

Enhancing Diagnostic Accuracy in Lung Cancer Detection Using Diverse Classification Models

Mohammed Najma Kulsoom.

Master's Student, Department of Computer Science and Engineering
Koneru Lakshmaiah Education Foundation, Andhra Pradesh, India – 522502

Mobile : 9154063330

Email: najma8201@gmail.com

Dr. RAJENDRA KUMAR GANIYA,

Professor,

Department of Computer Science and Engineering,
Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, A.P, 522302, India.

Mobile: +91 9393110375

E-mail: rajendragk@kluniversity.in

Abstract— Lung cancer is a leading cause of death and early identification of lung cancer is crucial for improved treatment and survival. The traditional diagnostic methods are mostly imaging and clinical assessment that can sometimes have drawbacks including late diagnosis, subjectivity, and a lack of predictive value. A comprehensive lung cancer risk dataset is used and pre-processed in this study, using normalisation methods (StandardScaler and MinMaxScaler) and feature selection (Recursive Feature Elimination with Support Vector Machines). The SMOTE was used to overcome class imbalance. RandomizedSearchCV with K-Fold cross validation was used to train and optimize various ML regression and classification models such as LR, KNN, Gaussian NB, Multinomial NB, DT, SVC, RF, XGBoost, GB and MLP. Implemented a Voting Classifier which is a combination of Random Forest, Decision Tree and Extra Trees. Improved the interpretability of the model using used explainable AI methods like LIME and SHAP. Deployed using Flask framework. The evaluation using accuracy, precision, recall, f1 score and confusion matrix indicated that VC achieved 98.1% on all the criteria compared to all the classifiers considered individually. Combining ensemble learning and explainable AI enhances diagnostic reliability, transparency, and interpretability, greatly boosting the accuracy of lung cancer detection compared to traditional methods.

Keywords— Lung Cancer, Magnetic Resonance imaging, Computed Tomography, Machine Learning, K-fold Cross validation.”

How to cite this article: Kulsoom MN, Ganiya RK., Enhancing Diagnostic Accuracy in Lung Cancer Detection Using Diverse Classification Models. Int J Drug Deliv Technol. 2026;16(86s): 1115-1120; DOI: 10.25258/ijddt.16.86s.111

I. INTRODUCTION

Lung cancer continues to be one of the most important global health problems due to high incidence and mortality rates with a heavy clinical, social and economic cost worldwide [1]. It is one of the most common causes of cancer-related death, mainly because it is detected at an advanced stage of the disease when treatment choices are restricted [2]. Medical imaging, clinical screening and diagnostic technology have improved and this has resulted in better detection capacities. However, the survival rate is still low, relative to other types of cancer [3]. Prompt detection of lung cancer is a key factor for better prognosis and facilitates timely treatment and overall better outcomes of patients [4].

Although significant improvements have been made using existing lung cancer detection methods, there are significant limitations. The use of traditional diagnostic methods may be very reliant on the skill of the physician and the advanced imaging interpretation, which results in a subjective assessment and delayed diagnosis [5]. Furthermore, a great variety of clinical conditions, patient characteristics and comorbidities increases the extent of complexity, and that decreases the generalisability of the existing diagnostic frameworks [6]. But most modern systems are lacking in robustness, interpretability, and consistency with heterogeneous datasets, with a significant gap between technological ability and clinical relevance [7].

In this paper, the purpose is to develop a reliable and systematic predictive paradigm for the diagnosis of lung cancer that is able to overcome the limitations of existing methods. The proposed work aims at improving the diagnostic accuracy, supporting the early stage detection and improving the reliability of decision making by using the full patient risk

information and advanced analytical methodologies [8]. The research adds value by focusing on systematic data processing, thorough validation, and clinically relevant outcomes, paving the way for reliable, non-complex lung cancer screening tools.

The significance of this research is that it could contribute to improving clinical decision support systems and health care efficiency by providing accurate, consistent and interpretable predictions. This proposed method can help in reducing the cost of health care, unnecessary procedures and improving the survival rate and reduce diagnostic uncertainty [10]. In summary, this work is intended to raise awareness of the value of data-driven intelligence for improved lung cancer detection, while helping to develop more effective, scalable and patient-centric health care solutions.

II. RELATED WORK

The recent advances in ML and DL have a great impact on medical diagnostics, especially in lung cancer detection. In order to enhance the classification performance, Sharma et al. [11] proposed CNN containing transfer learning features which are pre-trained on other CNNs. This approach can be applied to image-based diagnosis, but is resource-intensive and may not be scalable in low-resource clinical environments. Band et al. [12] utilized classic ML methods on clinical data that can be a cost-effective option to deep architectures. However, the challenges of interpretability of the models and high-dimensional data remained.

Sneha et al. [13] developed a DL model to classify histopathological images, with higher accuracy for classifying malignant from benign tissue. They were limited by the lack of labeled training data and possible overfitting when learning

complex geographical relationships. Khan et al. [14] systematically explored ML based diagnostic approaches and pointed out that performance differences arise due to the differences in preprocessing and handling of imbalance; the need for a common evaluation framework and common datasets were emphasised. Deep architectures have been applied to medical imaging tasks with high accuracy prediction, largely for image-level classification tasks, and particularly in the absence of demographic or clinical variables, to boost reliability.

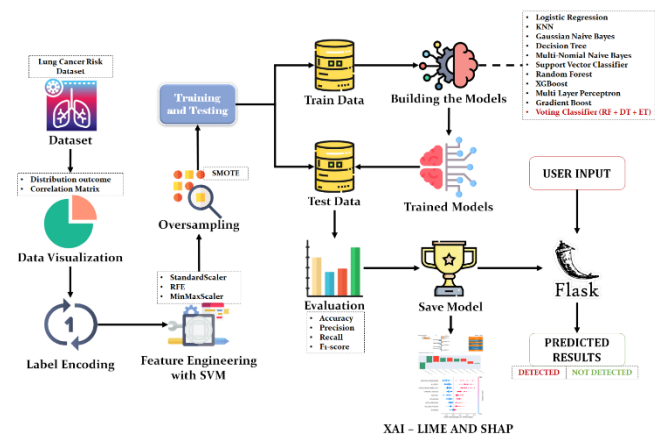
Indumathi et al. [16] performed an evaluation of various ML algorithms for lung cancer detection, which confirmed the potential of data-driven strategies for early diagnosis but they did not include proper feature optimisation or comparative analysis and validation over large datasets. Anbarasi [17] conducted a survey of the ML and DL techniques and discussed some of the problems including data heterogeneity, model overfitting, and cross validation and explained the importance of hybrid approaches to achieve a balance between model complexity and interpretability. Nakarmi et al. [18] developed a structured ML framework based on enhanced preprocessing and feature engineering leading to improved detection accuracy, but the transferability of the model across different populations and imaging modalities was not assessed.

Kumar et al. [19] surveyed the EDA and ML techniques and drew attention to the importance of visualizing data and correlation analysis to improve the performance of models, but failed to provide a complete ML pipeline. Bhuiyan et al. [20] combined numerous prediction algorithms to improve the early screening of lung cancer, hence, boosting the sensitivity and specificity. They lacked optimisation methods to reduce the computational burden and increase the interpretability, however.

All of these studies show the potential of ML and DL for lung cancer detection but also indicate the challenges currently faced such as data heterogeneity, lack of interpretability, high computing cost and poor generalisability. Such restrictions motivate the creation of effective and interpretable predictive frameworks which can potentially leverage clinical and imaging information to provide reliable, scalable and cost-effective diagnostics.

III. MATERIALS AND METHODS

The proposed approach aims to offer a smart prediction system for the risk assessment of lung cancer using Lung Cancer Risk Dataset. The pipeline consists of important data preprocessing, data transformation and model training, clinical evaluation and deployment. We explored different ML models to set a benchmark for comparison. These are RF, LR, KNN, SVM, DT, XGBoost, Gaussian NB, MLP and Multinomial NB. A Gradient Boosting model with Voting ensemble of RF, DT and ET classifiers was used for more resilience and accuracy. Feature selection and sampling methods were used to handle the class imbalance and maximise the prediction accuracy. Moreover, the methods of XAI like LIME and SHAP were applied to the interpretability of the model outputs. The user interface, designed using Flask, delivers clear and live risk forecasting that fosters clinical approachability, decision support and enhanced patient results.



“Fig. 1. System Architecture”

The steps of the procedure begin by visualizing the data, encoding the labels, then encoding features via RFE (with SVM) and normalization (StandardScaler and MinMaxScaler). To deal with the class imbalance, SMOTE is used prior to training various ML algorithms. The trained models are evaluated on standard metrics for performance, saved and deployed via a Flask interface. This enables user input and supports LIME and SHAP for explainability in prediction based on user input.

a) Dataset Collection

The Lung Cancer dataset [1] has been used for this study with 309 patient records: 14 numerical features (age, smoking, fatigue, alcohol intake) and 2 categorical features (gender, LUNG_CANCER label). It is well structured, there are no missing values. It provides all the important demographic, behavioural and clinical features to support rigorous analysis and prediction by modelling. It has been developed to cover a wide range of risk factors, symptoms and lifestyle habits, thus being suitable for early lung cancer risk assessment and classification, unlike generic datasets.

	GENDER	AGE	SMOKING	YELLOW_FINGERS	ANXIETY	PEER_PRESSURE	LUNGUNIL_DISEASE	FATIGUE	ALLERGY
0	M	69	1	2	2	1	1	2	1
1	M	74	2	1	1	1	2	2	2
2	F	59	1	1	1	2	1	2	1
3	M	63	2	2	2	1	1	1	1
4	F	63	1	2	1	1	1	1	1

“Fig. 2. UNSW-NB15 Dataset”

b) Pre-Processing

For the lung cancer dataset, preprocessing is the crucial step to make it useful for ML, because it enhances the quality of the data, standardizes the data features, and corrects the imbalance in the data, which is important for building a robust, accurate, and reliable model training.

1. *Data Visualization*: To gain an initial impression of the data and its underlying patterns, a visualisation of data was carried out. Analysing the outcome distribution was used to establish the percentage of lung cancer and non-cancer patients and potential class imbalance problems. A correlation matrix has also been used to look at the relationships between the input variables and the dependencies and influential variables have been obtained. This is crucial step in exploratory data analysis for informed decision making on data pre-processing, understanding of features and application of appropriate modelling and balancing methods.

2. *Label Encoding*: Categorical attributes were encoded as numbers using label encoding to be used in the ML algorithms. This encoding preserves the semantic significance of categorical variables, but also allows them to be provided as inputs to most algorithms that take a numeric input, thus ensuring consistency of features and avoiding the misinterpretation of features that are not numeric, so that all the patient attributes are useful for the model training and evaluation procedure.

3. *Normalization*: The purpose of normalisation was to scale the features to a common scale such that the mean of each feature was zero and its variance was unity. This will alleviate the impact of the large range number features and ensure that every feature has an equal impact. For algorithms which are sensitive to features size, standardisation is crucial as it enhances the overall prediction performance, gradients/steps convergence and numerical stability of the algorithm.

4. *Feature Engineering with SVM*: The feature engineering was done by iteratively discarding the least relevant features for lung cancer detection problem. The model learns more discriminative features over numerous iterations of the elimination of uninformative features. This not only shrinks the dimension, but also makes the model more interpretable, less prone to overfitting, less complex to compute and less likely to overlook the important features, leading to more effective and reliable learning.

5. *Oversampling*: To overcome the problem of class imbalance in the data set we used the oversampling technique and generated additional synthetic samples for the minority class. This technique balanced the representation of cancer and non-cancer cases and thus minimized the bias in the model towards the majority class. Treatment of imbalance enhances sensitivity and helps to achieve a fair classification and to help in the correct diagnosis of lung cancer cases.

6. *Feature Scaling*: Min-max scaling was performed on features to make them lie within a specific numerical range so as to maintain homogeneity in all the predictors. This step is useful for improving the numerical stability and for algorithms sensitive to the amplitude of features. The scaling increases the stability of the training, enables a fair comparison between the features contained in the system and enables better interoperability between multiple categorisation models used in the system.

c) Training and Testing:

The processed dataset was divided into a training set and a testing set, which gives an unbiased estimation of the model performance. This separation ensures that predictive models are tested on unseen data and gives an accurate measurement of the generalisation performance. Dataset splitting is important to avoid information leakage, confirm dependability and ensure thorough assessment of the lung cancer detection system.

d) Algorithms

RF is a technique that enhances the classification by combining a number of classifiers that are decision trees, trained with different subsets of the data. The ensemble learning and majority voting [21] can improve the robustness, alleviate overfitting and capture complex feature interaction, thus ensuring the stability and reliability of prediction.

$$"Gini = 1 - \sum_{i=1}^C (P_i)^2 \quad (1)"$$

LR is a valid method for probabilistic binary classification and its decision limits are easily interpretable [22]. It models the association of input variables with the probability of result, allows baseline comparison, helps assess generalisation and transparently analyses the risk factors which have the most influence.

$$"\hat{y}_i = \sigma(w^T x + b) = \frac{1}{1 + e^{-(w^T x + b)}} \quad (2)"$$

KNN is a method that allows classification based on the similarity with neighbouring points in the feature space [23]. It works instance by instance, making it highly flexible and adaptable to local data patterns, which leads to intuitive decision making and flexible working with non-linear data distributions.

$$"distance(x, X_i) = \sqrt{\sum_{j=1}^d (x_j - X_{ij})^2} \quad (3)"$$

SVM Makes optimum decision boundaries to separate classes as much as possible to improve accuracy. It can effectively handle high-dimensional feature spaces and complex interactions, has excellent generalisation ability and is able to withstand noise and outliers [24].

$$"\text{minimize } \frac{1}{2} ||W||^2 + C \sum_{i=1}^n \xi_i \quad (4)"$$

DT recursively partitions data based on feature values into rules that can be interpreted [25]. The hierarchical structure provides for unambiguous decision-making, efficient processing of several and diverse sources of input and for basic implementation of ensemble learning methodologies.

$$"I(i) = 1 - \sum_{i=1}^k p_i^2 \quad (5)"$$

An XGBoost model is built sequentially, with each model correcting the errors of the previous ones, to enhance the accuracy of predictions. The regularisation techniques and optimised learning process lead to good generalisation [26] and are capable of dealing with complex feature interactions, and are suitable for structured data.

Gaussian NB is a quick and easy classification method based on probability. It assumes a distribution to approximate the probability of a class. It enables rapid learning, is relatively easy to incorporate continuous qualities and offers good base for comparison evaluation [27].

MLP is built out of multiple layers of neurones for learning non-linearities. It has learning power to capture the complex feature relationships, improve the classification accuracy and facilitate generalisation in scenarios where linear models are not applicable [28].

The Multinomial NB is a probability-based classifier for categorical/coded variables. It is simple and inexpensive to process, which makes it possible to make fast inferences, compare with baseline, and make simple estimates of class membership probabilities [29].

$$\hat{y} = f(W^L f(W^{L-1} \dots f(W^1 X + b^1) + b^{(L-1)}) + b^L) \quad (6)$$

GB is a sequential approach that sequentially learns from the weak learners to reduce the error of classification [30] and enhance prediction ability. It sequentially optimises to improve accuracy, reduce bias and reveals complex relations in heterogeneous data sets.

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x) \quad (7)$$

The VC is an ensemble of multiple classifiers to improve the prediction stability and accuracy. The majority vote method decreases the bias and variation of individual models leading to robust, reliable and well generalised classification results.

$$\hat{y} = \underset{c}{\operatorname{argmax}} \left(\sum_{i=1}^n \mathbb{I}(\hat{y}_i = c) \right) \quad (8)$$

e) Integration of XAI and Flask Framework:

The system uses XAI methods such as LIME and SHAP to enhance the interpretability of lung cancer prediction. LIME provides local explanations by examining the effect of individual characteristics on a specific prediction, and presenting the results in an interactive visualisation, such as a waterfall chart. This makes it possible for clinicians to observe the impact of risk factors such as smoking, fatigue and coughing on the model's results. SHAP also provides global and local feature attributions, which gives a complete picture of feature relevance for numerous occurrences and allows for transparent and trustworthy decisions in clinical settings.

The XAI modules are used with a Flask based web framework for real-time predictions explanations for user provided inputs. The Flask interface serves as an easy-to-use input interface for patient data to be analysed and the relevant feature contributions to be visualised interactively. This synergy makes it possible to make the findings accessible, interpretable, and therapeutically useful, combining powerful ML algorithms with applications that are easily explained and implemented in healthcare.

IV. EXPERIMENTAL RESULTS

Accuracy: A test's capacity to be right in distinguishing between patient and health cases. The percent of true positive and true negative in all cases evaluated should be used to evaluate the accuracy of a test. This can be written mathematically as:

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

Precision: Precision is the number of correct positive predictions or samples divided by the number of positive predictions. Hence, the formula to compute precision is as follows:

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \quad (10)$$

Recall: Recall is a ML statistic which assesses how well a model can find all the relevant examples of a specific class. It is the ratio of correctly predicted positive observations to

the number of total positive observations and provides a sense of the completeness of a model in learning positive observations of a class.

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

F1-Score: F1 score is a ML evaluation statistic used to measure the accuracy of a model. It is a combination of a model's precision and recall scores. The accuracy statistic is the fraction of the total number of times a model has made a correct prediction.

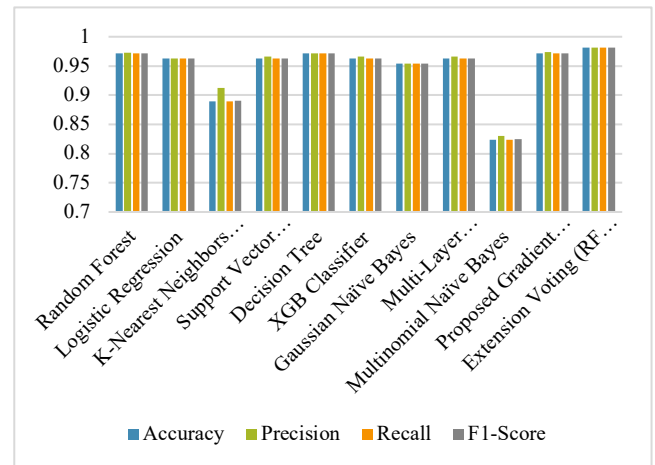
$$\text{F1 Score} = 2 * \frac{\text{Recall} * \text{Precision}}{\text{Recall} + \text{Precision}} * 100 \quad (12)$$

“Table. 1. Performance Evaluation”

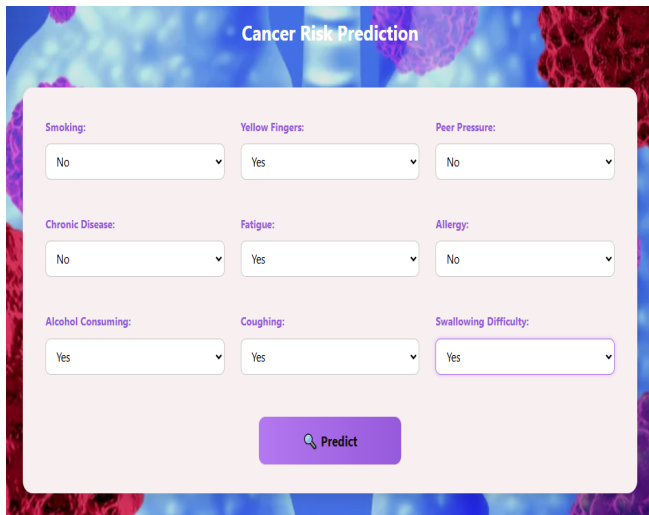
ML Model	Accuracy	Precision	Recall	F1-Score
Random Forest	0.972	0.973	0.972	0.972
Logistic Regression	0.963	0.963	0.963	0.963
K-Nearest Neighbors (KNN)	0.889	0.912	0.889	0.890
Support Vector Machine (SVM)	0.963	0.966	0.963	0.963
Decision Tree	0.972	0.972	0.972	0.972
XGB Classifier	0.963	0.966	0.963	0.963
Gaussian Naïve Bayes	0.954	0.954	0.954	0.954
Multi-Layer Perceptron (MLP)	0.963	0.966	0.963	0.963
Multinomial Naïve Bayes	0.824	0.830	0.824	0.825
Proposed Gradient Boosting	0.972	0.974	0.972	0.972
Extension Voting (RF + DT + ET)	0.981	0.981	0.981	0.981

The accuracy, precision, recall and f1-score of the different standard ML models are compared in Table.1. From the results obtained it can be concluded that the model Extension Voting (RF + DT + ET) is better performing as it has the highest value in all the metrics.

“Fig. 3. Comparison Graph”



The accuracy, precision, recall and F1-Score results of all the models are summarized in a graphical representation in Figure 3. The Extension Voting ensemble algorithm consistently gets the highest scores across all the algorithms evaluated and hence is the best.



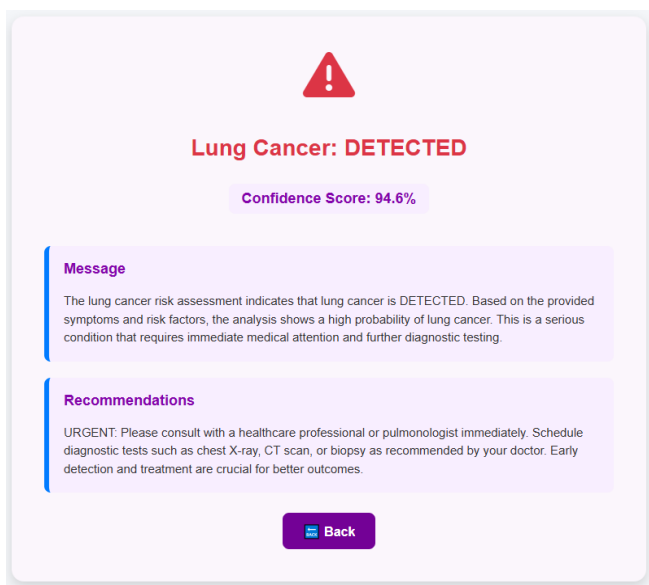
Cancer Risk Prediction

Smoking: No
 Yellow Fingers: Yes
 Peer Pressure: No
 Chronic Disease: No
 Fatigue: Yes
 Allergy: No
 Alcohol Consuming: Yes
 Coughing: Yes
 Swallowing Difficulty: Yes

Predict

“Fig.4 Upload the Data”

The input interface allows patients to enter their health and lifestyle data in dropdown menus for lung cancer risk prediction, which is 4.



Lung Cancer: DETECTED

Confidence Score: 94.6%

Message

The lung cancer risk assessment indicates that lung cancer is DETECTED. Based on the provided symptoms and risk factors, the analysis shows a high probability of lung cancer. This is a serious condition that requires immediate medical attention and further diagnostic testing.

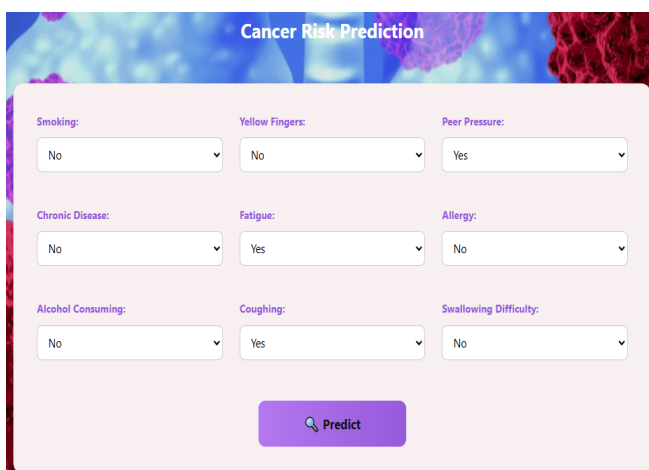
Recommendations

URGENT: Please consult with a healthcare professional or pulmonologist immediately. Schedule diagnostic tests such as chest X-ray, CT scan, or biopsy as recommended by your doctor. Early detection and treatment are crucial for better outcomes.

Back

“Fig.5 Predicted Result”

Fig. 5 displays the output interface in which the lung cancer is projected categorised with a confidence score of 94.6% for the users to check.



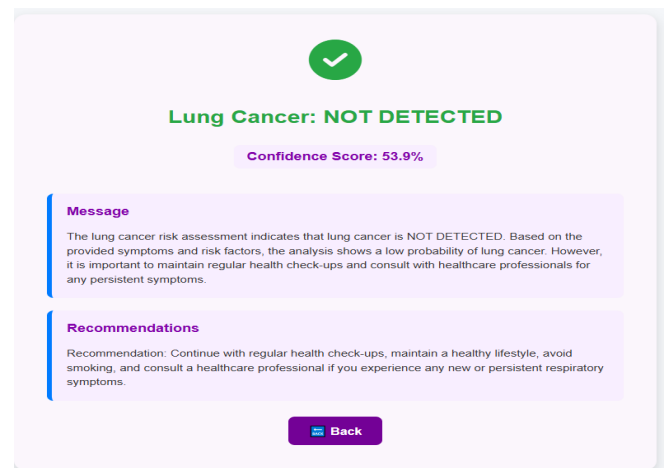
Cancer Risk Prediction

Smoking: No
 Yellow Fingers: No
 Peer Pressure: Yes
 Chronic Disease: No
 Fatigue: Yes
 Allergy: No
 Alcohol Consuming: No
 Coughing: Yes
 Swallowing Difficulty: No

Predict

“Fig.6 Upload the Data”

As shown in Fig. 6. Input interface to enable the user to input patient health facts and symptoms to be able to predict the risk of lung cancer accurately.



Lung Cancer: NOT DETECTED

Confidence Score: 53.9%

Message

The lung cancer risk assessment indicates that lung cancer is NOT DETECTED. Based on the provided symptoms and risk factors, the analysis shows a low probability of lung cancer. However, it is important to maintain regular health check-ups and consult with healthcare professionals for any persistent symptoms.

Recommendations

Recommendation: Continue with regular health check-ups, maintain a healthy lifestyle, avoid smoking, and consult a healthcare professional if you experience any new or persistent respiratory symptoms.

Back

“Fig.7 Predicted Results”

In Fig., Fig. 7 shows the output interface as NOT DETECTED with a confidence rate of 53.9% for user reference with the predicted lung cancer.

V. CONCLUSION

The main objective of this system was to enhance the accuracy and reliability of lung cancer detection for early diagnosis and clinical decision making. A structured ML process was used to analyse a Lung Cancer Risk Dataset comprised of clinical and demographic characteristics. The techniques applied were data normalisation using StandardScaler and MinMaxScaler, feature engineering using Support Vector Machine-based Recursive Feature Elimination and handling class imbalance using Synthetic Minority Oversampling Technique. RandomizedSearchCV with K-Fold cross-validations was used to train and optimize various classifiers like LR, KNN, NB (Gaussian and Multinomial), DT, SVC, RF, XGBoost, Gradient Boosting, MLP. The accuracy, precision, recall and F1-score were used to measure the performance of the system and a VotingClassifier with RF, DT and ET achieved an overall accuracy of 98.1%. We applied XAI techniques like LIME and SHAP, and implemented it with Flask for transparency and usability. In conclusion, the developed system is a reliable, interpretable and automated solution for the diagnosis of lung cancer with great potential to assist clinical care and decision-making in the real world.

Future study can be expanded to real-time detection and adaptive learning for the evolving DDoS patterns for achieving long-term robustness. Enhanced integration with cloud-based infrastructure and low power, low profile designs will expand scalability and resource efficiency. Collaborative monitoring agents, automated response systems and the associated reduced reaction time can increase situational awareness and thus enhance the defence against large-scale coordinated attacks in dynamic, heterogeneous network environments.

REFERENCES

- [1] Ramachandra, B., & Surekha, T. P. (2024). Development and evaluation of a network intrusion detection system for DDoS attack detection using machine learning. *Bulletin of Electrical Engineering and Informatics*, 13(6), 4207-4213.
- [2] Amrish, R., Bavapriyan, K., Gopinath, V., Jawahar, A., & Kumar, C. V. (2022). DDoS detection using machine learning techniques. *Journal of IoT in Social, Mobile, Analytics, and Cloud*, 4(1), 24-32.
- [3] Hossain, M. A., & Islam, M. S. (2024). Enhancing DDoS attack detection with hybrid feature selection and ensemble-based classifier: A promising solution for robust cybersecurity. *Measurement: Sensors*, 32, 101037.
- [4] Jasim, M. N., & Gaata, M. T. (2022). K-Means clustering-based semi-supervised for DDoS attacks classification. *Bulletin of Electrical Engineering and Informatics*, 11(6), 3570-3576.
- [5] Sawah, M. S., Elmannai, H., El-Bary, A. A., Lotfy, K., & Sheta, O. E. (2025). Distributed denial of service (DDoS) classification based on random forest model with backward elimination algorithm and grid search algorithm. *Scientific Reports*, 15(1), 19063.
- [6] R. Mall, K. Abhishek, M. S., A. Shankar, and A. Kumar, "Stack ing ensemble approach for DDoS attack detection in software defined cyber-physical systems," *Computers and Electrical Engineer ing*, vol. 107, p. 108635, Apr. 2023, doi: 10.1016/j.compeleceng.2023.108635.
- [7] M. Alkasassbeh, G. Al-Naymat, A. B. A. Hassanat, and M. Almseidin, "Detecting Distributed Denial of Service Attacks Using Data Mining Techniques," *International Journal of Advanced Computer Science and Applications*, vol. 7, no. 1, 2016, doi: 10.14569/IJACSA.2016.070159.
- [8] M. A. H. Azmi, C. F. M. Foozy, K. A. M. Sukri, N. A. Abdullah, I. R. A. Hamid, and H. Amnur, "Feature Selection Approach to Detect DDoS Attack Using Machine Learning Algorithms," *JOIV: International Jour nal on Informatics Visualization*, vol. 5, no. 4, p. 395, Dec. 2021, doi: 10.30630/joiv.5.4.734.
- [9] M. Arunadevi and V. Sathya, "DDoS Attack Detection using Back Propagation Neural Network Optimized by Bacterial Colony Optimization," *IJIES*, vol. 16, no. 5, pp. 301–312, Oct. 2023, doi: 10.22266/ijies2023.1031.26.
- [10] M. Khare and R. Oak, "Real-Time Distributed Denial-of-Service (DDoS) Attack Detection Using Decision Trees for Server Performance Maintenance," in *Performance Management of Integrated Systems and its Applications in Software Engineering*, M. Pant, T. K. Sharma, S. Basterrech, and C. Banerjee, Eds., Singapore: Springer Singapore, 2020, pp. 1–9. doi: 10.1007/978-981-13-8253-6 1.
- [11] U. S. Chanu, K. J. Singh, and Y. J. Chanu, "A dynamic feature selection technique to detect DDoS attack," *Journal of Information Security and Applications*, vol. 74, p. 103445, May 2023, doi: 10.1016/j.jisa.2023.103445.
- [12] M. Mittal, K. Kumar, and S. Behal, "Deep learning approaches for detecting DDoS attacks: a systematic review," *Soft Comput*, vol. 27, no. 18, pp. 13039–13075, Sep. 2023, doi: 10.1007/s00500-021-06608-1.
- [13] Y. B. Sanap and P. Aher, "A Comprehensive Survey On Detection And Mitigation Of DDoS Attacks Enabled With Deep Learning Techniques In Cloud Computing," in *2023 6th International Conference on Advances in Science and Technology (ICAST)*, 2023, pp. 149–154. doi: 10.1109/ICAST59062.2023.10454990.
- [14] J. Gera and B. P. Battula, "Detection of spoofed and non-spoofed DDoS attacks and discriminating them from flash crowds," *EURASIP Journal on Information Security*, vol. 2018, no. 1, p. 9, Dec. 2018, doi: 10.1186/s13635-018-0079-6.
- [15] F. O. Catak and A. F. Mustacoglu, "Distributed denial of service attack detection using autoencoder and deep neural networks," *IFS*, vol. 37, no. 3, pp. 3969–3979, Oct. 2019, doi: 10.3233/JIFS-190159.
- [16] A. Bhardwaj, V. Mangat, R. Vig, S. Halder, and M. Conti, "Distributed denial of service attacks in cloud: State-of-the-art of scientific and commercial solutions," *Computer Science Review*, vol. 39, p. 100332, 2021, doi: <https://doi.org/10.1016/j.cosrev.2020.100332>.
- [17] Ismail et al., "A Machine Learning-Based Classification and Prediction Technique for DDoS Attacks," *IEEE Access*, vol. 10, pp. 21443–21454, 2022, doi: 10.1109/ACCESS.2022.3152577.
- [18] M. A. Bouke, A. Abdullah, S. H. ALshatebi, M. T. Abdullah, and H. E. Atigh, "An intelligent DDoS attack detection tree-based model using Gini index feature selection method," *Microprocessors and Microsystems*, vol. 98, p. 104823, Apr. 2023, doi: 10.1016/j.micpro.2023.104823.
- [19] A. A. Alqarni, "Majority Vote-Based Ensemble Approach for Distributed Denial of Service Attack Detection in Cloud Computing," *JCSANDM*, Mar. 2022, doi: 10.13052/jcsmd2245-1439.1126.
- [20] Hossain, M. A. (2023). Enhanced ensemble-based distributed denial-of-service (DDoS) attack detection with novel feature selection: a robust cybersecurity approach. *Artificial Intelligence Evolution*, 165-186.
- [21] Najar, A. A., & Naik, S. M. (2022, August). Applying supervised machine learning techniques to detect DDoS attacks. In *2022 2nd Asian Conference on Innovation in Technology (ASIANCON)* (pp. 1-7). IEEE.
- [22] Mishra, A., Gupta, N., & Gupta, B. B. (2023). Defensive mechanism against DDoS attack based on feature selection and multi-classifier algorithms. *Telecommunication Systems*, 82(2), 229-244.
- [23] Hadi, H. J., Hayat, U., Musthaq, N., Hussain, F. B., & Cao, Y. (2022, November). Developing realistic distributed denial of service (DDoS) dataset for machine learning-based intrusion detection system. In *2022 9th International Conference on Internet of Things: Systems, Management and Security (IOTSMS)* (pp. 1-6). IEEE.
- [24] Prima, F., Dylan, L., & Gunawan, A. A. S. (2023, October). Comparison of Machine Learning Models for Classification of DDoS Attacks. In *2023 5th International Conference on Cybernetics and Intelligent System (ICORIS)* (pp. 1-6). IEEE.
- [25] Manthena, R., & Vangipuram, R. (2024). Integrating Machine Learning and T-tests to Optimize Distributed Denial of Service Attacks Detection. *International Journal of Intelligent Engineering & Systems*, 17(6).
- [26] Mamatha, B., Rashmi, D., Tiwari, K. S., Sikrant, P. A., Jovith, A. A., & Reddy, P. C. S. (2023, August). Lung Cancer Prediction from CT Images and using Deep Learning Techniques. In *2023 Second International Conference on Trends in Electrical, Electronics, and Computer Engineering (TEECCON)* (pp. 263-267). IEEE.
- [27] Bhuvaneshwari, V., Jacklin, J., & Malini, R. (2023, June). Lung Cancer Detection by Artificial Intelligence Using Deep Learning Techniques. In *2023 3rd International Conference on Pervasive Computing and Social Networking (ICPCSN)* (pp. 831-839). IEEE.
- [28] Sonal, D., Pattanaik, S., Nayak, K., Vyas, R. R., Sharma, T., & Kohli, A. (2025, March). An Approach for Detecting and classification of Lung Cancer Using Deep Learning and Machine Learning. In *2025 IEEE 14th International Conference on Communication Systems and Network Technologies (CSNT)* (pp. 53-58). IEEE.
- [29] Shah, S. N. A., & Parveen, R. (2023). An extensive review on lung cancer diagnosis using machine learning techniques on radiological data: state-of-the-art and perspectives. *Archives of computational methods in engineering*, 30(8), 4917-4930.
- [30] Agarwal, A. K., Maheshwari, H., Tiwari, R. G., & Jain, V. (2023, December). A Comprehensive Review of Machine Learning Techniques for Lung Cancer Detection. In *2023 12th International Conference on System Modeling & Advancement in Research Trends (SMART)* (pp. 623-627). IEEE.