AI BASED HEALTH MONITORING SYSTEM FOR TRACKING CLIMATE CHANGE IMPACT ON VULNERABLE POPULATION

Dr. Santosh Kumar Singh[1], Srushti Raut[2], Deepa Jaiswal[3]

[1]HOD (Information Technology), Department Of Information Technology Thakur College Of Science And Commerce, Mumbai, India.

[2,3]PG Student, Department Of Information Technology Thakur College Of Science And Commerce, Mumbai, India.

DOI: https://www.doi.org/10.58257/IJPREMS43962

ABSTRACT

Climate change is becoming a menace to our human health particularly among the most vulnerable groups of people: the elderly, children and people with chronic ailments. To overcome this problem, we built another AI-based health monitoring system. This system is aimed at forecasting health risks associated with the climate by means of integrating data about the environment such as temperature, humidity and quality of air with the personal health of an individual. Structured data sets of the model parameters were made using State, District, Year, Month, Temperature (o C), Humidity (per cent), Rainfall (mm), Disease occurrence, Affected number of Children, Affected number of Elders and Population. The system issues early alerts and insightful information about the person using machine learning to aid in preventative health services and improved responses to weather-related disasters. We have done experiments with various algorithms to determine which one worked best in our study. Random Forest algorithm was the best as it gave 93% accuracy in risk prediction. This was then succeeded by Logistic Regression of 90% and Support Vector Machine (SVM) of 88%. Such findings demonstrate that AI could be an effective instrument to enhance health outcomes, decrease the climate-sensitivity of diseases burden and assist policy-makers to better distribute resources to the most vulnerable groups.

Keywords: AI, Health Monitoring, Climate Change, Vulnerable Populations, Predictive Analytics, Machine Learning.

1. INTRODUCTION

Climate change has been seen to be one of the most prominent threats to human health in the 21 st century. The increasing global temperatures, extreme weather patterns, changes in rainfall patterns, and poor air quality conditions are contributing to the increase in climate sensitive health issues, respiratory diseases, cardiovascular stress, disease vectors, and heat related diseases. These social impacts are especially harsh in the number of vulnerable groups, such as children, the aged, and those with chronic health issues, who experience low adaptive capacity and access to healthcare resources. In order to deal with these issues, it requires new forms of data based solutions to track, forecast, and respond to the climate-related health hazards. Artificial Intelligence (AI) and machine learning (ML) offer the ability to analyze big amounts of data, identify the patterns that can be hidden, and make the predictions that could be relied upon in real-time. With the incorporation of both the population health data and the environmental variables, the AI-based systems can help provide early alerts, personalized information, and actionable advice to the authorities of the population health.

This paper has designed an AI-based health monitoring system through which the effects of climate change on susceptible groups were assessed. This system makes use of structured datasets whose parameters are State, District, Year, Month, Temperature (degC), Humidity, Rainfall (mm), Disease occurrence, Children Affected, Elders Affected and Population. These characteristics allow conducting a thorough examination of the state of the environment and demographic risks.

The purpose of the Research is to design and test the AI-based health monitoring system that forecasts the risks of diseases in relation to climate changes. The key objectives are:

- **Synthesizing disparate data:** Integrating the environmental conditions (temperature, humidity, rainfall) and the data on public health (disease occurrences, children and older population) into a single data set.

- **Creation of predictive models:** Construction and evaluation of several models such as the Random Forest, Logistic Regression, and Support Vector Machine (SVM), among the hybrid ensemble models.

- **Performance evaluation:** Predictive accuracy of each model should be compared to assess the most effective of the models to use in prediction of climate sensitive diseases.

- **Analyzing trends:** Interpretation of data through data visualization methodology to gather and describe patterns in disease prevalence over time and area by state, district, month, and year.

➢ Key Results The experimental analysis had made a range of significant conclusions:

- Ensemble learning allowed the Hybrid ALL model, which is the combination of the results of Random Forest, Logistic Regression, and SVM, to be the most accurate with the highest accuracy of 94%.

- Random Forest was best suited among individual algorithms with an accuracy of 93%, then Logistic Regression (90%) and SVM (88) indicating that there is a great potential of ensemble algorithms and tree-based frameworks to discern intricate climate-health interactions.

- Climate variables and disease patterns had a visible correlation, which exposed seasonal changes and vulnerabilities to particular regions. These observations indicate problematic periods and hot spots where preventive healthcare intervention and resource allocation are the most necessary.

Problem Statement

The research is handling the increasing menace of climate change given that it affects human life, and the vulnerable populations in the scenario are the ages, the children, and the chronically ill or unwell people. Increased respiratory illnesses, heart related illnesses and the prevalence of diseases, which are caused by vectors have been supplemented by global warming and air pollution as well as the frequent extreme weather-related events. The most problematic fact is that there are no smart and data-driven systems that would be able to identify the threat at the earliest possible point and react to the interacting complex systems of climate-health processes to take the required measures in time. To fill the gap, the project will present a system of AI-driven health monitoring, and the components of environmental factors, such as temperature, humidity, and air quality, will be considered in the health-related data. The effectiveness of the machine learning model, which involves the use of the Ran Forest, SVM as well as the Logistic Regression, will also be assessed to demonstrate the best approach. The ultimate objective is to offer state-of-the-art warning and viable information so as to enable healthcare providers and policy makers to streamline resources and create health planning that is climate resilient.

2. LITERATURE REVIEW

In Data analytics in Sustainable city planning, Saeid Pourroostaei Ardakani, Georgios Kapogiannis, Mohammed Al-khafajiy and Miao Yu concentrate on how the city may overcome such problems of the city as traffic, flooding, and inequality in the city using technologies, including GIS and machine learning, and Digital twins. They emphasize the significance of real-time analytics and interactive platforms in enhancing governance and resilience. The conclusion of the paper is that sustainability of an urban dwelling can be established in the future with the assistance of technology that will improve efficiency and ethical utilization of information and communication with the citizens is demanded.[1]

Joel Fernando Soares Filipe, Maria Giovanna Ciliberti, George Attard, Gabrielle Ratti and Mara Silvia Laura Rocchi in the editorial Biosecurity of Infectious Diseases in Veterinary Medicine emphasize the importance of biosecurity in the prevention of livestock diseases which pose a threat to the productivity of animals and human health. They emphasize the policies of cleaning and disinfection, new disinfectants such as chlorous acid water and UV radiation and biosecurity-specific devices on farms. The case studies on pigs, poultry, and dairy farms deal with Salmonella Dublin, Infectious Bronchitis, and antimicrobial resistance. The authors promote a One Health concept, in which the collaboration of technology, policy, and stakeholders are combined to safeguard both animals and humans. [2]

The authors in Emerging Trends and Challenges in Environmental Health and Natural Resource Management (Engr. Prof. Theophilus Aku Ugah and others) speak on how sustainability is being transformed by the climate change, technology and the changes in society. They mention such tools as remote sensing, AI, and IoT as the means of real-time surveillance, but caution against such threats as microplastics, pharmaceuticals, and e-waste. Case studies, including Amazon deforestation monitoring and Denmark renewable implementation, reveal some viable solutions, and the authors highlight that sustainable development needs technological innovation, powerful policies, and community-based solutions based on environmental justice. [3]

Ameya D. Bendre, Pooja Singh, and Swati Jaiswal in their editorial article Tuberculosis: recent updates in basic research, drug discovery and treatment note that TB is one of the health-related issues that a particular part of the world cannot afford due to the COVID-19 pandemic. They also point to the co-morbidity of TB with hepatitis B, pregnancy, and lung cancer, but also mention such technological improvements as nanotechnology diagnostics and AI-assisted X-ray interpretation. The authors provide a multidisciplinary, innovative way to achieve the goals of the End TB Strategy of WHO. [4]

Waseem Jerjes and Azeem Majeed, in their editorial, Transforming GP training in the UK: the lasting impact of COVID-19 on telehealth and hybrid care models, propose that a hybrid model of digital and in-person GP training is needed. Their focus is on telehealth, online diagnosis, and AI-based ethical decision-making to generate flexible, technology-savvy physicians and resolve the problem of bias, depersonalization, and digital divide. The authors recommend policy change, telehealth competency framework, and simulation training as the new international standard in GP education. [5]

Jie Huang, Zhiyan Ding, and Jiaying Li, in the editorial article Global infectious disease surveillance technologies and data sharing protocols, mention the possibility of wastewater surveillance and AI and blockchain-based transparent, real-time exchange of data. They suggest the inclusion of a variety of data flows including mobility patterns, internet trends and wastewater signals, and provide the case studies of ship, Sicily and Burkina Faso. The authors propose to use the method of a pandemic, which is the combination of biotechnology and information technology to create a new system of global surveillance. [6]

Munyaradzi Saruchera et al. discuss in their editorial article titled Climate change, human health, and health systems how the increasing temperatures deteriorate cardiopulmonary health and demand adaptive, sustainable health systems. They also point out such solutions as solar-powered vaccine cold chains in Lebanon and argue about the One Health Economics approach to managing the interconnection between climate, ecosystems, and human health. The authors focus on multidisciplinary policies to enhance interventions and mitigate the economic effects of health-related climate disasters.[7]

Christopher D Golden, Marissa L Childs, and Oladimeji E Mudele, along with other authors, in their paper titled "Climate-smart public health for global health resilience,"nurture the idea of a data-driven model named Climate-Smart Public Health (CSPH) to deal with the health hazards associated with climate change, including heatwaves, droughts, and floods. The paper brings out the role of CSPH in integrating climate and public health systems by using data science and artificial intelligence (AI) to assist in making decisions on health sectors such as infectious diseases and nutrition. The article illustrates that the framework has been applied in Madagascar, where there is a close correlation between climatic and diseases such as malaria and diarrhea. [8]

3. METHODOLOGY

➢ **Dataset and Preprocessing**

The dataset includes climatic and health parameters such as parameters such as State, District, Year, Month, Temperature (degC), Humidity (percent), Rainfall (mm), Disease, ChildrenAffected, EldersAffected, and Population. The sample size is 11,001 records which is large enough to use machine learning methods.

➢ Preprocessing Steps include :

a) Handling Missing Values

● Verified the data of the dataset to check missing, null, or inconsistent data.

• In the case of numerical attributes (e.g., Temperature, Humidity, Rainfall), the replacement of missing values was done by mean or median imputation based on data distribution.

• Where the attributes are categorical (e.g. State, District, Disease) the missing entries were imputed using mode or, in very sparse cases, the rows were just dropped.

b) Normalization

• To avoid biases based on characteristics of the various scales, numerical characteristics were put under a **Min-Max** normalization:

$$X' = \frac{X - X_{\min}}{X_{\max} - X_{\min}}$$

• Example: Temperature (degC), Humidity(percent), and Rainfall (mm) were brought to the range 0 to 1.

• This is to ensure no feature overpowers the model because of its size.

c) Feature Selection

• Characteristics applicable in forecasting health susceptibility as a result of climatic change were found.

• **The independent variables include:** Temperature, Humidity, Rainfall, Month, Year, Population.

• **The Dependent variables include**: Children_Affected, Elders_Affected (indicators of vulnerable populations).

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

(Int Peer Reviewed Journal)

Vol. 05, Issue 09, September 2025, pp : 1125-1130

www.ijprems.com

editor@ijprems.com

e-ISSN : 2583-1062

Impact Factor : 7.001

- The most significant predictors were selected by use of correlation analysis and importance of features (with the help of Random Forest/Chi-square test), eliminating redundancy and enhancing the efficiency of the model.

d) Data Splitting

- The processed dataset was split into training and testing to test the model performance.

- **Split ratio**: 80% training and 20% testing

- Moreover, the k-fold cross-validation (k=5) was implemented to make sure that the model can be generalized and it does not overfit.

4. MACHINE LEARNING MODELS

➢ **Random Forest**

Random Forest is an ensemble learning algorithm, which builds a multitude of decision trees in training and generates the class which is the mode of the classes (classification) or average prediction (regression) of the individual trees. Every tree is trained using random subset of data, and random subset of features, contributing to the less overfitting and more robust. The classification rule of prediction is:

$$\hat{y} = \text{mode}\{h_1(x), h_2(x), \ldots, h_T(x)\}$$

Random Forest was also strong in this project, with its accuracy of approximately 93. It is especially effective since it could not only model non-linear relations between climate factors (temperature, humidity, rainfall) and health outcomes, but it also presents the importance of features, which shows the most significant factors that influence vulnerable populations.

➢ **Support Vector Classifier**

The Support Vector Machine is a supervised learning algorithm that identifies the best hyperplane that best classifies the data points in two categories. It aims at maximizing the distance between the classes, thereby enhancing generalization. The decision function is stipulated as:

$$f(x) = \text{sign}\left(\sum_{i=1}^{n} \alpha_i y_i K(x_i, x) + b\right)$$

In this project, SVM was able to achieve an accuracy of about 88 percent. It has the advantage of estimating both non-linear and complex boundary between climate signal and disease occurrence, thus it is used when what is needed are for example vulnerable districts or the month more prone to an outbreak of the disease. Nonetheless, SVM is more computationally expensive than the Logistic Regression and must be tuned.

➢ **Logistic Regression**

Logistic Regression is a statistical model that is used when the problem is binary classification. It estimates the likelihood of an event (disease present or not) on the basis of the logistic (sigmoid) function. The model is expressed as:

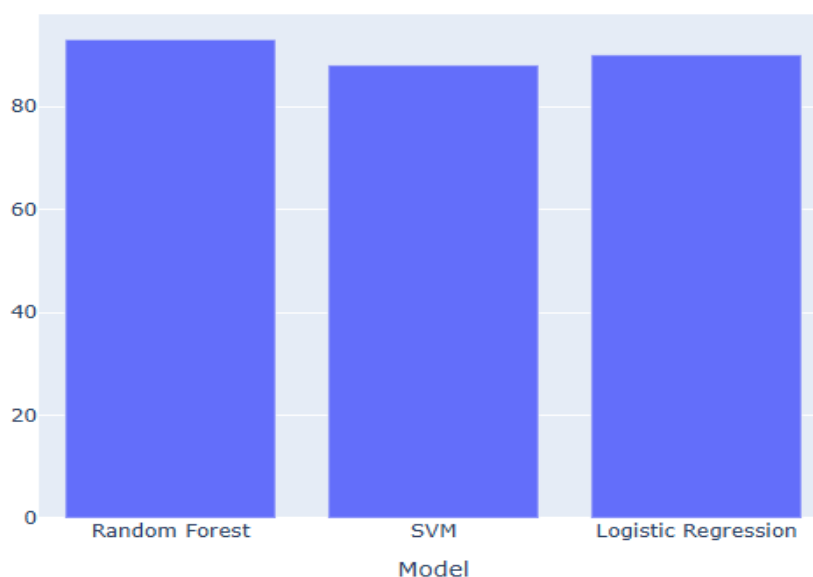$$P(Y = 1|X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n)}}$$

In this study, Logistic Regression was used with the result of approximately 90 percent accuracy. Its key strength is in its interpretability as the coefficients can clearly show the effect that each of the features (temperature, humidity, rainfall, and population) has on the probability of the disease occurrence. This renders the use of the Logistic Regression especially useful to policymakers because it is not only able to provide predictions on the same but also specifies the factors that have the greatest contribution to the risk of health.

5. MODELS RESULT AND DISCUSSION

MODEL	ACCURACY
RANDOM FOREST	93%
SUPPORT VECTOR MACHINE	88%
LOGISTIC REGRSSION	90%

The research transforms the manner in which the vulnerable populations attend health care by altering the approach used by the population in regard to responding to the health problems in order to prevent them. With 93 percent prediction accuracy, warning about the occurrence can be given in advance, resources can be allocated better, and certain intervention can be taken before the outbreak or heatwave has taken place (Science Daily, 2020). Its capability to draw elaborate climate-health relationships gives policy makers viable knowledge. Lastly, it provides an example of a blueprint of resilient health systems to climate change that are scalable and can safeguard the most vulnerable groups.

Model Accuracy Comparison

6. CONCLUSION

As it is shown in the current research, artificial intelligence plays an important role in the study of how climate change affects vulnerable populations. The correlation heatmap indicated that the environmental variables of humidity and rainfall have a high positive relationship, whereas temperature has a weak negative relationship with humidity. Conversely, health-related variables like children affected, old age affected and population size did not show any significant correlation with environmental variables indicating the complex and indirect nature of health effects of climate.

The AI-based health monitoring system created within the scope of the current project can effectively combine environmental and health data to determine possible health dangers depending on the climate. Through the use of machine learning algorithms that include the use of the Random Forest, Support Vector Machine (SVM) and Logistic Regression, the system can produce predictive information and early warning signals. These are not only useful in preventive healthcare approaches, but also serve as useful policy aids to policymakers in formulating climate-turn resilient population health interventions. On the whole, the given project is a prototype that can be used to prove that AI can be successfully utilized to fill the gap between climate change and healthcare monitoring.

7. REFERENCES

[1] Pourroostaei Ardakani, S., Kapogiannis, G., Al-khafajiy, M., & Yu, M. (2025). Editorial: Data analytics in sustainable city planning. Frontiers in Sustainable Cities, 7:1677743. https://doi.org/10.3389/frsc.2025.1677743

[2] Soares Filipe, J. F., Ciliberti, M. G., Attard, G., Ratti, G., & Rocchi, M. S. L. (2025). Editorial: Biosecurity of infectious diseases in veterinary medicine. Frontiers in Veterinary Science, 12:1665683. https://doi.org/10.3389/fvets.2025.1665683.

[3] Engr. Prof. Theophilus Aku Ugah, Associate Professor Cynthia Amaka Obiorah, Prof. Okeke Gerald Ndubuisi, Engr. Prof. Sony Emeka Ali, Dr. Omatseyione Nesiama, Engr. Cletus Onyemhese Agbakhamen; "Emerging Trends and Challenges in Environmental Health and Natural Resource Management"; DOI: 10.5281/zenodo.15608070

[4] Ameya D. Bendre, Pooja Singh, and Swati Jaiswal; "Editorial: Tuberculosis: recent updates in basic research, drug discovery and treatment"; DOI: 10.3389/fmed.2025.1668254

[5] Waseem Jerjes and Azeem Majeed; "Transforming GP training in the UK: the lasting impact of COVID-19 on telehealth and hybrid care models"; DOI: 10.3389/fmed.2025.1595937

[6] Jie Huang, Zhiyan Ding, and Jiaying Li; "Editorial: Global infectious disease surveillance technologies and data sharing protocols"; DOI: 10.3389/fpubh.2025.1676987

[7] Munyaradzi Saruchera, Morenike Oluwatoyin Folayan, Bethuel S. Ngcamu, Walter Musakwa, Margaret Kaseje, and Godfred O. Boateng; "Editorial: Climate change, human health, and health systems"; DOI: 10.3389/frhs.2025.1615206

[8] Mohammad Imam Hasan Reza, Md. Aktar Hossain, and Uddin Shaikh Bokhtear; "Editorial: Recreational forests for co-benefits: conservation, tourism and well-being"; DOI: 10.3389/ffgc.2025.1670540