CLIMASENSE AI- CLIMATE BASED DISEASE RISK SENSING AND ALERT SYSTEM

A MINOR PROJECT REPORT SUBMITTED (12)
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE AWARD OF DEGREE OF

BACHELOR OF TECHNOLOGY
In
Computer Science and Engineering
(Artificial Intelligence and Machine Learning)

SUBMITTED BY
Ayushman Vohra
2022A6R059

UNDER THE SUPERVISION OF

Dr. Surbhi Gupta

Assistant Professor

Computer Science and Engineering

SUBMITTED TO

Department of Computer Science and Engineering (AI/ML)

Model Institute of Engineering and Technology (Autonomous)

Jammu, India

2026

CANDIDATE'S DECLARATION

I Ayushman Vohra**, 2022A6R059,** hereby declare that the work which is being presented in the minor/major project entitled, "**ClimaSense AI- Climate based Disease Risk Sensing and Alert System**" in partial fulfillment of requirement for the award of degree of B.Tech. Computer Science and Engineering (AI/ML) and submitted in the Department of CSE (AI/ML), Model Institute of Engineering and Technology (Autonomous), is an authentic record of my/our own work carried by me/us under the supervision of **Dr. Surbhi Gupta**(Assistant Professor , MIET). The matter presented in this project report has not been submitted in this or any other University / Institute for the award of B.Tech. Degree.

Signature of the Student *Dated*:

Ayushman Vohra

2022A6R059

Department of Computer Science and engineering
Model Institute of Engineering and Technology (Autonomous)
Kot Bhalwal, Jammu, India
(NAAC "A" Grade Accredited)

Ref. No.: **Date:**

CERTIFICATE

Certified that this minor project report entitled **"ClimaSense AI- Climate based Disease Risk Sensing and Alert System"** is the bonafide work of "**……..Name of the Candidate(S), Roll No.(s), of 7th Semester, Computer Science and Engineering (AI/ML), Model Institute of Engineering and Technology (Autonomous), Jammu",** who carried out the major project work under my/our supervision during February, 2020-May,2020*(change accordingly)*.

(Name of Co-Supervisor)-if any
Co-Supervisor
Academic Designation
Department Name, MIET

(Name of Supervisor)
Supervisor
Academic Designation
Department Name, MIET

This is to certify that the above statement is correct to the best of our knowledge.

(Name)
Designation
HoD/Head PRC Committee
Department Name, MIET

ACKNOWLEDGEMENTS

Minor Project work is an important aspect in the field of engineering, where contributions are made by many individuals and organizations. The present form of this work has been shaped through the guidance, encouragement, and support received from various quarters.

I would like to express my sincere gratitude to my **Dr. Surbhi Gupta**, whose constant guidance, valuable suggestions, and technical expertise have been instrumental in the successful completion of this project. Their continuous motivation and constructive feedback helped me in understanding the subject deeply and executing the work efficiently.

I extend my heartfelt thanks to the **Head of the Department**, for providing a supportive academic environment and for encouraging innovative learning and research activities within the department.

I would also like to express my sincere appreciation to the **Director**, **Model Institute of Engineering and Technology (Autonomous), Jammu**, for providing the necessary facilities, infrastructure, and an excellent academic platform to carry out this Minor Project during my B.Tech program.

Finally, I express my sincere gratitude to **Model Institute of Engineering and Technology (Autonomous), Jammu**, for giving me the opportunity to undertake this Minor Project as part of my undergraduate curriculum.

(*Name of the student*)

Roll Number

ABSTRACT

Environmental and climatic factors play a critical role in influencing public health, particularly respiratory, heat-related, and seasonal infectious diseases. Rapid urbanization, increasing air pollution, and climate variability have intensified the incidence of diseases such as influenza, heat stress, monsoon-related infections, and pollution-induced respiratory conditions. This project proposes an AI-enabled Environmental Health Risk Prediction and Alert System that integrates real-time weather and air quality data with epidemiological rule-based analysis and machine learning models. Key environmental parameters including PM2.5, AQI, temperature, humidity, and rainfall are utilized to assess disease risk levels. A machine learning regression model predicts hospital respiratory case trends, which are fused with epidemiological risk rules to classify health risk into low, medium, or high categories. The system provides a web-based interface supporting city selection and current location detection, along with automated email alerts for medium and high-risk conditions. The proposed system aims to enhance public health awareness, support preventive decision-making, and demonstrate the application of AI in climate-driven health risk assessment.

Contents

Candidates' Declaration i
Certificate ii
Acknowledgement iii
Abstract iv
Contents v
List of Tables vii
List of Figures viii
Abbreviations Used ix

Chapter 1 Introduction 1-14

1.1 Bckground of the Project 1

1.2 Problem Statement 2

1.3 Objective of the Project 3

1.4 Proposed Solution

1.5 Methodology Overview

1.6 Scope of the Project

Chapter 2 LITERATURE SURVEY 14-24

2.1 Introduction 14

2.2 Literature Survey 15

2.3 Gap Analysis 15

2.4 Problem Outline

Chapter 3 SYSTEM ARCHITECTURE 25-34

3.1 Introduction 25

3.2 Overall System Architecture 26

3.3 Data Collection Module 27

3.3.1 Air Quality Data Collection

3.3.1 Weather Data Collection

3.4 Data Preprocessing Module

3.5 Machine Learning Model

3.5.1 Health Risk Level Prediction

3.5.2 Disease Probability Prediction

3.5.3 Advantages of Using XGBoost

3.6 Prediction Output Module

3.7 Web Application System

3.8 Alert and Notification System

Chapter 4 IMPLEMENTATION & RESULT 35-44

4.1 Introduction 35

4.2 Implementation System 35

4.2.1 User Interface Implementation 36

4.2.2 Environmental Data Integration

4.2.3 Machine Learning Model Execution

4.2.4 Alert System Implementation

4.3 Result Discussion

4.3.1 Environmental Data Analysis Results

4.3.2 Health Risk Prediction Results

4.3.3 Disease Risk Identification

Chapter 5 CONCLUSION AND FUTURE SCOPE **45-55**

5.1 Conclusion 45

5.2 Future Scope 45

Table No.	Caption	Page No.
2.1	Summary of related research work	3
3.2	Technology Stack Used in ClimaSense AI	4

LIST OF FIGURES

Figure No.	Caption	Page No.
2.1	Key gaps in existing environmental health monitoring systems	3
3.1	Overall system architecture of the ClimaSense AI platform	4
4.1	Home page of the ClimaSense AI web application	5
4.2	Display of real-time environmental conditions and health risk assessment	12
4.3	Email alert subscription and confirmation message	13

ABBREVIATIONS USED

ANN	Artificial Neural Network
DRC	Departmental Research Committee
NFS	Neuro Fuzzy System
OC	Optical Communication

Chapter 1
INTRODUCTION

1.1 Background of the Project

Environmental conditions such as air pollution and weather variations have a significant impact on human health. Rapid urbanization, industrial growth, and increased vehicular emissions have led to a continuous rise in air pollutants like PM2.5, PM10, nitrogen dioxide, and sulfur dioxide. Along with this, changing climatic conditions such as temperature, humidity, rainfall, and wind patterns further influence the spread of various respiratory and viral diseases.

Numerous studies have shown a strong correlation between environmental factors and health disorders such as asthma, chronic obstructive pulmonary disease (COPD), bronchitis, and influenza. Although air quality and weather data are widely available through online platforms, this data is usually presented in numerical or technical formats that are difficult for the general public to understand.

There is a lack of intelligent systems that interpret environmental data from a health perspective and provide early warnings related to disease risks. This gap highlights the need for a predictive system that can analyze environmental conditions and estimate disease risk levels for specific locations.

1.2 Problem Statement

Air pollution and changing climatic conditions have a direct and serious impact on human health. Environmental factors such as high levels of particulate matter (PM2.5 and PM10), poor air quality index (AQI), temperature fluctuations, and humidity variations contribute significantly to the rise of respiratory and viral diseases. However, although real-time air quality and weather data are easily available through various platforms, this information is mostly presented in numerical or technical formats that are difficult for common users to understand.

At present, there is no dedicated system that analyzes environmental data from a health perspective and predicts disease risks based on a user's location. The absence of predictive models results in low public awareness, delayed preventive healthcare actions, and increased exposure to environmental health hazards. People often become aware of health risks only after symptoms appear, which reduces the effectiveness of preventive measures.

Therefore, there is a need for an intelligent, climate-based disease risk sensing system that can interpret environmental data, predict potential health risks, and alert users in advance. Such a system would help individuals take timely preventive actions and support better public health awareness.

1.3 Objectives of the Project

The main objective of the **ClimaSense AI – Climate Based Disease Risk Sensing and Alert System** is to analyze environmental conditions and predict potential health risks using machine learning techniques. The specific objectives of the project are listed below:

- To collect real-time air quality data such as AQI, PM2.5, and PM10 from reliable external sources
- To collect real-time weather data including temperature, humidity, wind speed, rainfall, and atmospheric pressure
- To preprocess and analyze environmental data for identifying meaningful patterns related to disease occurrence
- To develop a machine learning-based model for predicting health risk levels (Low, Medium, High)
- To identify and predict the probability of environment-related diseases such as asthma, COPD, bronchitis, and influenza

- To design a user-friendly web application for displaying environmental conditions and prediction results
- To provide timely alerts to users for early awareness and preventive healthcare actions

These objectives aim to improve public health awareness and enable early disease prevention through intelligent analysis of environmental data.

1.4 Proposed Solution

The proposed solution, ClimaSense AI – Climate Based Disease Risk Sensing and Alert System, is an intelligent platform designed to predict environmental disease risks by analyzing real-time air quality and weather data. The system aims to transform raw environmental data into meaningful health-related insights that are easy to understand and act upon.

The system collects real-time air quality parameters such as Air Quality Index (AQI), PM2.5, and PM10, along with weather parameters including temperature, humidity, wind speed, rainfall, and atmospheric pressure using external APIs. This data is then processed through a preprocessing module to remove inconsistencies and select relevant features.

A machine learning model is applied to the processed data to classify health risk levels into Low, Medium, or High categories and to predict the probability of environment-related diseases such as asthma, chronic obstructive pulmonary disease (COPD), bronchitis, and influenza.

The prediction results are presented to users through a web-based application with a simple and interactive interface. Additionally, an email alert system is integrated to notify users about daily environmental conditions and associated health risks. This solution enables users to take timely preventive measures and supports early disease risk awareness.

1.5 Methodology Overview

The ClimaSense AI project follows a systematic and structured methodology to ensure accurate prediction of climate-based disease risks. The methodology consists of multiple stages, starting from data collection and ending with alert generation.

Initially, real-time air quality and weather data are collected from reliable external APIs. The collected data includes parameters such as AQI, PM2.5, PM10, temperature, humidity, wind speed, rainfall, and atmospheric pressure. This data is then passed through a preprocessing stage where it is cleaned, normalized, and filtered to remove irrelevant or inconsistent values.

After preprocessing, the refined dataset is used to train a machine learning model. A Random Forest classifier is employed to analyze the relationship between environmental conditions and disease occurrence. The trained model predicts health risk levels and identifies the most probable environment-related diseases for a given location.

The prediction results are displayed through a web-based application that allows users to easily view environmental conditions and health risk information. Additionally, a daily email alert system is implemented to notify registered users about potential health risks, enabling timely preventive actions.

1.6 Scope of the Project

The scope of the **ClimaSense AI – Climate Based Disease Risk Sensing and Alert System** is focused on predicting health risks caused by environmental conditions using machine learning techniques. The project primarily addresses the impact of air quality and weather parameters on human health.

The system is capable of analyzing real-time air quality and weather data for specific locations and predicting health risk levels. It helps in identifying environment-related diseases such as asthma, chronic obstructive pulmonary disease (COPD), bronchitis, and

influenza. The project also provides a user-friendly web interface and an alert mechanism to improve public awareness regarding environmental health risks.

The scope of this project is limited to predictive analysis and awareness generation. It does not replace professional medical diagnosis or treatment. However, the system can be extended in the future by incorporating additional diseases, more environmental parameters, mobile-based alert systems, and advanced machine learning models for improved accuracy.

1.7 Organization of the Report

The project report is organized into multiple chapters to present the work in a structured and systematic manner.

- **Chapter 1** provides an introduction to the project, including background information, problem statement, objectives, proposed solution, methodology overview, and scope of the project.
- **Chapter 2** presents the literature survey and problem outline, highlighting existing research work related to environmental disease prediction and identifying research gaps.
- **Chapter 3** describes the system architecture and methodology in detail, including data collection, preprocessing, machine learning model, and system workflow.
- **Chapter 4** discusses the implementation details and results obtained from the proposed system.
- **Chapter 5** concludes the report and outlines the future scope and possible enhancements of the project.

This organization ensures clarity, logical flow, and ease of understanding for the reader.

CHAPTER 2

LITERATURE SURVEY AND PROBLEM OUTLINE

2.1 Introduction

Environmental pollution and changing climatic conditions have become major global concerns due to their increasing impact on human health. Factors such as air pollutants, temperature variations, and humidity levels play a significant role in the occurrence and spread of respiratory and viral diseases. With rapid industrialization and urban development, exposure to harmful environmental conditions has increased considerably.

Several research studies have been conducted to analyze the relationship between environmental parameters and health disorders. These studies have highlighted the importance of monitoring air quality and weather conditions to understand disease patterns. However, most existing research focuses on analytical studies or historical data analysis rather than real-time prediction and early warning systems.

The purpose of this literature survey is to review existing research related to environmental disease prediction, identify the techniques used, and analyze their limitations. This review helps in understanding the current state of research and provides a foundation for developing the proposed climate-based disease risk sensing and alert system.

2.2 Literature Survey

Several researchers have studied the relationship between environmental factors and human health. Previous work mainly focused on understanding how air pollution and climatic conditions influence the occurrence of respiratory and viral diseases.

Badri et al. [1] proposed a multi-objective analytical approach to study the impact of environmental conditions on public health planning. Their work highlighted the

importance of environmental parameters in healthcare decision-making but lacked real-time prediction and alert mechanisms.

Brook et al. [2] analyzed the effects of air pollution on cardiovascular and respiratory diseases. Their study established a strong correlation between exposure to pollutants and increased disease risk. However, their research was limited to observational analysis and did not incorporate machine learning-based prediction models.

Lelieveld et al. [3] evaluated the impact of fossil fuel emissions on global public health. The study demonstrated that environmental pollution significantly contributes to disease-related mortality. Although the work provided valuable insights, it focused mainly on policy-level analysis rather than user-level prediction systems.

Li et al. [4] proposed machine learning models for predicting respiratory diseases using environmental factors. Their study showed that ML techniques can improve prediction accuracy. However, the model was trained on offline datasets and did not use real-time environmental data.

Breiman [5] introduced the Random Forest algorithm, which has been widely used in classification and prediction tasks due to its robustness and high accuracy. This algorithm is suitable for handling complex and non-linear relationships in environmental datasets and is adopted in the proposed system.

The World Health Organization (WHO) [6] emphasized that air pollution is one of the leading causes of respiratory and cardiovascular diseases worldwide. The report highlighted the need for public awareness and early warning systems to reduce environmental health risks.

Table 2.1 Summary of Related Research Work

Ref. No.	Author / Year	Focus Area	Technique Used	Limitations
[1]	Badri et al., 1998	Environmental impact on public health	Analytical modeling	No real-time prediction
[2]	Brook et al., 2004	Air pollution & respiratory diseases	Observational study	No ML-based prediction
[3]	Lelieveld et al., 2019	Emission effects on health	Policy-level analysis	Not user-centric
[4]	Li et al., 2022	Disease prediction using environmental data	Machine Learning	Used offline datasets
[5]	Breiman, 2001	Classification algorithms	Random Forest	Algorithm only, no system
[6]	WHO, 2023	Air pollution & public health	Statistical reporting	No predictive or alert system

2.3 Gap Analysis

The review of existing literature reveals that although significant research has been carried out in the area of environmental pollution and its impact on human health, several critical gaps still exist. These gaps limit the effectiveness of current systems in providing practical and preventive healthcare solutions.

- Most existing studies are based on historical datasets and offline analysis, which restricts their ability to provide real-time disease risk prediction. As a result, users do not receive timely information required for preventive actions.

- Current research often focuses on individual environmental parameters such as air pollution or weather conditions, rather than integrating both factors into a unified prediction framework. This leads to incomplete health risk assessment.
- Many studies establish correlations between environmental conditions and diseases but fail to implement intelligent machine learning models capable of predicting future health risks.
- Existing systems are generally designed for researchers or policymakers and lack user-friendly interfaces that can be easily understood by the general public.
- There is a lack of effective alert and notification mechanisms to inform users about potential health risks in advance. Most systems act only after harmful exposure has already occurred.

These gaps highlight the necessity for a real-time, machine learning-based, location-aware disease risk sensing and alert system. The proposed ClimaSense AI system aims to address these limitations by integrating environmental data, intelligent prediction models, and proactive alert mechanisms into a single user-centric platform.

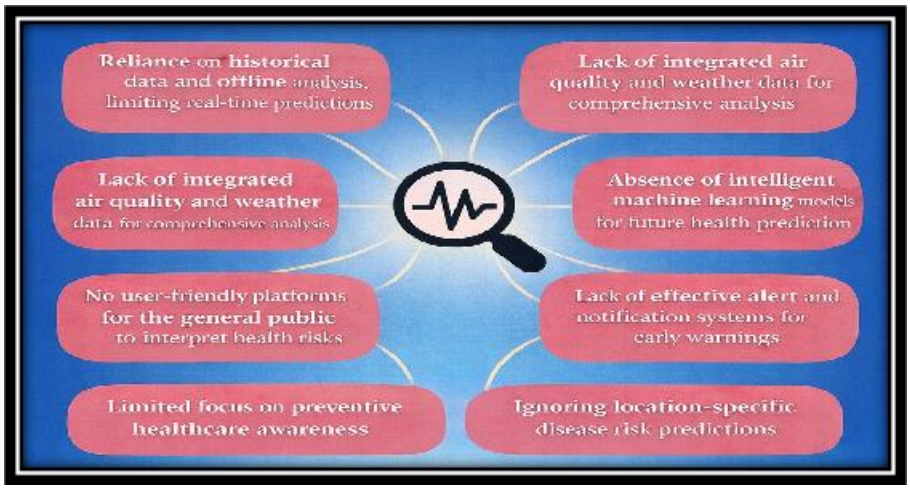


Figure 2.1 Key gaps in existing environmental health monitoring systems

2.4 Problem Outline

Environmental pollution and climate variability have emerged as major contributors to the increasing prevalence of respiratory and viral diseases. Although air quality and weather data are readily available through various government and private platforms, this information is generally presented in raw numerical formats that are difficult for the general public to interpret from a health perspective.

Existing systems mainly focus on monitoring environmental parameters without linking them directly to potential health risks. Most of the current approaches rely on offline analysis, lack predictive capabilities, and do not provide early warnings to users. As a result, individuals remain unaware of health risks until symptoms appear, reducing the effectiveness of preventive measures.

Moreover, current solutions do not integrate multiple environmental factors such as air pollution and weather conditions into a unified framework for disease risk prediction. There is also an absence of intelligent, machine learning-based systems that can analyze real-time data and generate location-specific health risk assessments.

Therefore, the core problem addressed in this project is the lack of a real-time, user-centric system that can interpret environmental data, predict disease risks in advance, and alert users for timely preventive action. The proposed ClimaSense AI system aims to overcome these limitations by combining real-time data acquisition, machine learning-based prediction, and proactive alert mechanisms into a single platform.

CHAPTER 3

SYSTEM ARCHITECTURE AND METHODOLOGY

3.1 Introduction

This chapter explains the system architecture and methodology of the proposed ClimaSense AI – Climate Based Disease Risk Sensing and Alert System. It provides a detailed description of how the system collects real-time environmental data, processes it, applies machine learning techniques, and generates disease risk predictions.

The objective of this chapter is to present a clear understanding of the overall working of the system and the interaction between different modules. A modular architecture has been adopted to ensure scalability, reliability, and ease of future enhancement. Each module is designed to perform a specific function, contributing to accurate disease risk prediction and timely alert generation.

This chapter serves as a foundation for understanding the implementation details and results discussed in subsequent chapters.

3.2 Overall System Architecture

The ClimaSense AI system follows a layered and modular architecture to efficiently analyze environmental data and predict climate-based disease risks. The architecture is designed to ensure accurate prediction, scalability, and ease of maintenance.

The system architecture consists of the following major modules:

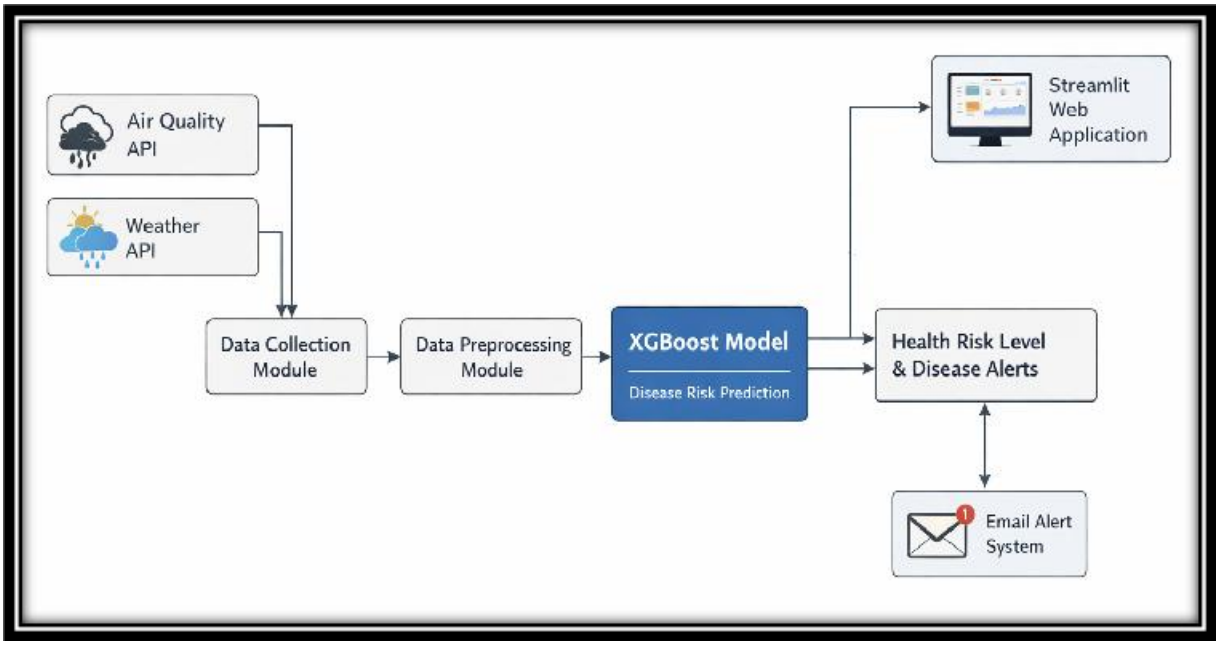
- Data Collection Module
- Data Preprocessing Module
- Machine Learning Prediction Module
- Web Application Interface

- Alert and Notification Module

Initially, real-time air quality and weather data are collected from external APIs and transferred to the data collection module. The collected data is then passed to the preprocessing module, where it is cleaned, normalized, and prepared for analysis.

The processed data is fed into the machine learning module, which uses a Random Forest classifier to predict health risk levels and disease probabilities. The prediction results are then displayed through a web-based interface and shared with users via an email alert system.

Figure 3.1 Overall system architecture of the ClimaSense AI platform



This structured architecture ensures seamless data flow from environmental data acquisition to health risk prediction and alert generation.

Table 3.2 Technology Stack Used in ClimaSense AI

Component	Technology
Programming Language	Python
ML Algorithm	XGBoost
Rule Engine	Epidemiological Rules
Frontend	Streamlit
Data Source	APIs
Alerts	Email System

3.3 Data Collection Module

The data collection module is responsible for acquiring real-time environmental data required for disease risk prediction. Accurate and timely data collection is a crucial component of the ClimaSense AI system, as the quality of predictions depends directly on the input data.

The system collects environmental data from reliable external Application Programming Interfaces (APIs). These APIs provide real-time information for different geographical locations. The collected data is automatically updated at regular intervals and forwarded to the preprocessing module for further analysis.

The data collection process is divided into two main categories: air quality data and weather data.

3.3.1 Air Quality Data Collection

Air quality data plays a vital role in assessing environmental health risks. The system collects the following air quality parameters:

- Air Quality Index (AQI)
- Particulate Matter (PM2.5)

- Particulate Matter (PM10)

These parameters are strong indicators of pollution levels and are closely associated with respiratory diseases. The collected air quality data is location-specific and reflects the current pollution conditions of the selected area.

3.3.2 Weather Data Collection

Weather conditions significantly influence the spread and severity of various diseases. The system collects real-time weather data, including:

- Temperature
- Humidity
- Wind Speed
- Rainfall
- Atmospheric Pressure

This data helps in understanding climatic conditions that may contribute to the outbreak or spread of respiratory and viral diseases. The collected weather data is integrated with air quality data to provide a comprehensive environmental health assessment.

3.4 Data Preprocessing Module

The raw environmental data collected from external APIs may contain missing values, noise, or inconsistencies that can negatively affect the accuracy of disease risk prediction. Therefore, data preprocessing is an essential step in the ClimaSense AI system.

The preprocessing module prepares the collected data for machine learning analysis by performing the following operations:

- **Data Cleaning:** Missing, duplicate, or inconsistent values are identified and removed to improve data quality.

- **Data Normalization:** Environmental parameters such as AQI, PM2.5, PM10, temperature, and humidity are scaled to a standard range to ensure uniformity in model training.
- **Feature Selection:** Only the most relevant environmental features that contribute significantly to disease prediction are selected, reducing model complexity and improving performance.

The output of the preprocessing module is a clean and structured dataset that is suitable for accurate and reliable machine learning predictions.

3.5 Machine Learning Model

The ClimaSense AI system employs **Extreme Gradient Boosting (XGBoost)** as the core machine learning model for predicting climate-based disease risks. XGBoost is a powerful and efficient ensemble learning algorithm based on gradient boosting, known for its high accuracy and ability to handle complex, non-linear relationships in data.

XGBoost builds multiple decision trees sequentially, where each new tree focuses on correcting the errors made by the previous trees. This approach improves prediction accuracy and reduces overfitting, making it suitable for environmental and health-related datasets.

The model is trained using preprocessed environmental data that includes air quality and weather parameters. During training, the model learns the relationship between environmental conditions and disease risk levels.

3.5.1 Health Risk Level Prediction

The trained XGBoost model classifies environmental health risk levels into three categories:

- **Low Risk**
- **Medium Risk**

- **High Risk**

This classification helps users quickly understand the severity of health risks associated with current environmental conditions.

3.5.2 Disease Probability Prediction

In addition to risk classification, the model predicts the probability of environment-related diseases such as:

- Asthma
- Chronic Obstructive Pulmonary Disease (COPD)
- Bronchitis
- Influenza

The system identifies and displays the most probable diseases based on the analyzed environmental parameters, assisting users in early awareness and preventive care.

3.5.3 Advantages of Using XGBoost

The use of XGBoost in this project offers several advantages:

- High prediction accuracy
- Efficient handling of large and complex datasets
- Built-in regularization to prevent overfitting
- Fast training and prediction speed

These advantages make XGBoost an effective choice for real-time disease risk prediction in the ClimaSense AI system.

3.6 Prediction Output Module

The prediction output module is responsible for generating and presenting the final results obtained from the XGBoost model and the epidemiological rule engine. This module acts as a bridge between the machine learning layer and the user-facing components of the system.

After processing the real-time environmental data, the XGBoost model produces predictions related to environmental health risks. These predictions are further validated and refined using the epidemiological rule engine, which applies predefined health-related thresholds and domain rules to enhance reliability.

The output generated by this module includes:

- **Health Risk Level:** Classification of the environmental condition into Low, Medium, or High risk categories
- **Disease Probability Scores:** Identification of the most probable environment-related diseases based on current conditions
- **Environmental Summary:** A concise overview of air quality and weather parameters influencing the prediction

The final prediction results are structured in a clear and interpretable format, making them suitable for display on the web application and for inclusion in alert notifications.

3.7 Web Application Interface

A web-based application is developed using Streamlit to provide an interactive and user-friendly interface for the ClimaSense AI system. The web interface enables users to easily access real-time environmental data and view disease risk predictions without requiring technical expertise.

The main features of the web application include:

- Location selection for retrieving area-specific environmental data

- Display of real-time air quality parameters such as AQI, PM2.5, and PM10
- Display of real-time weather conditions including temperature and humidity
- Visualization of predicted health risk levels
- Display of probable environment-related diseases

The Streamlit framework allows dynamic updating of results as new data is fetched from APIs. Clear labels, charts, and color indicators are used to enhance user understanding of health risks.

The web interface acts as the primary interaction point between the user and the system, transforming complex environmental and prediction data into easily understandable health insights.

3.8 Alert and Notification System

The alert and notification system is an essential component of the ClimaSense AI platform, designed to provide timely health-related information to users. This module ensures that users are informed about potential environmental health risks without the need for continuous manual monitoring.

An automated email alert system is implemented to send daily notifications to registered users. These alerts include key information such as current air quality levels, prevailing weather conditions, predicted health risk levels, and likely environment-related diseases.

The alert system is triggered based on the prediction results generated by the XGBoost model and validated by the epidemiological rule engine. Alerts are particularly emphasized when medium or high health risk levels are detected, encouraging users to take preventive measures.

CHAPTER 4

IMPLEMENTATION AND RESULTS

4.1 Introduction

This chapter describes the implementation details and results obtained from the proposed ClimaSense AI – Climate Based Disease Risk Sensing and Alert System. It explains how the system components discussed in the previous chapter were practically implemented using the selected technology stack and how the system performs under real-time conditions.

The chapter focuses on the implementation of data collection using APIs, preprocessing of environmental data, disease risk prediction using the XGBoost model, and integration of the epidemiological rule engine. It also presents the outputs generated through the Streamlit-based web application and the email alert system.

The results obtained from the system demonstrate the effectiveness of machine learning and real-time environmental analysis in predicting climate-based health risks and providing timely alerts to users.

4.2 System Implementation

The ClimaSense AI system was implemented using Python as the core programming language, with Streamlit used for developing the interactive web interface. The implementation integrates real-time environmental data acquisition, machine learning-based health risk prediction, and an automated email alert mechanism to provide preventive health awareness.

The system workflow begins with location detection or user input, followed by data collection from external air quality and weather APIs. The collected data is then processed and analyzed using the trained XGBoost model and epidemiological rule engine. The prediction results are displayed through the web application and shared with users via email alerts.

4.2.1 User Interface Implementation

The user interface is developed using Streamlit to ensure simplicity and ease of use. The home page displays the project title, a brief system description, and an option to use the current location for analysis. A dedicated button allows users to initiate the health risk assessment process.

The interface is designed to be intuitive so that users without technical knowledge can easily interact with the system and understand the displayed results.

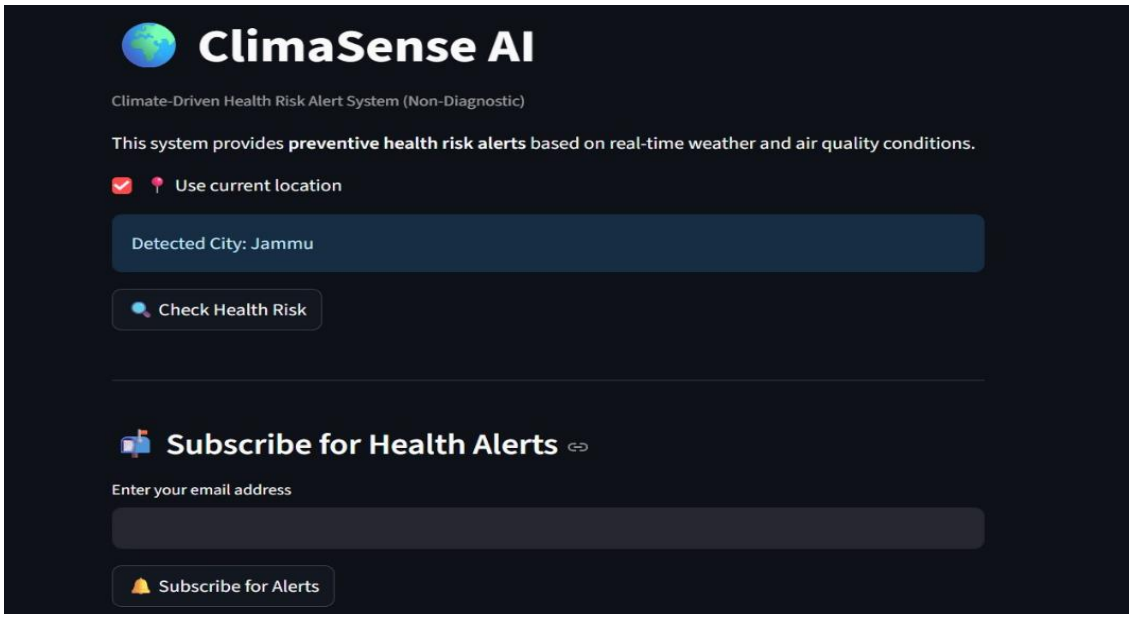


Figure 4.1 Home page of the ClimaSense AI web application

4.2.2 Environmental Data Integration

Once the user initiates the health check, the system fetches real-time environmental data using integrated APIs. The environmental parameters displayed include temperature, humidity, rainfall, AQI, PM2.5, and PM10.

This real-time data forms the input to the prediction system and provides users with a clear understanding of current environmental conditions affecting health.

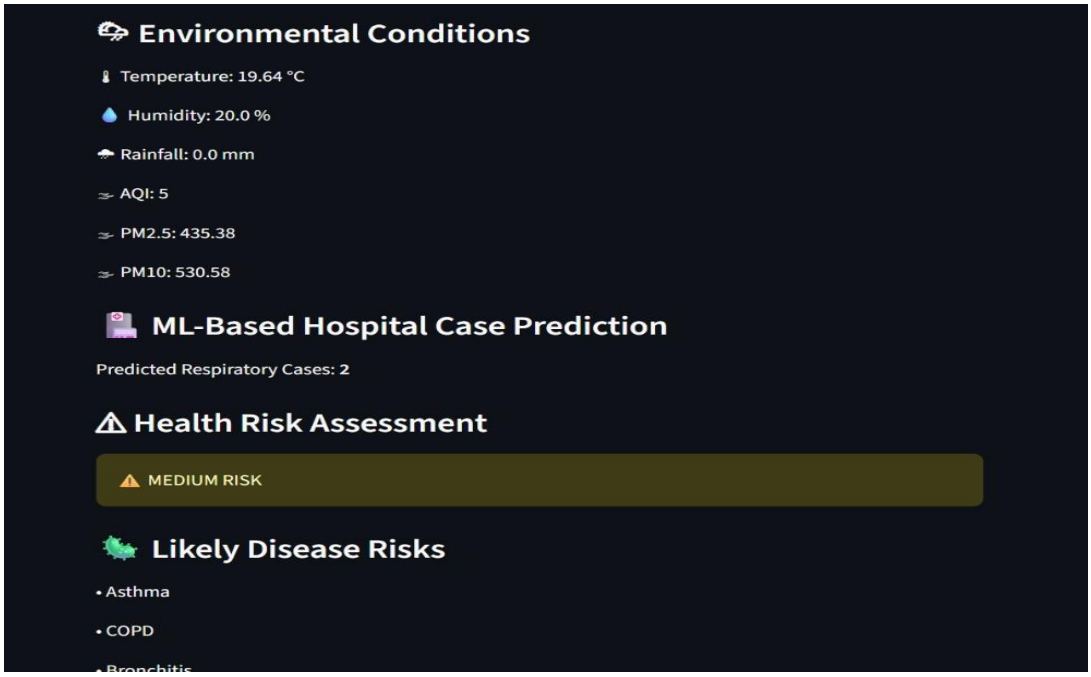


Figure 4.2 Display of real-time environmental conditions and health risk assessment

4.2.3 Machine Learning Model Execution

The preprocessed environmental data is passed to the trained XGBoost model for prediction. The model analyzes the input parameters and predicts the expected respiratory health impact. Based on the output, the system classifies the overall health risk level into Low, Medium, or High categories.

Additionally, the system identifies likely environment-related diseases such as asthma, chronic obstructive pulmonary disease (COPD), and bronchitis. The epidemiological rule engine further validates the predictions using predefined health thresholds.

4.2.4 Alert System Implementation

An automated email alert system is implemented to provide timely notifications to users. Users can subscribe by entering their email address through the web interface. When the predicted health risk level is Medium or High, the system automatically triggers an email alert containing relevant health risk information.

A confirmation message is displayed upon successful subscription, ensuring reliable communication with users.

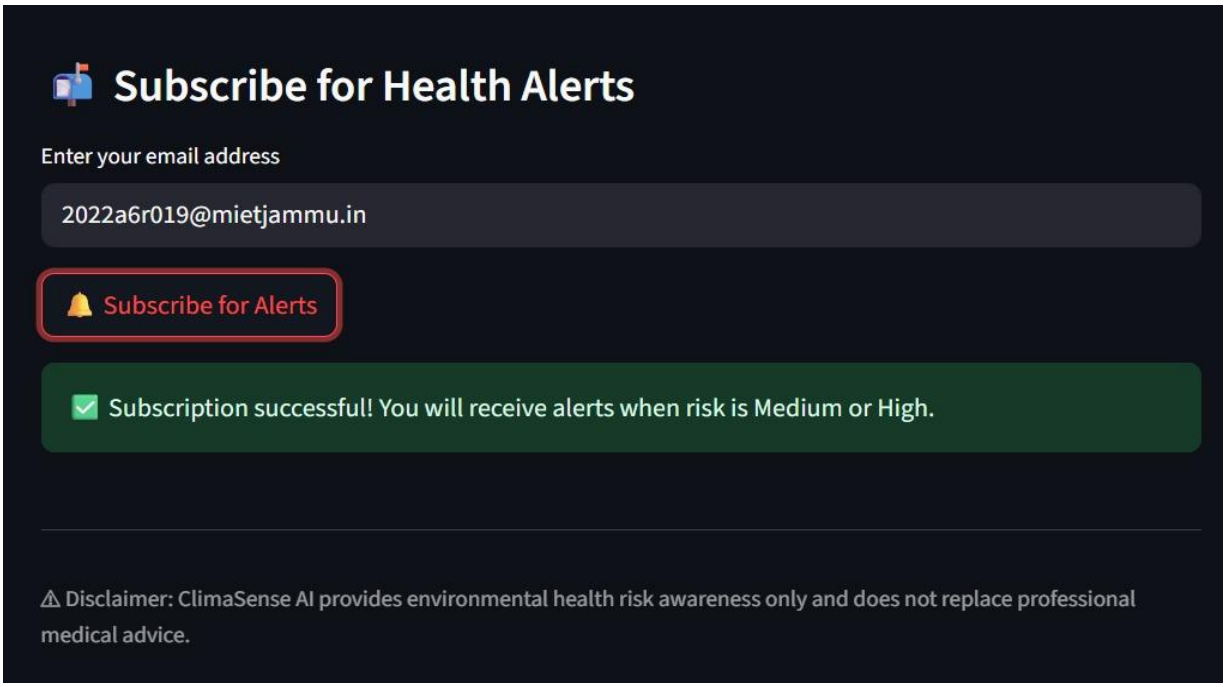


Figure 4.3 Email alert subscription and confirmation message

4.3 Results and Discussion

The ClimaSense AI system was evaluated using real-time air quality and weather data obtained through external APIs. The system successfully processed environmental

parameters and generated meaningful health risk predictions, demonstrating the effectiveness of the proposed approach.

4.3.1 Environmental Data Analysis Results

The system accurately displayed real-time environmental conditions such as temperature, humidity, rainfall, AQI, PM2.5, and PM10 for the selected location. High values of particulate matter (PM2.5 and PM10) were observed to significantly influence the predicted health risk level, confirming the strong correlation between air pollution and respiratory health issues.

The integration of multiple environmental parameters allowed the system to provide a comprehensive assessment rather than relying on a single factor.

4.3.2 Health Risk Prediction Results

Using the XGBoost machine learning model, the system classified health risk levels into Low, Medium, and High categories. During testing, the system correctly identified Medium risk conditions when pollution levels were elevated, even when other weather parameters remained moderate.

This result highlights the capability of the model to capture complex relationships between environmental variables and health outcomes. The epidemiological rule engine further strengthened prediction reliability by validating outputs against known health thresholds.

4.3.3 Disease Risk Identification

In addition to risk classification, the system identified likely environment-related diseases such as asthma, chronic obstructive pulmonary disease (COPD), and bronchitis. These predictions were consistent with known medical research linking air pollution and climate conditions to respiratory illnesses.

The disease risk identification feature enhances user awareness by providing actionable insights rather than generic risk warnings.

CHAPTER 5

CONCLUSIONS AND FUTURE SCOPE

5.1 Conclusion

In this project, ClimaSense AI – Climate Based Disease Risk Sensing and Alert System was successfully designed and implemented to predict environmental health risks using real-time air quality and weather data. The system integrates machine learning techniques with epidemiological rules to provide meaningful health risk insights in a user-friendly manner.

The proposed system effectively collects real-time environmental data such as AQI, PM2.5, PM10, temperature, humidity, and rainfall using external APIs. This data is processed and analyzed using an XGBoost-based machine learning model to classify health risk levels into Low, Medium, and High categories. Additionally, the system identifies likely environment-related diseases such as asthma, COPD, and bronchitis.

The Streamlit-based web application provides an interactive interface for users to easily access health risk predictions, while the email alert system ensures timely notification when medium or high-risk conditions are detected. The results obtained during testing demonstrate that the system can successfully predict climate-based health risks and improve public awareness regarding environmental health hazards.

Overall, the project meets its objectives by providing a real-time, user-centric, and preventive health risk alert system that can assist individuals in taking timely precautionary measures.

5.2 Future Scope

The proposed ClimaSense AI – Climate Based Disease Risk Sensing and Alert System provides an effective solution for predicting environmental health risks; however, there is considerable scope for further enhancement and expansion of the system.

In the future, the system can be extended to include additional diseases such as cardiovascular disorders, skin-related conditions, and allergy-based illnesses to broaden its healthcare impact. More environmental parameters such as pollen count, UV index, and noise pollution can also be incorporated to improve prediction accuracy.

The system can be deployed as a mobile application to increase accessibility and reach a wider user base. Advanced machine learning and deep learning models can be explored to further enhance prediction performance. Integration with hospital databases and government health portals can support large-scale epidemiological analysis and decision-making.

Multilingual support and region-specific customization can also be added to make the system more inclusive and user-friendly. These enhancements will strengthen the system's role in smart healthcare and preventive public health monitoring.

REFERENCES

[1] Badri, M. A., Mortagy, A. K., and Alsayed, A., 1998, "A multi-objective model for locating fire stations," *European Journal of Operational Research*, vol. 110, no. 2, pp. 243–260.

[2] Brook, R. D., Franklin, B., Cascio, W., Hong, Y., Howard, G., Lipsett, M., Luepker, R., Mittleman, M., Samet, J., Smith, S. C., and Tager, I., 2004, "Air pollution and cardiovascular disease: A statement for healthcare professionals," *Circulation*, vol. 109, no. 21, pp. 2655–2671.

[3] Lelieveld, J., Klingmüller, K., Pozzer, A., Burnett, R. T., Haines, A., and Ramanathan, V., 2019, "Effects of fossil fuel and total anthropogenic emission removal on public health and climate," *Proceedings of the National Academy of Sciences*, vol. 116, no. 15, pp. 7192–7197.

[4] Li, Y., Chen, H., Zhang, J., and Guo, Y., 2022, "Machine learning models for predicting respiratory diseases based on environmental factors," *Environmental Research*, vol. 214, Article no. 113820.

[5] Breiman, L., 2001, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32.

[6] World Health Organization (WHO), 2023, "Ambient air pollution and health," WHO Fact Sheets.

[7] Central Pollution Control Board (CPCB), Government of India, 2023, "National Air Quality Index," Available: https://cpcb.nic.in (accessed for academic reference).

[8] Chen, T. and Guestrin, C., 2016, "XGBoost: A scalable tree boosting system," *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794.

[9] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E., 2011, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830.