IJERST International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

Peer Reviewed, Referred & Indexed Journal

Research Paper

Lung Cancer Detection: A Comparative Approach Using Regressor Methods

Jamdade Shivani
PG Scholar, Department of Computer Science and Engineering, Sree Dattha Institute of Engineering and Science, Sheriguda, Ibrahimpatnam, Hyderabad, Telangana, India – 501510.
*Email:**[jamdadepothanna@gmail.com]

K. Mounika
Assistant Professor, Department of Computer Science and Engineering, Sree Dattha Institute of Engineering and Science, Sheriguda, Ibrahimpatnam, Hyderabad, Telangana, India – 501510.
*Email:**[kmounikacse@sreedattha.ac.in]

***Abstract*— Early and accurate detection of lung cancer is critical due to its high mortality rate and the potential for improved patient outcomes through timely intervention. Traditional diagnostic methods, relying primarily on imaging and clinical assessment, often face challenges such as delayed detection, subjective interpretation, and limited predictive accuracy. This investigation employs a comprehensive lung cancer risk dataset, preprocessed through normalization using StandardScaler and MinMaxScaler, along with feature selection via Recursive Feature Elimination with Support Vector Machines. Class imbalance was addressed using Synthetic Minority Oversampling Technique (SMOTE). Multiple regression and classification algorithms, including Logistic Regression, K-Nearest Neighbors, Gaussian and Multinomial Naive Bayes, Decision Tree, Support Vector Classifier, Random Forest, XGBoost, Gradient Boosting, and Multi-Layer Perceptron, were trained and optimized through RandomizedSearchCV with K-Fold cross-validation. A Voting Classifier combining Random Forest, Decision Tree, and Extra Trees was implemented, and model interpretability was enhanced using explainable AI techniques such as LIME and SHAP, with deployment facilitated via the Flask framework. Evaluation using accuracy, precision, recall, F1-score, and confusion matrices demonstrated that the Voting Classifier outperformed all individual models, achieving 98.1% across all metrics. The integration of ensemble learning with explainable AI provides enhanced diagnostic reliability, transparency, and interpretability, significantly improving lung cancer detection accuracy over conventional approaches.**

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

I. Introduction

Lung cancer remains one of the most critical global health challenges due to its high incidence and mortality rates, imposing substantial clinical, social, and economic burdens worldwide [1]. It is consistently ranked among the leading causes of cancer-related deaths, largely because the disease is often diagnosed at an advanced stage when treatment options are limited [2]. Advances in medical imaging, clinical screening, and diagnostic technologies have contributed to improved detection capabilities; however, survival rates remain low compared to other cancer types [3]. Early identification of lung cancer plays a decisive role in improving prognosis, enabling timely intervention, and enhancing overall patient outcomes [4].

Despite ongoing progress, several limitations persist in existing lung cancer detection approaches. Conventional diagnostic practices often rely heavily on clinician expertise and complex imaging interpretation, which may lead to subjective assessments and delayed diagnoses [5]. Furthermore, variability in patient demographics, clinical conditions, and comorbidities introduces additional complexity, reducing the generalizability of current diagnostic frameworks [6]. Many existing systems lack sufficient robustness, interpretability, or consistency when applied across heterogeneous datasets, highlighting a critical gap between technological capability and practical clinical applicability [7].

The objective of this study is to develop a reliable and systematic predictive framework for lung cancer detection that addresses the limitations of existing approaches. By utilizing comprehensive patient risk information and advanced analytical strategies, the proposed work aims to improve diagnostic accuracy, support early-stage detection, and enhance decision-making reliability [8]. The study contributes by emphasizing structured data handling, robust validation, and clinically meaningful outcomes, thereby establishing a foundation for dependable lung cancer screening solutions without increasing diagnostic complexity [9].

The significance of this work lies in its potential to strengthen clinical decision support systems and improve healthcare efficiency through accurate, consistent, and interpretable predictions. By enabling earlier detection and reducing diagnostic uncertainty, the proposed approach can help lower healthcare costs, minimize unnecessary interventions, and improve patient survival rates [10]. Overall, this work seeks to advance the role of data-driven intelligence in lung cancer diagnosis and contribute toward more effective, scalable, and patient-centric healthcare solutions.

II. Related Work

Recent advancements in machine learning (ML) and deep learning (DL) have substantially impacted medical diagnostics, particularly in lung cancer detection. Sharma et al. [11] proposed a convolutional neural network (CNN) framework integrated with transfer learning, improving classification accuracy through feature reuse from pretrained models. While effective in image-based diagnosis, this approach demands significant computational resources, limiting scalability in low-resource clinical settings. Band et al. [12] employed traditional ML algorithms using clinical data, offering a cost-effective alternative to deep architectures; however, challenges remained in handling high-dimensional datasets and ensuring model interpretability.

Sneha et al. [13] introduced a deep learning model for histopathological image classification, demonstrating superior precision in differentiating malignant and benign tissue. Despite capturing complex spatial patterns, their model was constrained by limited labeled data and potential overfitting. Khan et al. [14] conducted a systematic review of ML-based diagnostic methods, emphasizing performance variability due to inconsistent preprocessing and imbalance handling, underscoring the need for unified evaluation frameworks and standardized datasets. Devarajan et al. [15] applied deep architectures to medical imaging, achieving high predictive accuracy but focusing primarily on image-level classification, with minimal integration of demographic or clinical features to enhance reliability.

Indumathi et al. [16] evaluated multiple ML algorithms for lung cancer detection, confirming the potential of data-driven techniques in early diagnosis but lacking comprehensive feature optimization and comparative analysis across large datasets. Anbarasi [17] reviewed ML and DL techniques, identifying challenges such as data heterogeneity, model overfitting, and insufficient cross-validation, and emphasized the need for hybrid approaches balancing model complexity and interpretability. Nakarmi et al. [18] proposed a structured ML framework with enhanced preprocessing and feature engineering, achieving improved detection accuracy, though model adaptability across diverse populations and imaging modalities remained unaddressed.

Kumar et al. [19] surveyed exploratory data analysis and ML methods, highlighting the importance of effective data visualization and correlation analysis for model performance, yet without demonstrating a complete predictive modeling pipeline. Bhuiyan et al. [20] integrated multiple predictive algorithms to enhance early lung cancer screening, improving sensitivity and specificity; however, their study lacked optimization strategies to reduce computational overhead and enhance interpretability.

Collectively, these studies illustrate the promise of ML and DL in lung cancer detection while revealing persistent challenges, including data heterogeneity, limited interpretability, high computational requirements, and inadequate generalizability. These gaps motivate the development of integrative, efficient, and interpretable predictive frameworks capable of leveraging both clinical and imaging data for reliable, scalable, and cost-effective diagnostic solutions.

III. Materials And Methods

The proposed system aims to develop an intelligent predictive framework for lung cancer risk assessment using the Lung Cancer Risk Dataset. The workflow encompasses essential data preprocessing, transformation, and model training phases, followed by evaluation and deployment for clinical usability. Multiple machine learning models—including Random Forest, Logistic Regression, K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Decision Tree, XGBoost, Gaussian Naive Bayes, Multi-Layer Perceptron (MLP), and Multinomial Naive Bayes—were employed to establish comparative performance benchmarks. To enhance robustness and accuracy, a Gradient Boosting model and a Voting ensemble combining Random Forest, Decision Tree, and Extra Trees classifiers were implemented. Feature selection and sampling techniques were integrated to address class imbalance and optimize predictive performance. Furthermore, Explainable AI (XAI) methods such as LIME and SHAP were incorporated to provide interpretability of model outputs. A Flask-based user interface enables real-time, transparent risk prediction, promoting clinical accessibility, decision support, and improved patient outcomes.

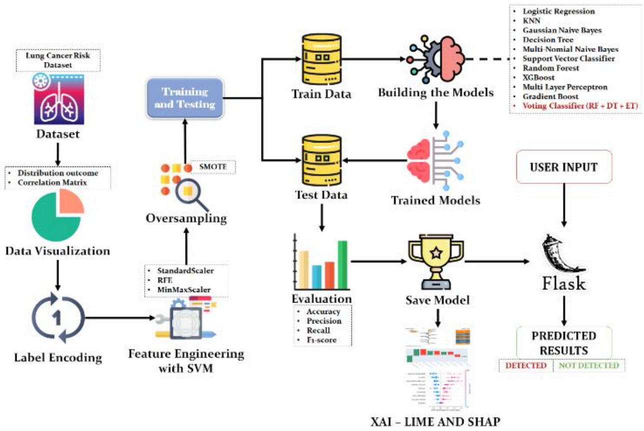


Fig. 1. System Architecture

The system architecture illustrates an end-to-end lung cancer detection pipeline using a risk-based dataset as shown in fig.1. The workflow begins with data visualization and label encoding, followed by feature engineering using SVM-based RFE and normalization with StandardScaler and MinMaxScaler. Class imbalance is handled through SMOTE before training multiple machine learning models. Trained models are evaluated using standard performance metrics, saved, and deployed through a Flask interface, enabling user input–based prediction with explainability supported by LIME and SHAP.

a) Dataset Collection

The Lung Cancer dataset used in this study [1] comprises 309 patient records with 16 features, including 14 numerical attributes (age, smoking, fatigue, alcohol consumption) and 2 categorical attributes (gender, LUNG_CANCER label). Well-structured with no missing values, it captures key demographic, behavioral, and clinical indicators, enabling robust analysis and predictive modeling. Its comprehensive representation of risk factors, symptoms, and lifestyle behaviors makes it ideal for early lung cancer risk assessment and classification compared to generic datasets.

	GENDER	AGE	SMOKING	YELLOW_FINGERS	ANXIETY	PEER_PRESSURE	CHRONIC 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

The preprocessing phase prepares the lung cancer dataset for effective machine learning by improving data quality, ensuring feature consistency, addressing imbalance, and enabling robust, accurate, and reliable model training.

1. Data Visualization: Data visualization was performed to gain an initial understanding of the dataset and its underlying patterns. Outcome distribution analysis helped identify the proportion of lung cancer and non-cancer cases, revealing potential class imbalance issues. Additionally, a correlation matrix was used to analyze relationships among input features, highlighting dependencies and influential variables. This step is essential for exploratory data analysis, as it supports informed preprocessing decisions, improves feature understanding, and guides the selection of appropriate modeling and balancing strategies.

2. Label Encoding: Label encoding was applied to convert categorical attributes into numerical form, enabling their effective utilization in machine learning models. Since most learning algorithms require numerical inputs, this transformation ensures compatibility while preserving the semantic meaning of categorical variables. Proper encoding maintains consistency across features, prevents misinterpretation of non-numeric data, and allows all patient attributes to contribute meaningfully during model training and evaluation.

3. Normalization: Normalization was carried out to standardize feature values by adjusting them to a common scale with zero mean and unit variance. This process minimizes the influence of features with larger numeric ranges and ensures equal contribution from all attributes. Standardization is particularly important for algorithms sensitive to feature magnitude, as it enhances numerical stability, accelerates convergence, and improves overall predictive performance.

4. Feature Engineering with SVM: Feature engineering was performed using a recursive elimination strategy to identify the most relevant predictors for lung cancer detection. By systematically removing less informative attributes, the model focuses on features with higher discriminative importance. This step reduces dimensionality, improves interpretability, decreases computational complexity, and mitigates overfitting, resulting in a more efficient and reliable learning process.

5. Oversampling: Oversampling was employed to address class imbalance within the dataset by generating additional synthetic samples for the minority class. This technique ensures balanced representation of both cancer and non-cancer cases, reducing model bias toward the majority class. Proper handling of imbalance enhances sensitivity, improves classification fairness, and supports reliable detection of lung cancer cases.

6. Feature Scaling: Min–max scaling was applied to rescale features into a fixed numerical range, ensuring uniformity across all predictors. This step improves numerical consistency and is beneficial for algorithms that are sensitive to feature magnitude. Scaling enhances training stability, supports fair comparison across features, and improves compatibility among multiple classification models used in the system.

c) Training and Testing:

The processed dataset was divided into training and testing subsets to enable unbiased evaluation of model performance. This separation ensures that predictive models are assessed using unseen data, providing a realistic measure of generalization capability. Dataset splitting is essential for preventing information leakage, validating reliability, and ensuring robust assessment of the lung cancer detection framework.

d) Algorithms

Random Forest enhances classification performance by aggregating multiple decision trees trained on diverse data subsets. Through ensemble learning and majority voting [21], it improves robustness, reduces overfitting, and captures complex feature interactions, making predictions more stable and reliable.

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

Logistic Regression contributes by providing probabilistic binary classification and interpretable decision boundaries [22]. It models the relationship between input variables and outcome likelihood, supporting baseline comparison, generalization assessment, and transparent analysis of influential risk factors.

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

K-Nearest Neighbors supports classification by assigning labels based on similarity to neighboring samples in the feature space [23]. Its instance-based learning enables adaptation to local data patterns, contributing intuitive decision-making and flexible handling of non-linear distributions.

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

Support Vector Machine improves accuracy by constructing optimal decision boundaries that maximize class separation. It effectively handles high-dimensional feature spaces and complex relationships, contributing strong generalization capability and robustness against noise and outliers [24].

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

Decision Tree provides interpretable, rule-based classification by recursively partitioning data according to feature values [25]. Its hierarchical structure enables transparent decision-making, efficient handling of mixed data types, and foundational support for ensemble learning strategies.

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

XGBoost enhances predictive accuracy through sequential ensemble learning, where each model corrects prior errors. Its regularization mechanisms and optimized learning process improve generalization [26], handle complex feature interactions, and deliver high performance on structured data.

Gaussian Naive Bayes contributes fast and efficient probabilistic classification by estimating class likelihoods under distributional assumptions. It supports rapid learning, effective handling of continuous attributes, and serves as a strong baseline for comparative evaluation [27].

Multilayer Perceptron models non-linear relationships through interconnected neural layers. Its learning capability captures complex feature dependencies [28], improves classification accuracy, and supports generalization in scenarios where linear models are insufficient.

Multinomial Naive Bayes provides efficient probabilistic classification for discrete or encoded features. Its simplicity and low computational cost enable fast inference, baseline comparison, and clear estimation of class membership probabilities [29].

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

Gradient Boosting strengthens predictive performance by iteratively combining weak learners to minimize classification [30] error. Its sequential optimization reduces bias, enhances accuracy, and captures intricate relationships across heterogeneous feature sets.

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

The Voting Classifier improves stability and accuracy by aggregating predictions from multiple diverse classifiers. Majority voting reduces individual model bias and variance, leading to robust, consistent, and well-generalized classification outcomes.

$$\hat{y} = argmax_c \left(\sum_{i=1}^{n} II\ (\hat{y}_i = c) \right) \quad (8)$$

e) Integration of XAI and Flask Framework:

The system incorporates Explainable Artificial Intelligence (XAI) techniques, including LIME and SHAP, to enhance the interpretability of lung cancer predictions. LIME provides local explanations by evaluating the contribution of individual features to a specific prediction, presenting results through interactive visualizations such as waterfall plots. This enables clinicians to understand the influence of risk factors like smoking, fatigue, and coughing on model decisions. SHAP complements this by computing global and local feature attributions, offering a comprehensive overview of feature importance across multiple instances, supporting transparent and trustworthy decision-making in clinical contexts.

The XAI modules are integrated within a Flask-based web framework, allowing real-time prediction explanations for user-provided inputs. The Flask interface serves as a user-friendly portal where input patient data can be analyzed, and corresponding feature contributions are visualized interactively. This combination ensures accessible, interpretable, and clinically relevant insights, bridging advanced machine learning models with practical, explainable, and deployable healthcare applications.

IV. Experimental Results

Accuracy: The accuracy of a test is its ability to differentiate the patient and healthy cases correctly. To estimate the accuracy of a test, we should calculate the proportion of true positive and true negative in all evaluated cases. Mathematically, this can be stated as:

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

Precision: Precision evaluates the fraction of correctly classified instances or samples among the ones classified as positives. Thus, the formula to calculate the precision is given by:

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

Recall: Recall is a metric in machine learning that measures the ability of a model to identify all relevant instances of a particular class. It is the ratio of correctly predicted positive observations to the total actual positives, providing insights into a model's completeness in capturing instances of a given class.

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

F1-Score: F1 score is a machine learning evaluation metric that measures a model's accuracy. It combines the precision and recall scores of a model. The accuracy metric computes how many times a model made a correct prediction across the entire dataset.

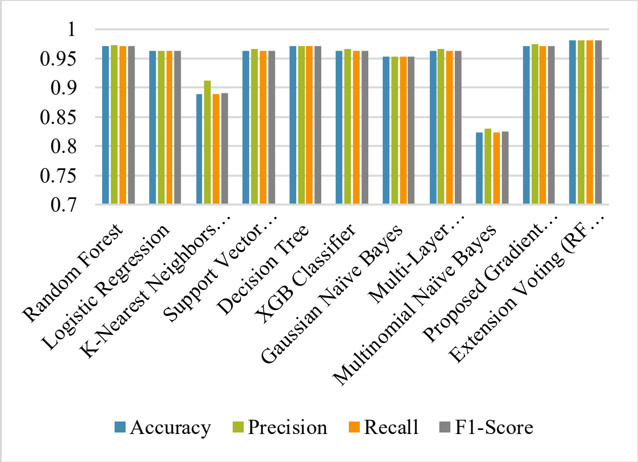
$$F1\ Score = 2 * \frac{Recall \times Precision}{Recall + Precision} * 100 \quad (12)$$

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

In Table.1 compares multiple machine learning models using accuracy, precision, recall, and F1-score, showing that the Extension Voting ensemble (RF + DT + ET) outperforms all others with highest values across all metrics.

Fig. 3. Comparison Graph



In Fig.3 visually represents accuracy, precision, recall, and F1-score across all models, highlighting the superior performance of the Extension Voting ensemble, which consistently achieves the highest values among all evaluated algorithms.

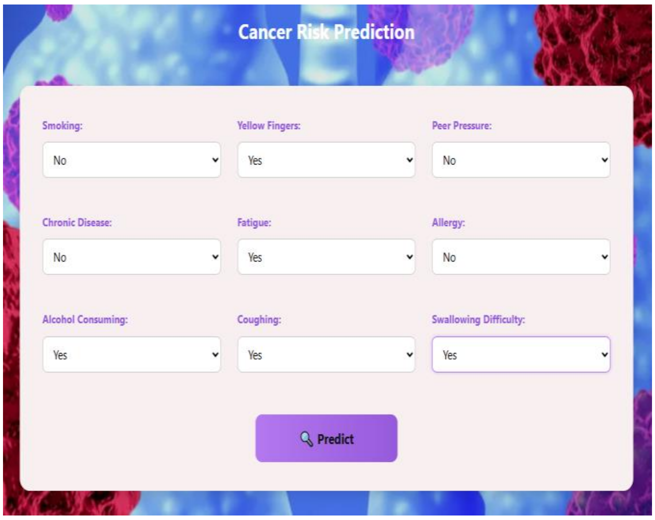


Fig.4 Upload the Data

In Fig. 4, the input interface allows users to provide patient health and lifestyle information through dropdown menus for lung cancer risk prediction.

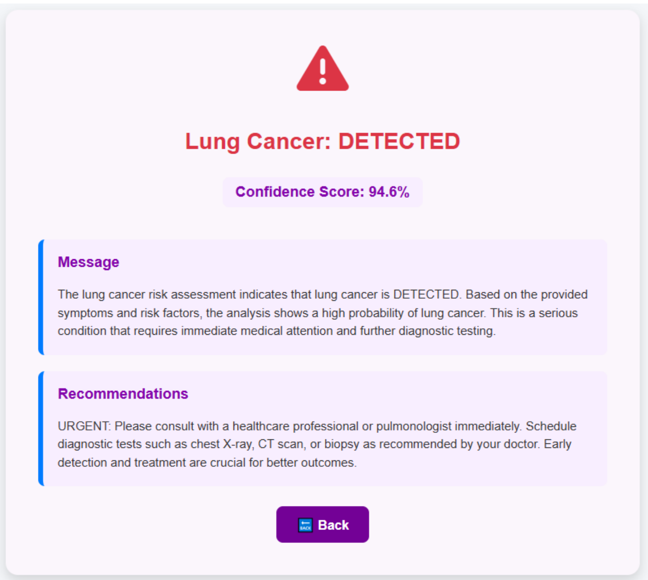


Fig.5 Predicted Result

In Fig. 5, the output interface displays predicted lung cancer classification with a confidence score of 94.6%, enabling user validation.

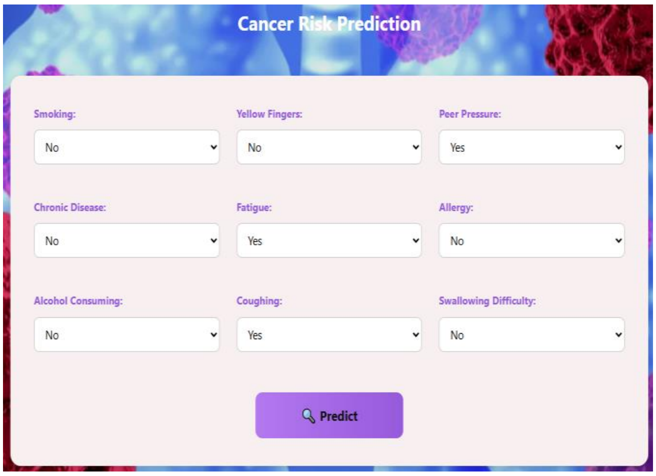


Fig.6 Upload the Data

In Fig. 6, the input interface allows users to enter patient health details and symptoms for accurate lung cancer risk prediction.

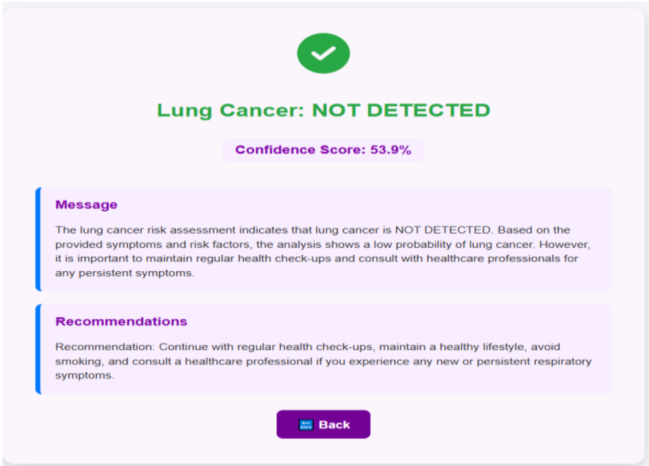


Fig.7 Predicted Results

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

V. Conclusion

In conclusion, the primary goal of this system was to improve the accuracy and reliability of lung cancer detection to support early diagnosis and informed clinical decision-making. The approach utilized a Lung Cancer Risk Dataset comprising clinical and demographic attributes, analyzed through a structured machine learning pipeline. Core techniques included data normalization using StandardScaler and MinMaxScaler, feature engineering through Support Vector Machine–based Recursive Feature Elimination, and class imbalance handling using the Synthetic Minority Oversampling Technique. Model training and optimization were performed using RandomizedSearchCV with K-Fold cross-validation across multiple classifiers, including Logistic Regression, K-Nearest Neighbors, Gaussian and

Multinomial Naive Bayes, Decision Tree, Support Vector Classifier, Random Forest, XGBoost, Gradient Boosting, and Multi-Layer Perceptron. The effectiveness of the system was validated through performance evaluation using accuracy, precision, recall, and F1-score, where a VotingClassifier combining Random Forest, Decision Tree, and Extra Trees achieved an overall accuracy of 98.1%. To enhance transparency and usability, explainable artificial intelligence techniques such as LIME and SHAP were incorporated, and deployment was facilitated through the Flask framework. Overall, the developed system provides a reliable, interpretable, and automated lung cancer detection solution with strong potential for real-world clinical support and decision-making.

Future work can extend the model to real-time detection with adaptive learning for evolving DDoS patterns, enhancing long-term robustness. Integration with cloud-based infrastructures and lightweight edge architectures will improve scalability and resource efficiency. Automated response systems and collaborative monitoring agents can reduce reaction time and expand situational awareness, strengthening defense against large-scale coordinated attacks across 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., Gopinaath, 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.2023.1 08635.

[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 Prop agation Neural Network Optimized by Bacterial Colony Optimization," IJIES, vol. 16, no. 5, pp. 301–312, Oct. 2023, doi: 10.22266/ijies2023.1 031.26.

[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.2023.10 3445.

[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/s00500-021-066 08-1.

[13] Y. B. Sanap and P. Aher, "A Comprehensive Survey On Detection And Mitigation Of DDoS Attacks Enabled With Deep Learning Tech niques 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 Microsys tems, vol. 98, p. 104823, Apr. 2023, doi: 10.1016/j.micpro.2023.1048 23.

[19] A. A. Alqarni, "Majority Vote-Based Ensemble Approach for Dis tributed Denial of Service Attack Detection in Cloud Computing," JCSANDM, Mar. 2022, doi: 10.13052/jcsm2245-1439.1126.

[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.