The impact of Climate Change on the Spread of Infectious Diseases Using Predictive Analysis

SWATHIKA R
Department of Information Technology
Sri Sivasubramaniya Nadar College of Engineering
TamilNadu
swathikar@ssn.edu.in

RADHA N
Department of Information Technology
Sri Sivasubramaniya Nadar College of Engineering
TamilNadu
radhan@ssn.edu.in

KAMALIKA M
Department of Information Technology
Sri Sivasubramaniya Nadar College of Engineering
TamilNadu
kamalika2310617@ssn.edu.in

KAVIYA R
Department of Information Technology
Sri Sivasubramaniya Nadar College of Engineering
TamilNadu
kaviya2310092@ssn.edu.in

KRISHITHA V
Department of Information Technology
Sri Sivasubramaniya Nadar College of Engineering
TamilNadu
krishitha2310088@ssn.edu.in

2025 International Conference on Emerging Technologies in Computing and Communication (ETCC) | 979-8-3315-2476-0/25/$31.00 ©2025 IEEE | DOI: 10.1109/ETCC65847.2025.11108619

***Abstract* — Climate change plays a growing role in the spread of infectious diseases by altering environmental conditions that influence disease vectors, host interactions, and exposure risks. Rising temperatures, shifting humidity levels, declining air quality, and increased rainfall contribute to the expanded reach and seasonal shifts of vector-borne diseases such as malaria and dengue. Extreme weather events—including floods and cyclones—also create favorable environments for the emergence of waterborne and respiratory infections. This study explores the connection between climate variability and disease outbreaks using weather-based risk modeling. Meteorological factors such as temperature, humidity, air quality index (AQI), and precipitation are analyzed to identify patterns and correlations with disease incidence. Machine learning techniques, including Random Forest models, are employed to forecast disease risks based on these climate indicators. The results emphasize the importance of climate-informed health strategies for enhancing early warning system**

Keywords* — *Climate change, disease risk modeling, extreme weather events, infectious disease spread, predictive analytics, vector-borne diseases.

I. INTRODUCTION

The global climate is undergoing rapid and unpredictable changes, significantly impacting human health. Among the most pressing concerns is the increasing prevalence and spread of infectious diseases driven by shifting environmental factors. Temperature fluctuations, humidity variations, and extreme weather events directly influence the survival and transmission rates of pathogens and their vectors. For example, higher temperatures accelerate mosquito breeding cycles, intensifying the risk of malaria and dengue fever in regions previously unaffected. Similarly, heavy rainfall and flooding enhance the spread of waterborne diseases such as cholera and leptospirosis.

In addition to vector-borne and waterborne diseases, climate conditions play a significant role in respiratory illnesses. Rising air pollution levels, coupled with temperature shifts, exacerbate respiratory infections, including influenza and pneumonia. As air quality deteriorates due to climate change-induced factors such as wildfires and increased urban pollution, the incidence of respiratory ailments is expected to rise.

To better understand the link between climate and disease prevalence, this study utilizes weather-based disease risk assessment models. By analysing historical weather data across different climatic zones—ranging from tropical to temperate regions—we quantify disease probabilities based on factors such as temperature fluctuations, precipitation intensity, air pollution levels, and humidity. This study analyzes real-world datasets to establish correlations between environmental changes and disease spread across regions. The dataset includes meteorological records from major Indian cities (Delhi, Mumbai, Chennai, Bangalore, and Kolkata), incorporating real-world variations in environmental conditions.

A key component of this study is the application of predictive analytics in disease forecasting. The analysis demonstrates a higher probability of dengue in humid conditions and increased respiratory illness risk in areas with severe air pollution. By integrating climate science with epidemiological modeling, this research provides insights to help public health authorities implement proactive disease prevention measures. As climate change continues to influence disease patterns, data-driven approaches will be essential for early warning systems and effective health policies.

II. RELATED WORK

The relationship between climate change and the spread of infectious diseases has been the subject of extensive research, highlighting the multifaceted ways in which environmental factors influence disease dynamics.

A comprehensive study [1] analyzed the impact of climatic hazards on human pathogenic diseases, revealing that 58%

of such diseases have been aggravated by climatic hazards. The study identified 1,006 unique pathways through which climatic hazards, via different transmission types, led to pathogenic diseases.

Review [2] focusing on how variations in temperature, precipitation, and humidity affect the transmission and distribution of infectious diseases. The study emphasized that the life cycles of many infectious agents are closely linked to climate, with changes in these climatic variables influencing disease prevalence.

In another study [3] the complex interplay between climate change, ecosystems, and human activities, was discussed which can create conditions favourable for the emergence and spread of infectious diseases. The research underscored the challenges posed to public health by these dynamics.

In study [4], the paper explains that both disease-causing infectious agents and the organisms that spread them (like mosquitoes) are very sensitive to changes in climate, especially temperature and humidity. Warmer weather speeds up mosquito breeding and helps the virus spread faster, while more rain creates stagnant water for mosquitoes to breed. In India, dengue cases increased five times from 2010 to 2019, showing how serious the problem has become.

Paper [5] focuses on climate change makes diseases spread more easily in different ways. It helps mosquitoes grow in new areas, which can increase diseases like malaria. Changes in weather, like temperature and humidity, can also affect breathing problems, causing more infections. All these changes make it easier for diseases to spread and harm people.

In study [6], the paper explains how climate change is causing higher global temperatures, extreme weather, and increased air pollution, allergens, and airborne pathogens. These changes are worsening air quality and respiratory health.

Study [7] explores machine learning models to study the impact of climate change on West Nile virus distribution, revealing how shifting environmental conditions expand high-risk zones.

Study [8] examines how applying causal inference methods can improve our understanding of how climate influences the spread of infectious diseases. By analysing the relationship between climate variables and disease transmission, researchers can make more accurate predictions about future outbreaks. The study emphasizes the need for collaboration across multiple disciplines, combining climate science, epidemiology, and data analysis to better anticipate and mitigate the health impacts of climate change.

Another study [9] utilized Neural Ordinary Differential Equations (Neural ODEs) to forecast Black Sigatoka infection risks, demonstrating the effectiveness of deep learning in modeling disease spread.

[10] demonstrated how climate variability influences infectious disease patterns, while [11] highlighted the role of warming in emerging pathogens. Subsequent work by [12] and [13] quantified health risks from heat extremes and projected climate-driven disease spread noting vulnerabilities in urban populations.

Respiratory health risks from climate-aggravated air pollution were explored in [14],[15] and [19], with focusing on Indian cities, where rapid urbanization exacerbates exposure. The impacts on vector-borne diseases were further detailed in [17], which analysed tick-borne disease risks, and [18], which mapped expansions of dengue, malaria, and Lyme disease.

Methodological advances emerged in parallel. [16] reviewed global research trends.

Above studies collectively underscore the critical need for integrating climate variables into disease risk modelling to enhance early warning systems and public health interventions.

Unlike previous studies that primarily focus on statistical correlations or single-disease models, this research integrates machine learning-based predictive modeling with real-time weather data to forecast multiple infectious diseases, including dengue, malaria, respiratory illnesses, flu, and pneumonia. By leveraging 10 years of historical weather data and a Random Forest classifier, this approach enhances predictive accuracy and provides dynamic, location-specific disease risk assessments. The incorporation of real-time weather APIs sets this study apart, enabling automated early warning systems that can assist public health officials in mitigating climate-driven disease outbreaks.

III. DATASET DESCRIPTION

The Spread of infectious diseases is determined using the generated data set. The dataset is in CSV format for processing. The dataset contains 10 years of daily weather data for five major Indian cities: Delhi, Mumbai, Chennai, Bangalore, and Kolkata.

TABLE I. DATASET OVERVIEW

Category	Details
Total Records	18256 (one record per day per city)
Period	March 2015 - March 2025
Data Collected For	Chennai Kolkata Mumbai Bangalore Delhi
Years of data	10 Years

TABLE II. DATASET DESCRIPTION

Parameters	Description
Date	Format DD-MM-YYYY
City	Delhi, Mumbai, Chennai, Bangalore, Kolkata
temperature	daily average temperature in degrees Celsius (°C)
humidity	daily average relative humidity in percentage (%)
aqi	daily air quality index
precipitation	daily precipitation in millimeters (mm)
season	summer, winter, monsoon, autumn

dengue_probability	range from 0.1 – 1.0 calculated based on levels of humidity and precipitation
malaria_probability	range from 0.1 – 1.0 calculated based on levels of temperature and precipitation
respiratory_issues_ probability	range from 0.2 – 1.0 calculated based on air quality index (AQI)
cold_flu_probability	range from 0.1 – 1.0 calculated based on levels of temperature and humidity
pneumonia_probability	range from 0.1 – 1.0 calculated based on levels of temperature and humidity

The weather parameters included are temperature, humidity, aqi (air quality index), precipitation along with 4 seasons Summer, Winter, Monsoon, Autumn. The Probabilities for various diseases are generated as numerical values between 0 and 1, indicating the likelihood of health issues under given environmental conditions.
The disease risk probabilities are calculated based on **weather parameters**. The conditions below are for high-risk probability of diseases:

Dengue: High humidity (>75%) and precipitation (>20 mm).

Malaria: High temperatures (>30°C) and moderate precipitation (>15 mm).

Respiratory Issues: High AQI (>200) increases the probability of respiratory infections due to poor air quality.

Flu and Pneumonia: Low temperatures (<18°C) and high humidity (>65%).

The probabilities are further calculated based on the **season** to reflect seasonal variations in disease transmission:

In **Monsoon (**higher precipitation and humidity) the probability of dengue and malaria is increased by 0.2.

In **Autumn (**seasonal changes in air quality) the probability of respiratory issues is increased by 0.2.

In **Winter (**lower temperatures and higher humidity) the probability of flu and pneumonia is increased by 0.2.

IV. SYSTEM ARCHITECTURE

In Fig. 1, the architecture of the proposed system is illustrated, depicting the end-to-end process of disease prediction based on real time weather data.

The components of the diagram are as follows:

A. User Input

The system starts with user interaction where the individual provides the name of the city. This acts as the reference to retrieve real time weather conditions.

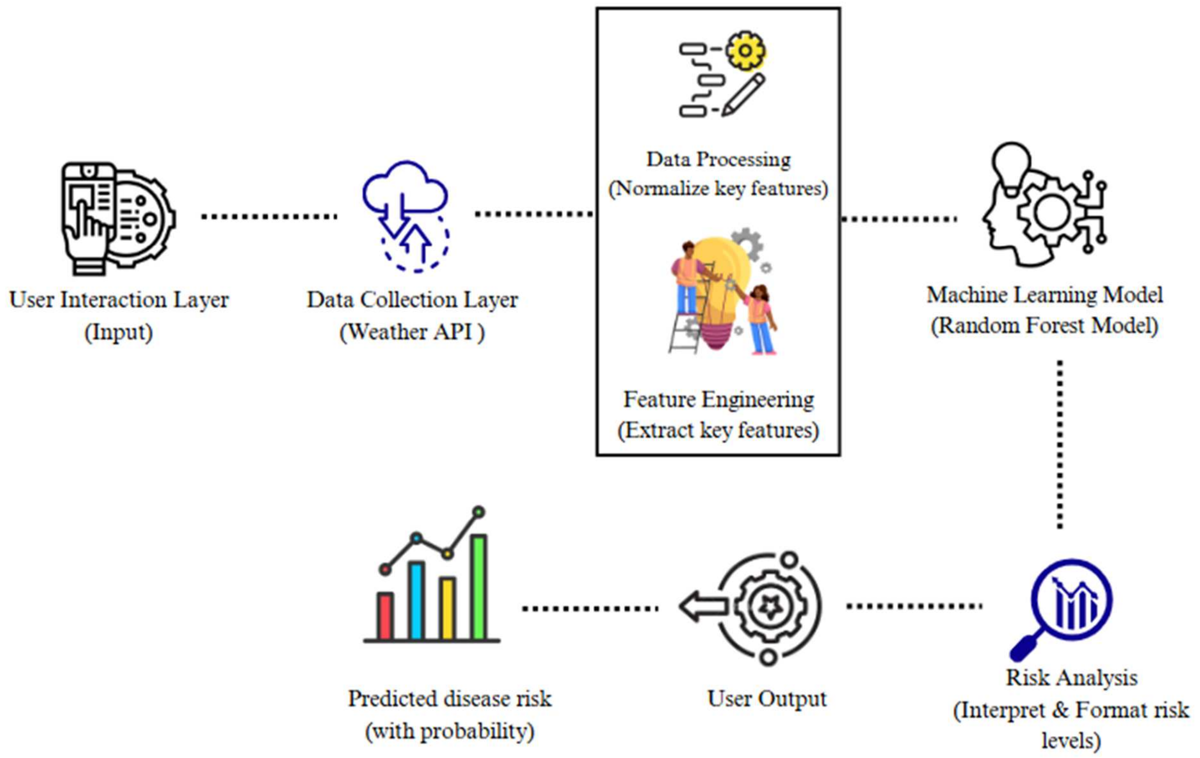

Fig. 1. System Architecture for Weather Based Disease Probability Prediction

B. Weather Data Collection (API Integration)

Once the location is provided, the system queries the OpenWeatherMap API to collect real-time weather attributes including temperature, humidity, Air Quality Index (AQI) [20], and precipitation. These factors influence disease occurrence.

C. Data Preprocessing & Feature Engineering

The retrieved weather data is cleaned and structured for model input, ensuring that only the most relevant parameters affecting disease risk are considered.

D. *Machine Learning Model*

A trained Random Forest Model classifier processes the weather data to compute probability scores for each disease predicting the likelihood of each disease.

E. Prediction Output

Instead of binary classification, the model provides probability scores for each disease, offering a quantitative risk assessment to users.

F. User Visualization & Alerts

The results are displayed in an effective way enabling users o take proactive measures based on their probability scores.

V.PROPOSED METHODOLOGY

Most disease predictions focus on just one or two factors, like temperature and humidity or air quality, to predict the risk of a disease. However, diseases are influenced by a combination of factors such as weather, environment, and seasons.

This study considers multiple factors like temperature, humidity, AQI, precipitation, and the season to understand how they work together to spread diseases.

The methodology employed in this study includes the integration of real-time weather data to improve the prediction of disease spreading.

Traditional disease prediction system relies on reported cases, which may cause delays in response. This methodology uses Machine Learning algorithm, specifically the Random Forest Classifier, to detect patterns in environmental conditions that correlate with disease outbreaks.

ALGORITHM FOR PREDICTING SPREAD OF INFECTIOUS DISEASES

1. Dataset generation
2. Loading dataset
3. Preprocessing dataset
4. Selecting features required to predict disease risk
5. Defining target variables (disease probabilities)
6. Converting target variables to binary labels (1 = high risk, 0 = low risk)
7. Splitting dataset into training and testing sets (80-20 split)
8. Training RandomForestClassifier model for each disease
9. Fetching Real time weather data by getting city as input
10. Using trained models to predict disease probabilities
11. Returning predicted disease risk probabilities

A. Input

The dataset used in this study contains 10 years of daily weather data for five major Indian cities: Delhi, Mumbai, Chennai, Bangalore, and Kolkata. It includes **18,256 entries** with features such as temperature, humidity, AQI, precipitation, season, and disease probabilities. These features are used to predict the risk of infectious diseases based on environmental conditions.

B. Data Preprocessing

The dataset is pre-processed to improve performance. This involves:

1. **Handling Missing Values**: Missing data is either filled or removed.
2. **Normalization**: Features like temperature, humidity, and AQI are scaled to a range (e.g., 0 to 1) to avoid bias in the models.
3. **Encoding Categorical Features**: The "season" feature is encoded into numerical values (e.g., Summer = 1, Winter = 2, Monsoon = 3, Autumn = 4).

C. Model Processing

To predict the risk of infectious diseases, machine learning models such as Random Forest Classifier were used, with separate models for dengue, malaria, respiratory issues, flu, and pneumonia. These models analyzed weather factors like temperature, humidity, AQI, precipitation, and seasons, using one-hot encoding to process seasonal data.

D. Classified Output

The model used here provides a classified output, predicting the risk level (High or Low) for each of the five diseases - dengue, malaria, respiratory issues, cold & flu, and pneumonia.

E. Comparison and Final Prediction

The performance of model was evaluated using key metrics like accuracy, precision, recall, and F1-score. The malaria and pneumonia models achieved perfect accuracy, while the dengue and respiratory issues models also showed high reliability. However, the cold & flu model had a lower recall, suggesting it needs improvement.

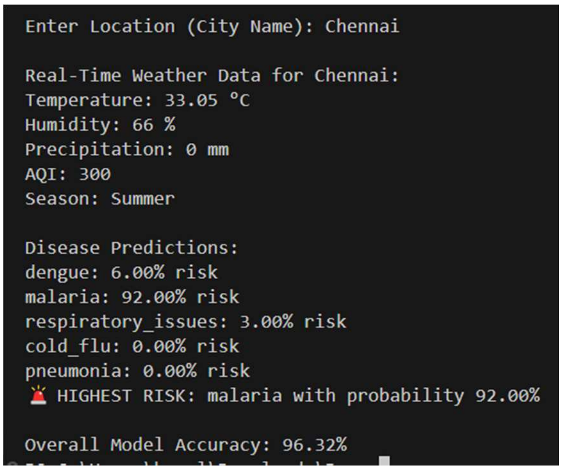

Fig. 2. Disease Prediciton

VI. EXPERIMENTAL RESULT

A. Overview of Model Performance

The predictive disease risk model was trained using 10 years of historical weather data from five major Indian cities (Delhi, Mumbai, Chennai, Bangalore, and Kolkata). The dataset was preprocessed by applying one-hot encoding to represent seasonality (Monsoon, Autumn, Summer, Winter). The features used for training included temperature, humidity, precipitation, air quality index (AQI), and the encoded seasonal variables.

The dataset was split into 80% training and 20% testing subsets using a fixed random state for reproducibility. Separate Random Forest classifiers were trained for each disease with balanced class weights to address potential data imbalance. The models were evaluated using accuracy, precision, recall, and F1-score. The performance metrics are summarized in Table III.

TABLE III. MODEL PERFORMANCE METRICS FOR DISEASE RISK PREDICTION

Disease	Accuracy (%)	Precision	Recall	F1-Score
Dengue	93.51	0.94	0.90	0.92
Malaria	100.00	1.00	1.00	1.00
Respiratory Issues	96.25	0.96	0.97	0.96
Cold & Flu	91.81	0.89	0.73	0.79
Pneumonia	100.00	1.00	1.00	1.00

Malaria and pneumonia predictions achieved 100% accuracy, indicating highly distinct feature distributions. Dengue prediction (93.51% accuracy) showed high reliability, with humidity and precipitation being strong indicators. Cold & flu had the lowest recall (0.73), suggesting that environmental factors alone may not fully explain its occurrence.

B. Correlation Heatmap: Weather Conditions vs. Disease Probabilities

To understand the relationship between environmental factors and disease prevalence, we analyzed correlations between temperature, humidity, AQI, precipitation, and disease probabilities. Fig. 3 presents a heatmap depicting these relationships.

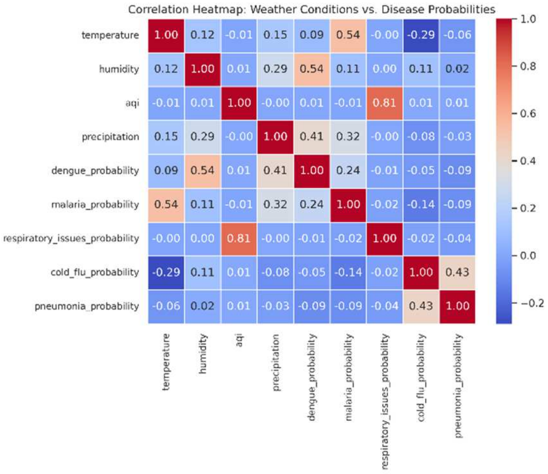

Fig. 3. Correlation Heatmap

Key findings from the correlation analysis include:

- Dengue probability is strongly correlated with humidity and temperature, confirming that warm, humid conditions favor transmission.
- Malaria probability has a significant correlation with precipitation, highlighting that rainfall contributes to stagnant water breeding grounds.
- Respiratory issues probability is highly correlated with AQI levels, reinforcing the impact of air pollution on public health.

These insights validate the selection of temperature, humidity, AQI, and precipitation as primary features for training the predictive model.

C. *Real-Time Disease Prediction (Chennai, India Case Study)*

To validate the model, real-time weather data was fetched for Chennai via the OpenWeatherMap API [20]. The retrieved weather parameters were used to predict disease risks, as shown in Table IV

TABLE IV. REAL-TIME WEATHER DATA FOR CHENNAI

Parameter	Value
Temperature	27.96°C
Humidity	82%
Precipitation	0 mm
AQI	300 (Severe)
Season	Summer

Using these environmental conditions, the model generated the following predictions, as shown in Table V.

TABLE V. PREDICTED DISEASE PROBABILITIES FOR CHENNAI

Disease	Predicted Risk (%)
Dengue	98.00 □
Malaria	1.00
Respiratory Issues	9.00
Cold & Flu	0.00
Pneumonia	0.00

D. *Observations*

To better illustrate the risk distribution, Fig. 4 presents a chart comparing disease probabilities for Chennai's real-time weather conditions.

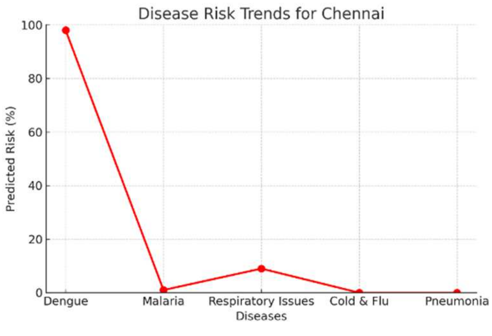

Fig. 4. Disease Risk Probability for Chennai

Dengue risk is extremely high (98.00%) due to high humidity (82%) and warm temperatures. Malaria risk remains low (1%), suggesting a stronger dependency on precipitation levels. Respiratory issues have a moderate risk (9%), influenced by AQI (300 - severe pollution). Cold & flu and pneumonia risks are negligible, likely due to the absence of cold weather conditions.

The model demonstrates reliable performance on test data and shows practical utility through real-time predictions. However, lower recall for cold & flu suggests environmental features alone may be insufficient to capture all influencing factors for this disease. Future work could integrate additional data sources or explore advanced tuning methods to improve model robustness.

VII CONCLUSION AND FUTURE WORK

This research explores how real-time weather data can be used to predict the probability of weather-sensitive diseases using Machine Learning. By leveraging a trained Random Forest Model, the system provides probability scores for various diseases.

With real-time insights, users can take protective measures to safeguard their health based on changing environmental conditions. This system is highly relevant in today's world where climate change is causing unpredictable weather patterns and increasing health risks. The study highlights the important role of ML in addressing weather-driven health concerns and promoting proactive healthcare. Future work could focus on improving model accuracy by incorporating additional factors such as population density, healthcare access, and historical disease outbreaks. Integration with mobile health applications or public health warning systems could enhance timely dissemination of risk information to affected communities.

[1] Mora, C., McKenzie, T., Gaw, I.M., et al. (2022). Over half of known human pathogenic diseases can be aggravated by climate change. Nature Climate Change, 12, 869–875.
[2] Wu, X., Lu, Y., Zhou, S., Chen, L., Xu, B. (2023). Climate change and infectious diseases: a review of evidence and perspectives from China. Infectious Diseases of Poverty, 12, 23.
[3] Carlson, C.J., Albery, G.F., Merow, C., et al. (2024). Climate change, its impact on emerging infectious diseases and implications for control. International Journal of Infectious Diseases, 113, 268–273.
[4] Hussain, S. S. A., & Dhiman, R. C. (2022). Distribution expansion of dengue vectors and climate change in India. *Geohealth*, *6*(6), e2021GH000477.
[5] Semenza, J. C., Rocklöv, J., & Ebi, K. L. (2022). Climate change and cascading risks from infectious disease. *Infectious diseases and therapy*, *11*(4), 1371-1390.
[6] Chung, F., Wong, G., Salvi, S., & Carlsten, C. (2024). Climate Change and Air Pollution: How Healthcare Providers Can Help Mitigate the Risks to Respiratory Health. *EMJ*.
[7] C. Lorenz, T. S. de Azevedo, and F. Chiaravalloti-Neto, "Impact of climate change on West Nile virus distribution in South America," Heliyon, vol. 8, no. 5, May 2022.
[8] Zhang, X., Li, Y., & Thompson, R. (2024). How Causal Inference Concepts Can Guide Research into the Effects of Climate on Infectious Diseases.
[9] Y. Wang et al., "Forecasting Black Sigatoka infection risks with latent neural ODEs," presented at the Tackling Climate Change with Machine Learning Workshop at the 38th

International Conference on Machine Learning (ICML), Jul. 2021.
[10] Patz, J. A., Githeko, A. K., McCarty, J. P., Hussein, S., Confalonieri, U., & De Wet, N. (2003). Climate change and infectious diseases. *Climate change and human health: risks and responses*, *2*, 103-132.
[11] Epstein, P. R. (2001). Climate change and emerging infectious diseases. *Microbes and infection*, *3*(9), 747-754.
[12] Ebi, K. L., Capon, A., Berry, P., Broderick, C., de Dear, R., Havenith, G., ... & Jay, O. (2021). Hot weather and heat extremes: health risks. *The lancet*, *398*(10301), 698-708.
[13] D'amato, G., Vitale, C., De Martino, A., Viegi, G., Lanza, M., Molino, A., ... & D'amato, M. (2015). Effects on asthma and respiratory allergy of Climate change and air pollution. *Multidisciplinary respiratory medicine*, *10*(1), 1-8.
[14] Singh, A. B., & Kumar, P. (2022). Climate change and allergic diseases: an overview. *Frontiers in allergy*, *3*, 964987.
[15] Kumar, P. (2021). Climate change and cities: challenges ahead. *Frontiers in Sustainable Cities*, *3*, 645613.
[16] Van de Vuurst, P., & Escobar, L. E. (2023). Climate change and infectious disease: a review of evidence and research trends. *Infectious Diseases of Poverty*, *12*(1), 51.
[17] Gilbert, L. (2021). The impacts of climate change on ticks and tick-borne disease risk. *Annual review of entomology*, *66*(1), 373-388.
[18] Parums, D. V. (2024). Climate change and the spread of vector-borne diseases, including dengue, malaria, lyme disease, and west nile virus infection. *Medical Science Monitor: International Medical Journal of Experimental and Clinical Research*, *29*, e943546-1.
[19] Kaur, R., & Pandey, P. (2021). Air pollution, climate change, and human health in Indian cities: a brief review. *Frontiers in Sustainable Cities*, *3*, 705131.
[20] OpenWeather API https://openweathermap.org/api