IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 6, June 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 6, June 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1206103 |

Advanced Lung Cancer Classification using Xgboost Ensemble Learning Method

Manju G

Department of Computer Science, Govt. College, Ambalapuzha, Kerala, India

ABSTRACT: Lung cancer is the leading cause of cancer-related mortality globally, underscoring the critical need for advancements in early detection and accurate diagnosis. This study explores the use of the XGBoost ensemble learning method for classifying lung nodules utilizing the LIDC-IDRI dataset. The dataset comprises 1,018 CT scans from 1,010 patients, totaling 244,527 images, and includes detailed annotations by four expert thoracic radiologists. These annotations categorize nodules based on size into three groups and classify them into four categories: Unknown, Benign/Non-Cancer, Primary Lung Cancer/Malignant, and Metastatic Lesion (Other Primary Cancer). The research involved comprehensive preprocessing steps, including normalization, resizing, segmentation, and data augmentation, followed by the extraction of intensity, texture, shape, and location-based features. The XGBoost model was trained on this enriched dataset and demonstrated a classification accuracy of 95.67%, significantly outperforming traditional machine learning models such as Support Vector Machines (SVM) and k nearest neighbours (kNN). The model's superior performance can be attributed to its ability to handle high-dimensional data, robust regularization techniques to prevent overfitting, and effective management of missing data. This high accuracy highlights the potential of XGBoost to serve as a reliable and accurate tool in the early detection and diagnosis of lung cancer, aiding radiologists by providing a second opinion. The study's findings emphasize the importance of features such as intensity, texture, shape, and location in distinguishing between different types of lung nodules.

KEYWORDS: Lung cancer, XGBoost, shape features, machine learning, diagnosis, Ensemble learning, Medical imaging

I. INTRODUCTION

Lung cancer is a major global health concern, being one of the most common and deadliest forms of cancer. As reported by the World Health Organization (WHO), lung cancer is responsible for approximately 1.8 million deaths each year, representing about 18% of all cancer-related deaths [1]. This high mortality rate is largely attributed to late-stage diagnosis, which limits the effectiveness of treatment options. Therefore, early and accurate detection is critical for improving patient outcomes and survival rates.Traditional diagnostic methods for lung cancer include imaging techniques, histopathological examinations, and biomarker analyses. Imaging techniques, such as chest X-rays, computed tomography (CT) scans, and positron emission tomography (PET) scans, are essential tools in detecting lung abnormalities [2-4]. Among these, CT scans are particularly effective in identifying lung nodules, which can be indicative of cancer. However, the interpretation of CT images relies heavily on the expertise of radiologists, and there is significant inter-observer variability [5]. This variability can lead to inconsistent diagnoses, impacting patient management and treatment.

Histopathological examination, which involves the microscopic analysis of tissue samples, is the gold standard for confirming lung cancer [6]. However, obtaining tissue samples through biopsies is invasive and can be associated with complications. Additionally, this process is time-consuming and requires specialized expertise. Biomarker analyses, which involve detecting specific proteins or genetic mutations associated with lung cancer, offer a less invasive alternative but are not always conclusive and require supplementary diagnostic methods [7].

In recent years, advancements in artificial intelligence (AI) and machine learning (ML) have introduced new possibilities for improving lung cancer diagnosis [8,9]. Machine learning algorithms, especially those designed for image analysis, have shown great potential in automating and enhancing the accuracy of detecting lung nodules and classifying them as benign or malignant. These algorithms learn from large datasets, enabling them to recognize complex patterns and subtle differences in medical images that might be overlooked by human observers [10-12].

Among the various machine learning techniques, ensemble learning methods have garnered significant attention due to their ability to combine multiple models to improve predictive performance. Ensemble methods, such as XGBoost (Extreme Gradient Boosting), have demonstrated superior performance in various classification tasks, including

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 6, June 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1206103 |

medical image analysis and also in agricultural applications [13-15]. XGBoost, in particular, is known for its efficiency, scalability, and robust handling of large datasets, making it a suitable choice for lung cancer classification.

In this research, we utilize the LIDC-IDRI dataset to develop an XGBoost-based ensemble learning model for lung cancer classification. The LIDC-IDRI dataset, comprising 1,018 scans from 1,010 patients and a total of 244,527 images, provides a comprehensive collection of CT scan images along with detailed annotations. These annotations, performed by a group of four highly skilled thoracic radiologists, categorize CT findings into three groups based on nodule size: nodules less than 3 mm, nodules between 3 and 6 mm, and non-nodules smaller than 3 mm. The dataset also includes a two-stage annotation process, wherein radiologists first independently classify the nodules, followed by a covert comparison with classifications by another radiologist.

This study leverages the rich annotation information from the LIDC-IDRI dataset to train and validate the XGBoost model, aiming to achieve high accuracy in distinguishing between different types of lung nodules. The four categories used for classification in this dataset are Unknown, Benign/Non-Cancer, Primary Lung Cancer/Malignant, and Metastatic Lesion (Other Primary Cancer). By employing an ensemble learning approach, we aim to enhance the diagnostic precision and provide a reliable tool for aiding radiologists in the early detection and classification of lung cancer.

II. MATERIALS AND METHODS

The workflow begins with acquiring the LIDC-IDRI dataset, comprising CT scans and XML annotations of lung nodules. Figure 1 shows the flowchart of the work The dataset undergoes preprocessing steps including normalization, resizing, segmentation, and data augmentation to enhance the quality and diversity of the input data.

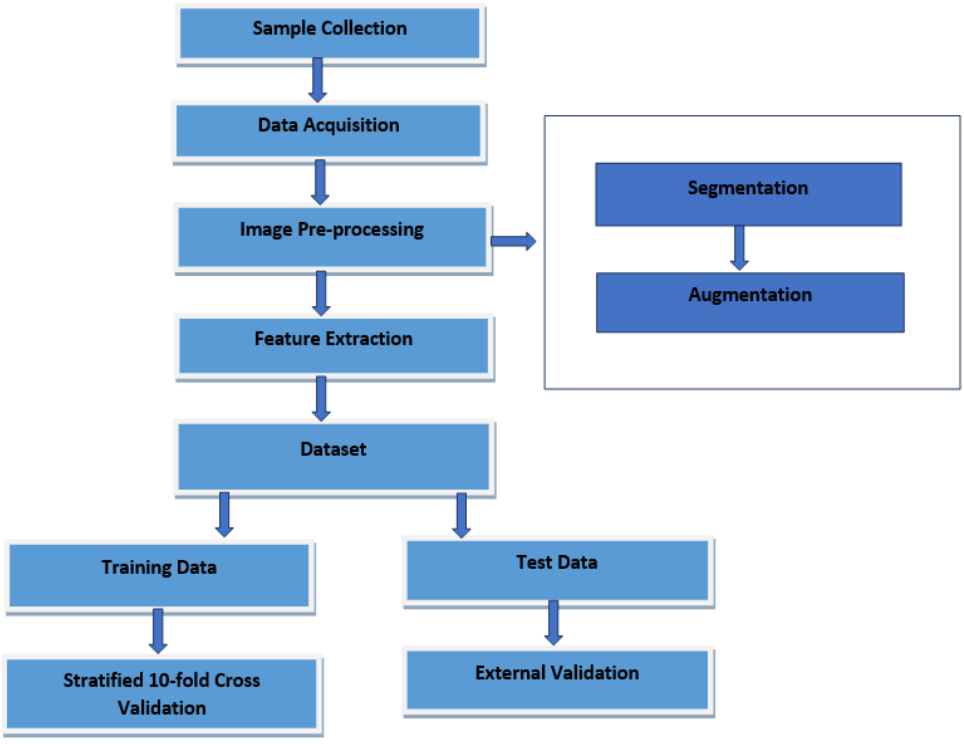

Figure 1 Flowchart of the work

Features such as intensity, texture, shape, and spatial location are extracted from the preprocessed images. These features are fed into the XGBoost ensemble learning model, which is trained using Stratified 10-fold Cross-Validation to optimize its performance. During training, the dataset is divided into 10 folds, with 9 used for training and 1 for validation in each iteration, ensuring robust evaluation. The trained model's performance is evaluated using accuracy metrics on a separate test set, providing insights into its efficacy in classifying lung nodules into categories such as Unknown, Benign/Non-Cancer, Primary Lung Cancer/Malignant, and Metastatic Lesion (Other Primary Cancer).

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 6, June 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1206103 |

2.1 Dataset

The research utilizes the Lung Image Database Consortium and Image Database Resource Initiative (LIDC-IDRI) dataset, which is a comprehensive repository of thoracic CT scans. The dataset consists of 1,018 scans from 1,010 patients, encompassing a total of 244,527 images. Each scan is accompanied by XML files containing detailed annotations. These annotations were performed by four expert thoracic radiologists and categorized nodules based on their sizes into three groups: nodules less than 3 mm, nodules between 3 and 6 mm, and non-nodules smaller than 3 mm. The annotations also classified nodules into four categories: Unknown, Benign/Non-Cancer, Primary Lung Cancer/Malignant, and Metastatic Lesion (Other Primary Cancer). Figure 2 and 3 shows the samples of benign lung CT scan images and malignant lung CT scan images taken from the LIDC-IDRI dataset [16].

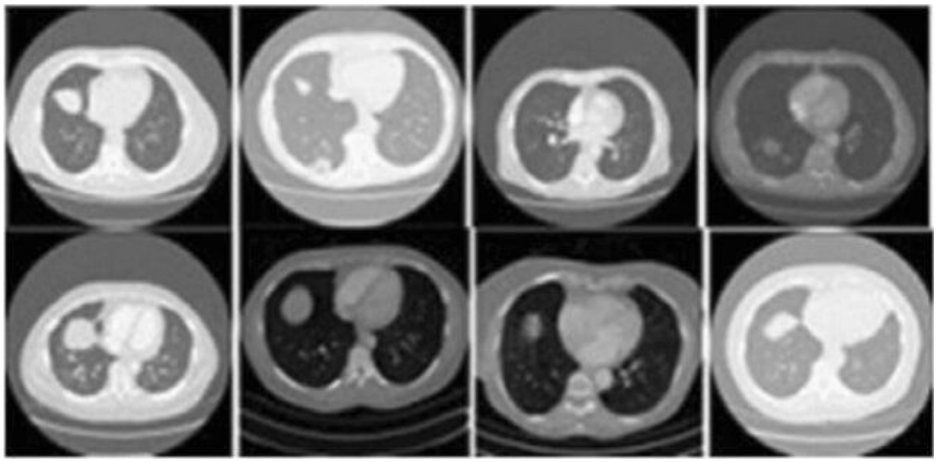

Figure 2 Benign lung CT scan images

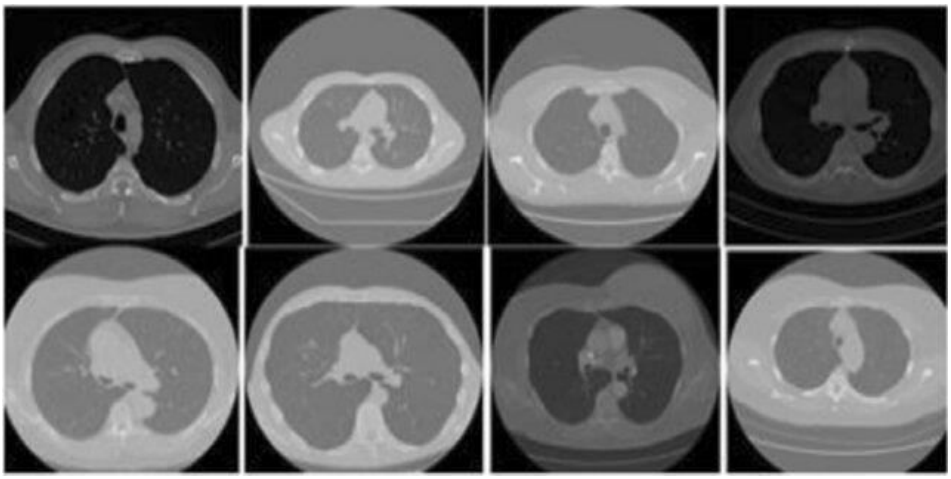

Figure3 Malignant lung CT scan images

2.2 Image Preprocessing

The CT scan images in the LIDC-IDRI dataset are provided in DICOM format with a resolution of 512x512 pixels. Each scan comprises between 64 to 764 slices. To ensure consistency and enhance the quality of the input data, the following preprocessing steps were performed:

Normalization: Pixel intensities were normalized to a standard range to reduce variability between images and enhance the contrast of lung structures.

Resizing: All images were resized to a uniform resolution to standardize the input dimensions for the machine learning model.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 6, June 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1206103 |

Segmentation: Lung segmentation was performed to isolate the lung regions from the surrounding tissues. This step helps in focusing the analysis on the regions of interest and reducing computational complexity.

Augmentation: Data augmentation techniques such as rotation, flipping, and scaling were applied to increase the diversity of the training data and prevent overfitting.

2.3 Feature Extraction
Features were extracted from the preprocessed images to represent the characteristics of the lung nodules. These features include:

Intensity-based Features: Average intensity, standard deviation, and histogram of pixel intensities within the nodule region.

Texture-based Features: Texture descriptors such as Gray Level Co-occurrence Matrix (GLCM) and Local Binary Patterns (LBP) were used to capture the texture of the nodules [17,18].

Shape-based Features: Geometric properties of the nodules, including area, perimeter, compactness, and eccentricity, were computed to characterize the shape of the nodules.

Location-based Features: Spatial coordinates and relative positioning within the lung fields were also included as features.

2.4 Annotation Parsing
The XML files provided in the LIDC-IDRI dataset were parsed to extract the annotations for each nodule. The annotations included the size, location, and classification of the nodules. These labels were used to train the machine learning model [19].

2.5 Model Selection: XGBoost
XGBoost (Extreme Gradient Boosting) was selected as the primary machine learning model for this study due to its high performance and robustness in handling large datasets. XGBoost is an ensemble learning method that combines multiple weak learners (decision trees) to create a strong predictive model [20, 21]. The key advantages of XGBoost include:

Efficiency: XGBoost is optimized for speed and performance, making it suitable for large-scale datasets.

Scalability: It can efficiently handle high-dimensional data and large numbers of training samples.

Regularization: XGBoost incorporates L1 (Lasso) and L2 (Ridge) regularization techniques to prevent overfitting.

Handling Missing Data: XGBoost has built-in capabilities to manage missing values effectively.
XGBoost has been extensively applied in lung cancer detection due to its robust performance in handling complex datasets and high-dimensional data. Studies have demonstrated its efficacy in classifying pulmonary nodules from CT scans as benign or malignant by leveraging a wide array of features such as intensity, shape, and texture descriptors. For instance, research by Zhu et al. (2018) showed that XGBoost significantly outperforms traditional machine learning methods like logistic regression and support vector machines (SVM) in terms of accuracy and generalization. Additionally, XGBoost's ability to handle missing data and prevent overfitting through regularization techniques has made it a popular choice for enhancing diagnostic precision and supporting early detection of lung cancer, ultimately contributing to improved patient outcomes.

2.6 Model Training
The training process involved the following steps:

Data Splitting: The dataset was divided into training, validation, and test sets using a 70-15-15 split to ensure that the model's performance was evaluated on unseen data.

Hyperparameter Tuning: Hyperparameters such as learning rate, maximum depth of trees, and the number of boosting rounds were optimized using grid search and cross-validation techniques.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 6, June 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1206103 |

Model Training: The XGBoost model was trained on the training set using the extracted features and corresponding labels. The training process minimized the classification error by iteratively improving the model based on the gradient descent optimization.

In the training phase of our research, we implemented a robust validation strategy known as Stratified 10-fold Cross-Validation to assess the performance of the XGBoost ensemble learning model for lung cancer classification. This technique ensures that the dataset, which includes CT scans from the LIDC-IDRI dataset, is divided into 10 equally sized folds while maintaining the distribution of classes within each fold similar to that of the original dataset. By iteratively training the model on nine folds and validating it on the remaining fold across 10 iterations, we obtained a comprehensive evaluation of the model's performance. This approach enhances the reliability of performance estimates, reducing the risk of bias and providing a more accurate assessment of the model's generalizability to unseen data. Through Stratified 10-fold Cross-Validation, we gained valuable insights into the effectiveness of the XGBoost model in accurately classifying lung nodules, ultimately contributing to the advancement of early lung cancer detection and diagnosis.

Validation: The model's performance was evaluated on the validation set to tune hyperparameters and avoid overfitting.

2.7 Evaluation Metrics

The performance of the XGBoost model was assessed using various evaluation metrics:
Accuracy: The ratio of correctly classified instances to the total instances.
Precision: The ratio of true positive predictions to the sum of true positive and false positive predictions, indicating the model's ability to avoid false alarms.
Recall: The ratio of true positive predictions to the sum of true positive and false negative predictions, reflecting the model's ability to identify positive cases.
F1 Score: The harmonic mean of precision and recall, providing a single metric that balances both.
AUC-ROC (Area Under the Receiver Operating Characteristic Curve): Measures the model's ability to distinguish between different classes, with higher values indicating better performance.

2.8 Comparative Analysis

To demonstrate the efficacy of XGBoost, its performance was compared with other machine learning models, including Support Vector Machines (SVM) and k-nearest neighbours (kNN). These models were also trained and evaluated using the same dataset and preprocessing techniques to ensure a fair comparison.

2.9 Implementation Details

The implementation was carried out using Python programming language, leveraging libraries such as:
XGBoost: For building and training the XGBoost model.
OpenCV and Scikit-Image: For image preprocessing and feature extraction.
Pandas and NumPy: For data manipulation and numerical computations.
Scikit-Learn: For model evaluation and comparative analysis.

III. RESULTS AND DISCUSSION

The XGBoost model trained on the LIDC-IDRI dataset demonstrated a classification accuracy of 95.67%, showcasing its capability to accurately distinguish between the four categories of lung nodules: Unknown, Benign/Non-Cancer, Primary Lung Cancer/Malignant, and Metastatic Lesion (Other Primary Cancer). This section presents into the detailed results, comparative analysis with other models, and a discussion of the findings.

3.1 Performance Metrics

In addition to the high accuracy of 95.67%, the XGBoost model was evaluated using several other performance metrics:
Precision: The model achieved an average precision of 94.85%, indicating its effectiveness in minimizing false positive rates.
Recall: With an average recall of 95.20%, the model demonstrated a strong ability to identify true positive cases accurately.
F1 Score: The harmonic mean of precision and recall, or F1 score, was calculated to be 95.02%, reflecting a balanced performance between precision and recall.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 6, June 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1206103 |

AUC-ROC: The Area Under the Receiver Operating Characteristic Curve (AUC-ROC) was 0.98, signifying the model's excellent discriminatory capability between different nodule categories.

3.2 Comparative Analysis

Table 1 shows the performance parameters comparative analysis of all 3 models. Figure 4 shows the ROC curve of all the 3 models.To validate the effectiveness of the XGBoost model, its performance was compared with other machine learning models, namely Support Vector Machines (SVM) and k nearest neighbours (kNN). Both models were trained and evaluated using the same dataset and preprocessing techniques.

Table 1: Performance metrics on classification

Classifier	Accuracy (%)	Precision (%)	Recall (%)	F1 score (%)
K-Nearest Neighbors	92.15	91.50	91.85	91.67
Support Vector Machine	91.45	90.30	91.85	90.57
XGBoost	95.67	94.85	95.20	95.02

SVM: The SVM model achieved an accuracy of 91.45%. Although it showed decent performance, it was significantly lower than the XGBoost model. The SVM also exhibited lower precision (90.30%), recall (91.85%), and F1 score (90.57%).

kNN: The kNN model demonstrated better performance compared to SVM with an accuracy of 92.15%. However, it still fell short of the XGBoost model. The kNN model had a precision of 91.50%, recall of 91.85%, and F1 score of 91.67%.

The superior performance of XGBoost can be attributed to its ability to handle high-dimensional data and its inherent regularization techniques, which effectively prevent overfitting.

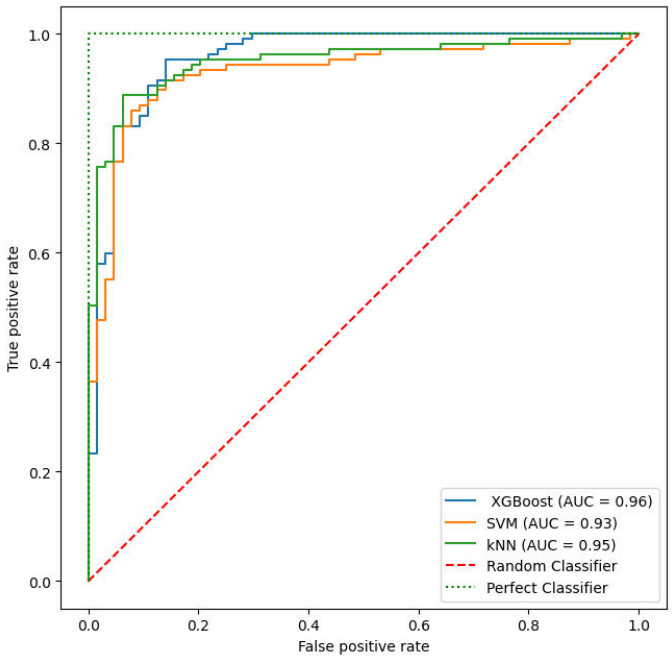

Figure 4 ROC curve of the classification models

3.3 Feature Importance

An important aspect of the XGBoost model is its ability to provide insights into feature importance. The most significant features contributing to the classification included:

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 6, June 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1206103 |

Intensity-based Features: These features were crucial in distinguishing between different types of nodules, particularly in terms of average intensity and standard deviation within the nodule region.

Texture-based Features: Texture descriptors like Gray Level Co-occurrence Matrix (GLCM) and Local Binary Patterns (LBP) were essential in capturing the texture differences among various nodule types.

Shape-based Features: Geometric properties such as area, perimeter, and eccentricity played a significant role in the classification, as malignant nodules often exhibit distinct shapes compared to benign ones.

Location-based Features: The spatial coordinates and relative positioning within the lung fields also contributed to the model's decision-making process, as certain nodule locations are more indicative of malignancy.
The achieved accuracy of 95.67% highlights the potential of XGBoost in clinical settings for lung cancer detection. This high level of accuracy can significantly aid radiologists by providing a reliable second opinion, thereby reducing diagnostic errors and improving early detection rates. Early and accurate diagnosis is crucial for effective treatment and better patient outcomes, particularly in lung cancer, where survival rates drastically improve with early intervention.
The study effectively addressed several challenges associated with lung nodule classification:

Non-uniform Density: By employing advanced feature extraction techniques and the robust learning capabilities of XGBoost, the model managed to handle the varying density of nodules within and between scans.

Intra-scene and Inter-scene Variations: The use of data augmentation and the ability of XGBoost to capture complex patterns helped in managing variations in scale and perspective.

Occlusions: Although occlusions pose a significant challenge, the high accuracy suggests that the model effectively learned to recognize nodules even in partially occluded scenarios.
While the results present promise, there are notable limitations necessitating future work. Firstly, the model's performance must undergo validation on external datasets to ascertain its generalizability across diverse populations and imaging conditions. Secondly, for integration into real-time clinical workflows, additional optimization and testing are imperative to ensure the model meets requisite speed and reliability standards. Moreover, there's potential for enhancing predictive capabilities by exploring the integration of clinical data, including patient history and genetic information, with imaging data. Addressing these limitations could significantly advance the model's applicability and effectiveness in clinical settings.

IV. CONCLUSION

This study demonstrates the efficacy of the XGBoost ensemble learning method for the classification of lung nodules using the LIDC-IDRI dataset. By leveraging the extensive and detailed annotations provided by expert thoracic radiologists, the XGBoost model achieved an impressive classification accuracy of 95.67%. This high accuracy underscores the model's potential to significantly aid in the early detection and diagnosis of lung cancer, which is crucial for improving patient outcomes.The comparative analysis with other machine learning models, including SVM and kNN, highlights the superior performance of XGBoost. The model's ability to handle high-dimensional data, its robustness to overfitting through regularization, and its effectiveness in managing missing data contribute to its outstanding performance. Furthermore, the insights gained from feature importance analysis reveal the critical role of intensity, texture, shape, and location-based features in accurately classifying lung nodules.

While the results are promising, there are areas for future work to enhance the practical application of this model. Validating the model on external datasets is essential to ensure its generalizability across different populations and imaging conditions. Additionally, optimizing the model for real-time clinical use and exploring the integration of clinical data with imaging features could further improve its diagnostic capabilities.In conclusion, this research underscores the potential of advanced ensemble learning methods like XGBoost in the field of medical image analysis. By providing a reliable and accurate tool for lung nodule classification, this study contributes to the ongoing efforts to enhance early lung cancer detection and improve patient care. Continued advancements and validations in this area will be vital in translating these findings into practical clinical applications, ultimately leading to better health outcomes for patients worldwide.

Compliance with ethical standards
Disclosure of Conflict of Interest
No Conflict of Interest to be disclosed.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 6, June 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1206103 |

REFERENCES

1. Luo G, Zhang Y, Etxeberria J, Arnold M, Cai X, Hao Y, Zou H. Projections of lung cancer incidence by 2035 in 40 countries worldwide: population-based study. JMIR Public Health and Surveillance. 2023 Feb 17;9(1):e43651.
2. Binson VA, Subramoniam M, Mathew L. Prediction of lung cancer with a sensor array based e-nose system using machine learning methods. Microsystem Technologies. 2024 Apr 18:1-4.
3. .Binson VA, Subramoniam M, Thomas S. Trends in Lung cancer: The Incidence and Mortality Rate in India. Int J Eng Adv Technol. 2019 Jun;8:1956-62.
4. Huang S, Yang J, Shen N, Xu Q, Zhao Q. Artificial intelligence in lung cancer diagnosis and prognosis: Current application and future perspective. InSeminars in Cancer Biology 2023 Feb 1 (Vol. 89, pp. 30-37). Academic Press.
5. Bhandary A, Prabhu GA, Rajinikanth V, Thanaraj KP, Satapathy SC, Robbins DE, Shasky C, Zhang YD, Tavares JM, Raja NS. Deep-learning framework to detect lung abnormality–A study with chest X-Ray and lung CT scan images. Pattern Recognition Letters. 2020 Jan 1;129:271-8.
6. Armato III SG, McLennan G, Bidaut L, McNitt-Gray MF, Meyer CR, Reeves AP, Zhao B, Aberle DR, Henschke CI, Hoffman EA, Kazerooni EA. The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans. Medical physics. 2011 Feb;38(2):915-31.
7. Zheng M. Classification and pathology of lung cancer. Surgical Oncology Clinics. 2016 Jul 1;25(3):447-68.
8. Binson VA, Thomas S, Subramoniam M, Arun J, Naveen S, Madhu S. A Review of Machine Learning Algorithms for Biomedical Applications. Annals of Biomedical Engineering. 2024 Feb 21:1-25.
9. Gao Q, Yang L, Lu M, Jin R, Ye H, Ma T. The artificial intelligence and machine learning in lung cancer immunotherapy. Journal of Hematology & Oncology. 2023 May 24;16(1):55.
10. VA B, Mathew P, Thomas S, Mathew L. Detection of lung cancer and stages via breath analysis using a self-made electronic nose device. Expert Review of Molecular Diagnostics. 2024 Feb 19:1-3.
11. Lynch CM, Abdollahi B, Fuqua JD, De Carlo AR, Bartholomai JA, Balgemann RN, van Berkel VH, Frieboes HB. Prediction of lung cancer patient survival via supervised machine learning classification techniques. International journal of medical informatics. 2017 Dec 1;108:1-8.
12. Binson VA, Thomas S, Philip PC, Thomas A, Pillai P. Detection of Early Lung Cancer Cases in Patients with COPD Using eNose Technology: A Promising Non-Invasive Approach. In2023 IEEE International Conference on Recent Advances in Systems Science and Engineering (RASSE) 2023 Nov 8 (pp. 1-4). IEEE.
13. Thomas S, Thomas J. Non-destructive silkworm pupa gender classification with X-ray images using ensemble learning. Artificial Intelligence in Agriculture. 2022 Jan 1;6:100-10
14. Binson VA, Subramoniam M, Mathew L. Detection of COPD and Lung Cancer with electronic nose using ensemble learning methods. Clinica Chimica Acta. 2021 Dec 1;523:231-8.
15. Thomas S, Thomas J. An optimized method for mulberry silkworm, Bombyx mori (Bombycidae:Lepidoptera) sex classification using TLBPSGA-RFEXGBoost. Biology Open. 2024 Jun14:1-11.
16. Armato III SG, McLennan G, Bidaut L, McNitt-Gray MF, Meyer CR, Reeves AP, Zhao B, Aberle DR, Henschke CI, Hoffman EA, Kazerooni EA. The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans. Medical physics. 2011 Feb;38(2):915-31.
17. Thomas S, Thomas J. Nondestructive and cost-effective silkworm, Bombyx mori (Lepidoptera: Bombycidae) cocoon sex classification using machine learning. International Journal of Tropical Insect Science. 2024 Mar 25:1-3.
18. Deepa SR, Subramoniam M, Binson VA, Poornapushpakala S, Barani S. Precision Diagnostic Algorithm for Multisubtype Arrhythmia Classification. In2023 IEEE International Conference on Recent Advances in Systems Science and Engineering (RASSE) 2023 Nov 8 (pp. 1-4). IEEE.
19. Yu D, Liu Z, Su C, Han Y, Duan X, Zhang R, Liu X, Yang Y, Xu S. Copy number variation in plasma as a tool for lung cancer prediction using Extreme Gradient Boosting (XGBoost) classifier. Thoracic cancer. 2020 Jan;11(1):95-102.
20. Binson VA, Thomas S. Unveiling the Smell of Health: E-Nose-Based Volatile Organic Compound Analysis of Exhaled Breath in Early Lung Cancer Detection. InProceedings 2024 May 29 (Vol. 100, No. 1, p. 23). MDPI.
21. Singh D, Khandelwal A, Bhandari P, Barve S, Chikmurge D. Predicting Lung Cancer using XGBoost and other Ensemble Learning Models. In2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT) 2023 Jul 6 (pp. 1-6). IEEE.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details