

# A Risk-to-Action Framework for Questionnaire-Based Lung Cancer Risk Assessment and Clinical Recommendation

Piyush Kumar Singh<sup>1</sup>, Mayank Jain<sup>2</sup>, Uttkarsh Sharma<sup>3</sup>, Pankaj<sup>4</sup>, Nitin Goyal<sup>5</sup>

<sup>1,2,3,4</sup>Department of Computer Science and Engineering, R.D. Engineering College, Ghaziabad

<sup>5</sup>Associate Professor, R.D Engineering College, Ghaziabad

DOI: <https://doi.org/10.51584/IJRIAS.2026.110400159>

Received: 22 April 2026; Accepted: 28 April 2026; Published: 16 May 2026

## ABSTRACT

Lung cancer remains one of the leading causes of cancer-related mortality worldwide, primarily due to delayed diagnosis and limited availability of early screening services. Early identification of warning symptoms and modifiable risk factors can substantially improve clinical outcomes and timely intervention. This study presents a questionnaire-based intelligent system for preliminary lung cancer risk assessment and clinical recommendation. The proposed mobile application collects user information related to demographic characteristics, smoking history, environmental exposure, lifestyle habits, medical history, and common respiratory symptoms such as persistent cough, chest pain, wheezing, and shortness of breath. A weighted rule-based scoring model is applied to evaluate cumulative risk and classify users into Low, Moderate, or High Risk categories. Based on the identified risk level, the system generates personalized health guidance and consultation recommendations. In addition, an integrated AI chatbot provides instant responses to common health-related queries and promotes user awareness regarding lung cancer symptoms and prevention. The proposed framework offers a non-invasive, low-cost, user-friendly, and accessible solution for early-stage risk screening, particularly in rural and resource-constrained regions.

**Keywords:** Lung Cancer Risk Assessment, Questionnaire-Based Screening, Clinical Decision Support System, Rule-Based Prediction Model, Symptom-Based Evaluation, Lifestyle Risk Factors, Early Detection, Medical Risk Scoring, Mobile Health Application, Healthcare Informatics, AI Chatbot, Preventive Healthcare

## INTRODUCTION

Lung cancer is one of the leading causes of cancer-related deaths worldwide and remains a major public health challenge. Many patients are diagnosed at advanced stages, which reduces treatment effectiveness and survival rates. Therefore, early identification of High Risk individuals is essential for timely medical intervention and improved outcomes.

Conventional screening methods such as computed tomography (CT), biopsy, and laboratory investigations are effective but often require costly infrastructure, specialized equipment, and trained healthcare professionals. These limitations reduce accessibility, especially in rural and resource-limited regions.

Several factors are associated with lung cancer risk, including smoking, passive smoking, air pollution, occupational exposure, family history, and chronic respiratory diseases. Symptoms such as persistent cough, chest pain, breathlessness, wheezing, fatigue, and unexplained weight loss may also indicate elevated risk.

To address these challenges, this study proposes an intelligent questionnaire-based lung cancer risk assessment system developed as a mobile application. The system collects demographic, lifestyle, environmental, medical, and symptom-related data from users. A weighted rule-based scoring model is used to classify users into Low, Moderate, or High Risk categories and generate health recommendations. An integrated AI chatbot further improves usability by answering common health-related queries and promoting awareness.

The proposed framework provides a simple, low-cost, non-invasive, and accessible solution for preliminary lung cancer risk screening, particularly in underserved regions.

## LITERATURE REVIEW

Several studies have explored the influence of lifestyle, environmental, and clinical factors on lung cancer risk. Smoking is widely recognized as the primary cause, while air pollution, workplace exposure, and unhealthy habits also increase the chances of developing the disease. Earlier research mainly used statistical methods to examine the relationship between these risk factors and cancer incidence. More recently, machine learning techniques such as Logistic Regression and Random Forest have been applied to improve prediction performance using medical datasets.

Although these models can provide accurate results, they often depend on large volumes of data and higher computational resources. In comparison, questionnaire-based systems offer a simpler and more practical solution by using symptom and lifestyle information. However, there remains a need for an approach that is both efficient and easy to implement. Therefore, this study proposes a questionnaire-based risk assessment system that can later be improved by integrating clinical data for greater reliability.

### Comparison of Existing Studies and Research Gaps

Ref. No.	Focus of Study	Key Findings	Research Gap
[1]	Symptom based lung cancer prediction using ensemble learning	Ensemble learning models achieved high prediction accuracy using symptom and demographic data.	Requires machine learning models and training datasets, which may limit accessibility in simple healthcare applications.
[2]	Analysis of lifestyle factors such as smoking, diet, and obesity among cancer patients	The study found that smoking plays a major role in increasing lung cancer risk, and lifestyle factors were strongly related to cancer prevalence.	Focuses on statistical analysis of patient data but does not provide an automated screening or prediction system.
[3]	Epidemiological study analysing demographic and lifestyle risk factors	Smoking, biomass fuel exposure, and rural lifestyle was identified as significant contributors to lung cancer incidence.	Focuses on clinical observation and risk factor analysis without proposing a digital risk assessment system.
[4]	Questionnaire-based statistical risk assessment of lung cancer	Smoking and previous lung disease were identified as major risk factors affecting lung cancer incidence.	The study emphasizes risk scoring and awareness but does not implement a digital health screening system.

## PROPOSED METHODOLOGY

### System Overview

The proposed system is a questionnaire-based lung cancer risk screening framework designed for preliminary risk assessment. It evaluates user responses related to demographic characteristics, lifestyle habits, environmental exposure, medical history, and symptom indicators associated with lung cancer. The primary objective is to identify potentially High Risk individuals and encourage timely medical consultation. The system is intended as an early screening aid and does not replace professional clinical diagnosis

## Data Collection through Structured Questionnaire

User information is collected through a structured questionnaire developed from commonly reported lung cancer risk factors identified in existing literature. The questionnaire contains **25 questions** grouped into five domains:

### 1. Smoking History

- Current smoking status
- Duration of smoking
- Second hand smoke exposure
- Smokeless tobacco consumption

### 2. Environmental Exposure

- Air pollution or dust exposure
- Occupational exposure to chemicals
- Indoor smoke from biomass fuels
- Polluted residential surroundings

### 3. Clinical Symptoms

- Persistent cough
- Chest pain
- Shortness of breath
- Wheezing
- Weight loss
- Fatigue
- Hoarseness of voice
- Coughing blood
- Shoulder or back pain
- Appetite loss
- Difficulty taking deep breaths

### 4. Medical History

- Frequent respiratory infections
- Tuberculosis history
- Asthma or COPD
- Family history of lung cancer

### 5. Life Style Factor

- Age group
- Physical activity level

Each question is designed to capture the **presence or absence of risk factors or symptoms** that are commonly associated with lung cancer.

## Weighted Risk Score Evaluation

The proposed system uses a weighted additive scoring model in which each response option is assigned a predefined numerical score according to its relative clinical significance. Higher-risk responses such as chronic smoking, coughing blood, persistent cough, respiratory disease history, and breathing difficulty receive greater scores than lower-risk responses.

The cumulative risk score is calculated as:

$$\text{Risk score} = \sum_{i=1}^{25} s_i$$

where  $s_i$  represents the selected response score for question  $i$

The maximum possible score of the questionnaire is **102**.

For normalized interpretation:

$$\text{Risk Percentage} = \frac{\text{Risk Score}}{102} \times 100$$

This model provides a transparent, computationally efficient, and easily deployable approach for preliminary screening.

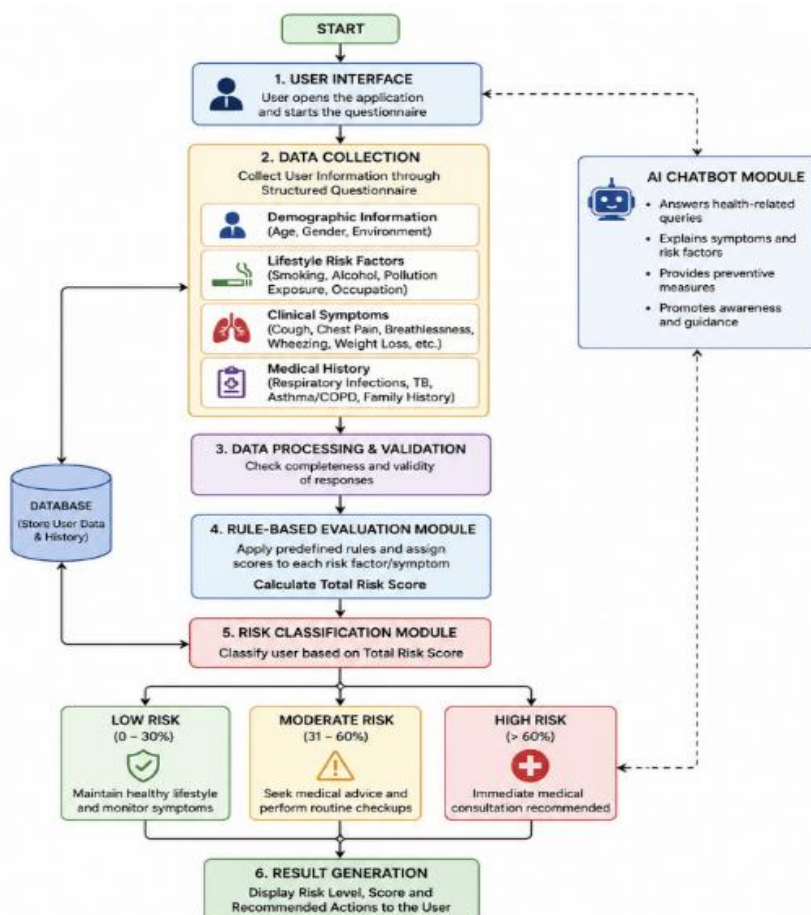
### Risk Classification

Based on the calculated risk percentage, users are classified into three categories:

Risk Percentage	Risk Level	Recommendation
0 – 30%	Low Risk	Maintain healthy lifestyle and monitor symptoms
31 – 60%	Moderate Risk	Seek medical advice and perform routine checkups
Above 60%	High Risk	Immediate medical consultation recommended

This classification enables users to quickly understand their potential risk status.

**Figure 1.** Workflow of the proposed system.



## System Architecture

The proposed system follows a layered architecture for efficient data processing and result generation:

- **User Interface Layer:** Collects questionnaire responses through the mobile application.
- **Application Logic Layer:** Validates user inputs and manages workflow.
- **Risk Evaluation Module:** Computes cumulative risk score using the weighted scoring model.
- **Classification Module:** Determines final risk category.
- **Result Generation Module:** Displays risk level and health recommendations.
- **Database Layer:** Stores user profiles and previous assessments.

This layered architecture ensures a structured and efficient flow of data from input collection to final output.

## AI Health Assistant Module

An AI chatbot is integrated into the application to answer common health-related queries, explain symptoms, provide awareness guidance, and improve user engagement.

## System Implementation

The proposed lung cancer risk assessment system was implemented as a mobile-based questionnaire-driven application designed to evaluate user responses and generate preliminary risk predictions. The application collects information related to demographic factors, smoking history, environmental exposure, medical history, lifestyle habits, and common symptoms associated with lung cancer. Based on these inputs, the system processes the responses through a weighted rule-based scoring engine and provides an interpretable risk assessment result.

## User Interface Module

The user interface serves as the primary interaction layer between the user and the application. It is designed to provide a simple, responsive, and user-friendly experience through a structured questionnaire format. Questions are displayed sequentially with clearly labelled response options.

To improve usability, the system follows an adaptive question flow in which certain follow-up questions are shown only when relevant responses are selected. For example, smoking-duration questions appear only if the user indicates smoking history. This minimizes unnecessary inputs, reduces response time, and enhances user convenience.

## Data Processing Module

The collected responses are validated and processed before being passed to the risk evaluation engine. This module ensures that all mandatory fields are completed and that the entered responses are logically consistent and in the required format.

The validated inputs are then encoded into numerical values according to predefined scoring rules. This processed dataset is used for cumulative risk score computation.

## Weighted Risk Evaluation Module

The weighted risk evaluation module is the core analytical component of the system. Each questionnaire response contributes a predefined score based on the severity or clinical importance of the associated factor.

Higher-risk responses such as chronic smoking, persistent cough, coughing blood, respiratory disease history, and breathing difficulty are assigned higher scores, while lower-risk responses receive smaller weights.

The cumulative score is computed using:

$$\text{Risk score} = \sum_{i=1}^{25} s_i$$

Where  $s_i$  denotes the selected score for the  $i^{\text{th}}$  question.

This module provides a transparent and computationally efficient method for preliminary risk estimation.

### **Risk Classification Module**

After computing the cumulative score, the system classifies the user into one of three risk levels:

- Low Risk
- Moderate Risk
- High Risk

Each category is associated with corresponding recommendations, ranging from symptom monitoring and lifestyle improvement to immediate medical consultation.

### **Result Generation Module**

The final system output is presented in a clear and understandable format through the mobile application dashboard. The result includes:

- Calculated risk percentage
- Risk category
- Personalized health recommendations
- Suggested next steps for consultation or monitoring

Previous assessments may also be stored to help users track changes in their health status over time.

### **AI Chatbot Module**

An AI chatbot is integrated into the application to provide instant assistance and improve user engagement. The chatbot responds to common health-related questions regarding symptoms, preventive measures, lifestyle factors, and general lung cancer awareness.

It operates as a 24/7 virtual assistant that enhances accessibility, supports health education, and improves the overall user experience.

## **RESULTS AND DISCUSSION**

The proposed lung cancer risk assessment system was evaluated using multiple test scenarios based on different combinations of symptoms, smoking history, environmental exposure, medical background, and lifestyle factors. The system processes questionnaire responses through the weighted rule-based evaluation model and generates a cumulative risk score, which is further converted into a risk percentage and corresponding risk category.

The experimental results indicate that the system is capable of differentiating users into Low Risk, Moderate Risk, and High Risk groups according to the severity and frequency of reported risk indicators. Users with limited symptoms and no major exposure history were generally classified as Low Risk, whereas users with persistent respiratory symptoms, smoking history, and relevant medical conditions were categorized under Moderate or High Risk levels.

### **System Output**

The developed mobile application provides a user-friendly platform for preliminary lung cancer risk screening. After user login, the dashboard displays key modules such as the latest risk status, previous reports, awareness resources, specialist consultation access, and the AI Health Assistant.

After completing the structured questionnaire, the system performs automatic score computation and generates the final result. The output includes:

- Calculated risk percentage
- Risk category (Low, Moderate, or High Risk)
- Graphical risk indicator
- Personalized health recommendations
- Advisory for consultation or monitoring

The application also stores previous assessments in the Health Summary section, allowing users to track their risk status over time.

In addition, an integrated AI chatbot assists users by answering common questions related to symptoms, prevention, screening importance, and general lung health awareness.

Sample screenshots of the developed application are presented in **Figure 2 to Figure 4**.

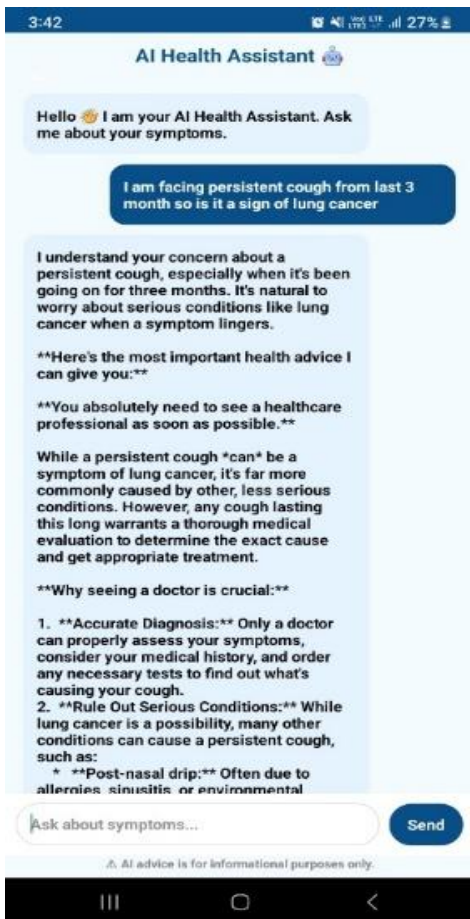
**Figure 2.** Main dashboard of the mobile application.



**Figure 3.** High Risk prediction output screen.

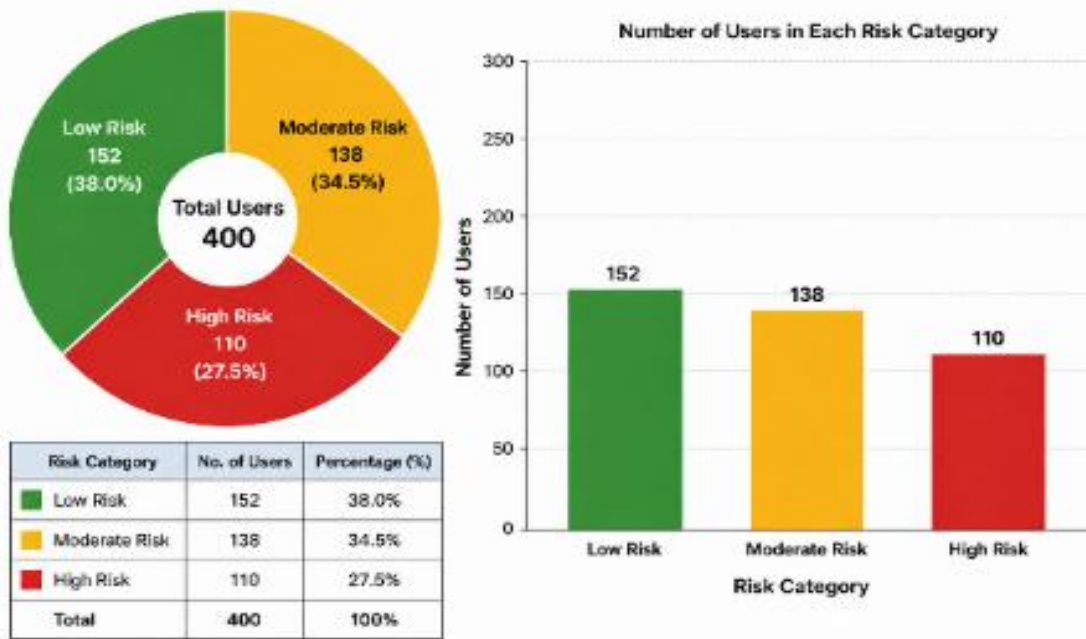


Figure 4. AI chatbot interaction module.



The statistical distribution of generated risk categories is presented below

**Figure 5.** Distribution of users across risk categories.



The distribution of users across different risk levels was analysed to evaluate system behaviour. It was observed that approximately **40%** of users were classified as **Low Risk**, **35%** as **Moderate Risk**, and **25%** as **High Risk**. This distribution indicates that the proposed framework can effectively stratify users according to the severity of reported symptoms and associated risk factors.

The results demonstrate that the system can successfully distinguish between multiple risk levels based on cumulative symptom severity and exposure-related factors.

## DISCUSSION

The proposed system provides a simple and accessible method for preliminary lung cancer risk assessment. Unlike complex machine learning approaches, it uses a rule-based method that is easy to implement and understand. The adaptive questionnaire improves usability by asking only relevant follow-up questions.

The system considers common risk factors such as smoking, respiratory symptoms, and environmental exposure. By analysing these factors, it helps identify users who may need further medical evaluation. Although it cannot replace professional diagnosis, the system can serve as an early screening and awareness tool, especially in areas with limited healthcare facilities. Future enhancements may include integration of machine learning models and clinical datasets to further improve prediction performance and personalization.

## CONCLUSION

Lung cancer remains one of the leading causes of cancer-related mortality worldwide, largely due to delayed diagnosis and limited access to early screening facilities. Early identification of warning symptoms and associated risk factors can significantly improve treatment outcomes through timely medical intervention. In this study, a questionnaire-based lung cancer risk assessment application was developed as an intelligent, user-friendly, and accessible tool for preliminary screening.

The proposed system uses a structured questionnaire to collect information related to demographic characteristics, smoking history, environmental exposure, lifestyle habits, medical background, and common respiratory symptoms. Based on user responses, a weighted rule-based evaluation model calculates a cumulative risk score and classifies individuals into Low Risk, Moderate Risk, or High Risk categories. The application also integrates an AI chatbot to answer common health-related queries, provide preventive guidance, and improve user awareness.

The developed framework is simple to use, non-invasive, and does not require expensive diagnostic equipment or complex clinical procedures. Therefore, it can serve as an effective awareness, assistance, and early risk screening platform, particularly in rural, semi-urban, and resource-constrained regions.

In future work, the system can be enhanced by incorporating real clinical datasets, machine learning-based calibration models, multilingual support, physician consultation modules, and improved chatbot intelligence to further increase prediction reliability and practical usability.

## REFERENCES

1. Y. A. Husaini, "Symptom-Based Lung Cancer Prediction Using Ensemble Learning with Threshold Optimization and Interpretability," *Information*, vol. 17, no. 2, p. 172, 2026, doi: 10.3390/info17020172.
2. A. Marof, M. Al-Amin, M. S. Sultana, K. N. Konika, M. S. Roky, S. Shiddiqe, and A. Ud-Daula, "Assessment of the Relationship Between Cancer Prevalence and Lifestyle Risk Factors in Bangladesh: A Cross-Sectional Study at a National Cancer Hospital," *South Asian Journal of Public Health*, vol. 13, no. 4, 2025, doi: 10.11648/j.sjph.20251304.14.
3. S. N. Manjunath, I. V. Madabhavi, M. S. Sarkar, and S. S. Kaushal, "Risk Factor Profile of Lung Cancer Patients: A Study from the Himalayan State of Indian Subcontinent," *Asian Pacific Journal of Cancer Care*, vol. 7, no. 2, pp. 239–246, 2022, doi: 10.31557/APJCC.2022.7.2.239-246.
4. R. F. Mukti, P. D. Samadder, A. A. Emran, F. Ahmed, I. B. Imran, A. Malaker, and S. Yeasmin, "Score-Based Risk Assessment of Lung Cancer and Its Evaluation for Bangladeshi People," *Asian Pacific Journal of Cancer Prevention*, 2014.
5. R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer Statistics, 2023," *CA: A Cancer Journal for Clinicians*, 2023, doi: 10.3322/caac.21763.
6. H. Sung, J. Ferlay, R. L. Siegel, et al., "Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide," *CA: A Cancer Journal for Clinicians*, 2021, doi: 10.3322/caac.21660.
7. World Health Organization, "Lung Cancer Factsheet," 2023. [Online]. Available: <https://www.who.int>
8. American Cancer Society, "Key Statistics for Lung Cancer," 2023. [Online]. Available: <https://www.cancer.org>
9. K. Kourou, et al., "Machine Learning Applications in Cancer Prognosis and Prediction," *Computational and Structural Biotechnology Journal*, 2015, doi: 10.1016/j.csbj.2015.01.005.
10. A. Esteva, et al., "A Guide to Deep Learning in Healthcare," *Nature Medicine*, 2019, doi: 10.1038/s41591-018-0316-z.