



Lung cancer detection from CT scans using modified DenseNet with feature selection methods and ML classifiers

Madhusudan G Lanjewar^{a,*}, Kamini G Panchbhai^b, Panem Charanarur^c

^a School of Physical and Applied Sciences, Goa University, Taleigao Plateau, Goa 403206, India

^b Goa College of Pharmacy, Panaji, Goa 403001, India

^c Department of Cyber Security and Digital Forensics, National Forensic Sciences University, Tripura, India



ARTICLE INFO

Keywords:

Lung Cancer
DenseNet201
MRMR
ETC
P-value

ABSTRACT

Lung cancer is a highly life-threatening disease worldwide, and detection is crucial. In this study, the Kaggle chest CT-scan images dataset was used to identify lung cancer in four categories: adenocarcinoma, large cell carcinoma, squamous cell carcinoma, and normal cell. A unique Deep Learning (DL) based method was suggested by modifying the DenseNet201 model and adding layers to the original DenseNet framework to identify lung cancer disease. Two feature selection methods were used to select the best features extracted from DenseNet201, which were then applied to various ML classifiers. The system's performance was evaluated using a confusion matrix, ROC curve, Cohen's Matthews Correlation Coefficient (MCC), Kappa score (KS), 5-fold method, and p-value. The proposed system achieved a high accuracy of 100%, an average accuracy of 95%, and a p-value of less than 0.001 after applying a 5-fold method. This study highlights the potential of using computer technology and ML methods to improve the accuracy of a lung cancer diagnosis from CT scans.

1. Introduction

The lungs are a vital organ in the respiratory system, responsible for breathing in oxygen and releasing carbon dioxide. Functioning lungs are essential for maintaining adequate oxygen levels in the body and promoting general health and well-being. Successful breathing is supported by healthy lungs that deliver enough oxygen and eliminate carbon dioxide. During physical activity, healthy lungs allow the body to use oxygen more efficiently, reducing fatigue and lowering the risk of respiratory disorders such as asthma, bronchitis, and pneumonia. In addition, healthy lungs contribute to overall health by increasing circulation, reducing heart strain, and boosting the immune system. Healthy lungs are less susceptible to injury and inflammation, which may reduce the risk of lung cancer. To maintain good lung capacity and lower the risk of respiratory disorders, it is crucial to quit smoking, reduce exposure to air pollution, and engage in regular physical activity (Young & Hopkins, 2014).

Lung cancers are illnesses that impact the lungs and affect the respiratory mechanism, one of the most common factors of death in humans worldwide. It is responsible for 13% of all recently identified cancer cases and 19% of cancer-related mortality globally (Raina &

Malik, 2015). Annually, 1.76 million individuals are killed by lung cancer. In India, lung cancer is responsible for 63,475 deaths, with a mortality rate of 5/100,000 yearly. It accounts for 7.5% of the country's cancer-related Disability-Adjusted Life-Years (Roy, 2020). 10 of the country's 27 population-based cancer registries have identified lung cancer as the most normal cancer in males, accounting for 10.4% of all cancers, and it ranks seventh in women, accounting for 4.4% of all cancers (Roy, 2020). Lung cancer is caused by the proliferation of abnormal cells that broaden and grow into a tumour. Cigarette smoking causes approximately 85% of male and 75% of female lung cancer cases. Lung cancer is one of the most mortal illnesses in developing nations, with a 19.4% mortality (Kalaivani, 2020). Annually, the global death toll from cancer and other chronic illnesses rises. People with lung cancer seemed to have the most death rate, followed by women with breast cancer, based on the WHO (World Health Organization, 2020). There are fundamental difficulties in recognizing lung tumours in patients over a couple of years, such as zero symptoms unrelated to age, patients with respiratory problems, and patients impacted by 30–40 years of cigarette smoking, all of which are essential to recognize in their early stages (Jonas et al., 2021; Wille et al., 2016).

There are various methods to detect lung cancer, such as medical

* Corresponding author.

E-mail addresses: madhusudan@unigoa.ac.in (M.G. Lanjewar), kaaminilanjewar@gmail.com (K.G. Panchbhai), panem.charanarur_tripura@nfsu.ac.in (P. Charanarur).

history and physical exam, imaging tests such as CT scans, magnetic resonance imaging (MRI) scans, positron emission tomography (PET) scans, and bone scans. Moreover, various tests are available to diagnose lung cancer, such as thoracentesis, sputum cytology, needle biopsy, and bronchoscopy (How to detect lung cancer, 2022). These tests are expensive, time-consuming, and involve a skilled individual to recognise cancer. As a result, checking for lung cancer is the primary essential step, with better detection methods that must enhance patients' health (Pradhan, Chawla, & Tiwari, 2023). Due to this, there is a need for a computerised system that can assist doctors in accurately identifying diseases. Furthermore, inaccurate care and monitoring increase the risk of death. Many scientists have tackled these difficulties in lung cancer detection using techniques known as segmentation, recognition, and prediction algorithms. Artificial intelligence (AI) has recently played an important role in computer vision (CV), big data, and healthcare applications because of its high-level recognition, forecasting effectiveness, and suitability for classification problems (Kasinathan & Jayakumar, 2022).

Medical imaging researchers are showing increasing interest in developing machine learning (ML) and deep learning (DL) techniques for diagnosing lung cancer. These technologies utilize algorithms that are trained to identify patterns in large datasets, enabling them to detect early signs of lung cancer in medical images such as CT scans, X-rays, and MRI scans. DL is a subset of ML that utilises artificial neural networks (ANN) to solve complex problems. The network learns to recognise patterns in data by identifying the relationships between variables, and it can identify the underlying factors that influence those relationships (Zhang, Yang, Lin, Ji, & Gupta, 2018). As a result, DL can extract meaningful information from complex datasets and make accurate predictions or classifications. There are various definitions of DL, which were discussed and clarify the concept of DL by Zhang et al., (2018). DL has shown great promise in detecting lung cancer from medical images, achieving high levels of accuracy and outperforming traditional machine learning algorithms. DL algorithms can identify subtle patterns in images that may not be visible to the human eye, which can help detect early signs of cancer. CNN's techniques have recently been used in medical image analysis tasks. Most interesting DL works aimed at improving the quality of medical images have either modified the architecture of established DL networks or suggested new ones. Convolutional neural networks (CNNs) and state-of-the-art pre-trained transfer learning (TL) models, such as VGG16, DenseNet201, etc., have been widely used among many DL techniques in CV applications. CNN and TL model findings have shown their reliability in object recognition localization in variations of pictures.

2. Literature survey

In this section, various kinds of literature reported ML and DL-based work to detect lung cancer will discuss. For lung image classification, Valluru and Jeya (2020) proposed a support vector machine (SVM) with a modified grey wolf optimization technique coupled with a genetic algorithm (GWO-GA). The work was divided into three stages: feature selection, parameter optimization, and optimal SVM. They used GWO-GA to optimize SVM parameters, which helped to enhance accuracy. When the current individual's optimal parameter outperforms the GA, the individual replaces the related GA, and GA resumes execution. Therefore, GA and GWO were executed alternately until the halting requirement was met. A standard image database of 50 low-dose and stored lung CT scans were used to assess the approach's efficacy and attained average accuracy of 93.54%. This study used a small dataset of 50 images and obtained an accuracy of 93.54% for three classes. The robustness was not tested using the K-fold cross-validation method. Serj, Lavi, and Hoff (2018) proposed deep CNN (DCNN)-based lung cancer detection with f1-scores, specificity, and sensitivity of 0.95, 0.991, and 0.87, respectively. They utilized the CT scans Kaggle Data Science Bowl 2017 (KDSB17) dataset. However, cross-validation was not performed

on their DCNN model. In contrast, we also used Kaggle datasets and evaluated the system's performance using several metrics. We further cross-validated our model using the K-fold approach. Song, Zhao, Luo, and Dou (2017) suggested SAE, DNN, and CNN for lung cancer detection from CT images, with changes for malignant and benign lung nodules. These networks were tested using the Lung Image Database Consortium and Image Database Resource Initiative (LIDC-IDRI) database, and 4581 photos of lung nodules were utilized for training, with 2311 and 2265 images for malignant and benign pulmonary nodules, respectively. The CNN network achieved 83.96% sensitivity, 84.15% accuracy and 84.32% specificity. The authors attempted a two-class task of recognizing benign and malignant lung nodules, but with only mediocre accuracy performance. Additionally, the K-fold approach was not employed in their study. In another study, Kuruvilla and Gunavathi (2014) used the feed-forward and feed-forward back propagation neural networks (BPNN) method to classify lung cancer from CT images. This method obtained 93.3% accuracy, 100% specificity, 91.4% sensitivity, and a mean square error (MSE) of 0.998. The K-fold approach was not employed in their study.

Tekade and Rajeswari (2018) suggested Computer-Aided Diagnosis (CAD) methods that use CT scans to recognize lung nodules and determine their malignancy level. The u-Net structure was used to segment CT scans. They used a 3D multipath VGG-like network tested on 3D cubes retrieved from the datasets: LIDC-IDRI, Lung Nodule Analysis 2016 (LUNA16), and Kaggle Data Science Bowl 2017. The outcome of u-Net and a 3D multipath VGG-like network were combined, achieving 95.60% accuracy. They utilized the Kaggle Data Science Bowl 2017 (KDSB17). However, cross-validation was not performed on their VGG-like network. In contrast, we also used Kaggle datasets and evaluated the system's performance using the K-fold approach. Kadir and Gleeson (2018) present a review article on ML-based lung cancer identification systems that help the doctor to manage incidental or screen-detected indeterminate pulmonary nodules. They also discussed the applicability of such methods to decrease variability in nodule categorization, enhance decision-making, and decrease the number of benign lesions accompanied or worked up unnecessarily. They present an overview of the primary lung cancer forecasting methods, their relative advantages and disadvantages with difficulties in developing and validating such techniques, and the path to clinical acceptance. Alakwa, Nassee, and Badr (2017) used a CT scan dataset from the Kaggle Data Science Bowl to present a CAD system for categorizing lung cancer. The modified u-Net learned on LUNA16 data was employed to identify nodule candidates in Kaggle CT scans. The output of u-Net was applied to 3D CNNs and achieved an accuracy of 86.6% on the test subset. Sun, Zheng, and Qian (2016) proposed three methods: Deep Belief Networks (DBNs), CNN, and Stacked Denoising Autoencoder (SDAE) for lung cancer diagnosis with the cases from the LIDC database and achieved accuracies of 81.19%, 79.76%, and 79.29%, respectively. Hawkins et al. (2018) suggested a random forest (RF) based malignant nodules prediction method from CT scans with an accuracy of 80%. The authors created their own dataset for detecting lung cancer, but their model had poor accuracy performance, and they did not perform cross-validation to validate their results.

Wang et al. (2018) demonstrated a 3D CNN method for identifying various types of lung cancer. Fudan University Shanghai Cancer Centre provided CT scans of 1,545 pre-invasive or invasive lung cancer patients. The scans were pre-processed using lung mask retrieval and image restoration. They achieved a sensitivity, specificity, accuracy, and AUC (area under the curve) values of 88.5%, 80.1%, 84.0%, and 89.2%, respectively. They also compared CNN results with three experienced radiologists and found that the AUC of 0.892 was greater than that of all radiologists. The authors used data from a cancer center to develop a lung cancer detection model, but they used only one metric to analyze its performance. Their model had poor accuracy performance, and they did not perform cross-validation to validate their results. Chaunzwa et al. (2021) presented a CNN-based system for recognizing non-small cell

lung cancer (NSCLC) tumour histology from a non-invasive standard-of-care CT dataset of 311 NSCLC patients in the early stages. The CNNs predicted tumour histology with an AUC of 0.71 ($p = 0.018$). They applied k-nearest neighbours (KNN) and SVM on CNN-derived quantitative features and achieved AUCs as high as 0.71 ($p = 0.017$). The paper presents the AUC and p-value metrics for evaluating the performance of the model. However, relying on these metrics alone is insufficient for evaluating the performance of a classification model. Other metrics, such as precision, recall, and F1-score, should have been included to provide a more comprehensive evaluation of the model's performance. Additionally, the model should have been cross-validated using the K-fold method to ensure the generalizability of the results. Pradhan et al. (2023) presented a CNN-based framework for lung cancer identification. Principal Component Analysis (PCA) and t-Distributed Stochastic Neighbor Embedding (t-SNE) were applied to remove features from CNN-removed features. The Best Fitness-based Squirrel Search Algorithm (BF-SSA) was used to pick the crucial components. High Ranking Deep Ensemble Learning (HR-DEL) was applied to five types of detection methods in the final phase. The developed HR-DEL had a 93.15% accuracy rate.

Similar to the proposed method, several researchers have developed hybrid techniques for detecting lung cancer. Qin et al. (2020) presented a DenseNet-based system that combines fine-grained elements from PET and CT scans. By retrieving fine-grained features from each imaging modality, the multidimensional attention mechanism was applied to efficiently minimize feature noise. The two features of feature fusion and attention mechanism were compared using quantitative assessment metrics and visualization of the DL process. They achieved an AUC of 0.92. They used only the AUC metric for evaluating the performance of the model. However, relying on only one metric alone is insufficient for evaluating the performance of the model. Other metrics, such as precision, recall, and F1-score, should have been included to provide a more comprehensive evaluation of the model's performance. Additionally, the model should have been cross-validated using the K-fold method to ensure the generalizability of the results. In contrast, we used Kaggle datasets and evaluated the system's performance using several metrics such as confusion matrix, accuracy, precision, recall, F1 score, ROC-AUC curve, MCC, KS, and p-value. We further cross-validated our model using the K-fold approach. Masood et al. (2018) proposed CADx systems for lung cancer detection. The Internet of Things (IoT) has empowered universal internet connectivity to biomedical data sets and techniques, significantly advancing CADx. The proposed method was based on the deep, fully convolutional neural network (DF-CNN), which was employed to classify each detected pulmonary nodule into 4 phases of lung cancer. CNN and DFCNet had an overall accuracy of 77.6% and 84.58%, respectively. They did not perform cross-validation to validate their results. Talukder et al. (2022) proposed a hybrid ensemble approach for feature extraction to reliably diagnose lung and colon cancer. It combines deeper extraction of features and ML with more powerful filtering for histopathology (LC25000) lung and colon cancer pictures and achieved 99.05%, 100%, and 99.30% accuracy for lung, colon, and (colon and lung) cancer, respectively. This work is similar to our proposed work. Chen, Xiong, Wu, Zhuang, and Yu (2020) presented a hybrid of SVM and Artificial Neural Network (ANN) to identify NSCLC cases with 88% accuracy. Similarly, Suresh and Mohan's (2020) suggested CNN-based method has a high capacity for acquiring nodule meaningful data with a few layers. They obtained an accuracy of 93.9%, a sensitivity of 93.4%, average specificity of 93%, and an AUC of 0.934 utilizing GAN-generated images. When compared to our work, the accuracy performance is lower. Sori, Feng, Godana, Liu, and Gelmecha (2021) used the residual learning denoising model (DR-Net) and two-path CNN to recognize lung cancer and obtained 87.8% accuracy. Pang, Zhang, Ding, Wang, and Xie (2020) classified malignant tumours from photos using image processing methods and DenseNet. To increase classification performance, the adaptive boosting technique combined numerous performances and attained an accuracy of 89.85%. Young and

Hopkins (2014) used the LIDC-IDRI databases to develop a deep neural network (DNN) with an adaptive optimization technique for diagnosing lung cancer. They used the fast nonlocal means (FNLM) filter for pre-processing, and applied the Masi entropy-based multilevel thresholding salp swarm algorithm (MasiEMT-SSA) to segregate the cancer nodules from the lung images. To extract various characteristics, they used the grey-level run length matrix (GLRLM) and identified the best features using the binary grasshopper optimization algorithm (BGOA). The approach achieved a classification accuracy of 99.17%. Shafi et al. (2022) proposed a DL-enabled SVM technique using the LIDC/IDRI database to identify lung cancer. Their CAD model can recognize physiological and pathological variations in the soft tissue of lung cancer lesions in cross-section. The approach achieved an accuracy of 94%. Faruqui et al. (2021) suggested a hybrid LungNet DCNN with a 22-layer structure and wearable sensor-based medical IoT to improve the detection accuracy from CT scans. They classified five classes with a high accuracy of 96.81%. Qian et al. (2021) proposed a faster regional CNN (faster R-CNN)-based technique for image-based polyp detection. They used a pre-processing approach to characterize polyps and then integrated the results into a faster R-CNN algorithm. The pre-processing of colonoscopy was intended to reduce the effect of specular reflections, resulting in a better image for the faster R-CNN algorithm to process. They achieved an mAP of 91.43% with the faster R-CNN (Qian et al., 2021).

As per Table 1, most researchers attempted to detect lung cancer using various ML and DL methods based on medical history or images and delivered satisfactory results. Some researchers utilised TL pre-trained models such as AlexNet and other DCNN models to recognise illnesses. Few proposed a hybrid model by combining CNN or DCNN with feature extraction methods and ML classifiers. However, previous techniques have a few limitations, such as being insufficient in early diagnosis, less effective in accuracy rate, unsuitable for stage classification, and so on. Aside from that, only a few researchers cross-validated their proposed methods, which is critical in developing any system. The goal of cross-validation is to assess an ML/DL model's capacity to forecast new data. It is employed to identify overfitting and selection bias issues and provide insight into how the model will generalise to an

Table 1
List of published research in the literature.

Authors	Methods	Performance
Valluru and Jeya (2020)	SVM with GWO-GA	Accuracy = 93.54%
Serj et al. (2018)	CNN	Sensitivity = 87%
Song et al. (2017)	CNN, DNN, and SAE	Accuracy = 84.15%
Kuruvilla and Gunavathi (2014)	BPNN	Accuracy = 93.3%
Tekade and Rajeswari (2018)	U-Net and 3D multipath VGG-like network	Accuracy = 95.60%
Kuruvilla and Gunavathi (2014)	BPNN	Accuracy = 93.3%
Tekade and Rajeswari (2018)	U-Net and 3D multipath VGG-like network	Accuracy = 95.60%
Alakwa et al. (2017)	modified U-Net with CNN	Accuracy = 86.6%
Sun et al. (2016)	CNN, DBNs, and SDAE	Accuracy = 81.19%
Hawkins et al. (2016)	RF	Accuracy = 80%
Wang et al. (2018)	CNN	Accuracy = 84.0%, AUC = 0.71
Chauzwa et al. (2021)	CNN	AUC = 0.71
Song, Zhao, Luo, and Dou (2017)	DenseNet	AUC = 0.92
Masood et al. (2018)	CNN and DFCNet	Accuracy = 84.58%
Pradhan, Chawla, and Tiwari (2023)	CNN, PCA, t-SNE, and HR-DEL	Accuracy = 93.15%
Talukder et al. (2022)	Hybrid ML model	Accuracy = 99.30%
Chen et al. (2020)	Hybrid SVM and ANN	Accuracy = 93.9% and AUC = 0.934
Sori et al. (2021)	Hybrid (DR-Net) and two-path CNN	Accuracy = 87.8%
Pang et al. (2020)	DenseNet	Accuracy = 89.85%
Faruqui et al. (2021)	hybrid LungNet DCNN	Accuracy = 96.81%

independent dataset. The findings revealed that there is an opportunity to improve classification performance.

Currently, there is a lack of research on the effectiveness of using modified DenseNet with feature selection techniques and ML classifiers to detect lung cancer through CT images. There has been insufficient research on the impact of different feature selection techniques on the performance of ML classifiers. Indeed, future research in these areas can help to improve the accuracy and efficiency of CT-scan-based systems for lung cancer diagnosis. Further investigations can be conducted to validate the proposed method on a small or large-scale dataset to test its generalizability and robustness. Additionally, this research can explore the use of other DL architectures or feature selection techniques to improve performance. It is also crucial to investigate the model's interpretability and how to integrate it into the clinical workflow to maximise its potential for improving patient outcomes. This research gap can be filled by creating a reliable structure that will help reduce computational complexity. As a result, there is a need to develop an efficient system that can detect lung cancer more accurately.

The paper aims to raise awareness of the issue by investigating it using traditional models with some improvements. Manual practices for disease identification are used by doctors, but they are time-consuming and expensive. In the medical field, improvements can help physicians detect early signs of illness. It will be beneficial to begin treatment strategies as soon as possible. The numerous types of research aid in implicating method studies and trying to implement the most approved methods and outcomes for other academic research. In this study, the authors used the technology-assisted research method. This research aims to improve existing systems by modifying pre-trained TL model structures to efficiently identify lung diseases while lowering their parameters.

The sampling method is also important in the model evaluation. Sampling is a statistical strategy that involves selecting a group of persons or things from a larger population to calculate population characteristics. Researchers can employ a variety of sampling strategies depending on the research subject, demographic factors, and resource availability. The random sampling technique picks participants at random from the population, ensuring that each participant has an equal chance of being chosen. Random sampling is frequently employed in research projects because it ensures that the sample is representative of the general population, eliminates bias, and enables statistical inference. Stratified sampling entails categorizing the population into strata or groupings based on certain criteria and then choosing participants at random from each stratum. Stratified sampling is important when a community contains multiple subgroups, and investigators want to verify that every subgroup is included in the sample. This method can minimize sampling error while increasing prediction precision. Cluster sampling involves categorizing the population into clusters by location or other variables, then randomly selecting clusters and sampling all participants inside those clusters. Convenience sampling involves selecting persons who are conveniently accessible and beneficial in an exploratory research inquiry. However, it is not representative of the population and may result in sampling bias. In this work, the author decided to use random sampling because it ensures that each member of the group has an equal chance of being chosen, eliminates bias, and guarantees that the sample is representative of the population. Random sampling is often employed in research projects because it allows for statistical inference and can be applied to large, varied groups with uncertain features. Moreover, random sampling provides precision for producing reliable population estimations.

In this work, a new hybrid method is suggested to fill the research gap. A modified DenseNet201 strategy with feature selection methods is proposed to address most of the concerns above. Modified DenseNet201 was used in this study, but despite their impressive results, TL models face unique challenges, such as balancing efficiency and computational power. On the other hand, DCNN necessitates a large amount of RAM and incurs computing expenses for categorization and verifying, making

its integration into IoT networks and single-board processors unsustainable. The authors investigate various network structure hyper-parameter combinations to find the best computational complexity and prediction performance balance. Furthermore, Extra Tree Classifier (ETC and Maximum Relevance Minimum Redundancy (MRMR) feature selection methods were applied to select the DenseNet201 extracted features. 11 ML classifiers were used to classify the lung cancer from these features.

This work uses the chest CT-Scan images dataset available on Kaggle. The dataset contains four categories, including three chest cancer types CT scans: Adenocarcinoma, Large cell carcinoma, Squamous cell carcinoma, and one Normal. As for our information, no work has been published using this dataset. To detect lung cancer from CT scans, five scenarios are presented as follows:

- i. The DenseNet201 model was modified by adding a few layers to reduce the number of network parameters, which will help to reduce the processing time.
- ii. The modified DenseNet201 was used as a feature extractor and extracted features applied to the 11 ML classifiers.
- iii. The modified DenseNet201 extracted features selected using the ETC feature selection method and then applied to the ML classifiers.
- iv. The modified DenseNet201 extracted features selected using the MRMR feature selection method and then applied to the ML classifiers.
- v. The selected features of the ETC and MRMR were combined and then applied to the ML classifier to classify lung cancer.

Apart from this, the main contribution of this work includes:

- Creating a practical approach for lung cancer recognition from CT scans by applying a modified DCNN model (DenseNet201)
- The DenseNet201 extracted features were used for classification.
- Two feature selection methods (ETC and MRMR) were applied to select the features from modified DenseNet201 extracted features.
- The ETC and MRMR features combined.
- 11 ML classifiers were used to classify the lung cancer from DenseNet201 extracted features, ETC selected features, MRMR selected features, and combined (ETC and MRMR) features.
- A confusion matrix, AUC-ROC, MCC, KS and p-value were used to evaluate model performance.

3. Material and methods

This section covers image pre-processing, modified DenseNet201 feature extraction, feature selection, and model performance evaluations. Fig. 1 shows the flow diagram of the suggested framework, and its steps are as follows:

- I. **Step-1:** The authors select lung cancer chest CT scans datasets to conduct the experiment. Four lung cancer CT scan images, including normal, are available in the dataset.
- II. **Step-2:** In this step, the pre-processing was done by re-dimensioning the dataset photos to 224x224. After that, the authors perform feature scaling using the image's generalisation method by dividing image array values by 255.0 and accomplish labelling by allocating a label (0, 1, 2, and 3) for each image using the label encoder.
- III. **Step-3:** The dataset was split into two parts: Train and Test, with a 90:10 ratio. The DenseNet201 basic model was modified by adding two layers: a Max pool with pool size 2 and a drop-out layer with a drop probability of 0.2. The Train subset was again divided into the train, and Val subsets and modified DenseNet201 were trained and validated these subsets. The 10-fold cross-

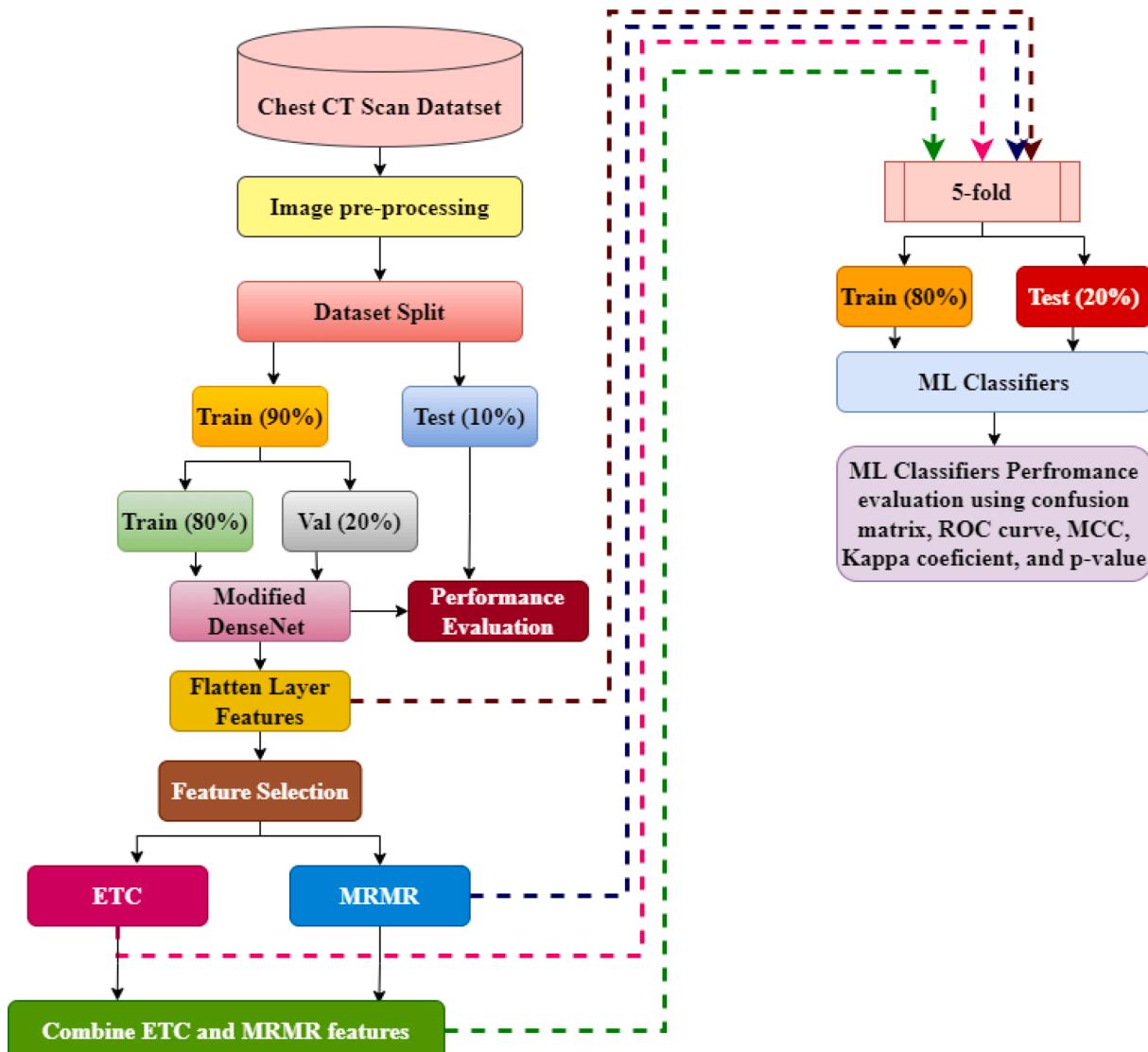


Fig. 1. Proposed system framework.

- validation technique was employed to cross-validate the model. The DenseNet201 extracted features collected for the next step.
- IV. *Step-4:* The modified DenseNet201 extracted features were applied to the 11 ML classifiers. After that, the performance metrics such as confusion matrix, f1-score, accuracy, recall, precision, MCC, KS, AUC score, ROC Curve and p-value were used to evaluate all 11 ML models. A 5-fold method was applied to cross-validate the 11 ML models.
- V. *Step-5:* The modified DenseNet201 extracted features were selected using the ETC feature selection method. These selected features were applied to the 11 ML classifiers, and performance was evaluated similarly to Step 4.
- VI. *Step-6:* The modified DenseNet201 extracted features selected using the MRMR feature selection method. These selected features were applied to the 11 ML classifiers, and performance was evaluated similarly to the Step-4.
- VII. *Step-7:* ETC and MRMR selected features combined and then applied to the 11 ML classifiers, and performance was evaluated similarly to Step-4.
- VIII. *Step 8*-The comparison analysis of Step3 to Step 7 was presented, and the performance was assessed based on each step's accuracy and selecting the best method.

3.1. Dataset description

The chest CT-Scan images dataset from Kaggle was used in this work ([Chest ct-scan images dataset, n.d.](#)) It was an initiative about detecting chest cancer utilising ML and DL to categorise and identify cancer patients. The CT scans were gathered from various sources and cleaned in preparation for ML or DL models. There were 1008 CT scans in .jpg and .png format, and they contained three types of chest cancer: adenocarcinoma (338 CT scans), large cell carcinoma (187 scans), squamous cell carcinoma (206 scans), and one normal cell (223 scans). [Fig. 2](#) depicts the CT scan images of all classes.

Adenocarcinoma: Adenocarcinoma is perhaps the most frequently diagnosed cancer, accounting for 30% of all case scenarios and approximately 40% of all non-small cell lung cancers. It can be discovered in various breast, prostate, and colorectal cancers. Adenocarcinomas of the lung are found in the glands that secrete mucus and assist us in breathing. Coughing, difficulty breathing, losing weight, and lack of strength are all symptoms.

Large cell carcinoma: This type of lung cancer starts to grow and spread rapidly and can be discovered anywhere within the lung and reports 10–15% of all cases.

Squamous cell carcinoma: This type of disease is observed in the centre of the lung and it accounts for approximately 30% of all non-small

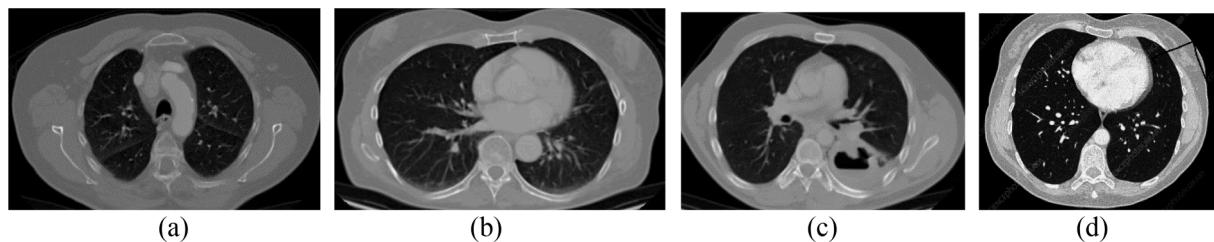


Fig. 2. CT scan images of a) adenocarcinoma, b) large cell carcinoma, c) Squamous cell carcinoma and d) normal.

cell cancers.

3.2. Pre-processing

The input image's size was resized to 224x224, and a scaling process was applied to normalize the value of image pixel intensities within the range of 0 to 1. The image was scaled by dividing the image array by 255.0 to minimize total complexity. Labelling folders with the label encoder was essential since it enabled us to recognize each class.

3.3. Modified DensNet201 feature extraction

Densely Connected Convolutional Networks (DenseNets) simplify the link structure developed by previous designs such as Highway Networks, Residual Networks, and Fractal Networks. DenseNets mix the image features of the layer with the receiving ones instead of summarising those. DenseNets are separated into Dense Blocks, in which the component sizes inside a block stay unchanged; however, the filters vary. Transition Networks in DenseNet manage downward sampling using batch normalisation, 1×1 convolution, and 2×2 pooling layers (Ruiz, 2018). TL applies knowledge gained from one problem to new, related concerns. The authors evaluate existing TL pre-trained models, select the best model that provides the best performance, and then modify it to meet our requirements. Instead of using simple DenseNet201 models, the authors attempted to change them by adding a few layers to the original DenseNet201 framework. As a result, the trainable parameters were reduced, which aids in reducing computational complexity and execution speed. In this paper, the authors use the DenseNet201 TL model to analyse lung cancer datasets. DenseNet201 extracted features were used in various ML models. DenseNet201 extracted features selected using two feature selection methods before applying to ML models. Furthermore, the feature selected by the two methods were merged and applied to the ML classifiers. Finally, evaluate the results of all scenarios and recommend the best strategy for detecting lung cancer from CT images.

The DenseNet201 model was enhanced by adding one Max-pooling layer (MPL) pool size of 2, one dropout layer, one flattened layer, and

one dense layer. Finally, using the TL method, this enhanced DenseNet201 model was trained. The MPL was a critical process that computed the maximum value in each patch of each feature map. Applying a pooling process is comparable to using a filter on feature maps. The pooling or filtering procedure is significantly smaller than the feature map. The utilization of MPL with pool size 2 in this work indicates that the MPL decreases the number of each feature map by a factor of two. Fig. 3 depicts the modified DenseNet201 feature extraction method.

The DensNet201 architecture consists of 33 components(blocks). Block0 contains zero padding (ZP), batch normalisation (BN) layer, convolutional layer (CL), zero padding (ZP), ReLu, and MPL, whereas Block1 contains BN, ReLu, CL, BN, ReLu, CL, BN, ReLu, CL, BN, ReLu, CL, and Concatenate layers. Block0 has extracted 56x56x64 features vectors from the input CT scans. The feature size retrieved by Block1 from the supplied components of Block0 was 56x56x96. Similarly, the retrieved feature size for Block2 was 56x56x128. This feature extraction process was maintained until Block32, with a total of 7x7x1920 extracted features. After Block 32, one MPL was introduced, resulting in a 3x3x1920 output feature vector. These removed MPL features were given to a drop-out layer; the resultant feature was 3x3x1920. These characteristics were then flattened at the flattened layer, yielding 17280 output features. These retrieved features were used for classification using the dense layers.

The advantage of modifying the existing DenseNet201 is that the original DenseNet201 has a total parameter of 18,698,308 and a trainable parameter of 376,328. The flattened layer output features were 94,080. On the other hand, in the modified DenseNet201, the total parameter was reduced to 18,391,108, and the trainable parameter to 69,124. The flattened layer output feature was also reduced to 17,280 from 94,080. Due to this, the authors used fewer flattened layer parameters to train and validate the ML classifiers.

3.4. Feature selection methods

In this work, two feature selection methods were applied to select the best features from the DenseNet201 extracted features.

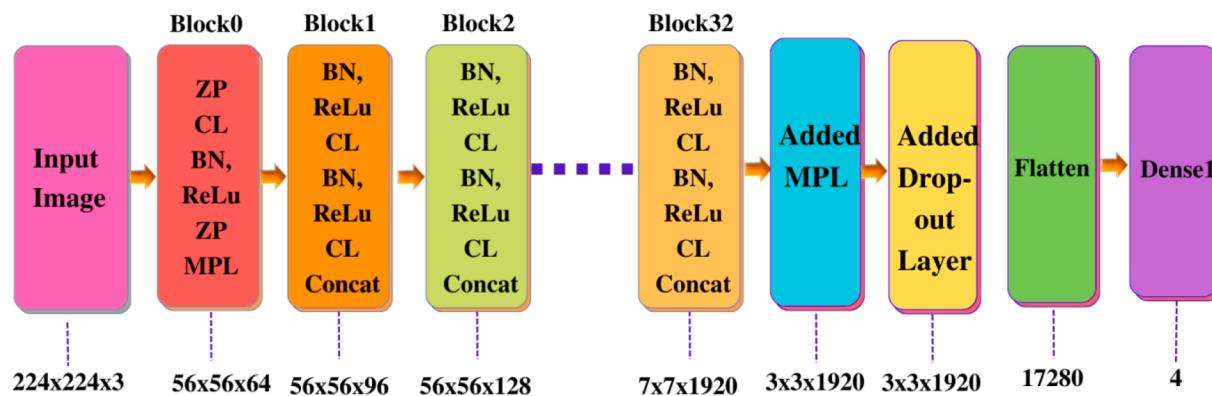


Fig. 3. DenseNet201 feature extraction process.

MRMR: The MRMR (Radovic, Ghalwash, Filipovic, & Obradovic, 2017) is an attribute selection method that favours variables with a strong correlation with the output but a weak correlation among themselves. The F-statistic can be employed to estimate correlation with the relevance for continuous characteristics, while the Pearson correlation coefficient can be utilised to estimate correlation across features. Next, attributes are chosen one at a time using a greedy algorithm to maximise the objective function, a component of redundancy and relevance (Radovic et al., 2017).

Extra Tree Classifier (ETC): The ETC is an approach that generates numerous tree models built randomly from the training data and separates out the best-voted-for attributes. Instead of a bootstrap replication, it makes every decision tree (DT) on the entire dataset and chooses a break position randomly to separate the nodes (Ceballos, 2020).

3.5. ML models

Various ML algorithms were used in this study to classify the lung cancer from the DenseNet201 extracted features or ETC or MRMR selected features, and a brief description is as follows:

Logistic Regression (LR): Despite the name, LR is a method of classification instead of a regression model (Subasi, 2020). LR is more effective process for the binary and linear multiclass type. The LR mode employs a statistical tool for binary classification that can be extended to multiclass groups. Scikit-learn includes a highly efficient LR version that handles multiclass classification tasks.

Support vector machine (SVM) (Sanaeifar, Bakhshipour, & de la Guardia, 2016): SVM is a supervised method that has lately gained popularity, widespread application, and investigation due to its capacity to forecast classification and regression. The SVM algorithm works by non-linearly mapping the input variables into a high-dimensional feature space in which they are linearly associated with the output variable.

Random Forest regressor (RF): The RF approach first generates a large number of decision trees (DTs), then these learners are used to vote on the test samples, and finally, the final choice is determined utilising the most dominant rule. The RF model selection process involves selecting a present number of good classifiers from the many DT classifiers developed and then using these classifiers to vote on test samples. The RF technique is divided into two primary phases. The first step is constructing many DT classifiers, which are then integrated into voting (Zhu & Zhang, 2022).

K-Nearest Neighbor (KNN): KNN is a supervised learning method that prepares attributes to forecast a new testing data category (Arslan & Arslan, 2021). The KNN is a non-parametric predictor and one of the easiest and laziest methods. In the forecasting procedure, the data class corresponding to new observations determines the shortest route from the observed samples to the K-nearest data (Lanjewar, Parab, Shaikh, & Sequeira, 2022). Different distance functions, such as Manhattan distance (MD), Minkowski distance (Md), and Euclidean distance (ED), can be used to estimate the closest distance.

Decision Tree (DT): A DT is a tree-like design architecture in which an internal node signifies features and a branch provides a decision function (Lanjewar, Parate, & Parab, 2022). This flowchart-like structure, it aids decision-making. The DT is a method that is independent of probability distribution assumptions, sometimes known as a nonparametric method. DT can manage high-dimensional information with high reliability. The first DT method divides the records by selecting the best attribute using the Attribute Selection Measures (ASM), making these chosen characteristics decision nodes, and breaking the dataset into subsets. Then, begin building a tree by repeating this procedure recursively on each child until one of the conditions is met, such as all tuples having the same attribute value.

Gaussian Naive Bayes (GNB): Nave Bayes is a stochastic ML technique based on the Bayes theorem that is utilised in many classifiers. The Bayes' theorem is a mathematical formula for estimating conditional

probabilities, which estimates the likelihood of one event happening provided that yet another event has happened (Naïve bayes algorithm: Everything you need to know, n.d.).

$$P(X|Y) = \frac{P(Y|X) \cdot P(X)}{P(Y)} \quad (1)$$

Where $P(X|Y)$ and $P(Y|X)$ - probability of X occurring give evidence Y has already occurred and Y occurring give evidence X has already occurred, respectively. $P(X)$ and $P(Y)$ probability of X and Y occurring, respectively,

4. Results

This section will compare the performance of modified DensNet201, DenseNet201 extracted features, ETC selected features, MRMR selected features and combined ETC with MRMR chosen features. The Adam optimizer and categorical-crossentropy loss function were employed with batches size 32.

4.1. Experimental setup and performance measures

The studies were conducted on a Google Colab PRO online platform with a Python3 Google compute engine backend (GPU) with 25.45 GHz RAM and 166.75 GB disc space. The Python programming language and various standard libraries, including NumPy, pandas, sklearn, utils, Matplotlib, Seaborn, TensorFlow, Keras, glob, etc., were used to build the proposed model.

The suggested system was evaluated using a variety of performance metrics, including accuracy, recall, precision, f1-score, AUC-ROC curve, MCC, KS, p-value, and cross-validation with the K-fold approach. The following metrics for evaluating performance were proposed:

Confusion matrix: The proposed system effectiveness was accessed using confusion matrix (Lanjewar, Morajkar, & Parab, 2022) and the stratified K-fold cross-validation techniques (Lanjewar & Panchbhai, 2023)

$$\text{Accuracy}(A_L) = \frac{\text{TP}_L + \text{TN}_L}{\text{TP}_L + \text{FP}_L + \text{TN}_L + \text{FN}_L} \quad (2)$$

$$\text{Precision } (P_L) = \frac{\text{TP}_L}{\text{TP}_L + \text{FP}_L} \quad (3)$$

$$\text{Recall } (R_L) = \frac{\text{TP}_L}{\text{TP}_L + \text{FN}_L} \quad (4)$$

$$\text{F1-score } (F1_L) = \frac{2}{\left(\frac{1}{R_L}\right) + \left(\frac{1}{P_L}\right)} \quad (5)$$

Where A_L -Accuracy, TN_L -Truly Negative, TP_L Truly Positive, FP_L -False Positive, and FN_L -False Negative.

AUC-ROC curve: The ROC curve visually depicts the ML/DL models' binary-class or multi-class testing abilities. For better classification results, the ROC curve should be closer or on to the upper left corner. It displays how well the model can distinguish between different classes. The true positive rate (TPR) is plotted on the y-axis, while the false positive rate (FPR) is plotted on the x-axis. ROC curves are charts frequently used to analyse and evaluate ML classifier efficacy (Fawcett, 2003). ROC diagrams clearly show a classifier's sensitivity or specificity swap for all possible classification thresholds, allowing for classifier evaluation and selection based on particular user needs, typically associated with changeable mistake costs and accuracy assumptions (Fawcett, 2003). The AUC represents the level of differentiation, whereas the ROC is a probability curve. AUC close to 1 suggests that a forecast model is effective at separability between class labels, while AUC close to 0 means a poor projected model.

K-fold Cross-Validation: K-Fold is a cross-validation strategy that helps check the ML model's robustness by dividing the data into K-

subsets and executing the holdout procedure K-times, with each K-subsets serving as the Val subset and the remaining K-1 serving as the training set. The average performance from all k attempts is then calculated, which is more trustworthy than the traditional holdout approach. The advantage of this method is there is no need to worry about how data is partitioned. More importantly, each CT scan image will participate in the training and validation process due to this process. In this work, for testing the DensNet201 model for the first scenario, the authors applied a 10-fold method because the original dataset was first split in Train and Test with a 90:10 ratio, and then the 10-fold method was used on the Train subset. So effectively, the dataset was divided into three parts with an 82:08:10 ratio. For the remaining scenarios, the 5-fold method was applied.

Mattheus Correlation Coefficient (MCC): It is a considerably more reliable statistical method that produces a high score only when the prediction performs satisfactorily for all 4 confusion matrix categories, as determined by the number of positive and negative items present in the dataset (Grandini, Bagli, & Visani, 2020; Lanjewar, Parab, & Shaikh, 2022).

Cohen kappa score (KS): It's a reliable measure that can be employed to evaluate interrater and intrarater accuracy (McHugh, 2012). Cohen suggested reading the Kappa value as 0 for no agreement, 0.01–0.20 for little to no agreement, 0.21–0.40 for reasonable agreement, 0.41–0.60 for moderate agreement, 0.61–0.80 for substantial agreement, and 0.81–1.00 for nearly total agreement (Grandini et al., 2020).

In this work, 11 ML models were applied to classify lung cancer from CT scans, as listed in Table 2.

4.2. Performance evaluation

4.2.1. Scanrio1: DenseNet201 performance on Val and Test subsets

The dataset was split into two parts: Train (90%) and Test (10%). To check the DenseNet201 performance, the 10-fold cross-validation method was applied to the Train subset. Fig. 4 shows the accuracy (Test_A), precision (Test_P), recall (Test_R), and f1-score (Test_F) performance of all models on the Test subset with validation accuracy (Val_A). The DenseNet201 achieved the highest Val_A of 97.9% and 96.6% average accuracy after applying the 10-fold method. The highest Test_A, Test_P, Test_R, and Test_F of 92%, 92.5%, 92.3%, and 91.8%, while average Test_A, Test_P, Test_R, and Test_F of 85.8%, 88.9%, 86.2%, and 86.3%, respectively on the Test subset. The performance of DenseNet201 was good on the Val sub-set, but performance on the Test sub-set was poor.

4.2.2. Scenario2: DenseNet201 extracted features with ML classifier

In this scenario, DenseNet201 was used as a feature extractor. DenseNet201 extracted features were applied to the 11 ML classifiers and checked the performance. The 5-fold method was applied to cross-validate all ML models. Table 3 depicts the performance of the ML classifiers with DenseNet201 extracted features. The LR1, LR2, LR3, and

Table 2

11 ML models with parameters and short name.

Models	Solver/Kernel	Short name
Logistic Regression	solver = saga	LR1
Logistic Regression	solver = newton-cg	LR2
Logistic Regression	solver = lbfgs	LR3
Logistic Regression	solver = liblinear	LR4
Support Vector Machine	kernel = rbf	SVM1
Support Vector Machine	kernel = linear	SVM1
Support Vector Machine	kernel = poly	SVM1
Decision Tree	—	DT
K nearest neighbour	—	KNN
Random Forest	—	RF
Gaussian Naive Bayes	—	GNB

SVM2 achieved the highest accuracy of 99%, followed by SVM1 (98%), LR4 (96%), SVM3 (92%), RF (92%), KNN (77%), GNB (72%), and DT (70%). The highest average accuracy (A_a), precision (A_p), recall (A_r), and F1-score (A_f) of 95%, MCC of 93% and KS of 93% were achieved by LR2, LR3, SVM1, SVM2. This performance indicates that the proposed method's performance is improved by using DenseNet201 extracted features with ML classifiers.

4.2.3. Scenario3: DenseNet201 with ETC extracted features

The DenseNet201 extracted features were selected using the ETC feature selection method in this scenario. In place of using all DenseNet201 extracted features, the author used ETC decided 1500 best features (Fig. 5), and these ETC selected features applied to ML classifiers for classification. This paper includes feature significance analysis, which scores input characteristics based on their significance in predicting the outcome. The higher the score, the more important the characteristic is in predicting the output. These scores have various applications in predictive modelling challenges, such as better understanding the data, interpreting a model, and reducing the number of input characteristics.

Table 4 depicts the performance of all models on ETC-selected features. The SVM1 and SVM2 achieved the highest accuracy (High_A) of 100% and the highest average accuracy, precision, recall, F1-score, MCC, and KS of 95%, 95%, 96%, 95%, 93% and 93%, respectively. The performance of all models with ETC selected features was slightly reduced compared to the previous scenario except for SVM1 and SVM2 models, but the highest accuracy improved to 100%. This performance indicates that the Scenario1 performance is almost similar to the SVM1 model.

The AUC-ROC curve of all models is shown in Fig. 6. The highest average AUC score of 0.9975 was achieved by LR2 (Fig. 6b), LR3 (Fig. 6c), LR4 (Fig. 6d), and SVM3 (Fig. 6g). These models have an AUC score of 1 for three classes large_cell_carcinoma (B), normal (C), and squamous_cell_carcinoma (D), while 0.99 AUC score for adenocarcinoma (A). All these models' performance is remarkable. The LR1 and SVM2 performance are also good an AUC of 0.98 for class A and 1 for the remaining three classes (B, C, and D).

All 11 ML models in this study were subjected to the p-value statistical technique. The p-value is the likelihood of obtaining analysis results based on the actual distribution. The p-value indicates how confidently the proposed model can respond to an applied CT scan image. The permutation test score was produced, which highly predicts the labels and the completely random features and dataset labels, implying that features and labels must be independent. The parameters used to compute the permutation test score were accuracy, stratified 2-fold, and 1000 permutations. At each round, the accuracy of all classifiers will be assessed. The permutation test score provides a NULL distribution by measuring the classifier's accuracy on 1000 distinct permutations of the ETC selected features, where features remain constant but label change. This is the null hypothesis distribution, which states no interdependence between the features and labels. The accuracy score percentage of permutations acquired is higher than the assigned score used to compute an estimated p-value.

A histogram plot of the permutation scores is shown in Fig. 7. The red line is the classifier's accuracy score on the ETC-selected features. SVM1 (Fig. 7e) and SVM2 (Fig. 7f) received the highest accuracy score of 0.95 with a p-value of 0.001, followed by LR1 (0.94), LR2 (0.94), LR3 (0.94), and LR4 (0.94). The score is significantly higher than that achieved utilizing permuted data, and the p-value is consequently very low (0.001). This suggests that it is unlikely that this score was reached solely by chance. It demonstrates that the ETC chosen features have a real relationship between features and labels, and the classifier was able to use this to achieve excellent results, particularly SVM1 and SVM2.

4.2.4. Scenrio4: DenseNet201 with MRMR selected features

In this scenario, the DenseNet201 extracted features were selected

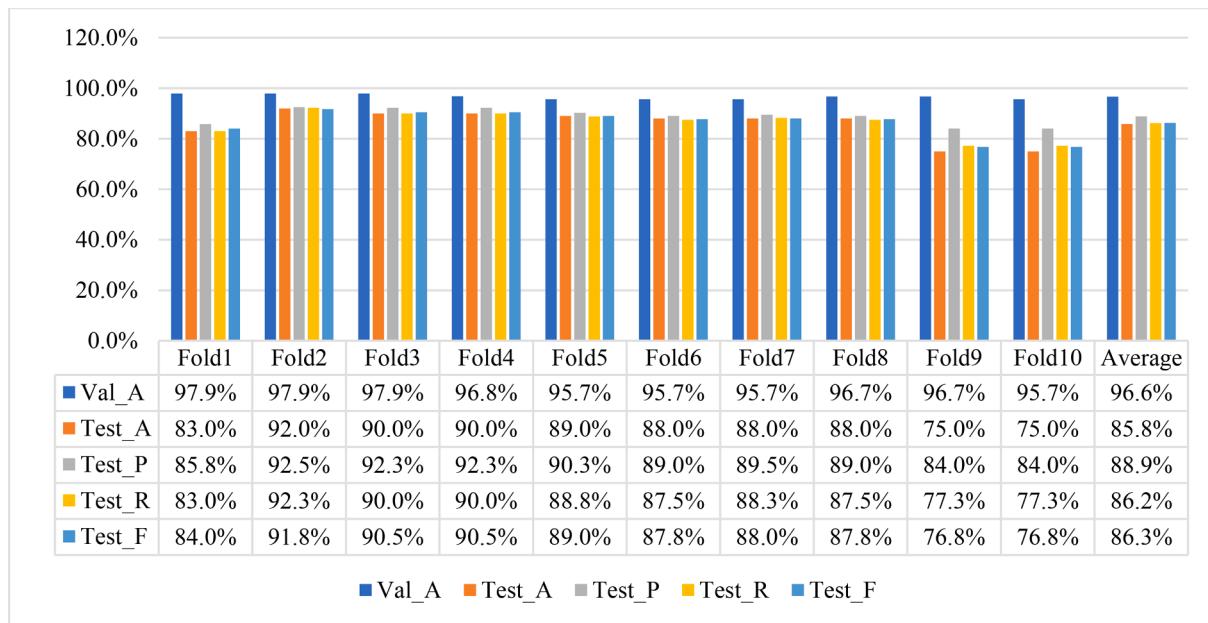


Fig. 4. DenseNet201 performance on Val and Test subsets.

Table 3
Performance of the ML classifiers with DenseNet201 extracted features.

Models	High_A	A _a	A _p	A _r	A _f	MCC	KS
LR1	99%	94%	95%	94%	94%	92%	92%
LR2	99%	95%	95%	95%	95%	93%	93%
LR3	99%	95%	95%	95%	95%	93%	93%
LR4	96%	92%	93%	93%	92%	90%	90%
SVM1	98%	95%	94%	95%	95%	93%	93%
SVM2	99%	95%	95%	96%	95%	93%	93%
SVM3	92%	85%	82%	92%	84%	81%	79%
DT	70%	68%	68%	68%	56%	56%	56%
KNN	97%	92%	92%	93%	92%	89%	89%
RF	92%	82%	80%	88%	81%	76%	75%
GNB	71%	65%	65%	68%	65%	53%	52%

Table 4
Performance of the ML classifiers with ETC selected features from DenseNet201 extracted features.

Models	High_A	A _a	A _p	A _r	A _f	MCC	KS
LR1	99%	93%	93%	93%	93%	90%	90%
LR2	99%	93%	93%	93%	93%	90%	90%
LR3	99%	93%	93%	93%	93%	91%	91%
LR4	97%	93%	93%	93%	93%	90%	90%
SVM1	100%	95%	95%	96%	95%	93%	93%
SVM2	100%	94%	94%	95%	94%	92%	92%
SVM3	98%	90%	88%	94%	90%	87%	86%
DT	72%	65%	66%	66%	66%	52%	52%
KNN	96%	91%	90%	93%	91%	88%	87%
RF	89%	83%	82%	88%	83%	78%	77%
GNB	84%	74%	73%	79%	74%	65%	64%

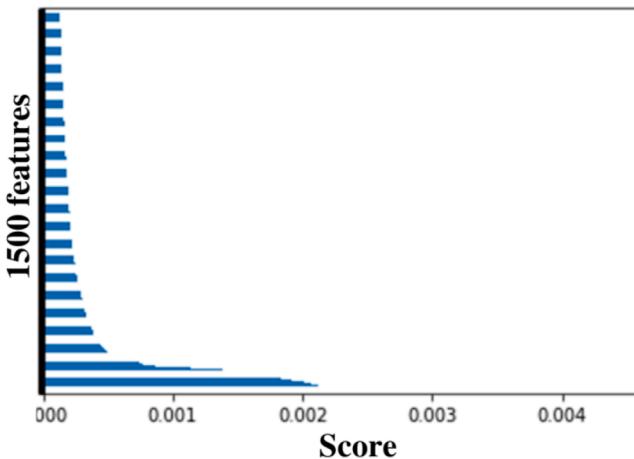


Fig. 5. ETC feature importance analysis.

using the MRMR feature selection method, just like in the previous scenario2. Instead of using all extracted features, MRMR selects 1500 best features (Fig. 8), which were applied to ML classifiers. Table 5 depicts the performance of all models on MRMR-selected features. The SVM1 achieved the highest accuracy of 95% and the highest average A_a, A_p, A_r, and A_f of 93%, MCC of 90% and KS of 90%. All other models'

average accuracy was less than 91%. The performance of all models with MRMR selected features is poor compared to the previous Scenario2 (ETC selected features). This performance indicates that scenario2 performance was better compared to scenario3.

The AUC-ROC curve of all models of this scenario3 is shown in Fig. 9. The highest average AUC score of 0.9925 was achieved by LR1 (Fig. 9a), LR2 (Fig. 9b), LR3 (Fig. 9c), and RF (Fig. 9j). These models have an AUC score of 1 for three classes B, C, and D, while 0.97 AUC score for class A. All these models' performance is good.

Fig. 10 shows a histogram plot of the permutation scores and p-value performance on MRMR-selected features. The red line indicates the accuracy score obtained by the classifier on the MRMR-selected features. The highest score of 0.92 with a p-value of 0.001 was obtained by SVM1 (Fig. 10e) and SVM2 (Fig. 10f), followed by LR2 (0.91), LR3 (0.91), and LR4 (0.91). The score is much better than those obtained using permuted data but lower than scenario2, and the p-value is very low (0.001), similar to the scenario2. It provides evidence that the MRMR-selected features have actual dependency between components and labels, and the classifier utilised this to get good results, especially SVM1 and SVM2.

4.2.5. Scenario5: DenseNet201 with combined ETC and MRMR features

The DenseNet201 extracted features were selected using MRMR and ETC feature selection methods in this scenario. In place of using MRMR and ETC selected features individually, the authors combined these 3000 (1550 + 1500) features and then applied them to ML classifiers.

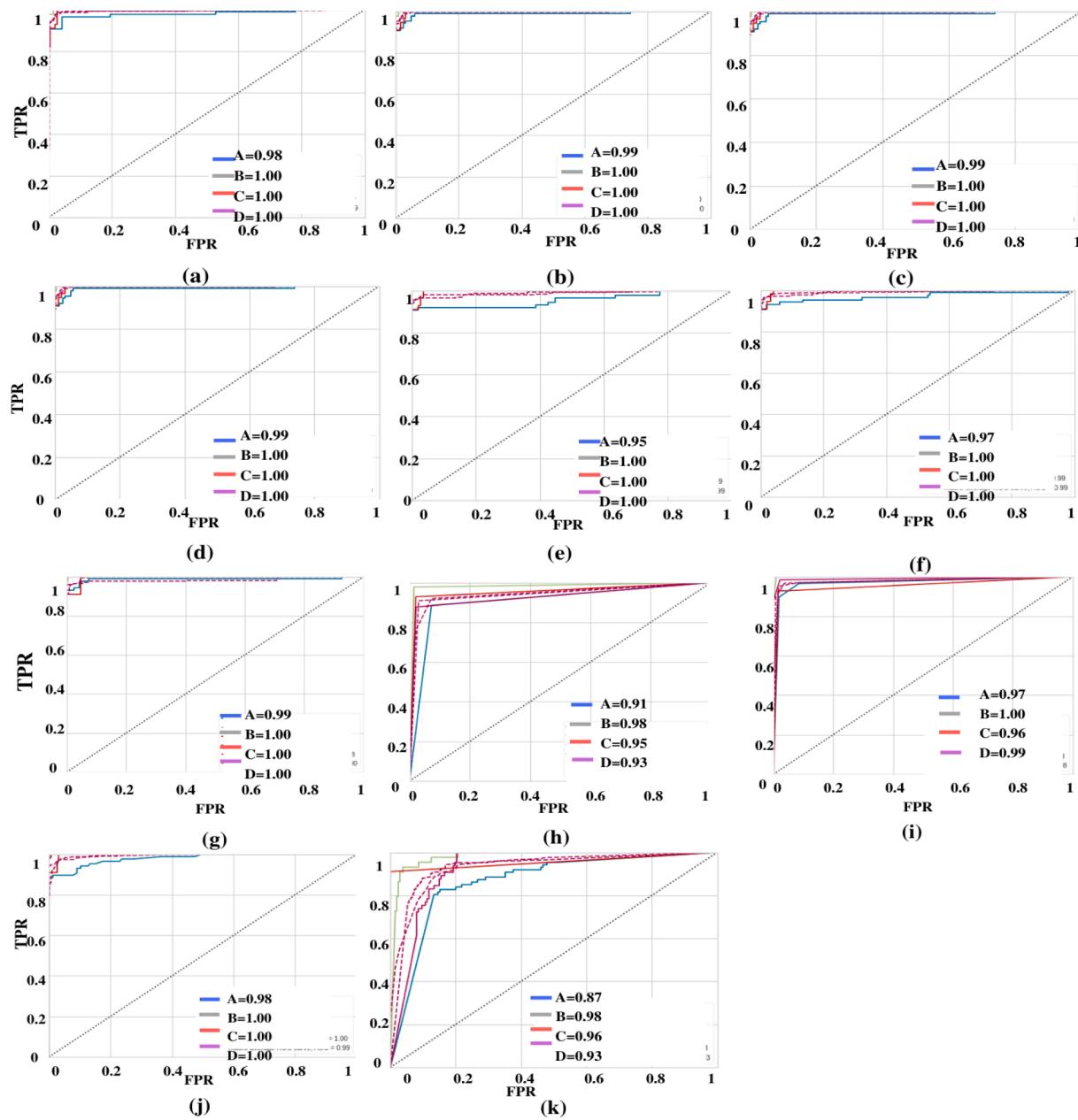


Fig. 6. AUC-ROC performance of ML classifiers ETC extracted features (A:adenocarcinoma, B: large_cell_carcinoma, C: normal, and D:squamous_cell_carcinoma).

Table 6 depicts the performance of all models on MRMR and ETC combined features. The SVM1 was the highest accuracy of 99% and the highest average A_a , A_p , A_r , and A_f of 95%, 95%, 96%, and 96%, respectively. The MCC of 93% and KS of 93% were achieved by SVM1 and SVM2. All other models' average accuracy was less than 94%. The performance of the SVM1 model with MRMR and ETC selected features is good, followed by SVM2.

The AUC-ROC curve of all models is shown in Fig. 11. The highest average AUC score of 1 was achieved by LR4 (Fig. 11d), and 99.75% was completed by LR2 (Fig. 11b) and LR3 (Fig. 11c). The LR4 model has an AUC score of 1 for classes. LR4 models' performance is better compared to other scenarios.

Fig. 12 shows a histogram plot of the permutation scores and p-value of all models' performance on MRMR and ETC combined features. The highest score of 0.96 with a p-value of 0.001 was obtained by SVM1 (Fig. 12e), followed by LR3 (0.95), LR4 (0.95), SVM2 (0.95), LR1 (0.94), and LR2 (0.94). The score is much better than those obtained using permuted data and higher than the other scenario, and the p-value is

very low (0.001). It provides evidence that the ETC and MRMR selected features to have actual dependency between components and labels, and the classifier utilised this to obtain good results, especially SVM1.

5. Discussion

Lung cancer is a severe form of cancer, and detection is crucial for successful treatment. Medical imaging techniques such as CT scans are commonly used for the early diagnosis of lung cancer. However, interpreting CT scans to identify lung cancer can be challenging, particularly in the early stages of the disease or when the scans have low quality or contain noise. The current study aims to investigate the potential of DL algorithms, specifically DenseNet, for identifying lung cancer using CT images. The CT scans underwent pre-processing using resizing and scaling methods before feeding them into the DenseNet model. To enhance the model's performance, the authors modified the DenseNet architecture by adding a few layers instead of using the state-of-the-art DenseNet201 model. In addition, the study explores the use of two

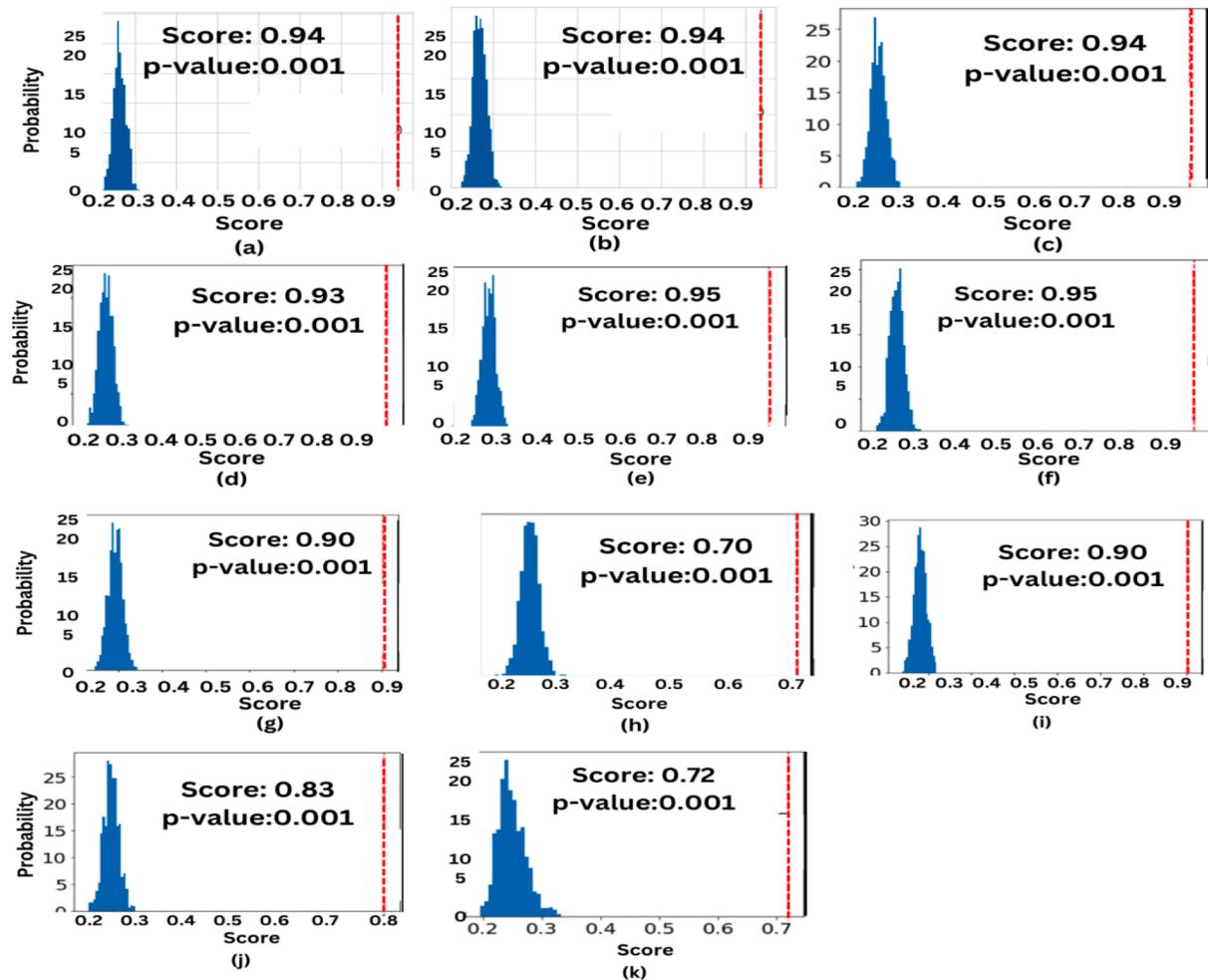


Fig. 7. Histogram plot of the permutation scores using ETC selected features.

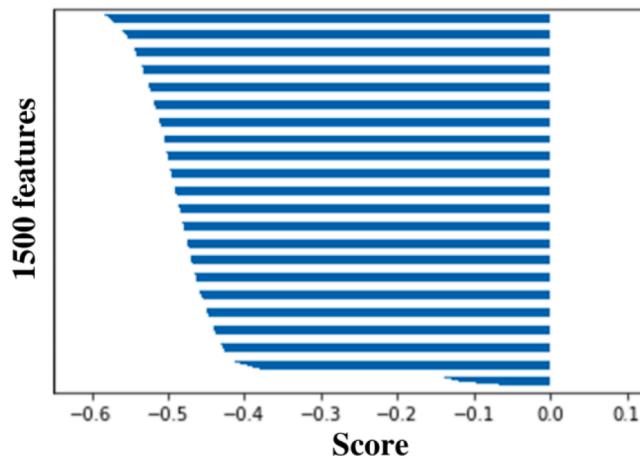


Fig. 8. MRMR feature importance analysis.

feature selection techniques (ETC and MRMR) to identify the most relevant features from the DenseNet extracted features. The selected features were used to train various ML classifiers to classify the CT scans. Furthermore, the authors combined the features selected by ETC and MRMR and trained ML models.

Table 5

Performance of the ML classifiers with MRMR selected features from DenseNet201 extracted features.

Models	High_A	Aa	Ap	Ar	Af	MCC	KS
LR1	94%	90%	90%	90%	90%	86%	86%
LR2	96%	91%	91%	91%	91%	88%	88%
LR3	96%	91%	91%	91%	91%	88%	88%
LR4	94%	91%	91%	91%	91%	87%	87%
SVM1	95%	93%	93%	93%	93%	90%	90%
SVM2	98%	91%	91%	92%	92%	88%	88%
SVM3	78%	70%	66%	84%	66%	63%	57%
DT	69%	65%	66%	67%	66%	53%	53%
KNN	94%	88%	88%	90%	89%	85%	84%
RF	88%	80%	77%	87%	79%	73%	72%
GNB	63%	60%	59%	66%	57%	47%	44%

5.1. Comparison of all scenarios

This section compares the performance of all scenarios and suggests the best possible method for lung cancer identification from CT scans with its advantages and disadvantages. Table 7 depicts the comparative analysis of all scenarios. The scenario2 performs very well and classifies lung cancer from CT scans with the highest accuracy of 99%, average accuracy of 95%, MCC of 93%, KS of 93%, AUC of 0.9975 and p-value of 0.001. This performance is excellent, but this scenario used all 17280 DenseNet201 extracted features. Scenario3 achieved the highest performance among all strategies by attaining 100% highest and 95% average accuracy with similar MCC, KS, AUC score, and p-value like

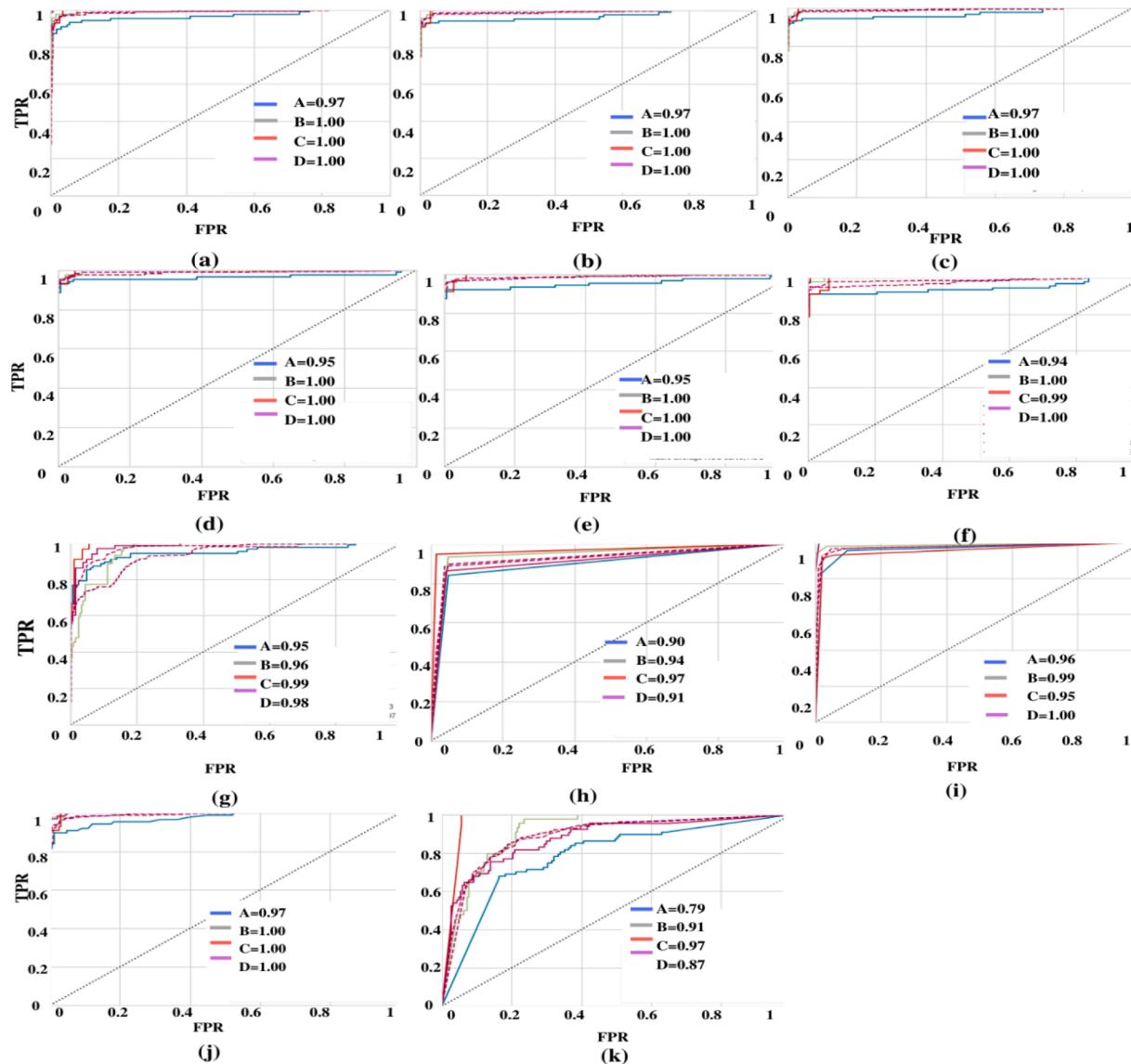


Fig. 9. AUC-ROC performance of ML classifiers ETC extracted features (A: adenocarcinoma, B: large_cell_carcinoma, C: normal, and D: squamous_cell_carcinoma).

sceanrio2. This scenario3 used only 1500 features selected from the DenseNet201 extracted features. Scenario4 performance is the lowest among all scenarios. Sceanrio5 used ETC, and MRMR selected 3000 features and achieved an excellent AUC score of 1, highest accuracy of 99%, and average accuracy of 95% with MCC and KS of 93% by applying the SVM1 model.

From the above, we can say that the performance of SVM1, SVM2, LR2 and LR3 is excellent for all scenarios. The sceanrio3 and scenario5 performance is better than others by using only 1500 and 3000 features, respectively, out of 17280.

The philosophy behind detecting lung cancer from CT scans using a modified DenseNet with feature selection techniques and ML classifiers is to establish reliable and efficient methods for identifying lung cancer, which can ultimately lead to improved patient outcomes. This method is rooted in the concept of evidence-based medicine, which emphasizes the importance of using the most reliable data available to inform clinical decision-making. In the case of lung cancer diagnosis, this involves analyzing and interpreting data from CT scans using ML algorithms. The philosophy behind this approach also highlights the importance of multidisciplinary collaboration among medical professionals, computer scientists, and statisticians. By bringing together experts from these fields, we can develop more precise and effective strategies for

diagnosing lung cancer, ultimately leading to better patient outcomes.

The concept of identifying lung cancer from CT scans using a modified DenseNet with feature selection techniques and ML classifiers is based on evidence-based medicine, interdisciplinary teamwork, and a commitment to improving patient outcomes. By applying these principles, we can develop more accurate and efficient methods for early detection of lung cancer, potentially saving lives and enhancing patients' quality of life.

5.2. Advantages of the proposed work

The current study suggests a modified TL (DenseNet201) framework with feature selection methods and ML classifiers for forecasting lung cancer disease. The following benefits of the planned work are highlighted:

- The suggested work employs a modified DenseNet201 by adding a one Max pooling layer with a pool size of 2 and a drop-out layer of 0.2. Due to this, the trainable features were reduced without affecting the performance. This model has effectively 17280 trainable parameters.

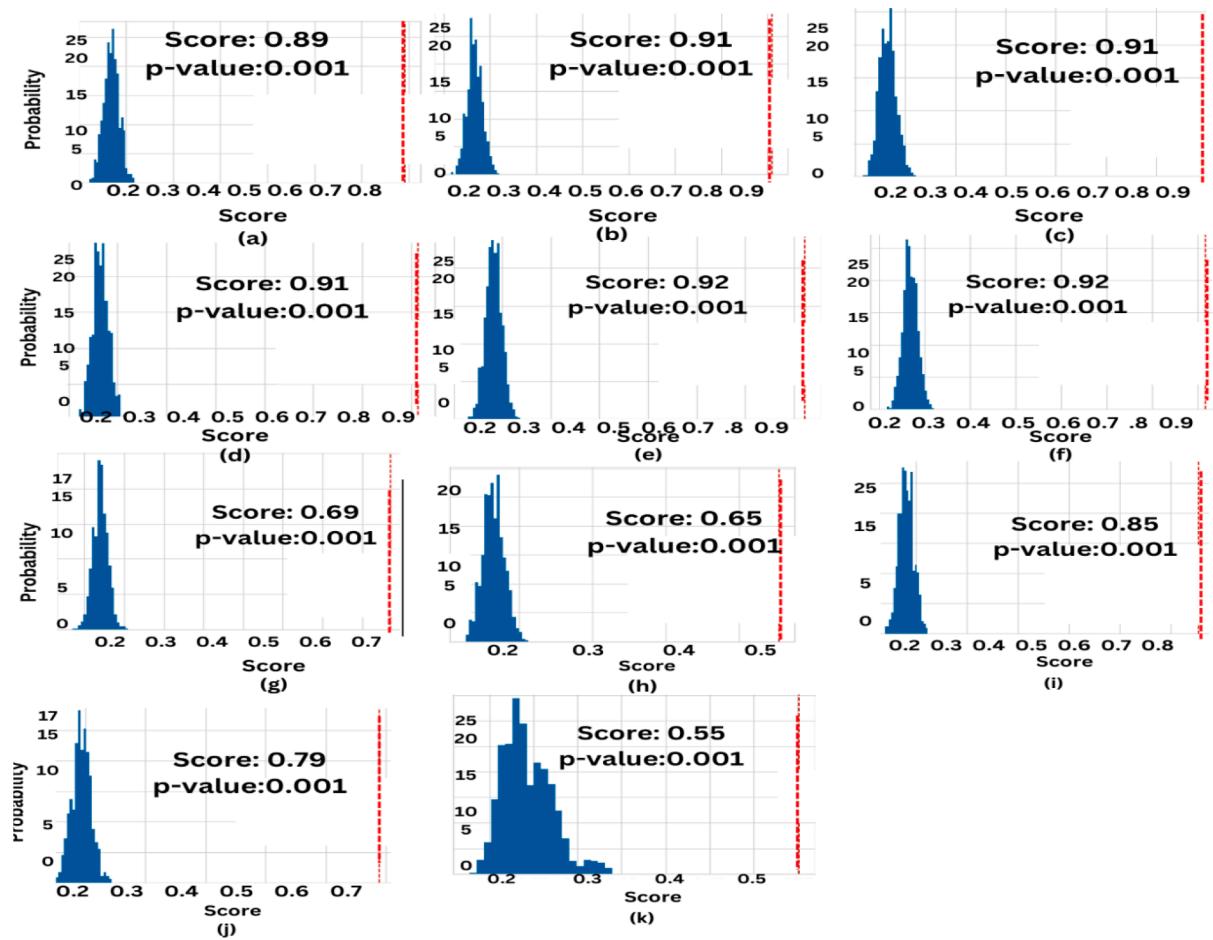


Fig. 10. Histogram plot of the permutation scores and p-value performance on MRMR-selected features.

Table 6
Performance of the ML classifiers with ETC and MRMR combined features.

Models	High_A	Aa	Ap	Ar	Af	MCC	KS
LR1	99%	94%	94%	94%	94%	91%	91%
LR2	99%	94%	94%	95%	95%	92%	92%
LR3	99%	94%	94%	95%	94%	92%	92%
LR4	98%	94%	94%	94%	94%	92%	92%
SVM1	99%	95%	95%	96%	96%	93%	93%
SVM2	99%	94%	94%	95%	95%	93%	93%
SVM3	95%	88%	86%	93%	88%	85%	84%
DT	71%	67%	68%	68%	68%	56%	56%
KNN	95%	90%	90%	92%	91%	87%	87%
RF	93%	84%	83%	88%	84%	79%	78%
GNB	74%	68%	67%	74%	68%	57%	56%

- The authors kept the size of the images 224×224 because the TL model used in this work has a standard image size of 224×224 and is better for comparison.
- A further benefit of the current study is that the dataset was Train and Test subset. The 10-fold method was applied to the Train subset. The models were trained on the Train set and validated on the Val set. The Test set did not use the training process, and tested model performance on Test set independently.
- The callback/Checkpoint class was used in this work. The model checkpoint callback class specifies where the model weights should be checked. Callback class to terminate the training process when the desired accuracy was achieved to avoid overfitting.

- The DenseNet201 extracted features (17280 features) were applied to 11 ML classifiers for classification and cross-validated with 5-fold method.
- In place of using all DenseNet201 extracted features, two feature selection method (ETC and MRMR) was used to select informative features from the DenseNet removed parts.
- First ETC selected features applied to the ML classifiers and checked the performance, and then MRMR selected features applied to the ML classifiers.
- The feature selected by ETC and MRMR were also combined and applied to the ML classifiers and checked the performance
- The confusion matrix, ROC-AUC curve, MCC, Kappa score, and p-value methods were applied to check the models' performance.
- These strategies help to prevent epidemics by detecting them early and implementing appropriate interventions.

6. Conclusion

This research provided a complete picture of commercially significant lung cancer infections, paving the path for future work in diagnosing lung cancer from CT scans and adding to the knowledge base in the research. Identification and confirmation of lung cancer must be performed appropriately, as this plays a significant role in the care and recovery of people affected by the clinical area, but it is a difficult task.

The authors presented five scenarios in which the first DenseNet201 framework was modified by adding a few layers. The reason for changing these models is to reduce trainable parameters and size of the models without affecting performance. Secondly, DenseNet201 was used as a feature extractor, and these features were applied to the 11 ML

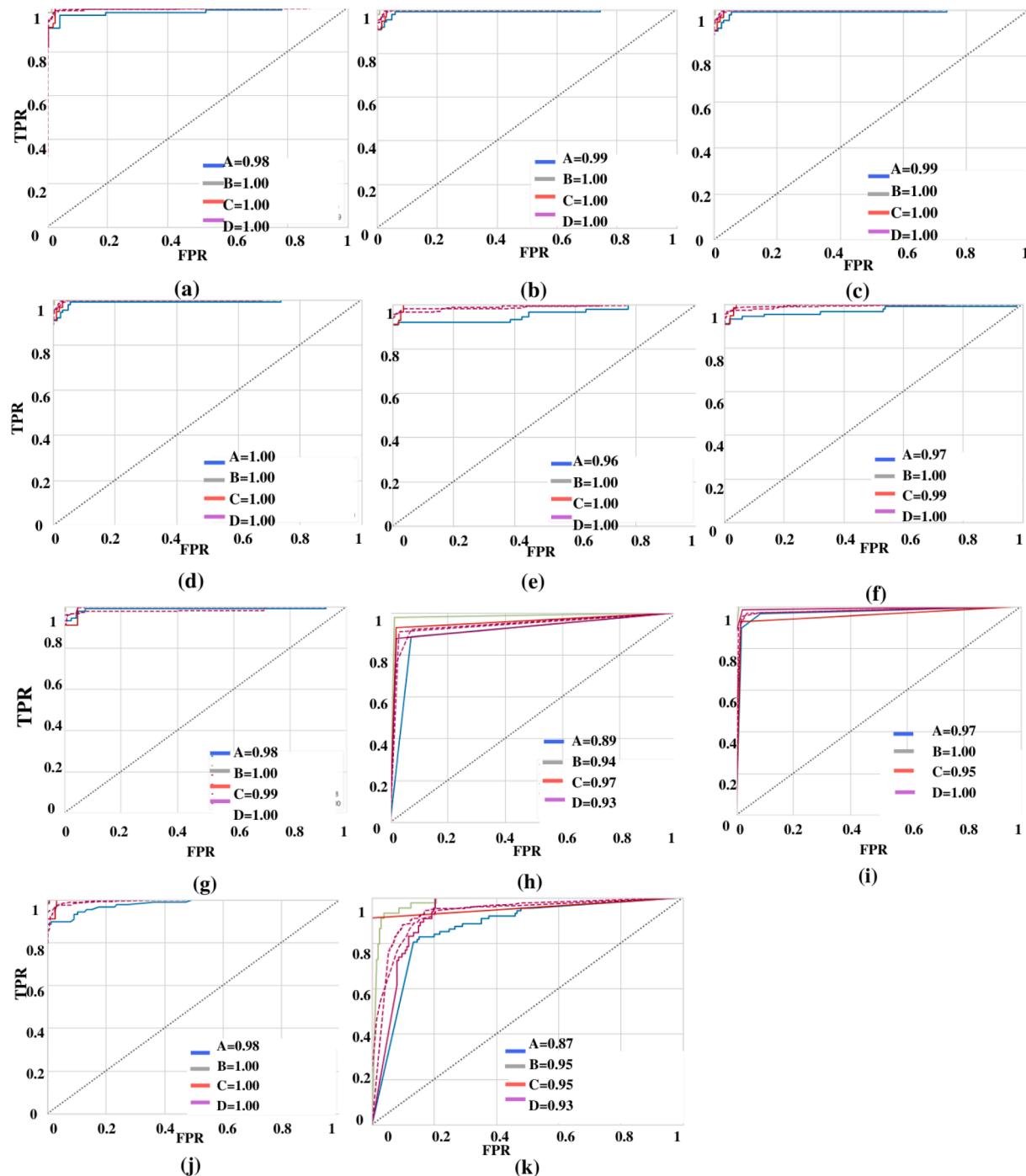


Fig. 11. AUC-ROC performance of ML classifiers ETC and MRRMR combined features (A: adenocarcinoma, B: large_cell_carcinoma, C: normal, and D: squamous_cell_carcinoma).

classifiers. DenseNet201 extracted features selected using ETC and MRRMR feature selection methods in the third and fourth scenarios, respectively. Finally, MRRMR and ETC features were combined and applied to the ML classifiers. All five scenarios' performances were evaluated and compared, and all scenarios performed better. The modified DenseNet201 with ETC selected 1500 features and achieved the highest accuracy, AUC score, MCC, KS, and p-value of 100%, 0.9925, 93%, 93% and 0.001, respectively, while combined ETC and MRRMR selected 3000 features achieved 99%, 1, 93%, 93%, and 0.001, respectively. Apart from this, we have applied the 5-fold cross-validation method and achieved the highest average accuracy of 95% for scenario2, scenario3, and scenario5. We also observed that the SVM and LR

classifiers were excellent in lung cancer identification compared to the RF, DT, GNB, and KNN.

This strategy is beneficial because it produces excellent results. Though the suggested model was effectively applied for lung cancer disease diagnosis from CT scans, the present study has various limitations and assumptions, which are as follows:

- CT scans quality: The quality and resolution of the CT scans can significantly impact the performance of the model. Scans that are of low quality or contain artifacts can introduce noise into the model, reducing its effectiveness.

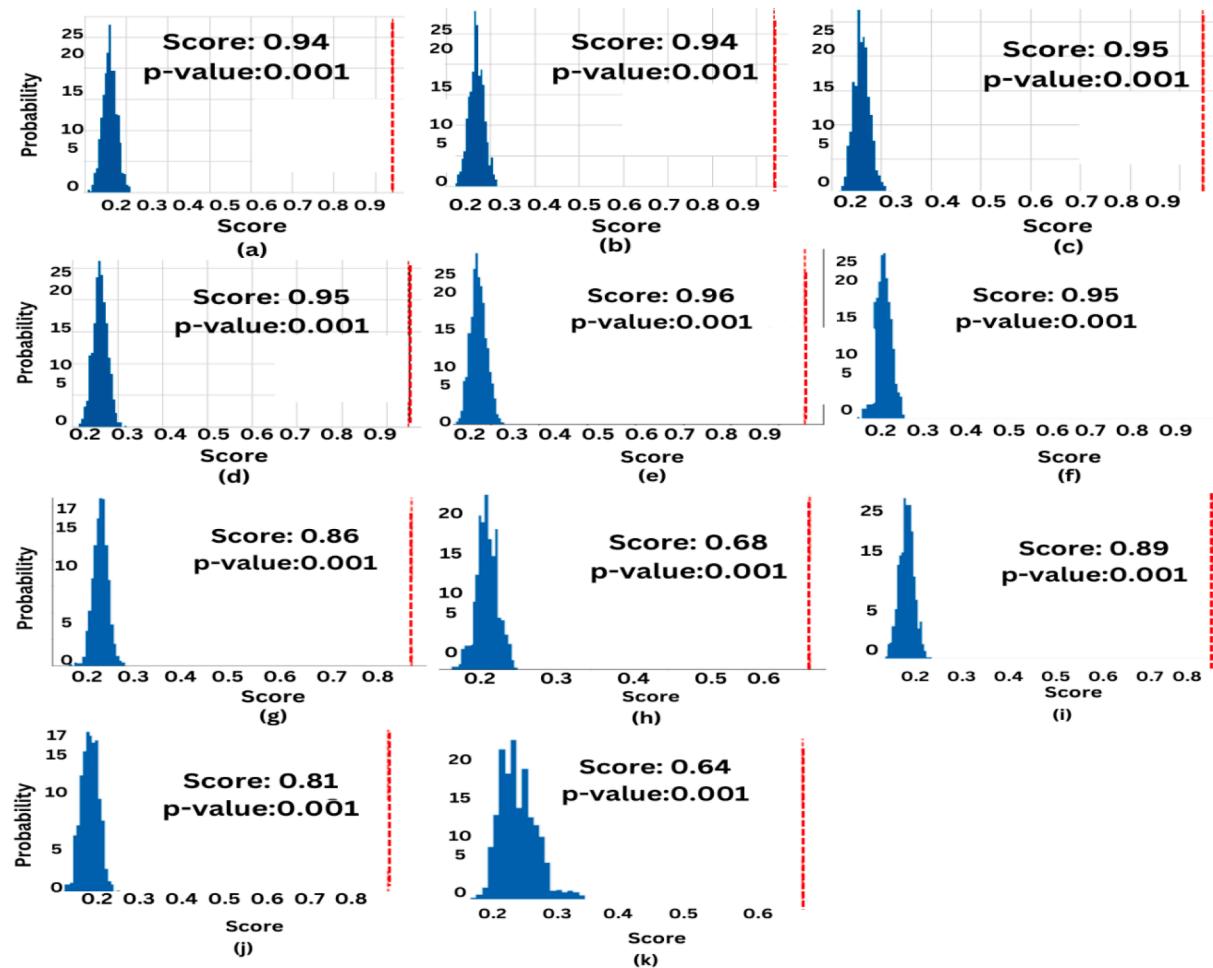


Fig. 12. Histogram plot of the permutation scores and p-value performance on ETC and MRMR combined features.

Table 7
Comparative analysis of all scenarios.

Scenario	Models	High_A	Aa	Ap	Ar	Af	MCC	KS	AUC score	p-value
2	LR2	99%	95%	95%	95%	95%	93%	93%	0.9975	0.001
	LR3	99%	95%	95%	95%	95%	93%	93%	0.9975	0.001
	SVM1	98%	95%	94%	95%	95%	93%	93%	0.9875	0.001
	SVM2	99%	95%	95%	96%	95%	93%	93%	0.9925	0.001
3	SVM1	100%	95%	95%	96%	95%	93%	93%	0.9925	0.001
	SVM2	100%	94%	94%	95%	94%	92%	92%	0.9825	0.001
4	SVM1	95%	93%	93%	93%	93%	90%	90%	0.99	0.001
5	SVM1	99%	95%	95%	96%	96%	93%	93%	0.99	0.001
	SVM2	99%	94%	94%	95%	95%	93%	93%	0.99	0.001
	LR4	98%	94%	94%	94%	94%	92%	92%	1	0.001

- Assumption of feature independence: It is important to note that the assumption that the selected characteristics are critical to identifying lung cancer may not always be accurate. Other variables or factors that contribute to lung cancer detection, but are not captured by the model, may exist.
- Generalizability: It is crucial to acknowledge that while the model may perform well on the dataset used to train it, its generalisation to new datasets may be limited.
- In this research, the authors use a chest lung cancer dataset from Kaggle, which is small in size and has 1008 photos. Generally trained models on small datasets are subject to overfitting, resulting in erroneous analysis.
- The accuracy of the feature selection technique can be affected by the algorithm used and the number of features selected.

- The authors also noted that the lack of relevant datasets in this discipline is a significant concern that must be addressed. As a result, it is required to apply this method to more real-time datasets that can be acquired by considering numerous characteristics such as background, reflection, etc.

In conclusion, the limitations and assumptions discussed highlight the importance of further research and validation on large-scale datasets to ensure the robustness and generalizability of the Lung Cancer detection model. Additionally, while analysing data and making therapeutic decisions, it is crucial to combine professional knowledge and clinical judgement with the model's predictions.

In the future, the authors may investigate further combination or hybrid DL approaches for forecasting lung cancer illnesses. The authors

will also explore various lung diseases and combinations of the TL models.

7. Data availability

Data will be made available on request. Dataset Available on <https://www.kaggle.com/datasets/mohamedhanyyy/chest-ctscan-images>.

CRediT authorship contribution statement

Madhusudan G Lanjewar: Conceptualization, Visualization, Investigation, Methodology, Software, Data curation, Writing – original draft, Validation. **Kamini G Panchbhai:** Data curation, Conceptualization, Writing – review & editing. **Panem Charanarur:** Conceptualization, Writing – original draft, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

- Alakwaa, W., Nassef, M., & Badr, A. (2017). Lung cancer detection and classification with 3d convolutional neural network(3d-cnn). *International Journal of Advanced Computer Science and Applications*, 8(8). <https://doi.org/10.14569/IJACSA.2017.080853>.
- Arslan, H., & Arslan, H. (2021). A new COVID-19 detection method from human genome sequences using CpG island features and KNN classifier. *Engineering Science and Technology, an International Journal*, 24(4), 839–847. <https://doi.org/10.1016/j.jestch.2020.12.026>
- Ceballos, F. (2020, April 6). *An intuitive explanation of random forest and extra trees classifiers*. Medium. <https://towardsdatascience.com/an-intuitive-explanation-of-random-forest-and-extra-trees-classifiers-8507ac21d54b>.
- Chauzwa, T. L., Hosny, A., Xu, Y., Shafer, A., Diao, N., Lanuti, M., ... Aerts, H. J. W. L. (2021). Deep learning classification of lung cancer histology using CT images. *Scientific Reports*, 11(1), 5471. <https://doi.org/10.1038/s41598-021-84630-x>
- Chen, H., Xiong, W., Wu, J., Zhuang, Q., & Yu, G. (2020). Decision-making model based on ensemble method in non-invasive medical system for non-small cell lung cancer. *IEEE Access*, 8, 171903–171911. <https://doi.org/10.1109/ACCESS.2020.3024840>
- Chest ct-scan images dataset*. (n.d.). Retrieved November 5, 2022, from <https://www.kaggle.com/datasets/mohamedhanyyy/chest-ctscan-images>.
- Faruqi, N., Yousuf, M. A., Whaiduzzaman, M., Azad, A. K. M., Barros, A., & Moni, M. A. (2021). LungNet: A hybrid deep-CNN model for lung cancer diagnosis using CT and wearable sensor-based medical IoT data. *Computers in Biology and Medicine*, 139, Article 104961. <https://doi.org/10.1016/j.combiomed.2021.104961>
- Fawcett, T. (2003). Roc graphs: Notes and practical considerations for data mining researchers. *Undefined*. <https://www.semanticscholar.org/paper/ROC-Graphs-3A-Notes-and-Practical-Considerations-for-Fawcett/b32852abb9e55424f2dfadef4da74cbe194059c>.
- Grandini, M., Bagli, E., & Visani, G. (2020). *Metrics for multi-class classification: An overview*. <https://doi.org/10.48550/ARXIV.2008.05756>.
- Hawkins, S., Wang, H., Liu, Y., Garcia, A., Stringfield, O., Kremer, H., ... Gillies, R. J. (2016). Predicting malignant nodules from screening ct scans. *Journal of Thoracic Oncology*, 11(12), 2120–2128. <https://doi.org/10.1016/j.jtho.2016.07.002>
- How to detect lung cancer | lung cancer tests. (n.d.). Retrieved November 5, 2022, from <https://www.cancer.org/cancer/lung-cancer/detection-diagnosis-staging/how-diagnosed.html>.
- Jonas, D. E., Reuland, D. S., Reddy, S. M., Nagle, M., Clark, S. D., Weber, R. P., ... Harris, R. P. (2021). Screening for lung cancer with low-dose computed tomography: Updated evidence report and systematic review for the us preventive services task force. *JAMA*, 325(10), 971. <https://doi.org/10.1001/jama.2021.0377>
- Kadir, T., & Gleeson, F. (2018). Lung cancer prediction using machine learning and advanced imaging techniques. *Translational Lung Cancer Research*, 7(3), 304–312. <https://doi.org/10.21037/tlcr.2018.05.15>.
- Kalaivani, N., Manimaran, N., Sophia, Dr. S., & Devi, D. (2020). Deep learning based lung cancer detection and classification. *IOP Conference Series: Materials Science and Engineering*, 994(1), 012026. <https://doi.org/10.1088/1757-899X/994/1/012026>.
- Kasinathan, G., & Jayakumar, S. (2022). Cloud-based lung tumor detection and stage classification using deep learning techniques. *BioMed Research International*, 2022, 1–17. <https://doi.org/10.1155/2022/4185835>
- Kuruvilla, J., & Gunavathi, K. (2014). Lung cancer classification using neural networks for CT images. *Computer Methods and Programs in Biomedicine*, 113(1), 202–209. <https://doi.org/10.1016/j.cmpb.2013.10.011>
- Lanjewar, M. G., Morajkar, P. P., & Parab, J. (2022). Detection of tartrazine colored rice flour adulteration in turmeric from multi-spectral images on smartphone using convolutional neural network deployed on PaaS cloud. *Multimedia Tools and Applications*, 81(12), 16537–16562. <https://doi.org/10.1007/s11042-022-12392-3>
- Lanjewar, M. G., & Panchbhai, K. G. (2023). Convolutional neural network based tea leaf disease prediction system on smart phone using paas cloud. *Neural Computing and Applications*, 35(3), 2755–2771. <https://doi.org/10.1007/s00521-022-07743-y>
- Lanjewar, M. G., Parab, J. S., & Shaikh, A. Y. (2022). Development of framework by combining CNN with KNN to detect Alzheimer's disease using MRI images. *Multimedia Tools and Applications*. <https://doi.org/10.1007/s11042-022-13935-4>
- Lanjewar, M. G., Parab, J. S., Shaikh, A. Y., & Sequeira, M. (2022). CNN with machine learning approaches using ExtraTreesClassifier and MRMRA feature selection techniques to detect liver diseases on cloud. *Cluster Computing*. <https://doi.org/10.1007/s10586-022-03752-7>
- Lanjewar, M. G., Parate, R. K., & Parab, J. S. (2022). Machine learning approach with data normalization technique for early stage detection of hypothyroidism. In *Artificial Intelligence Applications for Health Care*. CRC Press.
- Masood, A., Sheng, B., Li, P., Hou, X., Wei, X., Qin, J., & Feng, D. (2018). Computer-assisted decision support system in pulmonary cancer detection and stage classification on CT images. *Journal of Biomedical Informatics*, 79, 117–128. <https://doi.org/10.1016/j.jbi.2018.01.005>
- McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica*, 276–282. <https://doi.org/10.11613/BM.2012.031>
- Naïve bayes algorithm: Everything you need to know. (n.d.). *KDnuggets*. Retrieved November 5, 2022, from <https://www.kdnuggets.com/naive-bayes-algorithm-everything-you-need-to-know.html>.
- Pang, S., Zhang, Y., Ding, M., Wang, X., & Xie, X. (2020). A deep model for lung cancer type identification by densely connected convolutional networks and adaptive boosting. *IEEE Access*, 8, 4799–4805. <https://doi.org/10.1109/ACCESS.2019.2962862>
- Pradhan, K. S., Chawla, P., & Tiwari, R. (2023). HRDEL: High ranking deep ensemble learning-based lung cancer diagnosis model. *Expert Systems with Applications*, 213, Article 118956. <https://doi.org/10.1016/j.eswa.2022.118956>
- Qian, Z., Lv, Y., Lv, D., Gu, H., Wang, K., Zhang, W., & Gupta, M. M. (2021). A new approach to polyp detection by pre-processing of images and enhanced faster R-CNN. *IEEE Sensors Journal*, 21(10), 11374–11381. <https://doi.org/10.1109/JSEN.2020.3036005>
- Qin, R., Wang, Z., Jiang, L., Qiao, K., Hai, J., Chen, J., ... Yan, B. (2020). Fine-grained lung cancer classification from pet and CT images based on multidimensional attention mechanism. *Complexity*, 2020, 1–12. <https://doi.org/10.1155/2020/6153657>
- Radovic, M., Ghalwash, M., Filipovic, N., & Obradovic, Z. (2017). Minimum redundancy maximum relevance feature selection approach for temporal gene expression data. *BMC Bioinformatics*, 18(1), 9. <https://doi.org/10.1186/s12859-016-1423-9>
- Raina, V., & Malik, P. (2015). Lung cancer: Prevalent trends & emerging concepts. *Indian Journal of Medical Research*, 141(1), 5. <https://doi.org/10.4103/0971-5916.154479>
- Roy, M. P. (2020). Factors associated with mortality from lung cancer in India. *Current Problems in Cancer*, 44(4), Article 100512. <https://doi.org/10.1016/j.curprobcancer.2019.100512>
- Ruiz, P. (2018, October 18). *Understanding and visualizing DenseNets*. Medium. <https://towardsdatascience.com/understanding-and-visualizing-densenets-7f688092391a>.
- Sanaeifar, A., Bakhtipour, A., & de la Guardia, M. (2016). Prediction of banana quality indices from color features using support vector regression. *Talanta*, 148, 54–61. <https://doi.org/10.1016/j.talanta.2015.10.073>
- Serj, M. F., Lavi, B., Hoff, G., & Valls, D. P. (2018). *A deep convolutional neural network for lung cancer diagnostic*. <https://doi.org/10.48550/ARXIV.1804.08170>.
- Shafi, I., Din, S., Khan, A., Díez, I. D. L. T., Casanova, R. D. J. P., Pifarre, K. T., & Ashraf, I. (2022). An effective method for lung cancer diagnosis from ct scan using deep learning-based support vector network. *Cancers*, 14(21), 5457. <https://doi.org/10.3390/cancers14215457>
- Song, Q., Zhao, L., Luo, X., & Dou, X. (2017). Using deep learning for classification of lung nodules on computed tomography images. *Journal of Healthcare Engineering*, 2017, 1–7. <https://doi.org/10.1155/2017/8314740>
- Sori, W. J., Feng, J., Godana, A. W., Liu, S., & Gelmecha, D. J. (2021). DFD-Net: Lung cancer detection from denoised CT scan image using deep learning. *Frontiers of Computer Science*, 15(2), Article 152701. <https://doi.org/10.1007/s11704-020-9050-z>
- Subasi, A. (2020). Machine learning techniques. In *Practical Machine Learning for Data Analysis Using Python* (pp. 91–202). Elsevier. <https://doi.org/10.1016/B978-0-12-821379-7.00003-5>
- Sun, W., Zheng, B., & Qian, W. (2016). *Computer aided lung cancer diagnosis with deep learning algorithms* (G. D. Touassi & S. G. Armato, Eds.; p. 97850Z). <https://doi.org/10.1117/12.2216307>
- Talukder, M. A., Islam, M. M., Uddin, M. A., Akhter, A., Hasan, K. F., & Moni, M. A. (2022). Machine learning-based lung and colon cancer detection using deep feature extraction and ensemble learning. *Expert Systems with Applications*, 205, Article 117695. <https://doi.org/10.1016/j.eswa.2022.117695>
- Tekade, R., & Rajeswari, K. (2018). Lung cancer detection and classification using deep learning. *Fourth International Conference on Computing Communication Control and Automation (ICCCBEA)*, 2018, 1–5. <https://doi.org/10.1109/ICCCBEA.2018.8697352>

- Valluru, D., & Jeya, I. J. S. (2020). IoT with cloud based lung cancer diagnosis model using optimal support vector machine. *Health Care Management Science*, 23(4), 670–679. <https://doi.org/10.1007/s10729-019-09489-x>
- Wang, S., Wang, R., Zhang, S., Li, R., Fu, Y., Sun, X., ... Peng, W. (2018). 3D convolutional neural network for differentiating pre-invasive lesions from invasive adenocarcinomas appearing as ground-glass nodules with diameters ≤ 3 cm using HRCT. *Quantitative Imaging in Medicine and Surgery*, 8(5), 491–499. <https://doi.org/10.21037/qims.2018.06.03>.
- Wille, M. M. W., Dirksen, A., Ashraf, H., Saghir, Z., Bach, K. S., Brodersen, J., ... Pedersen, J. H. (2016). Results of the randomized danish lung cancer screening trial with focus on high-risk profiling. *American Journal of Respiratory and Critical Care Medicine*, 193(5), 542–551. <https://doi.org/10.1164/rccm.201505-1040OC>
- World Health Organization. (2020). *Coronavirus disease 2019 (COVID-19): Situation report*, 51. World Health Organization. <https://apps.who.int/iris/handle/10665/331475>.
- Young, R. P., & Hopkins, J. R. (2014). Primary and secondary prevention of chronic obstructive pulmonary disease: Where to next? *American Journal of Respiratory and Critical Care Medicine*, 190(7), 839–840. <https://doi.org/10.1164/rccm.201405-0883LE>
- Zhang, W. J., Yang, G., Lin, Y., Ji, C., & Gupta, M. M. (2018). On definition of deep learning. 2018 World Automation Congress (WAC), 1–5. <https://doi.org/10.23919/WAC.2018.8430387>.
- Zhu, Z., & Zhang, Y. (2022). Flood disaster risk assessment based on random forest algorithm. *Neural Computing and Applications*, 34(5), 3443–3455. <https://doi.org/10.1007/s00521-021-05757-6>