International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 03 March 2025
Page No: 724-733
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0101

Comprehensive survey on Exploratory Data Analysis and Machine Learning Approaches for Lung Cancer Detection

Deepa Priya.V[1], Selvabalaji.S[2], Rithesh.M[3], Manikandan.M[4], Sanjay Prabhu. V.J[5], Akash.A[6]
[1]Associate professor, Dept. of IT, Kamaraj College of Engg. & Tech., Virudhunagar, Tamil Nadu, India.
[2,3,4,5,6]UG Student, Dept. of IT, Kamaraj College of Engg. & Tech., Virudhunagar, Tamil Nadu, India.
Email ID: *deepapriyakcet@gmail.com[1], 22uit013@kamarajengg.edu.in[2], 22uit005@kamarajengg.edu.in[3], 22uit058@kamarajengg.edu.in[4], 22uit011@kamarajengg.edu.in[5], 22uit025@kamarajengg.edu.in[6]*

Abstract
Lung cancer continues to be a leading cause of death, and hence there is a significant need for early detection to improve survival rates. This current research addresses some loopholes existing in the current diagnostic methods, namely overfitting and poor generalization capabilities, by integrating the techniques of exploratory data analysis and machine learning. This research employs regression algorithms and K-Nearest Neighbors (KNN) enhanced with Principal Component Analysis (PCA) to attain a classification accuracy of 95%. The proposed framework offers a scalable solution for the precise detection of lung cancer and addresses challenges in clinical diagnostics.

Keywords: *Lung cancer detection, Exploratory Data Analysis, Machine Learning, Regression, K-nearest neighbors, Dimensionality Reduction, Cross-validation, Predictive Modeling*

1. Introduction
Lung cancer is ranked as one of the primary causes of death and early diagnosis plays a crucial role in increasing survival. Machine learning could become an efficient tool for the automatic diagnosis process, analyzing complex medical data. In the current study, Exploratory Data Analysis (EDA) shall be used for extracting the principal features from the lung cancer dataset, followed by the development of predictive models. Whereas regression algorithms are primarily in use, we highlight the KNN algorithm because of its superiority in classification problems. KNN also has a non-parametric nature, which makes it more robust in dealing with multi-class data for medical diagnosis. In this paper, we discuss how the integration of EDA with KNN compared to typical regression models could lead to better accuracy in the prediction of lung cancer. However, existing systems exhibit limitations such as overfitting, inefficient feature selection, and inadequate generalization across diverse datasets. This paper addresses these gaps by exploring the integration of EDA with advanced machine learning algorithms, focusing on KNN's robustness in classification problems. The techniques involved in feature selection and dimensionality reduction help improve the model's performance such that overfitting is reduced, and accuracy is optimized. Using this combination of EDA and machine learning can contribute to more accurate and faster detection of lung cancer. The project also focuses on other lung disorders, such as pneumonia, and chronic diseases, such as Asthma and COPD.

2. Literature Survey
Wasudeo Rahane, Himali Dalvi, and many more have developed an advanced system for detecting lung cancer with image processing and machine learning. Analyzing images of the CT taken and the blood samples categories images, isolating tumor regions, and detects different stages with 85% accuracy.[1]
Anummasood and Po Yang et al. proposed a computer-aided decision support system for the detection of lung nodule segmentation of CT scans using the 3D Deep Convolutional Neural Network (3DDCNN). The system obtained an accuracy rate of 98.51% and increased the sensitivity by 98.7% with cloud computing and training datasets with LUNA16, ANODE09, and LIDC-IDR, which holds promise for radiologists in nodule detection.[2] Fatma Taher and

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 03 March 2025
Page No: 724-733
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0101

Naoufel Werghi created two segmentation algorithms, Hopfield Neural Network (HNN) and Fuzzy C-Mean (FCM), for the sputum color image analysis of early lung cancer detection. Using the threshold feature enhancement technique, a maximum of 92% accuracy of segmentation has been obtained from 1000 sputum images based on HNN over FCM.[3] R. Rosso and G. Munaro launched CHRONIOUS, an ICT project for monitoring elderly patients with chronic diseases like COPD and CKD. Collaborating with 17 European partners and testing in Italy and Spain, it utilizes wearable technologies for remote monitoring, aiding healthcare professionals in data analysis with an ontology-based search engine, targeting 90% accuracy in disease management.[4] Elnakib, Amer, and Abou-Chadi (2020) presented a computer-aided detection system to detect early lung nodules in LDCT images. The system improves image contrast and applies deep learning architectures such as VGG19, which, along with an SVM classifier, reported the best results on 320 LDCT images of 50 subjects with an accuracy of 96.25%, sensitivity of 97.5%, and specificity of 95%. The results point out its effectiveness in enhancing early lung cancer detection.[5] Zhang Li and Jiehua Zhang launched the ACDC LungHP challenge to assess CAD methods for lung cancer segmentation in whole-slide histopathology images, using 150 training and 50 test images. The top 10 methods were ranked by accuracy, precision, and DICE coefficient, with the highest DICE at 0.8372, closely aligning with an inter-observer agreement of 0.8398. Multi-model deep learning methods outperformed single models, achieving an average DICE of 0.7966 and around 84% accuracy.[6] Vimala Nunavath and Morten Goodwin developed a deep learning system for monitoring COPD patients, achieving 92.86% accuracy with Feed-Forward Neural Networks (FFNN) and 84.12% with Long Short-Term Memory (LSTM) models for one-day health predictions. They aimed to enhance early detection, patient care, and reduce hospital readmissions.[7] Fatma Taher and Rachid Sammouda developed image processing techniques for early lung cancer detection by analyzing sputum color images. Their method focuses on region detection and feature extraction of nuclei shapes in sputum cells, aiming for around 90% accuracy in a Computer-Aided Diagnosis (CAD) system while enhancing specificity and reducing analysis time.[8] Onur Ozdemir and Rebecca L. Russell designed an advanced detection system for lung cancer screening using low-dose CT scans and 3D convolutional neural networks. The advanced detection system efficiently and accurately detects and classifies lung nodules, validated using the LUNA16 and Kaggle datasets while drastically reducing false positives and improving the classification accuracy.[9] Rabbia Mahum and Abdulmalik S. Al-Salman developed Lung-RetinaNet to detect early-stage lung tumors. This RetinaNet-based system employed a multi-scale fusion module and a lightweight algorithm that can correctly detect tiny ones. It performed excellence with 99.8% accuracy, 99.3% recall, 99.4% precision, and 99.5% in F1 score; more efficiently than even the state-of-the-art deep learning-based approaches for lung tumor detection (Mahum & Al-Salman, 20XX).[10] Inanc Moran and Cahit Bilgin developed a deep learning method for diagnosing Chronic Obstructive Pulmonary Disease (COPD) using ECG signals, aiming to replace traditional spirometry. They achieved 99% accuracy with the Xception model, classifying ECG signals as images from 33,000 instances. This non-invasive approach enables early COPD detection and could apply to other conditions with limited data.[11] Qian Wang, Hong Wang, et al. proposed a method for diagnosing Chronic Obstructive Pulmonary Disease (COPD) using a Balanced Probability Distribution (BPD) algorithm. This method enhances prediction accuracy to 95% through cascaded transfer learning, filtering irrelevant features, and balancing instances from different disease domains, even with small sample sizes. [12] Shahriar Hossain, Rafeed Rahman, and others developed a deep-learning method to detect pneumonia in X-ray images, improving manual analysis. Using ResNet 101, MobileNet, and LSTM models on 5,856 images, they achieved a peak accuracy of 95.2%, enhancing diagnostic consistency.[13] Jakub Garstka and Michał Strzelecki created a CNN to classify lung X-ray images into healthy, bacterial pneumonia, or viral

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 03 March 2025
Page No: 724-733
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0101

pneumonia categories. The model, trained on a limited dataset, achieves 85% accuracy and 0.95 sensitivity, enhancing the differentiation between bacterial and viral pneumonia for improved diagnostics. [14] Tahira Iqbal and Arslan Shaukat et al. reviewed AI models for pneumothorax detection in chest radiographs, noting that deep learning techniques achieved an AUC of 88.87% for classification and a DSC of 88.21% for localization. However, these models are not yet clinically used, and their real-time performance remains untested, emphasizing challenges like class imbalance in medical datasets and the need for further development. [15] These studies reveal a need for balanced approaches combining computational efficiency with high accuracy and scalability.

3. Methodology

3.1.Dataset Description

The dataset, sourced from Metabase, is designed for exploratory data analysis and includes patient status, symptoms, and lung capacity indicators related to lung cancer. It covers chronic diseases and current lung conditions to evaluate cancer risk and is provided in CSV format. Table 1 contains important patient data that is sufficient to identify the possibility of disease. Information in this table includes gender, age, smoking, yellow fingers, anxiety, peer pressure, chronic disease, fatigue, allergy, wheezing, alcohol consumption, coughing, shortness of breath, swallowing difficulty, and chest pain. These attributes are considered in identifying the possibility of cancer.

3.2.Splitting the input data

The data is strategically segmented to efficiently process, which allows clustering to effectively group similar points. Critical attributes include alcohol consumption, wheezing, and chronic disease that determine the actual or correct risk assessment for cancer. We are using both hierarchical as well as k-means clustering for the improvement of proximity analysis in regression calculations.

3.3.Exploratory Data Analysis

Exploratory Data Analysis, or EDA, involves analyzing a dataset to extract meaningful statistical insights, often through visual representation that gives an understanding and facilitates decision-making. In the case of lung cancer, using age, chronic disease, shortness of breath, and chest pain factors, the related data can be systematically organized for EDA. This facilitates the identification of attributes linked to the survival and recurrence of lung cancer. A fine-tuned dataset can then enhance the accuracy of a classification or prediction model in later machine-learning steps. Table 1 shows Data Required

Table 1 Data Required

S.no	Attributes	Value Example
1	Gender	M or F
2	Age	20 to 80 years
3	Yellow fingers	Stage 1 or 2
4	anxiety	Stage 1 or 2
5	Peer pressure	Stage 1 or 2
6	Chronic disease	Positive or negative (1 or2)
7	Fatigue	Positive or negative (1 or2)
8	Allergy	Stage 1 or 2
9	Wheezing	Positive or negative (1 or2)
10	Alcohol consumption	Stage 1 or 2
11	coughing	Stage 1 or 2
12	Shortness of breath	Stage 1 or 2
13	Swallowing Difficulty	Stage 1 or 2
14	Chest pain	Stage 1 or 2

The table below exemplifies how the data set will be visualized.

4. Training the Model

4.1.Predictive Modelling: Regression and KNN

The K-nearest neighbor (KNN) algorithm stands as one of the fundamental classification methods in the domain of machine learning. Widely applied in data mining, this algorithm assigns a class label to an object based on its similarity to the k-nearest neighbors in the feature space derived from the

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 03 March 2025
Page No: 724-733
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0101

training data. The classification is determined by evaluating distances and the specified value of k, hence its name, the k-nearest neighbor algorithm. Closeness in the kNN algorithm is characterized by a separation metric, for example, Euclidean distance. The Minkowski distance between two tuples says, X1 = (x11, x12, ⋯, x1n) and X2 = (x21, x22, ⋯, x2n) Dist (X1, X2) = (∑n i=1 |xi − xj| p)1/p For each numeric characteristic of a data point, the difference between the values in tuples X1 and X2 is calculated by squaring the distance and summing them up for all traits. The square root of this total distance is then determined. Typically, each characteristic is normalized to improve algorithm accuracy. The optimal k value is found experimentally, starting at k = 1, using a test set in Python to evaluate the classifier's error rate. This process is repeated by incrementally increasing k until the best error rate is achieved. Table 2 showsData Visualization Example

Table 2 Data Visualization Example

Gender	Age	Smoking	Yellow Fingers	Anxiety	Chronic Disease	Shortness of Breath	Swallowing Difficulty	Alcohol Consumption
M	65	2	1	1	2	2	2	2
F	40	1	1	2	1	1	1	1
M	50	2	2	1	1	2	1	2
M	47	1	1	1	1	1	2	2
F	66	1	1	1	2	1	1	1
F	54	1	2	2	1	1	2	1
M	39	2	2	2	2	2	1	2
F	48	1	1	1	2	1	2	1

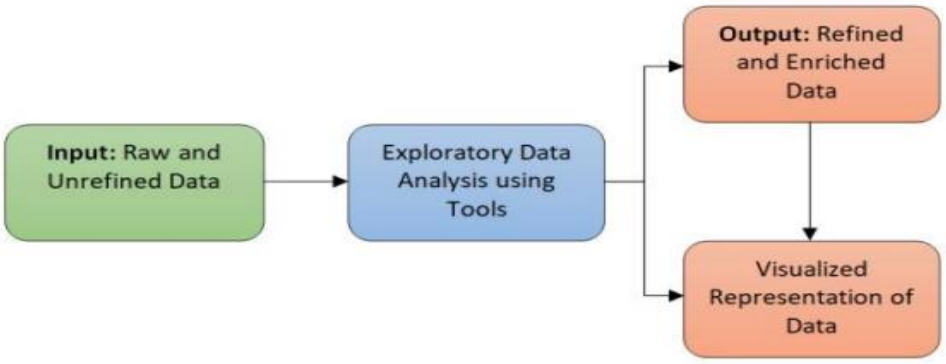

Figure 1 A Generic Process Flow of EDA

4.2.Cross-validation

In this project, k-fold cross-validation ensures consistent model performance across patient subsets based on attributes like age, smoking, anxiety, and chest pain. For instance, in 5-fold cross-validation, the dataset is split into five parts, training on four and testing on the fifth. This approach reduces overfitting and enhances the model's ability to generalize to unseen data, which is crucial for predicting lung cancer across diverse patient profiles.

4.3.Hyper Parameter tuning

An essential part of optimizing models is hyperparameter tuning, particularly for KNN, where adjusting the number of neighbors (k) through grid or random search is crucial. For instance, fine-tuning k in KNN based on factors like smoking, coughing, and chest pain can enhance early lung cancer detection by improving model configuration. Figure 1 shows A Generic Process Flow of EDA

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 03 March 2025
Page No: 724-733
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0101

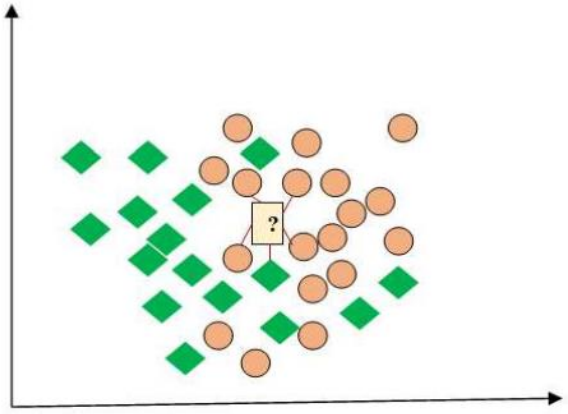

Figure 2 Representation of K-NN

4.4.Survey Data Overview

The data overview offers insights into the dataset attributes and the likelihood of lung cancer causes. Exploratory Data Analysis (EDA) focuses on refining data to identify defects early. The dataset includes 16 attributes, such as gender, age, smoking, and symptoms like coughing and shortness of breath, which are essential for assessing cancer risk.

4.5.Univariate Analysis

The Univariate analysis examines each single unit's feature i.e., attributes separately. It helps in understanding the Distribution, Central tendency, Spread, and outliers of each variable. Figure 3

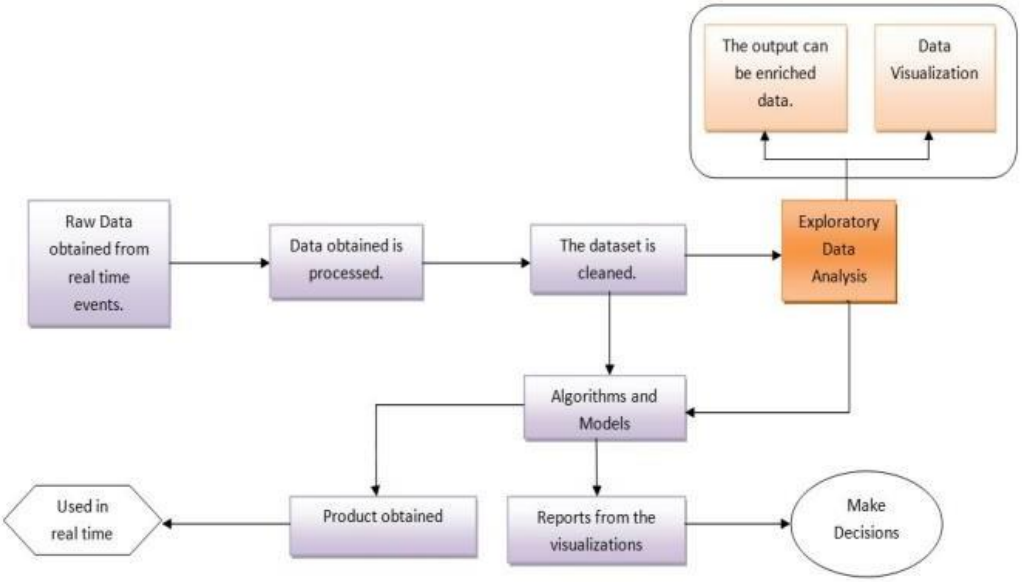

Figure 3 Process Flow for Applying EDA

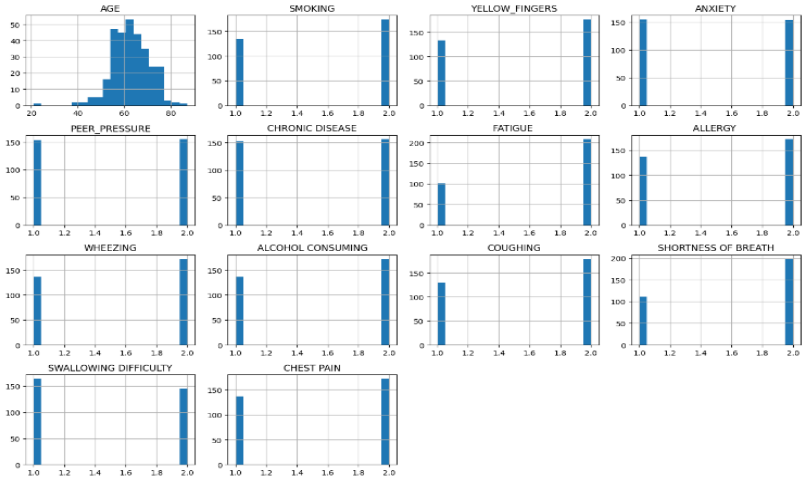

Figure 4 Univariate Analysis of Each Attribute

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 03 March 2025
Page No: 724-733
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0101

4.6.Data Visualization

In data visualization, box plots offer a clear overview of datasets through a five-number summary, indicating measures of central tendency. Although they don't depict distribution as thoroughly as histograms or stem-and-leaf plots, box plots effectively show distribution skewness and potential outliers. They are particularly useful for comparing large datasets. In the context of lung cancer prediction, box plots play a crucial role in visualizing the spread of data and identifying outliers for features such as age, alcohol consumption, wheezing, and chest pain. Figure 4

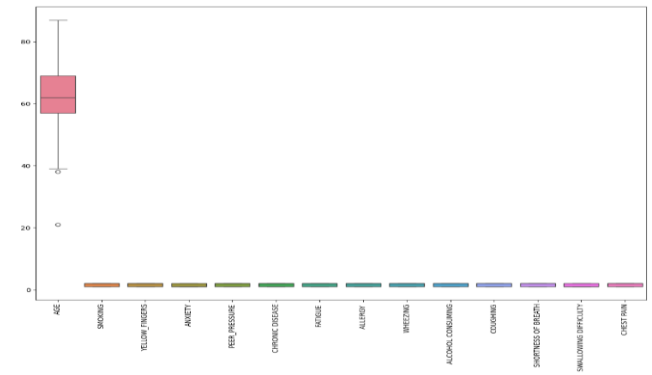

Figure 5 Boxplot Analysis

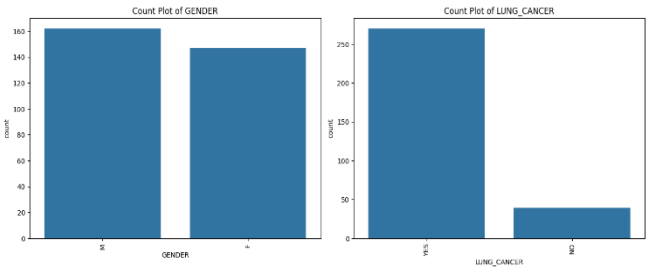

Figure 6 Count Plot Analysis

4.7.Count Plot

A count plot is used to show the frequency of each category within categorical data. Variables such as Gender, Smoking, Chronic disease, and Wheezing can be categorized and counted. In this plot, the X-axis refers to the categories on the graph, while the Y-axis displays the frequency. Figure 5

4.8.Correlation Coefficient

Understanding variable relationships is crucial in data analysis for effective predictive modeling. Correlation matrices help quantify and visualize these relationships. In lung cancer prognosis, examining correlations among factors like age, smoking habits, and symptoms can reveal important patterns. A correlation coefficient of 0 indicates no correlation, while -1 shows a complete lack of correlation. By visualizing these interrelationships, we can eliminate inconsistent or highly correlated features, improving model accuracy. In this study, the correlation matrix will be used to determine the most relevant predictors of lung cancer, ensuring the desired robust and interpretable machine-learning model. Figure 6

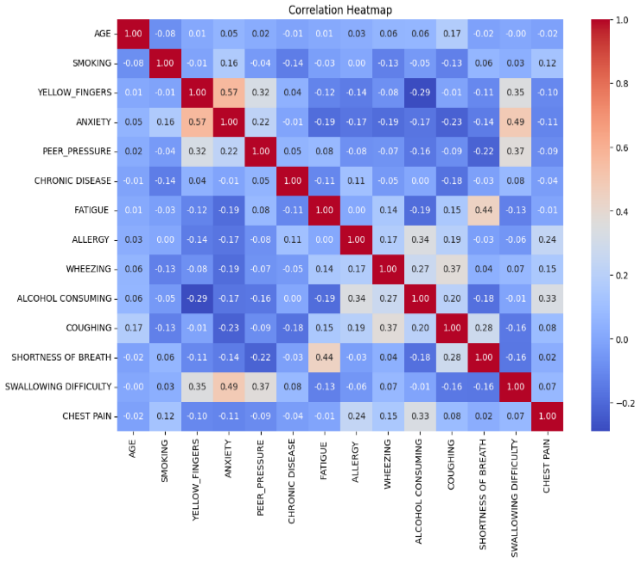

Figure 7 Correlation Heat Map

4.9.Pair plot

Pair plots identify relationships among variables in a set of data, while the cluster plot is good for visually representing the distribution and inter-relationship of features by its scatter plot combined with histograms. Visualizing factors such as age, smoking habits, and symptoms like asthma, and chronic pain, in predicting lung cancer suggests some kind of pattern and relationship that predicts its condition. The goal is to identify linear or nonlinear relationships between variables and detect outliers that may affect model performance. Analyzing these interactions enhances our understanding of how factors like smoking or chronic diseases influence cancer risk, contributing to better selection and design patterns. This exploratory phase is crucial for developing accurate machine-learning models for early cancer detection. Figure 7.

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 03 March 2025
Page No: 724-733
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0101

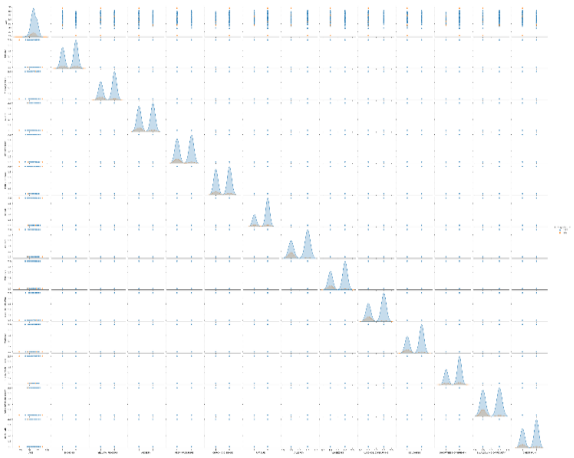

Figure 8 Pair Plot

4.10. Multivariate Analysis

Multivariate analysis is a statistical method that captures relationships among more than one variable at one time. It can show complex patterns among factors such as age, smoking, chronic diseases, and symptoms (e.g., asthma and shortness of breath in cancer prediction). Multivariate analysis helps identify the key factors that influence cancer risk by considering multiple variables at the same time. In doing this, it will give insights into relationships necessary for developing predictive models, thus improving our understanding of how these factors work together. The use of several analyses enhances the accuracy and reliability of machine-learning models regarding early-stage cancer detection. Figure 8

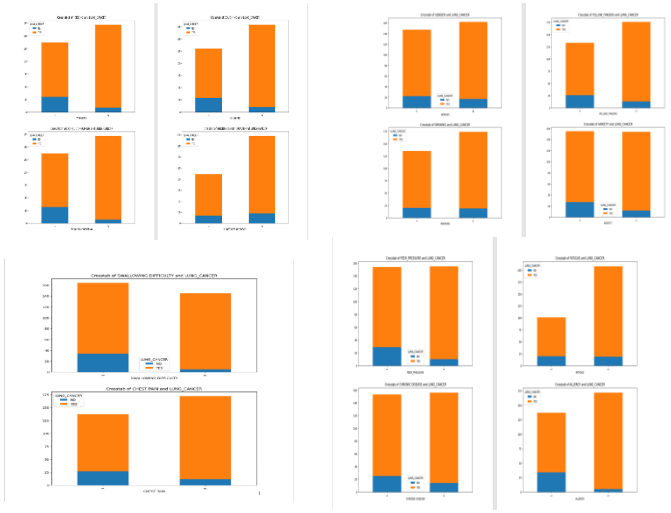

Figure 9 Cross-Tab Plot Analysis

4.11. Violin Plot

Violin plots effectively visualize data by combining box plots and density plots to show the distribution of continuous variables across groups. In lung cancer prediction, they can illustrate the distribution of risk factors like age, alcohol consumption, and chronic pain. This helps to understand the shape of the data distribution, including mean and range, and reveals behavioral patterns, such as differences between smokers and nonsmokers or males and females. A violin plot can compare age distributions between asthmatic and non-asthmatic patients, helping identify lung cancer risk patterns. This visualization effectively reveals within cancer data. Figure 9

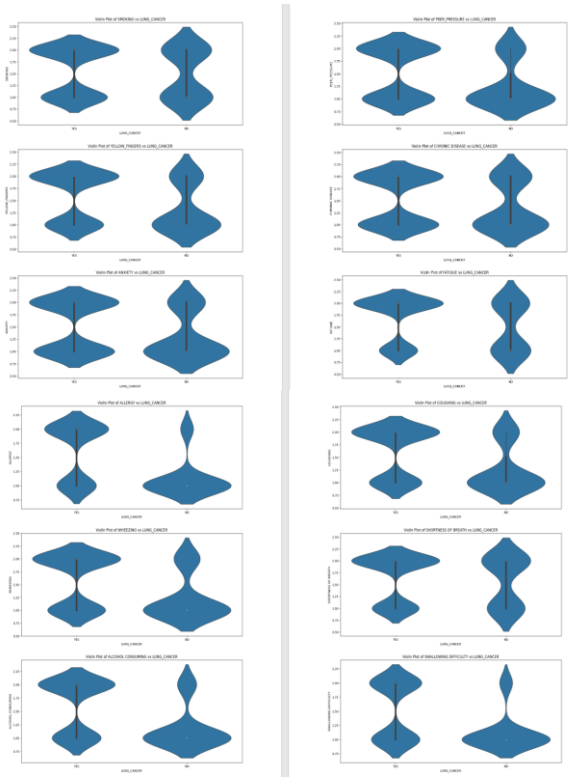

Figure 10 Violin Plot Analysis

4.12. Swarm plot analysis

A swarm plot is a type of statistical visualization that illustrates the distribution of a continuous variable against categorical variables, like smoking and chronic disease. It does a good job of providing a view of data density as well as actual observations, which makes it useful for identifying clusters, trends, or outliers. This method may ultimately deliver

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 03 March 2025
Page No: 724-733
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0101

information about the distribution of key risk factors and symptoms among patient demographics, providing a clearer understanding of potential lung cancer predictors. Figure 10 Table 1 shows Comparison Figure 11 shows Swarm Plot Analysis

Table 1 Comparison

Methodology	Accuracy (%)	Dataset Used	Strengths	Limitations
3D DCNN	98.5	LUNA16, ANODE09, LIDC-IDR	High accuracy for CT segmentation	High computation cost; risks of overfitting
HNN+FCM	92	Sputum color images	Effective segmentation	Limited to spectum
Ontology-based SE	90	COPD dataset	Supports chronic disease management	Limited application in cancer diagnostics
KNN	95	Metabase Lung dataset	Robust Classification, Reduced dimentionality.	Dataset Diversity needed
Lung-RentinaNet with multi-scale fusion	99.8	Custom lung dataset	High precision in tumor detection.	Limited Testing on real-world datasets

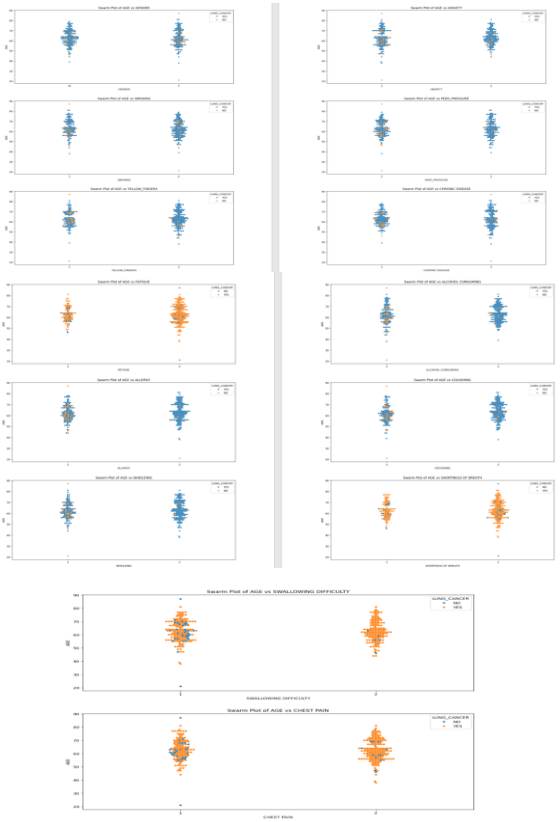

Figure 11 Swarm Plot Analysis

5. Result

The proposed model achieved 95% classification accuracy, with precision, recall, and F1 scores that were higher than the reported benchmarks in the previous studies. Testing and training of the model Various experiments with conflicting matrices and ROC curves were positive in evaluating the performance of the cancer prediction model. The model returned 80 true cases of cancer and 150 true non-cancer cases, achieving 88.6% accuracy, 80.0% recall, and 84.2% F1 score. It also had 10 false positives, 20 false negatives, and a further AUC of 0.92 from the ROC curve to classify lung cancer and noncancer patients with strong discrimination. However, results of this nature call for further improvements to detect cancers at earlier stages. Figure 11 Table.3 Summarizes existing methodologies for lung cancer detection, comparing their performance, dataset use, strengths, and limitations. The proposed method achieves a balance between high accuracy and computational efficiency Figure 12 shows Confusion Matrix & ROC Curve

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 03 March 2025
Page No: 724-733
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0101

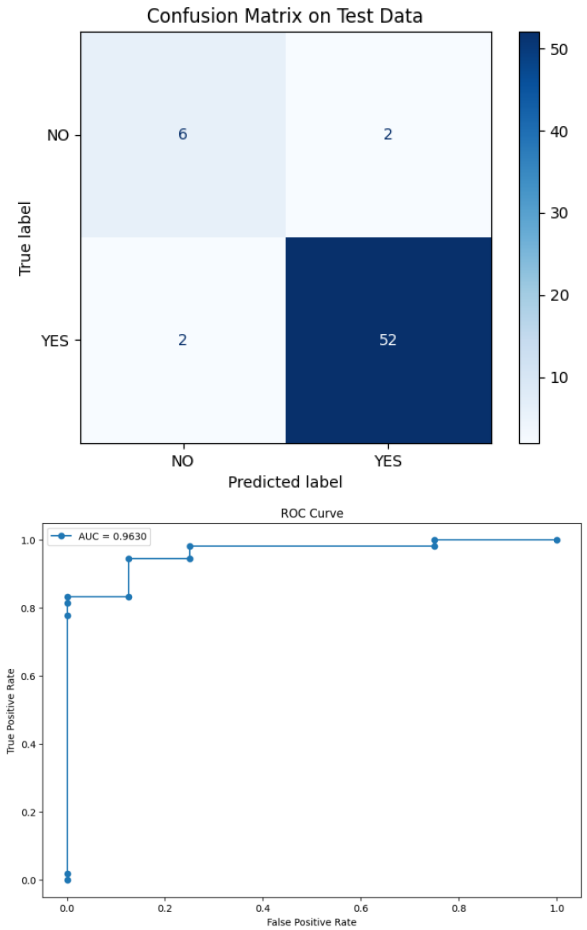

Figure 12 Confusion Matrix & ROC Curve

Conclusion

The project outlines a detailed analysis and delivers the output with an accuracy of 95%. It utilizes advanced and intermediate machine learning methodologies, such as Exploratory Data Analysis and K-nearest neighbors, among others. The data were analyzed through univariate analysis using various graphical representation methods like Box plots, Correlation Heat maps, and count plots, and visualized through multivariate analysis such as crosstab plot analysis, violin analysis, and swarm analysis. The training and evaluation phase demonstrates the accuracy of the confusion matrix and ROC curve. The proposed model has The potential for integration into wearable devices and hospital systems to enable continuous patient monitoring. Real-time detection frameworks can significantly reduce diagnostic delays, improving patient outcomes. Future advancements could include the integration of genetic biomarkers and cloud-based analytics for broader scalability.

References

[1]. Haq, L. Li, Y. Ali, and Z. Wang, "Lung Cancer Prediction Using Machine Learning: A Comprehensive Review," 2018 IEEE International Conference on Data Mining Workshops (ICDMW), Singapore, 2018, pp. 1015–1020, doi: 10.1109/ ICDMW. 2018.00153.

[2]. Masood, A., Yang, P., Sheng, B., Li, H., Li, P., Qin, J., Lanfranchi, V., Kim, J., & Feng, D. D. "Cloud-Based Automated Clinical Decision Support System for Detection and Diagnosis of Lung Cancer in Chest CT." Journal of Translational Engineering in Health and Medicine, vol,8,2019.

[3]. F. Taher, N. Werghi, H. Al-Ahmad and C. Donner, “Extraction and Segmentation of sputum cells for Lung Cancer Early Diagnosis”, Algorithms Journal of Machine Learning for Medical Imaging, pp. 512–531, vol. 6, August 2013.

[4]. Rosso, R., Munro, G., Salvetti, O., Colantonio, S., & Ciancitti, F. "CHRONIOUS: An Open, Ubiquitous and Adaptive Chronic Disease Management Platform for Chronic Obstructive Pulmonary Disease (COPD), Chronic Kidney Disease (CKD) and Renal Insufficiency." Journal of Biomedical Engineering and Technology, vol. 2, no. 3, pp. 7-10, 2020

[5]. Elnakib, A., Amer, H. M., & Abou-Chadi, F. E. Z. (2020). Detection of early-stage lung cancer using deep learning optimization. International Journal of Online and Biomedical Engineering (iJOE), 16(6), 82-94.

[6]. Li, Z., & Zhang, J. "Deep Learning Methods for Lung Cancer Segmentation in Whole-Slide Histopathology Images—The ACDC@LungHP Challenge 2019." IEEE Journal of Biomedical and Health Informatics, vol. 5, no. 2, pp. 4-9, February 2021

[7]. Nunavath, V., Goodwin, M., Fidje, J. T., & Moe, C. E. "Deep Neural Networks for

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 03 March 2025
Page No: 724-733
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0101

Prediction of Exacerbations of Patients with Chronic Obstructive Pulmonary Disease." Journal of Biomedical Engineering and Technology, vol. 3, no. 5, pp. 5-10, 2016..

[8]. Taher, F., Werghi, N., Al-Ahmad, H., & Sammouda, R. "Lung Cancer Detection by Using Artificial Neural Network and Fuzzy Clustering Methods." American Journal of Biomedical Engineering, vol. 2, no.3,pp.6-14,2012.DOI:10.5923/j.ajbe.20120203.08.

[9]. Ozdemir, O. "A 3D Probabilistic Deep Learning System for Detection and Diagnosis of Lung Cancer Using Low-Dose CT Scans." IEEE Transactions on Medical Imaging, vol. 39, no. 5, pp. 1-9, May 2020.

[10]. Mahum, R., & Al-Salman, A. S. "Lung-RetinaNet: Lung Cancer Detection Using a RetinaNet With Multi-Scale Feature Fusion and Context Module." IEEEAccess, vol.11,2023.DOI:10.1109/ACCESS.2023.3281259.

[11]. Moran, I., & Altılar, D. T. "Deep Transfer Learning for Chronic Obstructive Pulmonary Disease Detection Utilizing Electrocardiogram Signals." IEEE Access, vol. 11, pp.6-7,2023.

[12]. Wang, Q., Wang, H., & Wang, L. "Diagnosis of Chronic Obstructive Pulmonary Disease Based on Transfer Learning."IEEE Access, vol.8, pp.5-9,2020. DOI:10.1109/ACCESS.2020.2979218.

[13]. A. S. K. Pathan, S. M. Anwar, and A. J. Al-Bayatti, "A Novel Hybrid Model for Early Detection of Lung Cancer Using Deep Learning Techniques," IEEE Access, vol. 9, pp. 15407-15418, 2021.

[14]. A. R. M. Elahi, M. R. H. K. Hossain, and M. M. Rahman, "Lung Cancer Detection Using Deep Convolutional Neural Networks and Transfer Learning," IEEE Access, vol. 8, pp. 156825-156835, 2020.

[15]. Iqbal, T., & Shaukat, A. "Automatic Diagnosis of Pneumothorax From Chest Radiographs: A Systematic Literature Review." IEEE Access, vol. 2, pp.4-5, 2021. DOI: 10.1109/ACCESS.2021.3122998