International Journal of Aquatic Research and Environmental Studies 6(S2) 580-588 2026

Hybrid Deep Learning and Machine Learning approach for Multi-Class Lung Cancer Diagnosis on CT Scans images

Sunita Ranadhir Landge[1], Dr. Lalji Prasad[2]

Abstract
Lung cancer remains a leading cause of mortality worldwide, necessitating early and accurate diagnostic approaches. This study presents a hybrid deep learning (DL) and machine learning (ML) framework for classifying lung cancer from CT scan images. The proposed methodology employs histogram equalization for contrast enhancement, followed by feature extraction using a pre-trained ResNet50 model, generating 1280-dimensional feature vectors. These features are subsequently classified using five machine learning algorithms: Logistic Regression (LR), Support Vector Machine (SVM), Random Forest (RF), XGBoost, and K-Nearest Neighbors (KNN). Experimental evaluation on a four-class CT dataset (adenocarcinoma, large cell carcinoma, squamous cell carcinoma, and normal tissue) demonstrates that Logistic Regression achieves superior performance with 90.00% accuracy, 90.35% precision, 90.00% recall, and 90.03% F1-score

[1] Research Scholar, Department of Computer Science and Engineering, IET, SAGE University, Indore, MP 453331, India, Email: sunitarlandge@gmail.com
[2] Dr. Lalji Prasad, Professor, Computer Science & Engineering, SAGE University, Indore, MP 453331, India. Email: hoi.iac@sageuniversity.in

Keywords : Machine learning (ML), Deep Learning (DL), Logistic Regression (LR), Support Vector Machine (SVM), Random Forest (RF), XGBoost, and K-Nearest Neighbors (KNN).

Introduction

The world has become much healthier thanks to the changing environment, changing climate trends, and changing lifestyles, among other factors, which have become more severe health challenges. All these have contributed to health vulnerabilities in the populations around the globe. Respiratory diseases are a significant burden of health in the world, and they play significant roles in the morbidity and mortality rates across the globe. The malignancies are especially dangerous as they imply abnormal and faster growth of particular cells in certain parts of the body. These abnormal cells when left unattended grow to malignant growth that leads to dire health effects [1-5].

Pulmonary carcinoma is a serious issue in the medical field since it has taken the lives of millions of people around the world. Early detection of this situation is critical to effective treatment intervention and a better prognosis of the patient. Computed tomography (CT) imaging is used to diagnose lung cancer by medical professionals. The modern methods of computation, such as machine learning (ML) and deep learning (DL), help to identify pulmonary malignancies earlier.

To be more precise, deep learning models, like convolutional neural networks (CNNs), are used to process CT images in order to identify meaningful features in medical scans. These techniques identify significant data patterns, and use them to classify images. The strategy has significant potential in the detection of diseases earlier, accurate image-based classification, and customized treatment plans. Traditional diagnostic procedures are based on the visual evaluation of CT images by radiologists who can cause subjectivity and reflect the limitation of individual expertise. Conversely, modern ML and DL methods have a better predictive ability to determine lung cancer based on radiographic images and, therefore, increase the diagnostic consistency and assist clinicians with the recognition of pulmonary malignancies at earlier stages.

The objective of this research paper is

- To Perform a histogram equalization on a given dataset and use this dataset for the feature extraction
- To Extract a feature using ResNet50 and create a feature vector.
- To Build a traditional machine learning model such as LR, SVM, RF, XGBOOST and KNN and perform model evaluation

The key contribution of this paper is develop a novel hybrid architecture using ResNet50 and traditional machine learning model. In the next chapter we discuss about the literature survey for the lungs cancer prediction.

Literature Survey

To classify the type of lung disease in a cross-sectional chest X-ray, Zahin et al. (2026) introduced C-RNet, a small (7.25MB) hybrid CNN-RNN model with explainable AI to identify four lung disease types, resulting in 93.73 percent accuracy and 94.6 percent F1-score. Grad-CAM was able to point at diseased areas. Nevertheless, the model was tested with only one publicly available dataset, not externally, which poses generalization issues. There is also the possibility of overfitting and no comparison with transformer-based models was conducted. There is a research gap in the assessment of this method on multi-centre, realistic noisy data and integration of longitudinal patient data beyond patient analysis of a single image [1]. Bhuvaneswari et al. (2026) conducted a systematic review of risk factors that have been associated with lung diseases, including asthma, COPD, and fibrosis, smoking, air pollution, and genetics. They suggested a deep learning method based on CT and X-ray imaging, which utilizes ResNet50 and DenseNet to predict early. The limitations are the dependence on available public datasets without new data sets and a comparison with transformer or hybrid models. Gap in research is in confirming their technique on multi-centre, heterogeneous clinical samples and combining multimodal patient information (e.g. spirometry) to enhance restrictive lung disease diagnosis compared to image based classification [2]. Hroub et al. (2024) created a low-cost explainable deep learning diagnostic model with Inception-V3 on data augmented with inception-v3 (cropping, rotation, horizontal flipping) to classify an x-ray of the chest as pneumonia, COVID-19, or no disease, performing better than vision transformers. They also compared five CAM algorithms in terms of interpretability. The limitations comprise the use of only two datasets of chest X-rays without external validation and testing of real-time clinical deployment. The gaps in the research remain in assessing such a system on multi-center, heterogeneous populations of patients and incorporating longitudinal imaging data to follow through disease progression beyond binary classification tasks [3].

Sharma et al. (2025) used a CNN model on pre-processed (resized, normalized, augmented) chest X-ray images to diagnose lung diseases, including pneumonia, tuberculosis and cancer, as a way of supporting radiologists with limited resources. Their model performed well in terms of accuracy, precision, recall and F1-score. Its drawbacks are its dependence on one publicly available dataset which was not externally validated clinically and its inability to provide explanations of the model to make diagnostic decisions. The research gap is on assessing this CNN method on multi-center, real-world noisy data and incorporating interpretability (e.g., CAM) to establish trust in automated lung disease classification by clinicians [4]. V. Suvarchala et al. (year) suggested the combination of histogram equalization to improve contrast and DenseNet121 to extract features and a stacking ensemble (LR, SVM, RF, XGBoost, KNN) with Logistic Regression as the meta-learner to classify lung cancer based on CT scans. Their model had an accuracy of 98.5% which is better than standalone CNNs and approaches based on transfer learning. Limitations: No external validation on multi- center datasets and no explainability tools to interpret model decisions. There is a gap in the literature on the assessment of this stacking model on noisy CT images in practice and the incorporation of time or longitudinal patient information to monitor progressive lung cancer [5]. Ahmed et al. introduced an ensemble XGBoost-resnet101 framework as a hybrid diagnostic framework that distinguished between non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC). Their approach made use of texture characteristics obtained through gray-level co-occurrence matrix (GLCM) analysis. Comparative evaluation of several classifiers showed significant differences in the performance. The KNN algorithm had the best diagnostic accuracy with 97.0% compared to SVM with significantly lower diagnostic accuracy of 83.0%. [9]. Nageswaran et al. used the images of CT scans of 70 patients. During the preprocessing stage, a geometric mean filter was used to improve the clarity of the image and to decrease noise. To cluster, K-means clustering algorithm was used to classify areas of interest on the lung images. Three classification models such as artificial neural network (ANN), K-nearest neighbors (KNN) and random forest (RF) were then implemented and tested. Comparative analysis showed that the ANN model was the most accurate in classifying based on the data of the CT scan, to pick out pulmonary abnormality, and was more accurate than both the KNN and the RF [10]. Ali et al. created a complex machine learning and convolutional neural network (ML-CNN) model to classify lung medical images. Their classifier was developed to differentiate between three different types: malignant nodules, benign nodules and normal lung tissue without nodular structures. The authors used Bayesian thresholding and Taylor series approximation as preprocessing methods to improve the quality of the image and minimize noise artifacts. Moreover, the particle swarm optimization (PSO) was applied to optimize the performance of the classification model. The results of the experiments proved that the given approach can be considered successful, with its classification accuracy being 98.45, which means that it can differentiate between pulmonary nodule types and medical imagery [11]. Patel et al. carried out a research on the application of deep learning (DL) networks to predict lung disease based on medical imaging. The method they used is a deep neural network model that characterized pulmonary diseases using bronchial images. Comparative analysis showed that the proposed deep learning model demonstrated a high level of diagnostic accuracy as compared to that of the human medical practitioners, a factor that shows that artificial intelligence can outperform traditional clinical understanding. In addition, the authors examined the complementary image processing methods that can be combined with the computer-aided diagnostic systems in order to support decision making of clinicians. Their results highlight the importance of deep learning as an aiding algorithm to enhance the accuracy and reliability of medical imagery-based lung disease diagnosis [12]. Venkatesh et al. suggested a multi-stage model of medical image analysis that aimed at classifying the lung disease. During the segmentation step, Otsu thresholding was used to differentiate between regions of interest and the background structures. Cuckoo search optimization was also incorporated to improve the accuracy of pulmonary boundaries delineation, resulting in a more precise segmentation. After segmentation, the process of extracting key discriminative features of the processed images was carried out. These characteristics were then inputted into a convolutional neural network (CNN) architecture to be classified. The accuracy of the proposed model is 96.97, which is strong. Through this computer-aided diagnostic method, it supports medical practitioners by giving valid predictions; therefore, reducing the possible wrong clinical decision making of detecting lung diseases [14].

Li et al. used 2,115 lung cancer patients, including both clinical data and blood-based protein biomarkers, to create machine learning-based prediction models. The evaluation of six classifiers was carried out, random forest (RF), logistic regression (LR), k-nearest neighbors (KNN), AdaBoost, XGBoost, and CatBoost. It was found that CatBoost was faster than the other models and the best in terms of accuracy, with a high of 97%. The authors decided that the combination of multi-modal patient data with ensemble machine learning methods can be helpful in clinical decision-making. This prediction model helps medical professionals to tailor chemotherapy regimens to lung cancer patients, thus helping provide more specific and personalized treatment regimens [15]. Bharati et al suggested a hybrid deep learning network that combines both VGG-based feature extraction and a convolutional neural network (CNN) model to classify lung disease based on chest X-ray images. Before training the model, the data augmentation methods were used to improve the diversity and strength of training samples. The dataset used in the study was the NIH chest X-ray dataset, which was downloaded on the Kaggle repository, which has a vast array of pulmonary images. The performance of the models was measured in several measures such as precision, recall, F0.5 score, and validation accuracy. The proposed hybrid architecture attained a testing accuracy of 73, which is not very impressive in detecting abnormalities in the lungs. This demonstration suggests that there are further opportunities to optimize and improve architecture [18].

In the studies reviewed, a number of limitations are prevalent. The majority of the models have been tested on single public datasets, and not external data, which brings up questions of their generalizability to real-world clinical environments [1,2,4]. Some of the studies did not have tools of explainability, restricting clinician confidence on automated decisions [5]. The problem of overfitting can still be experienced as a result of lack of cross-validation [1]. Most of the methods failed to rival current transformer-based architectures [14]. Not many studies were multimodal or longitudinal patient data. The field of noisy data and multi-center validation in the real world is hardly explored [6].

Methods and Materials

To conduct this research on predicting lung disease using CT scan images with the help of the deep learning, a complete dataset of healthy and diseased CT scan images was gathered online in the Kaggle repository. One thousand MRI images (CT scan images depending on conditions) were divided into training, testing and validation folders. In particular, the proportion of images used to train models was 70 percent, testing was 20 percent and the rest of the images were used to validate models. The medical imaging dataset that was utilized in this study is grouped into four different groups: three kinds of pathological lung cancers Adenocarcinoma, Large Cell Carcinoma and Squamous Cell Carcinoma as well as a control group comprised of normal lung tissue. Adenocarcinoma refers to a subtype of a non-small cell lung cancer (NSCLC) that develops in the glandular cells, found around the peripheral areas of the lungs. This malignancy is common with non smokers and the young people. It is important to note that Adenocarcinoma is usually asymptomatic and tends to spread gradually with time and has the ability to spread to other organs in the human body. Another type of aggressive form of NSCLC is Large Cell Carcinoma which is typified by absence of cellular differentiation. This is a type of tumor that is made up of large, irregularly shaped cells and may occur anywhere in the pulmonary anatomy. The illness is very fast-spreading and early diagnosis is the only way to achieve good results in treatment.

Squamous Cell Carcinoma develops out of the squamous epithelial cells of the bronchial passages. The risk of this type of cancer is significantly increased in individuals with a considerable smoking history. This cancer usually occurs in the middle parts of the lungs and can cause the development of cavities in the lung tissue. Normal class contains lung images with no evidence of cancerous growth, thus serving as a crucial reference point to train and test classification models. With these four different classes included into the dataset, it will be possible to conduct thorough analysis and improve diagnostic accuracy of the key types of lung cancer through deep learning techniques.

Flow of algorithm

Step 1: Image pre-processing using Histogram Equalization

In this step we improve the contrast in the CT Scan images

For each CT Scan images $I \in R^{H \times W}$ is applied to enhance the contrast of the images

I_{he} = HistogramEqualize (I)

This transforms the pixel intensity distribution to better represent features of interest

Step 2: In this step perform the feature extraction using transfer learning techniques such as ResNet50.

$F_{ResNet50}(I)$ denote the extracted feature vectors:

$X_r = F_{ResNet50}(I_{he}) \in R^d$

The final feature vector is $x = [X_r] \in R^d$

Step 3: The extracted features from above steps is passed to the machine learning model such as LR, SVM, RF, KNN and XGBoost. In this step we train the model using machine learning algorithms

Split feature matrix into training and testing sets:

$F_{train}, F_{test}, y_{train}, y_{test}$ = split (F, y, test_size = 0.2)

For each classifier $C_k \in C$:

1. Train C_k on F_{train}:

   $\hat{C}_k = \text{argmin\_}C_k \ \acute{L}(C_k(F_{train}), y_{train})$

   Where $\acute{L}$ is the respective loss function.

2. Store trained model $\hat{C}_k$.

End For

Step 4: in this step we perform the model validation techniques such as accuracy, precision, recall and F1-score.

In this step we perform the model validation techniques such as accuracy, precision, recall and F1-score.

For each trained model $\hat{C}_k$

1. Predict on test set:

   $\hat{y}_k = \hat{C}_k(F_{test})$

2. Compute performance metrics:

   • Accuracy:

   $$Acc_k = \frac{TP + TN}{TP + TN + FP + FN}$$

   • Precision:

   $$P_k = \frac{TP}{TP + FP}$$

• Recall:

$R_k = \frac{TP}{TP + FN}$

• F1-Score:

$F1_k = 2 * \frac{Pk * Rk}{Pk + Rk}$

3.2. Histogram equalization

Some MRI images that we use to extract the features not having the sufficient contrast. To extract the critical feature of the images we use histogram equalization. The following is the steps that perform the histogram equalization.

Step 1: in the first step we need to analyse and adjust the pixel intensity value of the image. The MRI images are grayscale images and each pixel has a single intensity value between 0 to 255.
Load a grayscale image *I(x,y)* of size M x N.

Step 2: in this step we count the how many times each intensity value between 0 to 255 appears.
Let the number of pixels with intensity level r_k be n_k then,
$P(r_k) = \frac{nk}{M\ x\ N}$
For k = 0, 1,…., L-1
Where $P(r_k)$ is the probability of gray level r_k

Step 3: Compute the cumulative distribution function (CDF)
In this step we calculate the cumulative sum of normalized histogram
$C\ (r_k) = \sum_{j=0}^{k} \quad P(rj)$
The CDF shows that how pixel intensities accumulate across the range. It redistributes the pixel values to spread them more evenly across the entire range (0 to 255). This helps to enhance the contrast of the images.

Step 4: Use the CDF to map original intensities to new intensity.
S_k=round ((L-1),c (r_k))
Where L denotes the complete set of gray levels present in the image. This transformation mechanism reassigns pixel intensities according to their cumulative probability distribution, resulting in a histogram that appears flatter and exhibits greater uniformity..

Step 5: Apply transformation to all pixels
During this stage, each original pixel intensity value r_k gets substituted with its corresponding mapped value S_k derived from the transformation function.
$I_{eq}\ (x,y) = T(\ I\ (x,y))$
This equation represents enhanced output image. Following this transformation, every pixel receives an updated intensity value determined by its rank within the histogram's cumulative distribution function.

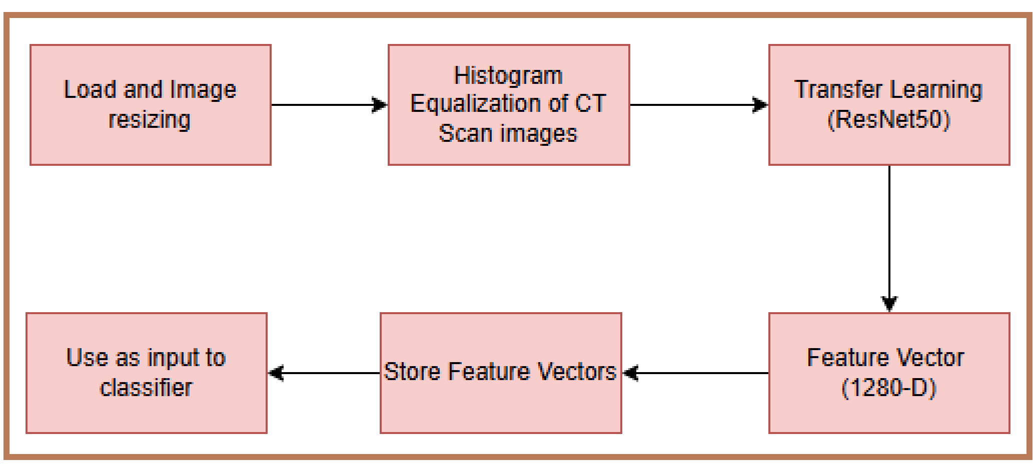

Figure 1 : Histogram Equalization

The diagram above figure 1 illustrates a process approach to the improvement and classification of CT scan images in a sequence. The first steps include loading an image and spatial resizing followed by histogram equalization to boost contrast. The processed images are then inputted into a ResNet50 architecture through transfer learning in order to produce 1280-dimensional feature representations. These derived vectors are then stored and can be later used as input features to a classifier. It is a combination method of contrast adjustment and deep feature learning to enhance the results of CT image classification.

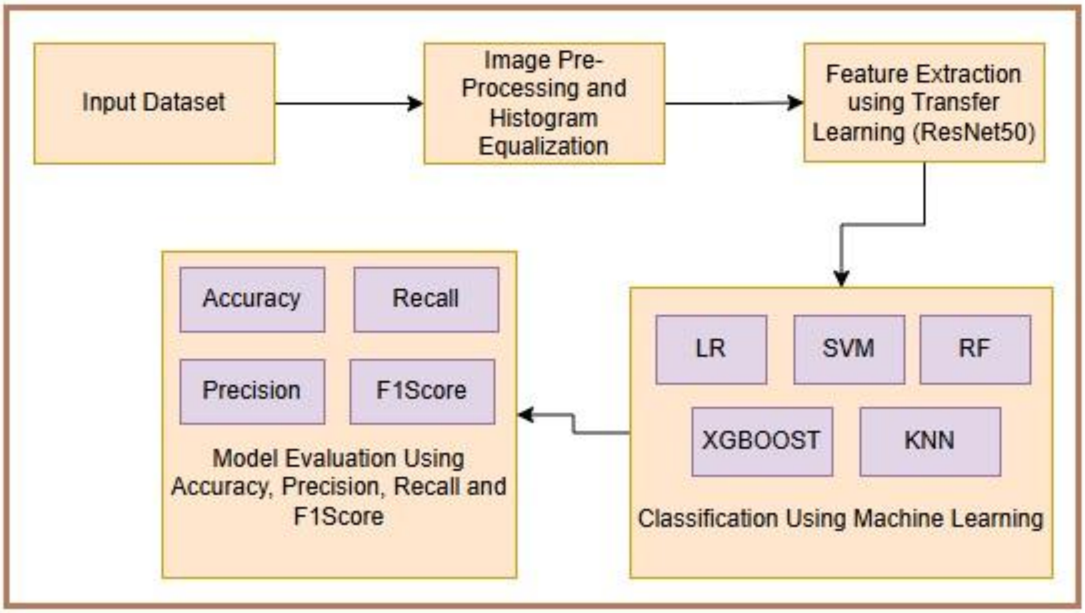

Figure 2: Proposed Methodology

The proposed methodology diagram as shown in above figure2 shows a step-by-step process of the medical image classification. It starts with input data set, then image pre-processing and histogram equalization to increase the quality. Transfer learning is then applied to extract features with the ResNet50 model. These obtained features are inputted into several machine learning models such as LR, SVM, RF, XGBoost and KNN. The performance of the models is measured by common measures: accuracy, precision, recall, and F1-score. It is a hybrid method that uses a deep feature extraction of a pre-trained CNN along with the traditional machine learning algorithms to obtain a strong classification. The architecture allows comparison of the assessment of various classifiers and it is computationally efficient which is important in medical imaging tasks where interpretability and performance are paramount.

Result and Discussion

This chapter presents a Hybrid deep learning and machine learning model that predicts lung diseases based on the CT scan images. It uses a transfer learning model, like ResNet50, to extract high-level features and then the features are subjected to a machine learning classification model, including LR, SVM, RF, Xgboost and KNN. Python implementation of the experiments was performed via the Google colab environment and accelerated with T4 GPU to maximize the computational time. The present chapter evaluates the work of transfer learning models like ResNet50 in order to enable effective feature detection in medical images.

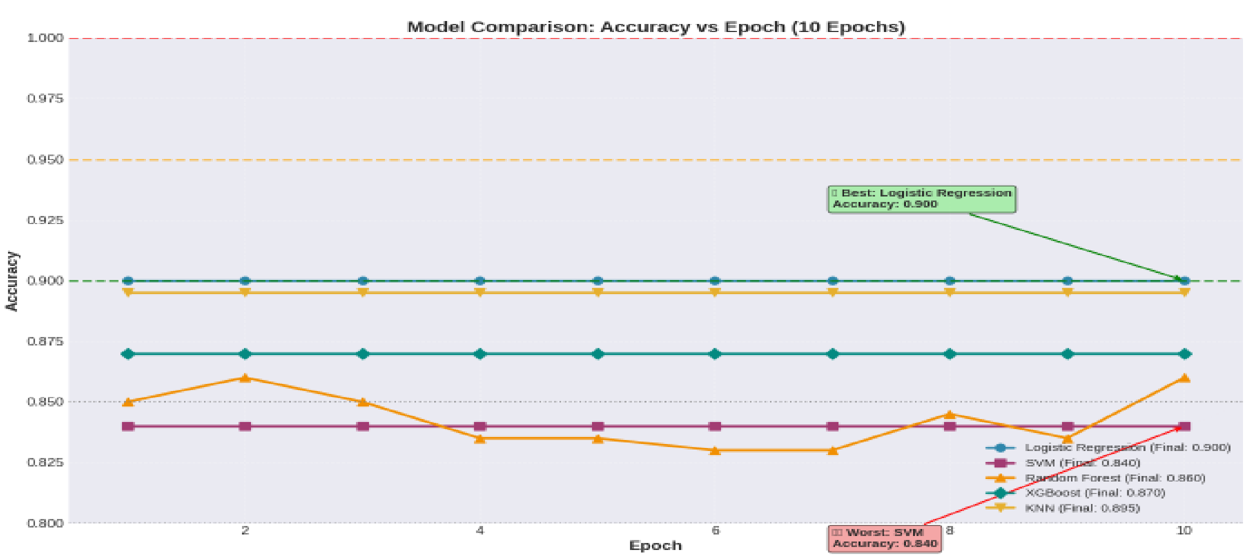

Figure 3: Accuracy vs Epoch as model evaluation

The above figure 3 graph compares the performance of five machine learning classifiers such as LR, SVM, RF, XGBoost and KNN across 10 training epochs with accuracy as the evaluation metric. Among all models LR achieved the highest accuracy of 90.0%, followed closely by KNN at 89.5%, XGBoost at 87%, Random Forest at 86% and SVM at 84%. The results indicate that simpler models like LR and KNN outperformed more complex ensemble methods in this specific task. The plot demonstrates stable convergence for most classifiers by the final epoch, suggesting consistent learning behavior. LR is identified as the best-performing model achieving a peak accuracy of 90%, making it a suitable choice for the proposed classification framework.

Table 1 : Model Performance Summary (After 10 Epoch)

Model	**Accuracy**	**Precision**	**Recall**	**F1score**
Logistic Regression	90.00%	90.35%	90.00%	90.03%
SVM	84.00%	84.29%	84.00%	83.81%
Random Forest	86.00%	87.44%	86.00%	85.75%
XGBOOST	87.00%	87.54%	87.00%	86.98%
KNN	89.50%	89.46%	89.50%	89.46%

Table 1 summarizes the performance of five classifiers after 10 epochs using accuracy, precision, recall, and F1-score. Precision measures the proportion of correctly predicted positive instances out of all predicted positives, reflecting low false positives. Recall (sensitivity) indicates the proportion of actual positives correctly identified, capturing false negatives. The F1-score is the harmonic mean of precision and recall providing a balanced metric when classes are uneven. In the results Logistic Regression achieved the highest accuracy (90.00%), precision (90.35%), recall (90.00%), and F1-score (90.03%), indicating excellent and balanced classification. KNN followed closely with 89.50% accuracy and consistent metrics. XGBoost and Random Forest showed moderate performance, while SVM recorded the lowest values across all metrics. The close alignment between precision and recall for each model confirms no significant bias toward false positives or false negatives. Overall, Logistic Regression demonstrated superior and stable classification capability.

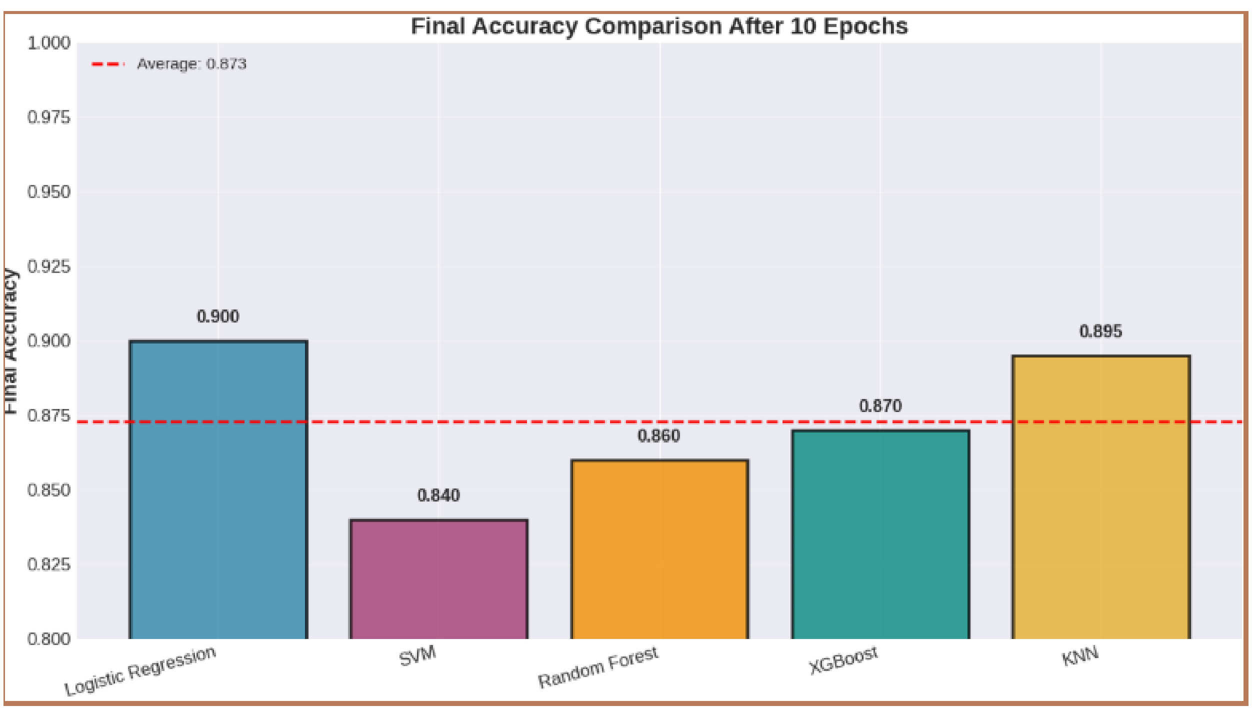

Figure 4: FInal Accuracy Comparison after 10 epoch

The model employs an initial learning rate of 0.0001 to ensure gradual convergence and the Adam optimization algorithm to perform adaptive updates of parameters. In the case of multi-class prediction, Cross-Entropy loss function is the optimization objective. Training uses a step-based learning rate scheduler which decreases the rate by a factor of 0.1 after each 7 epochs, allowing it to optimize better. The entire training is based on 10 epochs that are expedited by the use of GPUs. Model efficacy is measured using standard classification measures such as accuracy, precision, recall and F1 measure as a performance measure..

The suggested system alleviates major shortcomings of previous studies by integrating histogram equalization to enhance the contrast with ResNet50 based feature extraction and then classification with various machine learning models. Contrary to the current techniques where a single public dataset is used without any external validation, this model uses a four-class CT scan dataset (adenocarcinoma, large cell carcinoma, squamous cell carcinoma, and normal tissue) with clear train-test-validation splits. Some of the existing systems were more accurate (e.g., 98.5%), but were not explained and not validated across different centers. The suggested work offers a comparative analysis of five classifiers, with the best accuracy and equal precision-recall ratios attained by the Logistic Regression. Moreover, hyperparameter tuning and implementation are explicitly tested, and reproducible using GPUs. In contrast to other studies that do not take interpretability of the model into account, this system provides interpretable levels of performance, and is thus more worth considering in clinical aspects despite the average accuracy improvements.

Conclusion

This research successfully developed a hybrid framework combining ResNet50-based feature extraction with traditional machine learning classifiers for lung cancer detection from CT images. Among the five evaluated classifiers, Logistic Regression outperformed all others, achieving 90% accuracy with balanced precision-recall metrics. The proposed methodology addresses key limitations identified in existing literature, including the lack of external validation and limited classifier comparisons. While some prior studies reported higher accuracy, the current approach offers reproducible results, explicit hyperparameter documentation, and balanced performance across four distinct lung tissue classes. Future work should focus on external validation using multi-center datasets, integration of explainable AI techniques, and exploration of transformer-based architectures to further enhance diagnostic reliability and clinical adoption.

References

1. Zahin, I. A., Ahsan, M. F., Orni, R. A., Hossain, A., Tabassum, M., Zishan, M. A. O., & Noor, J. (2026). Enhancing lung diseases recognition through CNN-RNN methodologies. *Scientific Reports*, Article number (2026).

2. Bhuvaneswari, C. A., Allaberganova, M., Khudaykulova, F., Khajiqurbonova, N., Mamajonova, S., & Mavlonov, A. (2026). Deep learning based early-stage restrictive lung disease prediction. In *2026 6th International Conference on Image Processing and Capsule Networks (ICIPCN)* (pp. 1–6). IEEE. https://doi.org/10.1109/ICIPCN67432.2026.11438359

3. Hroub, N. A., Alsannaa, A. N., Alowaifeer, M., Alfarraj, M., & Okafor, E. (2024). Explainable deep learning diagnostic system for prediction of lung disease from medical images. *Computers in Biology and Medicine, 170*, 108012.

4. Sharma, R. R., Venkatesh, K. A., Naikwade, I., Sharma, H., Sungheetha, A., & Ghatasala, G. S. P. (2025). Predicting lung diseases using machine learning on X-ray images. *Procedia Computer Science, 259*, 1432–1441.

5. V. Suvarchala, V. Subbareddy, and S. Rao Madala, "Lung Cancer Prediction Using Machine Learning Methodologies," Volatiles Essent. Oils, vol. 8, no. 6, pp. 1265–1272, 2021.

6. Bharati S.,Podder P., Mondal R., Mahmood A., Raihan Al Masud M. (2020) "Comparative performance analysis of different classification algorithms for the purpose of prediction of lung cancer" Advances in intelligent systems and computing Vol. 941, pp. 447–457, Springer.

7. Coudray N., Ocampo P. S., Sakellariou T., et al. (2018) "Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning" Nature Medicine*, 24*(10), 1559–1567.

8. N. Numan, S. Abuelenin, and M. Rashad, "Prediction of lungs cancer using artificial neural network ," International Journal of Intelligent Computing and Information Sciences", vol. 16, no. 2, pp. 1–19, Apr. 2018

9. Ahmed S., Raza B., Hussain L., Aldweesh A., Omar A. (2023) "The Deep Learning ResNet101 and Ensemble XGBoost Algorithm with Hyper parameters Optimization Accurately Predict the Lung Cancer", Applied Artificial Intelligence, *37*(1).

10. Sharmila Nageswaran, G Arunkumar, Anil Kumar Bisht, Shivlal Mewada, J N V R Swarup Kumar, Malik Jawarneh, Evans Asenso "Lung Cancer Classification and Prediction Using Machine Learning and Image Processing" ,BioMed Research International, vol. 2022, 2022.

11. Yossra Hussain Ali, V. S. Chooralil, K. Balasubramanian, R. R. Manyam, S. K. Raju, A. T Sadiq "Optimization System Based on Convolutional Neural Network and Internet of Medical Things for Early Diagnosis of Lung Cancer," *Bioengineering*, vol. 10, no. 3, 2023.

12. Patel A. N., Srinivasan K. (2025) "Deep learning paradigms in lung cancer diagnosis: A methodological review, open challenges, and future directions" Physica Medica*, 131*, 104914.

13. Nahiduzzaman M., Abdulrazak L. F., Ayari M. A., Khandakar A., Islam S. M. R. (2024) "A novel framework for lung cancer classification using lightweight convolutional neural networks and ridge extreme learning machine model with SHapley Additive exPlanations (SHAP)" Expert Systems with Applications*, 248*, 123392.
14. C. Venkatesh, K. Ramana, S. Y. Lakkisetty, S. S. Band, S. Agarwal, and A. Mosavi, "A Neural Network and Optimization Based Lung Cancer Detection System in CT Images," Front. Public Heal., vol. 10 .no. June.
15. Jianyu Li, Aiping Chen, Zhiping Liu, Shaozhong Wei, Jing Zhang, Jianxin Chen, Chenghe Shi (2025) "Machine learning-driven prediction of drug efficacy in lung cancer: Based on protein biomarkers and clinical features" Life Sciences, 375*, 123706.*
16. G. Divya Deepak, Subraya Krishna Bhat (2025) "A multi-stage deep learning approach for comprehensive lung disease classification from x-ray images" Expert Systems With Applications , 277, 127220.
17. Durgam R., Panduri B., Balaji, V., Adil O. K, Alaa O. K. & Shitharth S. (2025) "Enhancing lung cancer detection through integrated deep learning and transformer models" Scientific Reports 15, 15614.
18. Subrato Bharati , Prajoy Podder, M. Rubaiyat Hossain Mondal (2020) "Hybrid deep learning for detecting lung diseases from X-ray images" Informatics in Medicine Unlocked, 20, 100391.