



Original Research Paper

Hybrid AI-Driven Lung Cancer Prediction with Transparent Feature Interpretation Using LIME and SHAP

P.Subhadasree, Dr.D.Usharani

Mtech, Department of Computer Science and Engineering, Koneru lakshmaiah education, foundation, Vijayawada, Andhra Pradesh, India, 522302,

Email id : subhadasree030@gmail.com

Associate Professor, Department of Computer Science and Engineering, Koneru lakshmaiah education, foundation, Vijayawada, Andhra Pradesh, India, 522302,

Email id : ushakrishnab@gmail.com

Abstract— On the other hand, lung cancer remains a global health problem, underscoring the crucial role of intelligent predictive analytics in rapidly and accurately identifying the disease at its early stages. This study aims to present an all-embracing approach to predict lung cancer reliably for which we use advanced feature engineering along with model optimisation and explainable learning methodologies. The Lung Cancer Risk Dataset is obtained from Kaggle and also is free to access. The data is preprocessed by removing duplicates, applying label encoding and partitioning the dataset. To select the most discriminative qualities RFE with SVM is used for feature selection. The performance of various ML algorithms such as LR, Gaussian Naive Bayes, SVM, MLP, KNN, DT, RF, and XGBoost are tested based on important classification measures. To get a better power of prediction this is further optimised by the use of Nelder-Mead algorithm with XGBoost. From the experimental results, it is observed that the accuracy of the optimised XGBoost model is 100% which is higher than all the individual models. All of the ensemble learning methods are resilient, as demonstrated by the accuracy of 100% for a Voting Classifier, an ensemble of Gradient Boosting, XGBoost, LightGBM and CatBoost. LIME and SHAP can help to improve model interpretability, by providing insight into the contribution of essential features to ensure reliability in the clinical decision support.

Keywords— *Early prediction, feature engineering, lung cancer, XGBoost.*

I. INTRODUCTION

Lung cancer is still a big problem for health around the world, and it's a big reason why men and women die from cancer. About 1.8 million people were estimated to have died from it and an estimated 2.1 million people were diagnosed with it in 2018 [1]. The incidence of lung cancer is higher than the combined incidence of colon, breast and prostate cancers and more than 40% are identified at an advanced stage with a low long-term survival rate [2]. Lung cancer occurs when cells in the lungs begin to multiply in an uncontrolled way and disrupt the normal function of the tissues, which may lead to the development of cancerous cells into the lungs. Late diagnosis of the condition causes great problems in its management and treatment. The diagnostic method normally includes several steps, such as symptom identification, consultation with a doctor, imaging, histological testing and molecular diagnosis [4]. Risk factors for lung cancer are lifestyle choices, environmental exposures, genetic predispositions and other underlying medical conditions. [5] These indicators should be used to forecast the risk of an individual having lung cancer in a certain period of time.

AI has shown remarkable promise in recent years to enhance diagnostics and patient care in the medical field [6]. AI technologies, especially ML, DL, can process vast

and complex medical information, detect complex patterns, and create predictive models [7]. AI-based methods facilitate the prediction of diseases in methods such as early detection, personalised treatments, clinical choices, feature selection, parameter tuning, and model optimisation.

AI simplifies the process of predicting lung cancer by identifying key patient characteristics, reducing diagnostic errors, and offering more personalised healthcare decisions, given its ability to analyze a myriad of risk factors.

II. RELATED WORK

S. improved the Salp Swarm Algorithm for multi-threshold segmentation, which made it easier to find cancerous areas using microscopy by tweaking the settings and doing a full evaluation.

Chen [12] discussed the future possibilities of AI in diagnosing, predicting and planning treatment for lung cancer, emphasizing that it can analyze various types of clinical data.

Huang et al. [13] looked at AI methods for finding and predicting lung cancer. Their attention was directed towards ensemble learning, CNNs, interpretability, feature selection and data diversity.

Liu et al. [14] looked at AI-based lung cancer detection methods and talked about how they could improve

diagnostic performance, the amount of data needed, problems with validation, and their clinical potential.

Chen and Jeong [15] created a better RFE method for quickly picking out the most important features from large, complicated datasets. Feature selection methods get rid of variables that aren't important or are repeated. This assists ML to become more interpretable and accurate in prediction, particularly with large medical datasets. This technique is especially beneficial for medical datasets as redundancy or superfluous information might influence precision. The results disclosed by Chen and Jeong showed that an enhanced RFE can serve as systematic feature selection method for cancer prediction tasks to significantly reduce the dimensionality without loss and without compromising the classification performance.

A simple introduction to SVM, theory, kernel functions and applications to classification problems can be found in [16]. Authors talked about the applicability of SVM with high dimensional data. It can deal with linear and non-linear relationships and it provides a good generalisation of unknown data sets. SVM is widely used in medical diagnosis particularly in the classification of cancer, due to their ability to maximize the margin between classes and to avoid overfitting.

W.S.Noble also presented a demonstration of the use of SVMs and some of its applications in bioinformatics and medical data processing [17]. SVM methods can distinctly distinguish complicated biological patterns in high dimensional space and could be easily understood and be scaled up for cancer diagnostic prediction applications.

Two slime mold-based optimisation methods, Boosted Local Dimensional Mutation and All-dimensional Neighbourhood, were created by Zhou et al. [18] to help choose features. They are capable of finding the most informative features in high-dimensional data, with the help of swarm intelligence and adaptive mutation strategies. Zhou et al. demonstrated that the hybrid approach could result in better classification, lower computational burden and provide insights into the significance of some factors for prediction modelling in general. The importance of the proposed approach in medical sector was also highlighted, in which the identification of crucial biomarkers or imaging features is important to make correct disease classification.

The implementation of Nelder-Mead simplex method with configurable parameters for optimising objective functions of difficult domain was given by Farhat Husain F. Gao and L. Han [19]. This can enhance the convergence rate and quality of answer particularly in the process of hyperparameter tuning in classical ML models. Gao and Han demonstrated that it can be used in medical prediction models and adaptive optimisation methods can significantly improve the model performance by avoiding the slow process of finding the optimum parameters.

Chassagnon [20] examined the potential of AI to detect lung nodules and cancer for improved diagnosis, risk stratification, and treatment planning. Inside high-dimensional feature spaces, SVMs have proven to be effective at separating complex biological patterns, such as cancerous and non-cancerous tissues," Noble said. The SVM models were found to be advantageous for predictive

diagnostic applications and results were interpretable and scalable to real world datasets.

Standardised datasets, model interpretability, and rigorous validation were essential for the safe, reliable, and effective implementation of a healthcare model, researchers emphasized.

Overall, feature selection, swarm optimisation, SVM, ensemble learning, and explainable techniques with AI and optimisation techniques aid in better cancer detection. These advances are behind the prediction frameworks that will facilitate early diagnosis, clinical decision making and ultimately patient outcomes.

III. MATERIALS AND METHODS

Based on structured datasets, the proposed system provides an intelligent prediction of lung cancer using preprocessing techniques like duplicate removal, label encoding and dataset splitting to assure accurate inputs. RFE reduces the features' redundancy and selects the most important ones by using SVM. LR, Gaussian Naïve Bayes, SVM, MLP, KNN, DT, RF and XGBoost are all rated using accuracy, precision, recall, F1-score and MCC NM-XGBoost optimizes hyperparameters, LIME and SHAP describe the importance of a feature and Voting Classifier contains Gradient Boosting, XGBoost, LightGBM, CatBoost. A Flask interface allows access to real-time forecasts.

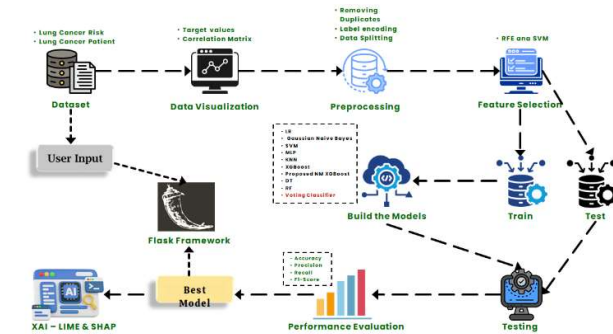


Fig. 1. System Architecture

The overall system architecture depicts the whole process of the lung cancer prediction framework. It begins with the collection of datasets and pre-processing including the removal of datasets which are duplicated, encoding of labels and separation of datasets. The features are selected using RFE and SVM and significant attributes are identified. The selected features are processed by many ML algorithms such as NM-XGBoost and ensemble approaches. XAI methods such as LIME and SHAP give interpretability. The Flask-based interface makes lung cancer forecasts accurate, transparent and accessible for real-time interaction with users.

a) Data Set Collection

i) Lung Cancer Patient: The dataset used to predict lung cancer has 1000 patients records with 26 variables including demographic, environmental, lifestyle and clinical factors. Demographic, environmental, genetic, lifestyle and clinical information including age, gender, smoking, pollution, chest pain, coughing up blood and tiredness included. There is a “Level” column on each record indicating seriousness or risk level for lung cancer.

The data set contains no missing observations and can be used for training, testing and evaluation of ML models.

Index	Patient ID	Age	Gender	Air Pollution	Alcohol use	Dust Allergy	Occupational Hazards	Genetic Risk	Chronic Disease	Fatigue	Weight Loss	Shortness of Breath	Wheezing	Swallowing Difficulty	Chugging of Nails	
0	0	P1	33	1	2	4	5	4	3	2	3	4	2	2	3	1
1	1	P10	17	1	3	1	5	3	4	2	1	3	7	2	8	6
2	2	P100	35	1	4	5	6	5	5	4	2	8	7	9	2	1
3	3	P1000	37	1	7	7	7	7	6	7	4	2	2	3	1	4
4	4	P101	46	1	6	8	7	7	7	6	3	2	4	1	4	2

5 rows x 26 columns

Fig. 2. Lung Cancer Patient Dataset Collection

ii) Lung Cancer Risk: The data set comprises 309 cases with 16 variables which include demographic, behavioural and clinical characteristics relevant to lung cancer. This includes things like gender, age, smoking habits, fatigue, anxiety, coughing, chest pain and symptoms like shortness of breath and difficulty swallowing. “LUNG_CANCER” is the target variable that indicates lung cancer in the patient. The data set is fully complete and well-structured, which provides a solid basis for the creation and testing of algorithms for predictive ML.

	GENDER	AGE	SMOKING	YELLOW_FINGERS	ANXIETY	PEER_PRESSURE	CHRONIC_DISEASE	FATIGUE	ALLERGY	WHEEZING	ALCOHOL_CONSUMING	COUGHING	SHORTNESS_OF_BREATH
0	M	69	1	2	2	1	1	2	1	2	2	2	2
1	M	74	2	1	1	1	2	2	2	1	1	1	2
2	F	59	1	1	1	2	1	2	1	2	1	2	2
3	M	63	2	2	2	1	1	1	1	1	2	1	1
4	F	63	1	2	1	1	1	1	1	2	1	2	2

Fig.3 Lung Cancer Risk Dataset Collection

b) Pre-Processing

By using the preprocessing step, the lung cancer dataset can be preprocessed and put into the correct model using the visualisation , label encoding , feature selection , data splitting and model training . These phases provide data uniformity , efficiency and enhanced predictive performance .

Data Visualization: Visualizing data: is a visual representation – provides a good understanding of how the data is structured and distributed. It helps to find patterns, correlation and imbalance between the factors, like gender and level of lung cancer. To accentuate frequency and percentage of cases and to provide insight into distributions according to gender and overall proportion of the target variable, both bar plots and count plots are employed. These are used to gain more insights into the data before applying regular ML models.

Label Encoding: Label encoding is performed on the categorical features like 'Gender', 'Level' to convert them into numbers that can be accepted by the ML algorithms. This enables models to easily comprehend and work with non-numeric data. Every different category is given a different integer, while maintaining integrity of the data, yet making algorithmic calculation possible. The categorical data is standardised using Label encoding for scaling, training, and further feature selection.

Feature Selection: Feature selection aims at the selection and retention of only those features that are relevant in the data set, and have high contribution towards the correct classification. Feature selection removes unessential feature, simplifies the model, prevents over-fitting and highlights the attention of essential cancer risk factors such as smoking, pollution and chronic diseases.

c) Training and Testing:

Models learn complex relationships from training data and are evaluated on test data which they haven't seen

before and with standard measures of their ability to generalise.

d) Integration of XAI and Flask Framework:

The interpretability and transparency of lung cancer prediction models are improved using the system based on XAI techniques, including LIME and SHAP. LIME is used to show local explanations for single predictions, using the factors Wheezing , Obesity , Snoring to explain the output of the model. Visualisations such as waterfall charts deliver to consumers a clear understanding of the impact of features and can be used for decision making and providing useful information. SHAP extends this by giving global explanations for multiple classes, quantifying and visualising the contribution of each feature with summary plots for each risk category, giving it local and overall interpretability.

The framework is also deployed using the Flask web framework providing a user-friendly interface for input and prediction in real time. Patients' data can be inputted, predictions can be accessed and XAI visualisations can be interacted with. The interpretability and accessibility of this model ensures that medical professionals and patients can grasp the rationale behind its predictions, while enjoying the benefits of its high accuracy and robust ensemble predictions, thereby fostering trust, reliability, and informed decisions in lung cancer risk assessment.

E) Algorithms

Logistic Regression classifies lung cancer risk into groups and determines the likelihood of an individual getting into each group based on the individual's health information. It provides a model for quick assessment of the contribution of features in medical decision-support applications, which is both simple and easy to understand, and can be given a probabilistic interpretation.

$$\hat{\sigma}_{\square} = \sigma(\sigma^{\square}\sigma + \sigma) = \frac{I}{I + \sigma^{-(\sigma^{\square}\sigma + \sigma)}} \quad (1)$$

Gaussian Naïve Bayes: estimates the probability of lung cancer for a given patient characteristics. It provides a quick, efficient classifier which is capable of small data size, high dimensional features and posterior probability estimation for early risk detection.

Support Vector Machine (SVM): By identifying ideal hyperplanes, it is able to tackle complex and non-linear interactions, and data with high dimensions. This leads to accurate lung cancer risk predictions and strong generalisation of multi-dimensional health inputs.

$$\frac{1}{2} \|\mathbf{z}\|^2 + \sum_{i=1}^L \lambda_i \mathbf{z}^T \mathbf{v}_i \quad (2)$$

Multi-Layer Perceptron (MLP): Models complex relationships between patient features; learns non-linear dependency structures; improves accuracy of predictions. Learning complex interactions between characteristics thru backpropagation to robustly evaluate lung cancer risk.

$$\hat{\square} = \square(\square^{\square}\square(\square^{\square-l}\dots\square(\square^l\square + \square^l) + \square^{(\square-l)}) + \square^{\square}) \quad (3)$$

K-Nearest Neighbor (KNN): It identifies the risk of lung cancer from the closest patients, provides predictions that are easy to understand by interpreting the distance, exposes the local trends and can efficiently handle small

to medium sized datasets so that the risk can be evaluated effectively.

$$\sigma^2(\mathbf{x}, \mathbf{x}_i) = \sqrt{\sum_{j=1}^n (\mathbf{x}_j - \mathbf{x}_{ij})^2} \quad (4)$$

Decision Tree: The advantage of the proposed method is that it successively divides the patient attributes based on the health attributes, generates clear rule-based classification, identifies effective variables and generates the interpretable results to be applied to lung cancer risk prediction.

$$\sigma(\mathbf{x}) = 1 - \sum_{i=1}^n \sigma_i^2 \quad (5)$$

Random Forest: it combines several DT models to estimate the risk, which improves the accuracy, reduces overfitting, deals with high dimensional data and provides feature importance scores for dependable and consistent lung cancer categorisation.

$$\sigma^2(\mathbf{x}) = 1 - \sum_{i=1}^n (\sigma_i)^2 \quad (6)$$

XGBoost: Sequentially builds the DT models which will minimize the errors, giving high accuracy, efficiency, robustness and scalable computation which will capture the complex patterns of the health data so that the lung cancer risk can be predicted accurately.

$$\hat{\sigma}_i = \sigma \left(\sum_{j=1}^n \sigma_j(\mathbf{x}_i) \right), \sigma_j \in \sigma \quad (7)$$

NM-XGBoost: It is a combination of Nelder-Mead optimisation and XGBoost that optimises hyperparameters automatically to gain improved accuracy and generalisation, and uncover fine-grained patterns for health features to predict lung cancer risk.

Voting Classifier: This is an ensemble of Gradient Boosting, XGBoost, Light GBM, Cat Boost to exploit the strength of every model to improve the overall accuracy, stability, robust, and consistent predictions to perform accurate lung cancer risk assessment.

$$\hat{\sigma} = \sigma(\mathbf{x}) \left(\sum_{i=1}^n \sigma_i(\hat{\sigma}_i = \sigma) \right) \quad (8)$$

IV. EXPERIMENTAL RESULTS

Accuracy: A test's accuracy is its ability to make proper discrimination between the patient and healthy patients. To assess how accurate a test is we should compute the proportion of all the cases analysed to be classified as true positive and the proportion to be classified as true negative. This can be expressed mathematically as:

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \quad (9)$$

Precision: Precision shows how accurate positive classifications made by models are by showing the percentage of expected positive cases that are actually positive.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (10)$$

Recall: Recall is the ability of a model to identify the relevant positive instances, and indicates the ability of the model to identify cases belonging to the goal class.

$$\text{Recall} = \frac{TP}{TP + FN} \quad (11)$$

F1-Score: F1-score is a single value that combines precision and recall for a fair assessment of the classification performance over the dataset used.

$$F1\text{-Score} = 2 * \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (12)$$

MCC: In ML, the MCC is a measure of the quality of binary classifications by a classifier. It considers all 4 elements of a confusion matrix and is a measure of the correlation of the expected and actual binary outcomes.

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (13)$$

Table. 1. Performance Evaluation– Lung Cancer

ML Model	Accuracy	Precision	Recall	F1-Score	MCC
Logistic Regression	0.941	0.941	0.941	0.941	0.911
GaussianNB	0.900	0.905	0.900	0.900	0.852
SVM	0.968	0.971	0.968	0.968	0.953
MLP	0.963	0.965	0.963	0.963	0.946
KNN	0.977	0.979	0.977	0.977	0.966
Decision Tree	0.922	0.930	0.922	0.922	0.887
Random Forest	0.936	0.938	0.936	0.936	0.905
XGBoost	0.950	0.950	0.950	0.950	0.925
Proposed NM-XGBoost	1.000	1.000	1.000	1.000	1.000
Extension Voting	1.000	1.000	1.000	1.000	1.000

Testing of different models are compared using the parameters accuracy, precision, memory, F1-score and MCC and the result is shown in Table 1. The Proposed NM-XGBoost and Extension Voting algorithms obtained the highest scores for all the measures, suggesting the best predictive performance.

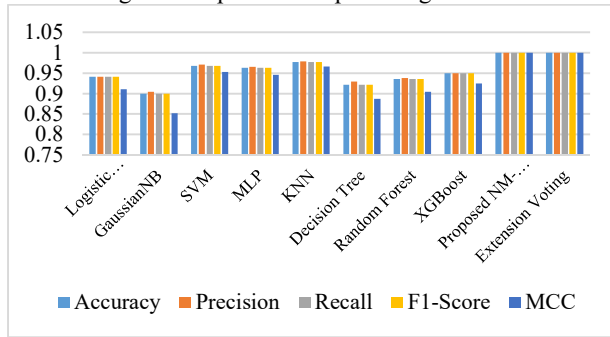
Table.2 Performance Evaluation – Lung Cancer

Model	Accuracy	Precision	Recall	F1-Score	MCC
Logistic Regression	0.972	0.974	0.972	0.972	0.945
GaussianNB	0.954	0.954	0.954	0.954	0.905
SVM	0.954	0.954	0.954	0.954	0.905
MLP	0.972	0.974	0.972	0.972	0.945
KNN	0.954	0.958	0.954	0.954	0.910
Decision Tree	0.981	0.981	0.981	0.981	0.962
Random Forest	0.981	0.981	0.981	0.981	0.962
XGBoost	0.981	0.981	0.981	0.981	0.962
Proposed NM-XGBoost	0.991	0.991	0.991	0.991	0.981
Extension Voting	0.991	0.991	0.991	0.991	0.981

When it comes to a number of different evaluation criteria, Table 2 demonstrates that overall NM-XGBoost

is better and has a stronger predictive stability as compared to other models.

Fig. 4. Comparison Graph– Lung Cancer Patient



As shown in Figure 4, accuracy, precision, recall, F1-score and MCC are the primary metrics for comparing the performance of different models. The overall best values for the metrics with a great prediction accuracy and reliability were obtained with the Extension Voting method.

Fig.5 Comparison Graph – Lung Cancer Risk

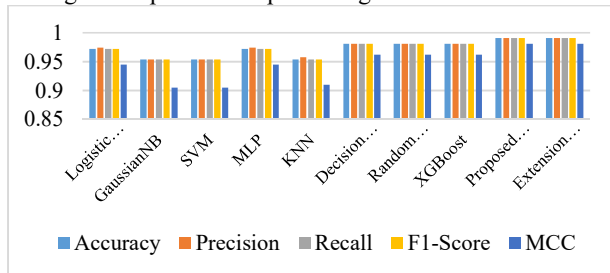


Figure 5 shows the comparison graph of the model success whereas in terms of accuracy, precision, recall, F1-score and MCC. The Extension Voting algorithm was the best way in classification accuracy, and had the highest overall metric values.

Figure 6 shows the 'Lung Cancer Risk Assessment' input form. It contains 12 dropdown menus for various symptoms: Alcohol Use, Chronic Lung Disease, Obesity, Passive Smoker, Coughing of Blood, Fatigue, Weight Loss, Shortness of Breath, Wheezing, Swallowing Difficulty, Clubbing of Finger Nails, Dry Cough, and Smoking. Each dropdown menu has a 'Select Level' option. A 'Predict Risk' button is located at the bottom right.

Fig.6 Enter the input data

When users enter all of the level of symptoms in all of the input boxes as shown in Fig.6 then the result of lung-cancer risk is generated.

Figure 7 shows the 'Prediction Results' screen. It displays a 'Risk Level' of 'MEDIUM' with a yellow warning icon and a 'Confidence Score' of '73.8%' with a green checkmark icon. Below the risk level, it states 'Some risk factors are present, regular monitoring is recommended.' Below the confidence score, it states 'Model confidence in this prediction based on advanced ML analysis.' The background is a light blue gradient.

Fig.7 Predicted Result

The result of the projection is displayed in Fig.7, with medium risk level and the advanced ML analysis computation score of “73.8% confidence score”.

Figure 8 shows the 'Lung Cancer Risk Assessment' input form, which is identical to Figure 6. It contains 12 dropdown menus for various symptoms: Alcohol Use, Chronic Lung Disease, Obesity, Passive Smoker, Coughing of Blood, Fatigue, Weight Loss, Shortness of Breath, Wheezing, Swallowing Difficulty, Clubbing of Finger Nails, Dry Cough, and Smoking. Each dropdown menu has a 'Select Level' option. A 'Predict Risk' button is located at the bottom right.

Fig.8 Enter the input data

Users fill in several fields with values of symptom intensities and can make a forecast of lung-cancer risks accurately by the evaluation model (Fig.8).

Figure 9 shows the 'Prediction Results' screen, which is identical to Figure 7. It displays a 'Risk Level' of 'HIGH' with a red warning icon and a 'Confidence Score' of '81.97%' with a green checkmark icon. Below the risk level, it states 'Multiple risk factors detected, immediate medical consultation advised.' Below the confidence score, it states 'Model confidence in this prediction based on advanced ML analysis.' The background is a light blue gradient.

Fig.9 Predicted Result

The Model as shown in Fig.9 predicts a high risk of lung-cancer with a confidence score of 81.97% showing severe symptoms and requires immediate evaluation.

V. CONCLUSION

In the present work the successful hybridisation between the feature engineering, optimisation and ensemble learning has been observed for the accurate prediction of lung cancer. RFE-SVM was used for feature selection to retain only the most relevant features, thus ensuring proper training of the model and the minimization of noise. Results showed that optimised XGBoost with Nelder-Mead had significantly higher predictive capability when compared with various ML algorithms evaluated. In addition, ensemble learning was used to improve the strength of the system. When Gradient Boosting, XGBoost, LightGBM, and CatBoost were all combined in the Voting Classifier, it got a perfect score on accuracy, precision, recall, and F1. The use of XAI techniques such as LIME and SHAP showed the importance of salient traits, thereby highlighting the need for transparency and interpretability for clinical use. The results indicate that ensemble learning and advanced optimisation techniques can significantly boost prediction accuracy and aid in medical decision-making. The approach offers high classification accuracy and good real-world healthcare deployment trustworthiness, allowing for the creation of very high accuracy lung cancer prediction models that are easily interpretable.

Future work will involve increasing the size and diversity of data that are representative of real world clinical variation. Some generalisation enhancement can be obtained by exploring other combination methods of the ensemble of advanced gradient boosting algorithms and DL architectures. A cross-domain validation on several healthcare datasets will be carried out in an effort to evaluate the flexibility and resilience of the system. Moreover, the approach can also be used for multi-class classification and thus might predict the various phases of

lung cancer instead of risk assessment. Future tests will consider clinical deployment, scalability, response time, reliability and adaptability to ensure that the models remain accurate as new data is incorporated.

REFERENCES

- [1] Gote, P. M., Kumar, P., Kumar, H., Verma, P., & Jiet, M. (2025). Integrating Machine Learning Algorithms: A Hybrid Model for Lung Cancer Outcome Improvement. *Applied Sciences*, 15(9), 4637.
- [2] Rashid, D., Tohin, M. I., Chowdhury, M. J. U., & Mony, M. J. I. (2024, December). Enhancing Lung Cancer Classification Performance on Diverse Datasets: Utilizing Advanced Feature Selection and Novel Ensemble Approaches. In *2024 27th International Conference on Computer and Information Technology (ICCIT)* (pp. 897-902). IEEE.
- [3] Pati, A., Panigrahi, A., Sahu, B., Patro, R., Pati, A. K., & Patro, R. R. (2025, May). EnLung: An Ensemble Learning-Based Approach for Enhancing Lung Cancer Prediction. In *2025 International Conference in Advances in Power, Signal, and Information Technology (APSIT)* (pp. 1-6). IEEE.
- [4] Jamil, D., Shah, S. M. A., & Al-Jarwani, F. M. M. (2024, September). A Comprehensive Survey of Techniques for Lung Cancer Diagnosis and Prediction. In *2024 Global Conference on Wireless and Optical Technologies (GCWOT)* (pp. 1-7). IEEE.
- [5] Kesiku, C. Y., & Garcia-Zapirain, B. (2024). AI-Enhanced Lung Cancer Prediction: A Hybrid Model's Precision Triumph. *IEEE Journal of Biomedical and Health Informatics*.
- [6] S. M. Varnosfaderani and M. Forouzanfar, "The role of AI in hospitals and clinics: Transforming healthcare in the 21st century," *Bioengineering*, vol. 11, no. 4, p. 337, Mar. 2024.
- [7] G. Chassagnon, C. D. Margerie-Mellon, M. Vakalopoulou, R. Marini, T.-N. Hoang-Thi, M.-P. Revel, and P. Soyer, "Artificial intelligence in lung cancer: Current applications and perspectives," *Jpn. J. Radiol.*, vol. 41, pp. 235–244, Nov. 2022.
- [8] A. P. Zhao, S. Li, Z. Cao, P. J. -H. Hu, J. Wang, Y. Xiang, D. Xie, and X. Lu, "AI for science: Predicting infectious diseases," *J. Saf. Sci. Resilience*, vol. 5, no. 2, pp. 130–146, Jun. 2024.
- [9] H. A. Al-Jamimi, "Synergistic feature engineering and ensemble learning for early chronic disease prediction," *IEEE Access*, vol. 12, pp. 62215–62233, 2024.
- [10] Ş. Ay, E. Ekinici, and Z. Garip, "A comparative analysis of meta-heuristic optimization algorithms for feature selection on ML-based classification of heart-related diseases," *J. Supercomput.*, vol. 79, no. 11, pp. 11797–11826, Jul. 2023.
- [11] S. Zhao, P. Wang, A. A. Heidari, H. Chen, W. He, and S. Xu, "Performance optimization of salp swarm algorithm for multi-threshold image segmentation: Comprehensive study of breast cancer microscopy," *Comput. Biol. Med.*, vol. 139, Dec. 2021, Art. no. 105015.
- [12] H.-Y. Chiu, H.-S. Chao, and Y.-M. Chen, "Application of artificial intelligence in lung cancer," *Cancers*, vol. 14, no. 6, p. 1370, Mar. 2022.
- [13] S. Huang, J. Yang, N. Shen, Q. Xu, and Q. Zhao, "Artificial intelligence in lung cancer diagnosis and prognosis: Current application and future perspective," in *Seminars in Cancer Biology*, vol. 89. Amsterdam, The Netherlands: Elsevier, 2023, pp. 30–37.
- [14] M. Liu, J. Wu, N. Wang, X. Zhang, Y. Bai, J. Guo, L. Zhang, S. Liu, and K. Tao, "The value of artificial intelligence in the diagnosis of lung cancer: A systematic review and meta-analysis," *PLoS ONE*, vol. 18, no. 3, Mar. 2023, Art. no. e0273445.
- [15] X.-W. Chen and J. C. Jeong, "Enhanced recursive feature elimination," in *Proc. 6th Int. Conf. Mach. Learn. Appl. (ICMLA)*, Dec. 2007, pp. 429–435.
- [16] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf, "Support vector machines," *IEEE Intell. Syst. Appl.*, vol. 13, no. 4, pp. 18–28, Apr. 1998.
- [17] W. S. Noble, "What is a support vector machine?" *Nature Biotechnol.*, vol. 24, no. 12, pp. 1565–1567, Dec. 2006.
- [18] X. Zhou, Y. Chen, Z. Wu, A. A. Heidari, H. Chen, E. Alabdulkreem, J. Escorcia-Gutierrez, and X. Wang, "Boosted local dimensional mutation and all-dimensional neighborhood slime mould algorithm for feature selection," *Neurocomputing*, vol. 551, Sep. 2023, Art. no. 126467.
- [19] F. Gao and L. Han, "Implementing the Nelder–Mead simplex algorithm with adaptive parameters," *Comput. Optim. Appl.*, vol. 51, no. 1, pp. 259–277, Jan. 2012.
- [20] C. De Margerie-Mellon and G. Chassagnon, "Artificial intelligence: A critical review of applications for lung nodule and lung cancer," *Diagnostic Interventional Imag.*, vol. 104, no. 1, pp. 11–17, Jan. 2023.
- [21] H. Park, J. Yun, S. M. Lee, H. J. Hwang, J. B. Seo, Y. J. Jung, J. Hwang, S. H. Lee, S. W. Lee, and N. Kim, "Deep learning—Based approach to predict pulmonary function at chest CT," *Radiology*, vol. 307, no. 2, Apr. 2023, Art. no. 221488.
- [22] T. I. A. Mohamed and A. E.-S. Ezugwu, "Enhancing lung cancer classification and prediction with deep learning and multi-omics data," *IEEE Access*, vol. 12, pp. 59880–59892, 2024.
- [23] S. K. Bhatt and S. Srinivasan, "Lung cancer detection using ai and different techniques of machine learning," *Int. J. Intell. Syst. Appl. Eng.*, vol. 12, no. 8, pp. 630–638, 2024.
- [24] P. Sathe, A. Mahajan, D. Patkar, and M. Verma, "End-to-end fully automated lung cancer screening system," *IEEE Access*, vol. 12, pp. 108515–108532, 2024.
- [25] A. Heidari, D. Javaheri, S. Toumaj, N. J. Navimipour, M. Rezaei, and M. Unal, "A new lung cancer detection method based on the chest CT images using federated learning and blockchain systems," *Artif. Intell. Med.*, vol. 141, Jul. 2023, Art. no. 102572.