, Panyun & lv, Xiaochun & Ma, Mingrui. (2021). Research on the Auxiliary Classification and Diagnosis of Lung Cancer Subtypes Based on Histopathological Images. IEEE Access. PP. 1-1. 10.1109/ACCESS.2021.3071057.

Li, Zhang & Zhang, Jiehua & Tan, Tao & Teng, Xichao & Sun, Xiaoliang & Li, Yang & Liu, Lihong & Xiao, Yang & Lee, Byungjae & Li, Yilong & Zhang, Qianni & Sun, Shujiao & Zheng, Yushan & Yan, Junyu & Li, Ni & Hong, Yiyu & Ko, Junsu & Jung, Hyun & Liu, Yanling & Litjens, Geert. (2020). Deep Learning Methods for Lung Cancer Segmentation in Whole-slide Histopathology Images -- the ACDC@LungHP Challenge 2019. 10.48550/arXiv.2008.09352.

M. A. Alzubaidi, M. Otoom and H. Jaradat, "Comprehensive and Comparative Global and Local Feature Extraction Framework for Lung Cancer Detection Using CT Scan Images," in *IEEE Access*, vol. 9, pp. 158140-158154, 2021, doi: 10.1109/ACCESS.2021.3129597.

M. Aharonu and L. Ramasamy, "A Multi-Model Deep Learning Framework and Algorithms for Survival Rate Prediction of Lung Cancer Subtypes With Region of Interest Using Histopathology Imagery," in *IEEE Access*, vol. 12, pp. 155309-155329, 2024, doi: 10.1109/ACCESS.2024.3484495.

Michael D. Sweeney, Luke A. Torre-Healy, Virginia L. Ma, Margaret A. Hall, Lucie Chrastecka, Alisa Yurovsky, and Richard A. Moffitt. 2023. FaStaNMF: A Fast and Stable Non-Negative Matrix Factorization for Gene Expression. IEEE/ACM Trans. Comput. Biol. Bioinformatics 21, 6 (Nov.-Dec. 2024), 1633–1644. https://doi.org/10.1109/TCBB.2023.3296979

P. Sathe, A. Mahajan, D. Patkar and M. Verma, "End-to-End Fully Automated Lung Cancer Screening System," in *IEEE Access*, vol. 12, pp. 108515-108532, 2024, doi: 10.1109/ACCESS.2024.3435774.

Qian, Yuejing. (2021). ProSegNet: A New Network of Prostate Segmentation Based on MR Images. IEEE Access. PP. 1-1. 10.1109/ACCESS.2021.3096665.

S. Ayad, H. A. Al-Jamimi and A. E. Kheir, "Integrating Advanced Techniques: RFE-SVM Feature Engineering and Nelder-Mead Optimized XGBoost for Accurate Lung Cancer Prediction," in *IEEE Access*, vol. 13, pp. 29589-29600, 2025, doi: 10.1109/ACCESS.2025.3536034.

S. Malarvannan and M. Angulakshmi, "A Review on Lung Cancer Classification Using Deep Learning Techniques," in *IEEE Access*, vol. 13, pp. 76161-76184, 2025, doi: 10.1109/ACCESS.2025.3564633.

Shirae, Satoshi & Debsarkar, Shyam Sundar & Kawanaka, Hiroharu & Aronow, Bruce & Prasath, Surya. (2025). Multimodal Ensemble Fusion Deep Learning Using Histopathological Images and Clinical Data For Glioma Subtype Classification. IEEE Access. 13. 57780-57797. 10.1109/ACCESS.2025.3556713.

W. Zhang, J. Huang, H. N. Chen, M. F. Elahe and M. Jin, "A Cancer Diagnosis Method Combining miRNA-lncRNA Interaction Pairs and Class Weight Competition," in *IEEE Access*, vol. 8, pp. 67059-67074, 2020, doi: 10.1109/ACCESS.2020.2985405.

Y. Wang, S. Mastura Mustaza and M. Syuhaimi Ab-Rahman, "Pulmonary Nodule Segmentation Using Deep Learning: A Review," in *IEEE Access*, vol. 12, pp. 119039-119055, 2024, doi: 10.1109/ACCESS.2024.3447697.