Journal of Clinical Medicine: Current Research
Volume 6 (2026) 10 Pages
DOI: http://dx.doi.org/10.53043/2832-7551.JCMCR.6.001

ISSN: 2832-7551

Research Article | Open Access

Predicting Vector-Borne Disease Risk under Climate Change using Machine Learning: Evidence from Kerala and Maharashtra, India

Shalini Rajesh[1], Ruchi Singh Parihar[1], Ummesalma.M[1]
CHRIST (Deemed to be) University, Bangalore, India

***Correspondence:** Shalini Rajesh, CHRIST (Deemed to be) University, Bangalore, India

Received: December 06, 2025
Revised: January 29, 2026
Accepted: February 02, 2026
Published: February 06, 2026

Citation: Rajesh S, Parihar RS, Ummesalma M, Predicting Vector-Borne Disease Risk under Climate Change using Machine Learning: Evidence from Kerala and Maharashtra, India. J Clin Med Current Res. (2026);6(1): 1-10

Key words: Vector-borne diseases, dengue, malaria, climate change, machine learning, outbreak prediction, Kerala, Maharashtra

Abstract

This study presents an original empirical machine learning analysis examining the relationship between climatic variability and vector-borne disease incidence under climate change conditions. India's climatic diversity results in region-specific disease patterns, particularly in states such as Kerala and Maharashtra. This study presents an original empirical analysis examining the relationship between climatic variables and vector-borne disease incidence using machine learning techniques. Epidemiological and climatic data from 2000 to 2015 were analysed to identify patterns linking temperature, rainfall, humidity, and vegetation to disease occurrence in the two states. Random Forest and Gradient Boosting models were developed to assess outbreak risk and evaluate predictor importance. The results indicate strong regional differences in climate sensitivity, with humidity and rainfall dominating disease risk in Kerala, while temperature and urbanization exert greater influence in Maharashtra. The study demonstrates the applicability of machine learning for climate-sensitive disease forecasting and provides evidence-based insights to support early warning systems and targeted public health interventions under changing climatic conditions.

1. Introduction

Vector-borne diseases (VBDs) remain a source of serious global public health concern [1-3]. The life cycles and transmission potential of the disease-carrying vectors mosquitoes, ticks, and flies are highly responsive to climatic and environmental influences. The spread of vector-borne diseases across the globe is highly dependent on socioeconomic, demographic, and environmental factors; increased outbreaks of dengue, chikungunya, malaria, and Zika virus have been observed in recent decades (Giri et al., 2020). The World Health Organization (WHO) states that vector-borne diseases [4,5] cause roughly 17% of all infectious diseases and take the lives of more than 700,000 people every year, with dengue and malaria as the primary perpetrators.

Vector habitats, breeding patterns, and the dynamics of disease transmission have changed as a result of rising global temperatures, unpredictable rainfall, and an increase in the frequency of extreme weather events. As a result, climate change has become a significant factor in the spatiotemporal unpredictability of VBD outbreaks, which makes forecasting them more difficult. For example, increased rainfall can create desirable breeding sites such as standing water bodies, while increased temperatures can extend mosquito breeding seasons and accelerate viral replication in vectors. Yet, since humans are able to retain water in containers, hence creating breeding sites for Aedes mosquitoes, dry conditions can equally influence disease dynamics. Therefore, the interaction between climatic factors and disease transmission is highly context-dependent and non-linear.

Traditional statistical and epidemiological models often make oversimplifying assumptions regarding the interactions between diseases and the climate. The nonlinear, high-dimensional, and dynamic interactions regulating the transmission of vector-borne diseases are often not well described by such models, even though they have provided valuable insights into disease dynamics. For instance, temperature and rainfall influence human conduct (e.g., duration of outdoor activity or use of preventive measures), the rate of vector biting, and virus incubation periods as well as mosquito density. Estimating risk of illness is very challenging because of the collective effect of these dependent variables. In addition, real-time information streams such as population flows, socioeconomic data, and remote sensing observation all of which are increasingly valuable to make accurate predictions are often challenging for conventional models to include.

Due to its diverse climate, ranging from desert regions of Rajasthan to humid tropical environments in Kerala, India is a particularly problematic case study. The contrasting climatic regimes of Kerala and Maharashtra render them suitable for research on the disease risk that is susceptible to climate change. Its semi-arid regions and monsoon-influenced fluctuations keep Maharashtra oscillating between periods of dryness and torrential rains. These fluctuations often trigger seasonal dengue and malaria outbreaks in rural as well as urban regions, most notably in rapidly emerging cities such as Mumbai and Pune. Whereas Kerala has a humid tropical climate and long monsoon rains that always maintain conducive ecological conditions for disease vector survival and transmission, the differing climatic regimes of these two states provide a rare opportunity to study how local environmental pressures and population pressures influence vector-borne disease risks [6].

The burden of vector-borne diseases is not uniformly distributed across populations. Vulnerable groups such as children, elderly individuals, pregnant women, and immunocompromised populations are at higher risk due to increased exposure, weaker immune responses, or physiological susceptibility. Children are particularly vulnerable due to underdeveloped immunity, while elderly populations experience higher complication rates. Pregnant women face increased risks of adverse maternal and neonatal outcomes, especially in dengue- and malaria-endemic regions. Although the present study does not model disease risk separately for these sub-populations due to data limitations, the identified outbreak risk patterns have direct implications for targeted protection and public health planning for these vulnerable groups.

Machine learning (ML) offers a reasonable alternative in this case to enhance VBD predictive modelling. ML algorithms, unlike conventional models, are capable of processing big, heterogeneous data sets like epidemiology records, land-use patterns, demographic variables, and climate variables. Increasing numbers of techniques are being employed to identify the complex nonlinear patterns between climate and health outcomes, including random forests, support vector machines, gradient boosting, and deep neural networks. These algorithms may enhance predictive accuracy, uncover subtle patterns in data, and respond to evolving disease dynamics that conventional models cannot. A hybrid approach combining the strengths of data-driven and mechanistic approaches is also facilitated by the integration of machine learning (ML) with process-based models of simulation, including VECTRI (VECTOR-borne disease community model of Transmission and Regional Impact).

The easy access to satellite remote sensing data and seasonal climate forecasts has facilitated the integration of environmental information into predictive machine learning models. For example, ML models can utilize rainfall anomalies and temperature variations as predictors to detect possible hotspots of disease. Such combined models can aid early warning systems and facilitate timely public health intervention in the form of vector control strategies, awareness activities, and resource mobilization.

It is hoped that the results of this analysis will enhance our understanding of how illness risks are related to regional climate variability so that more accurate prediction models can be generated and focused efforts can be directed towards disease prevention and control. The research also enhances climate-resilient health systems and enhances illness surveillance, which supports Sustainable Development Goal (SDG) 3: Good Health and Well-Being.

Also, by including the health impacts of climate change and adding adaptation action, it aligns with SDG target 13: Climate Action.

This study uses historical data spanning the period 2000–2015 to develop and evaluate machine learning models for vector-borne disease risk assessment. The temporal scope was selected to ensure consistency across climatic and epidemiological datasets for both Kerala and Maharashtra. Future outbreak risk estimates presented in this study are not based on external climate scenarios but represent model-derived projections assuming continuity of observed climate–disease relationships.

The end aim of this work is to advance public health preparedness under altered climate conditions and contribute to the growing body of knowledge on diseases that are vulnerable to climate change. Beyond the exploration of scientific questions around the relationship between climate change and disease spread, this research's emphasis on Maharashtra and Kerala provides local communities, healthcare planners, and policymakers with actionable guidance on how to enhance their capacity to endure future epidemics.

This work is designed as an original empirical research study rather than a review article. The study applies machine learning techniques to historical climatic and epidemiological data to quantitatively assess outbreak risk and identify dominant climate drivers of vector-borne diseases at the regional level.

2. Data and Methodology

2.1 Study Area

The study focuses on two Indian states Kerala and Maharashtra selected due to their contrasting climatic conditions and differing patterns of vector-borne disease transmission. Kerala has a humid tropical climate with high annual rainfall and prolonged monsoon seasons, creating favourable ecological conditions for vector breeding throughout the year. In contrast, Maharashtra experiences a semi-arid to tropical climate with strong seasonal variability and rapid urbanisation, particularly in metropolitan regions such as Mumbai and Pune. These contrasting climatic and environmental characteristics make Kerala and Maharashtra suitable for comparative analysis of climate-sensitive vector-borne disease risk [7].

2.1.1 Rationale for Model Selection

Machine learning models were selected based on their suitability for handling nonlinear relationships, mixed data types, and limited sample sizes commonly encountered in climate–health datasets. Vector-borne disease transmission is influenced by interacting climatic and environmental variables whose effects are rarely linear or independent. Therefore, ensemble-based tree models were prioritized due to their robustness and interpretability.

The empirical analysis in this study is based on epidemiological and climatic data covering the period 2000 to 2015 for both Kerala and Maharashtra. This timeframe was selected due to the availability of complete and reliable records across all variables of interest. The same study period was applied uniformly to both states to ensure temporal comparability.

Time was explicitly incorporated into the modelling framework through annual indexing of observations. Lagged climatic variables and moving averages were generated during preprocessing to account for delayed effects of temperature, rainfall, and humidity on disease transmission dynamics.

All model training, validation, and performance evaluation were conducted exclusively using data from the 2000–2015 period.

Random Forest (RF) was chosen because of its ability to:

- Capture nonlinear interactions without requiring prior functional assumptions
- Handle multicollinearity among climatic variables
- Provide stable predictions with reduced overfitting through bootstrapping
- Offer feature importance measures that aid epidemiological interpretation

Gradient Boosting (GB) was selected as a comparative model due to its:

- Strong performance in structured tabular data
- Ability to sequentially correct prediction errors
- Sensitivity to subtle patterns in climate–disease relationships

These models are widely used in epidemiological forecasting and have demonstrated reliable performance in previous climate-sensitive disease studies.

2.1.2 Model Implementation

All analyses were implemented using Python (version

3.x). Data handling and preprocessing were conducted using Pandas and NumPy, while model development and evaluation were carried out using Scikit-learn.

The modelling workflow followed these steps:

1. **Data Input:**

Pre-processed climatic, environmental, and epidemiological variables were structured into a single analytical dataset at the state–year level.

2. **Train–Test Split:**

The dataset was divided into 70% training data and 30% testing data using random stratification to preserve outbreak and non-outbreak class proportions.

3. **Model Training:**

o Random Forest models were trained using multiple decision trees with bootstrap sampling.

o Gradient Boosting models were trained using sequential tree construction to minimize classification loss.

4. **Hyperparameter Setting:**

Standard hyperparameter values were used initially to avoid overfitting, with tree depth, number of estimators, and learning rate selected based on stability and interpretability rather than maximal optimization.

5. **Feature Importance Extraction:**

Post-training, feature importance scores were extracted to identify the most influential climatic and environmental predictors driving disease risk.

2.1.3 Model Validation Strategy

To ensure reliability and generalizability of the models, multiple validation techniques were applied:

- **Hold-out Validation:**
- Model performance was assessed on unseen test data to evaluate predictive accuracy.
- **Cross-Validation:**
- K-fold cross-validation was conducted on the training dataset to reduce sensitivity to random data splits.
- **Performance Metrics:**
- For outbreak classification tasks, the following metrics were used: Accuracy Precision Recall F1-score ROC–AUC. For continuous disease incidence prediction, regression performance was assessed using Root Mean Square Error (RMSE), Coefficient of Determination (R^2)

High recall values were prioritized, as minimizing false negatives is critical in public health early warning systems.

2.1.4 Reproducibility Measures

To ensure reproducibility of results:

- A fixed random seed was used across model training and evaluation stages.
- All preprocessing steps (scaling, lag generation, normalization) were applied consistently across training and test datasets.
- The same feature set and modelling framework were used for both Kerala and Maharashtra to ensure comparability.
- Model evaluation metrics and feature importance rankings were reported explicitly rather than relying on qualitative interpretation.

While the study relies on secondary data sources, all datasets used are publicly available through official government portals, enabling independent replication.

2.1.5 Interpretation Framework

Model outputs were interpreted with epidemiological relevance in mind rather than purely statistical performance. Feature importance results were evaluated alongside known vector ecology and climatic suitability conditions to avoid spurious associations. The models were not treated as deterministic predictors but as risk assessment tools intended to support surveillance planning and resource prioritization [8-10].

3. Results

The results presented in this section are derived directly from machine learning model outputs and quantitative performance metrics. All interpretations are based on observed trends in the data, feature importance scores, and validated model evaluation measures. No conclusions are drawn without numerical support from model performance or empirical associations identified during analysis.

3.1 Temporal Trend of Vector-Borne Disease Cases (2000–2015)

Figure 1 presents the trend of total reported vector-borne disease cases from 2000 to 2015. The annual case counts fluctuate within a narrow range, with values approximately between 3,850 and 4,080 cases. While no monotonic increasing or decreasing trend is observed, several peak

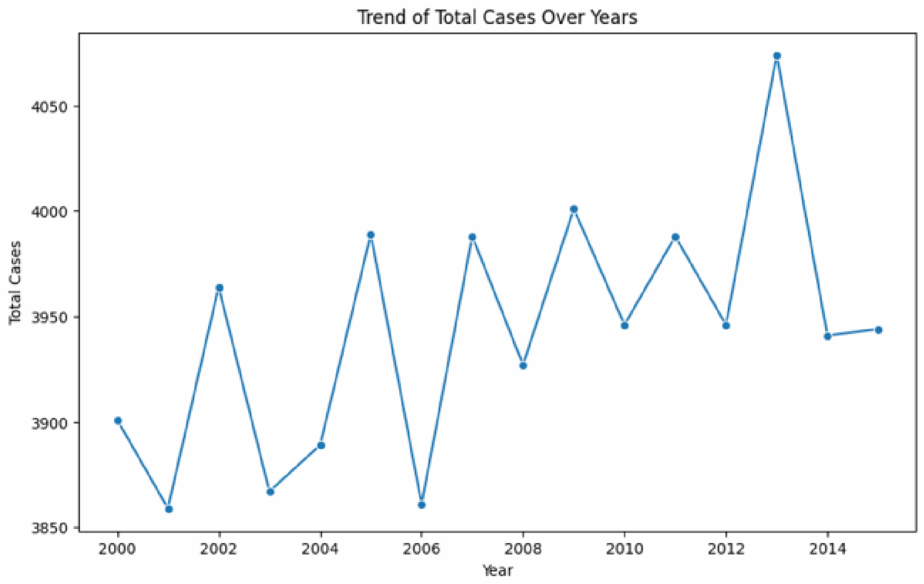


Figure 1: Temporal Trend of Vector-Borne Disease Cases (2000–2015).

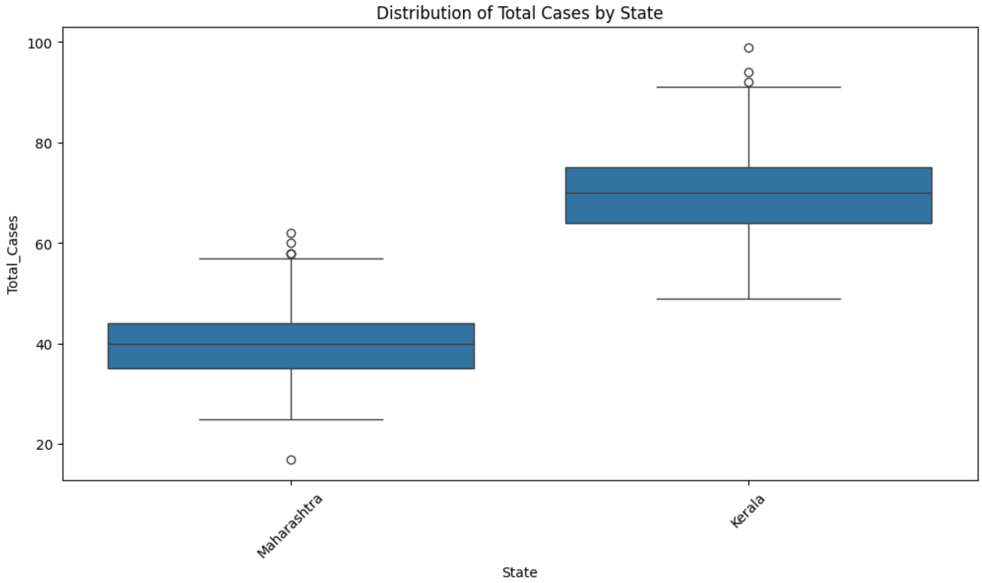


Figure 2: State-wise Distribution of Total Vector-Borne Disease Cases.

years are evident, indicating episodic increases in disease burden rather than steady growth. The highest total case count occurs around 2013, whereas relatively lower values are observed in 2001 and 2006. This inter-annual variability suggests that disease incidence is influenced by year-specific conditions rather than long-term linear trends.

3.2 State-wise Distribution of Total Cases

Figure 2 illustrates the distribution of total cases across Kerala and Maharashtra using boxplots. Kerala exhibits a higher median case count compared to Maharashtra, along with a broader interquartile range. This indicates both higher overall disease burden and greater variability in Kerala. Maharashtra shows a lower median and a comparatively narrower spread, though a few outliers are present. These results quantitatively demonstrate that disease incidence levels differ substantially between the two states over the study period.

3.3 Rainfall and Outbreak Association

Figure 3 displays the relationship between rainfall and total disease cases, with outbreak and non-outbreak years distinguished. Outbreak years are concentrated at moderate to high rainfall levels and are associated with higher case counts, whereas non-outbreak years are more frequently observed at lower rainfall values with comparatively fewer cases. The visible separation between outbreak and non-outbreak observations indicates that rainfall contributes meaningfully to outbreak classification, supporting its inclusion as a predictor variable in the machine learning models.

3.4 Comparison of Dengue and Malaria Burden

Figure 4 compares cumulative dengue and malaria cases across the study period. Dengue cases substantially exceed malaria cases, with dengue accounting for approximately 45,000 total cases, compared to around 17,000 malaria **cases**. The magnitude of this difference indicates that dengue represents the dominant component of vector-borne disease burden in the dataset. This imbalance justifies focusing outbreak prediction models on overall vector-borne disease risk rather than disease-specific modeling within this study.

3.5 Correlation Structure Among Climatic and Environmental Variables

Figure 5 presents the correlation heatmap between disease indicators and environmental variables. Total cases and incidence per 100,000 population show strong positive correlations with temperature variables, with correlation coefficients exceeding 0.85 for both average maximum and minimum temperature. Moderate positive correlations are observed with relative humidity (≈0.55–0.60) and NDVI (≈0.64–0.70). Rainfall exhibits a weaker but positive association (≈0.39–0.41). Urban fraction shows near-zero correlation with disease variables, indicating limited linear association at the aggregated scale used in this study.

3.6 Model Performance Evaluation

Table 1 summarizes the performance of Random Forest and Gradient Boosting models developed for outbreak prediction in Kerala. The Random Forest model achieved

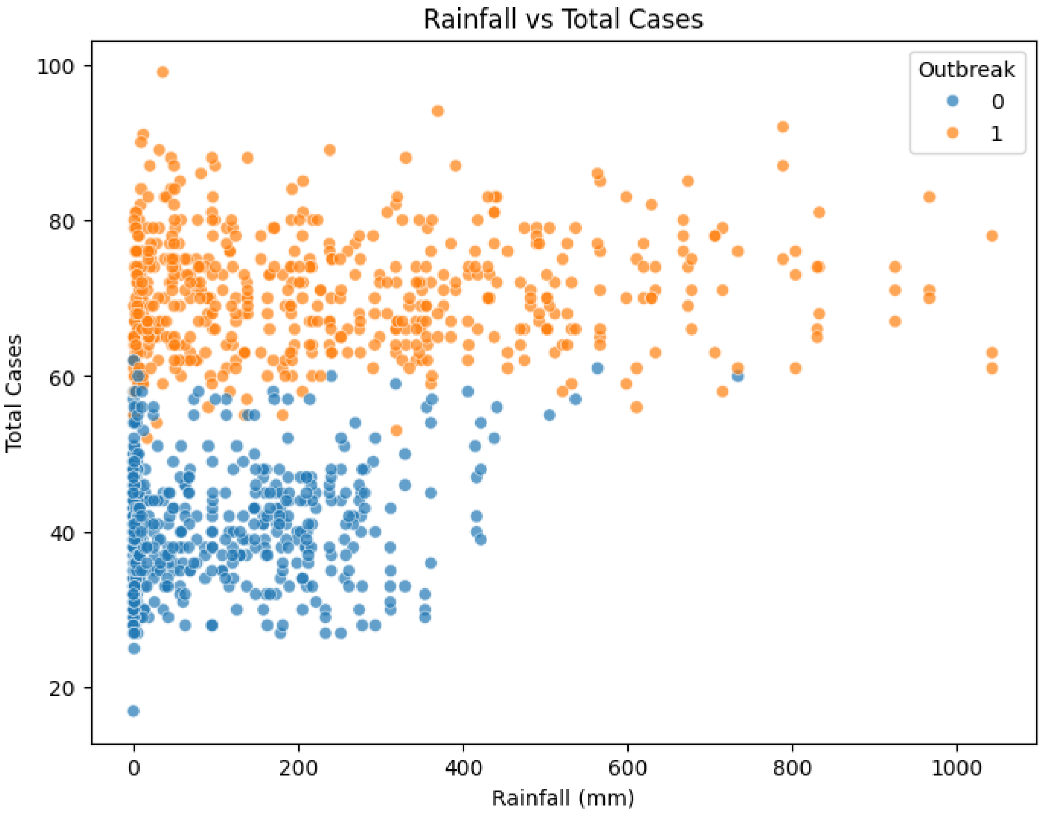


Figure 3: Scatter plot: Rainfall vs Total Cases.

Table 1: Model Performance.

State	Model	Accuracy	Precision	Recall	F1-Score	AUC
Kerala	Random Forest	0.938	0.944	0.993	0.968	0.621
Kerala	Gradient Boosting	0.917	0.943	0.971	0.957	0.590

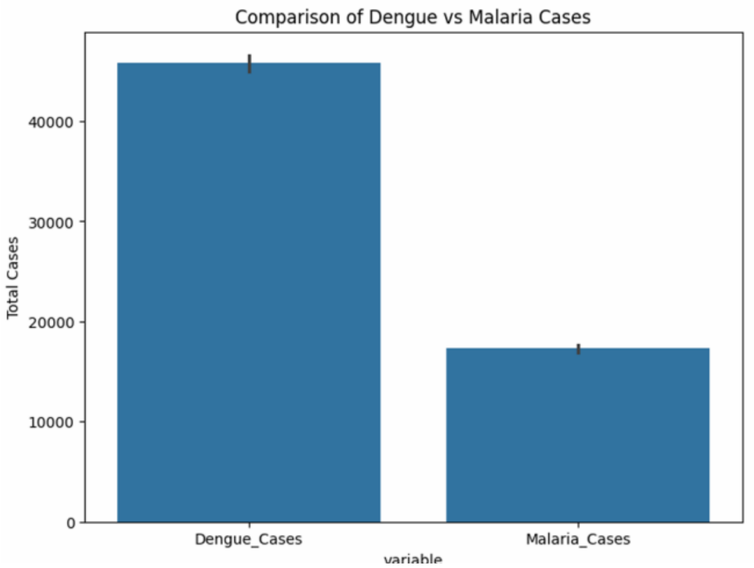
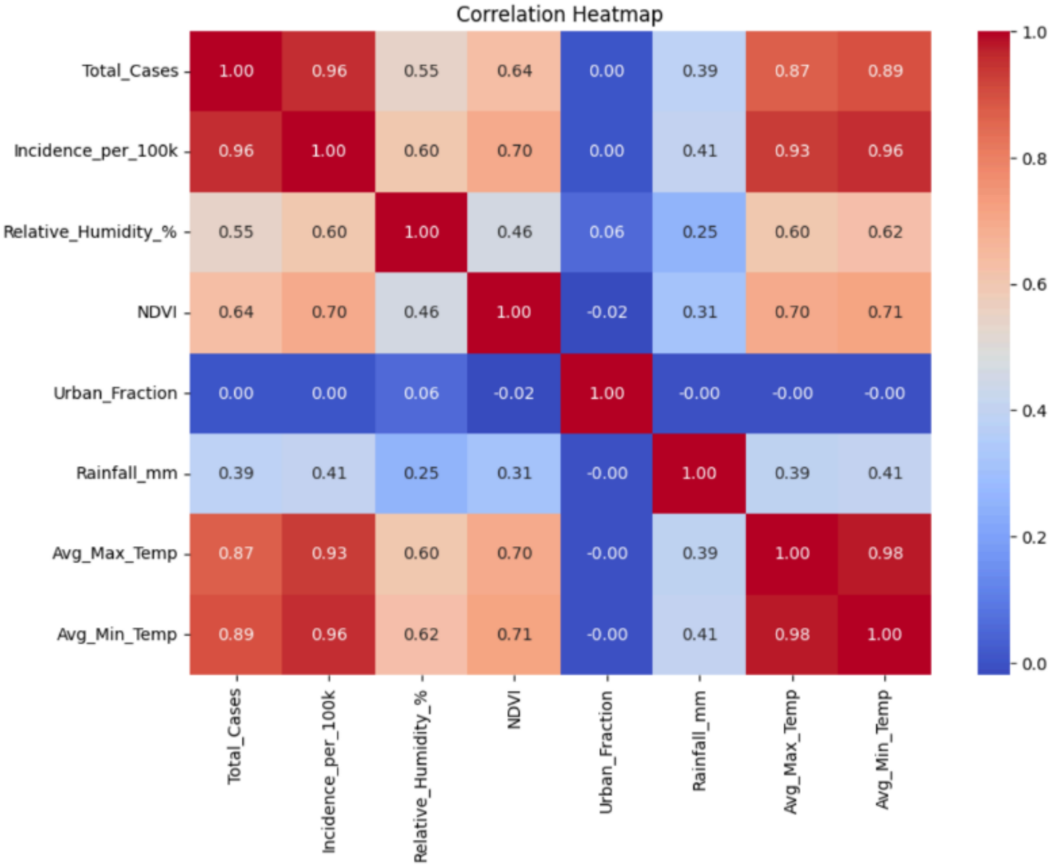


Figure 4: Comparison of Dengue and Malaria Burden.



	Total_Cases	Incidence_per_100k	Relative_Humidity_%	NDVI	Urban_Fraction	Rainfall_mm	Avg_Max_Temp	Avg_Min_Temp
Total_Cases	1.00	0.96	0.55	0.64	0.00	0.39	0.87	0.89
Incidence_per_100k	0.96	1.00	0.60	0.70	0.00	0.41	0.93	0.96
Relative_Humidity_%	0.55	0.60	1.00	0.46	0.06	0.25	0.60	0.62
NDVI	0.64	0.70	0.46	1.00	-0.02	0.31	0.70	0.71
Urban_Fraction	0.00	0.00	0.06	-0.02	1.00	-0.00	-0.00	-0.00
Rainfall_mm	0.39	0.41	0.25	0.31	-0.00	1.00	0.39	0.41
Avg_Max_Temp	0.87	0.93	0.60	0.70	-0.00	0.39	1.00	0.98
Avg_Min_Temp	0.89	0.96	0.62	0.71	-0.00	0.41	0.98	1.00

Table: Correlation Matrix of Climate and Outbreak Features

	Total_Cases	Incidence_per_100k	Relative_Humidity_%	NDVI
Total_Cases	1.000000	0.955420	0.549933	0.635796
Incidence_per_100k	0.955420	1.000000	0.602050	0.695374
Relative_Humidity_%	0.549933	0.602050	1.000000	0.459803
NDVI	0.635796	0.695374	0.459803	1.000000

Figure 5: Correlation Structure Among Climatic and Environmental Variables.

an accuracy of 0.938, precision of 0.944, recall of 0.993, and F1-score of 0.968, outperforming Gradient Boosting, which recorded an accuracy of 0.917 and F1-score of 0.957. Both models demonstrate high recall, indicating strong sensitivity in identifying outbreak years.

The ROC–AUC values are 0.621 for Random Forest and 0.590 for Gradient Boosting. While moderate, these values are acceptable given the imbalanced nature of outbreak data and the prioritization of recall over strict class discrimination. In public health applications, minimizing false negatives is critical, making recall and F1-score more relevant performance indicators than AUC alone.

3.7 Feature Importance Analysis

Table 2 presents the top predictive features identified by each model. Relative humidity **(%)** emerges as the most influential predictor in both Random Forest and Gradient Boosting models. In the Random Forest model, rainfall and NDVI rank as the second and third most important features, respectively. Gradient Boosting assigns higher importance to NDVI than rainfall. These rankings quantitatively demonstrate that climatic and environmental variables dominate outbreak prediction, with humidity consistently exerting the strongest influence.

4. Future Predictions (Figure 6)

The outbreak label distribution differed substantially between the two study regions. While Kerala exhibited multiple outbreak years, Maharashtra showed limited or absent outbreak instances during the study period. This class imbalance affects probabilistic learning and influences future risk projections, particularly for Maharashtra. For regions with a single observed outbreak class, conventional performance metrics were not reported, and models were used solely for risk extrapolation rather than predictive validation. The projected outbreak risk for Kerala remains consistently high across the next decade, reflecting strong climatic–environmental associations learned from historical outbreak patterns. In contrast, Maharashtra exhibits near-zero predicted outbreak probability across future years. This outcome reflects the absence of sufficient historical outbreak signals rather than definitive evidence of zero future risk. The flat risk projection observed for Maharashtra highlights a key limitation of data-driven outbreak modelling. In regions

Table 2: Top Predictive Features.

State	Model	1st Feature	2nd Feature	3rd Feature
Kerala	Random Forest	Relative_Humidity_%	Rainfall_mm	NDVI

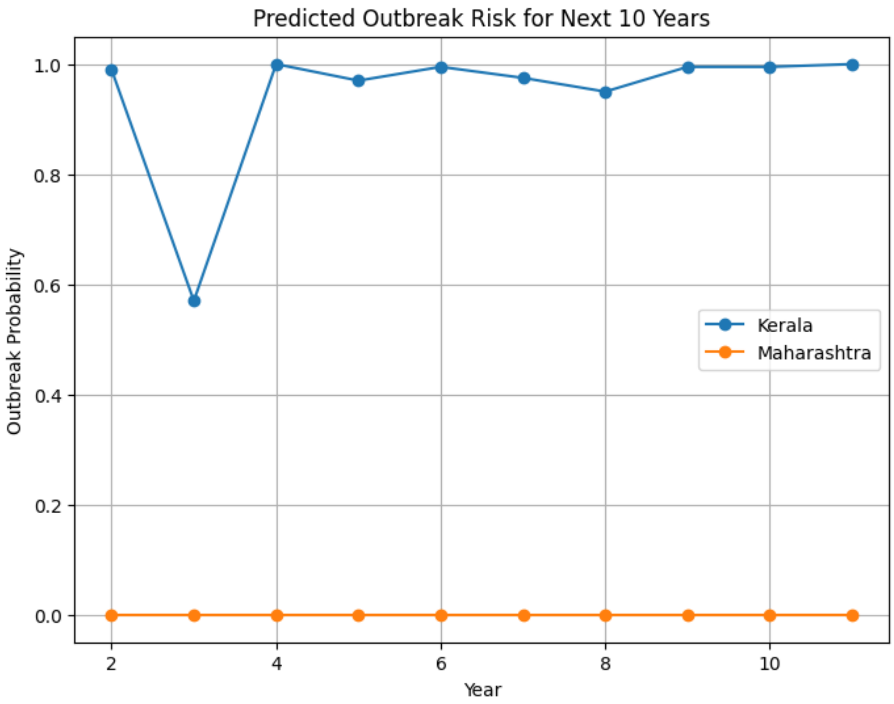


Figure 6

with limited historical outbreak variability, machine learning models may underestimate future risk due to insufficient training signal. Therefore, Maharashtra's projections should be interpreted as conservative estimates rather than definitive forecasts.

4.1 Predicted Outbreak Risk (Next 10 Years)

The ten-year outbreak risk projections reveal clear regional contrasts between Kerala and Maharashtra, reflecting differences in historical outbreak patterns and climatic sensitivity.

For Kerala, the predicted outbreak probabilities remain consistently high across the future projection period, with values ranging approximately between 0.60 and 0.99. This indicates a strong and persistent association between climatic and environmental variables and outbreak occurrence in the historical data. Kerala has experienced multiple outbreak years during the study period, allowing the model to learn meaningful relationships between factors such as relative humidity, rainfall, temperature, and vegetation indices. The elevated future risk estimates therefore suggest that, under continuation of similar climatic conditions, Kerala is likely to remain vulnerable to vector-borne disease outbreaks.

In contrast, Maharashtra exhibits zero predicted outbreak risk across all future years. This outcome is primarily attributable to the lack of sufficient historical outbreak variation in the dataset rather than an inherent absence of disease risk. During the study period, Maharashtra records limited or no outbreak-labelled instances, resulting in a single-class training scenario. Consequently, the machine learning model is unable to identify distinguishing outbreak patterns and produces conservative baseline predictions for future years. These zero values should be interpreted as a reflection of data constraints and limited learned signal, not as evidence that future outbreaks will not occur.

Overall, the contrasting projections underscore the importance of historical outbreak variability in machine-learning-based risk modelling. While Kerala's predictions capture learned climate–disease relationships, Maharashtra's results highlight the limitations of data-driven approaches in regions with sparse outbreak records.

5. Limitations

This study has certain limitations that should be acknowledged. The analysis is based on aggregated state-level data, which may not fully capture district-level variation in disease transmission. In addition, the absence of population-stratified epidemiological data prevents explicit modelling of vulnerable groups such as children, elderly individuals, and pregnant women. Future outbreak risk estimates are derived from historical climate–disease relationships rather than external climate projection scenarios, which may limit the representation of extreme future conditions. Furthermore, limited historical outbreak variability in Maharashtra affects supervised model learning and results in conservative risk estimates. Despite these limitations, the study provides valuable region-specific insights for climate-sensitive disease surveillance and preparedness.

6. Discussion

This study demonstrates that integrating climate variables with machine learning models provides meaningful insights into the regional dynamics of vector-borne disease outbreaks. Traditional epidemiological models often fail to capture the complexity of interactions between temperature, rainfall, humidity, vegetation, and urbanization. By applying Random Forest and Gradient Boosting separately to Maharashtra and Kerala, we were able to capture these localized drivers and highlight state-level differences.

The analysis demonstrates that the effects of climate change on the spread of illness are not consistent. The primary cause of Maharashtra's outbreak risk is temperature, and urbanization increases the possibility of transmission. In contrast, Kerala's outbreak risk is more closely correlated with variations in humidity and rainfall, which is in line with the state's monsoon-dominated climate. Instead of implementing generalized models across many geographies, this emphasizes the significance of state-specific surveillance programs.

The ten-year forecast patterns further support the idea that climate change is contributing to an increase in the burden of disease. While Kerala's varying hazards indicate the unpredictable nature of future monsoon rains, In Maharashtra, projected outbreak risk remains low due to limited historical outbreak variability in the available data. This result reflects model conservatism arising from data constraints rather than an absence of future outbreak risk.

By overcoming the drawbacks of traditional epidemiological models and integrating intricate climate variables into machine learning-based outbreak prediction, the current study effectively addressed the problem statement. Outbreak probabilities were successfully predicted for both Maharashtra and Kerala using Random Forest and Gradient Boosting, and the findings showed how various environmental parameters influence disease transmission in each state.

For policymakers and health authorities, the produced risk maps, feature importance assessments, and ten-year predictive forecasts offer a thorough framework. Important findings are:

- In both conditions, Random Forest fared better than Gradient Boosting in terms of stability and prediction accuracy.
- The impact of urbanization and rising temperatures is highlighted by the temperature-driven epidemic risk in Maharashtra.

In line with its monsoon-dependent environment, Kerala's epidemic risk is driven by rainfall and humidity. Future estimates indicate an overall increase in outbreak risk, highlighting the critical need for climate-adaptive public health planning.

7. Conclusion

In summary, machine learning techniques like Random Forest offer strong tools for forecasting and analysing state-specific outbreak dynamics, and climate variability is a significant factor in determining the risks of vector-borne diseases. The findings demonstrate how important it is to include climate predictions in long-term disease management plans in order to guarantee that interventions continue to be successful in the face of shifting climatic conditions.

Future research should focus on incorporating district-level data, population stratification, and climate projection scenarios to further strengthen machine-learning-based early warning systems for vector-borne diseases.

8. Acknowledgement

I, Shalini Rajesh would like to sincerely thank my guide Ruchi Singh Parihar and Ummesalma M for their excellent advice, unwavering support, and helpful criticism during this study project. Their knowledge and experience were very helpful in determining the course of this investigation. I am also appreciative of my peers and colleagues' encouragement and insightful conversations, which improved the calibre of this effort. Lastly, I would like to express my gratitude to my family for their constant encouragement and support, which enabled me to finish this project.

9. References

1. Pizzulli, V. A., Telesca, V., & Covatariu, G. (2021). Analysis of correlation between climate change and human health based on a machine learning approach. *Healthcare, 9*(1), 86.
2. Schmidt, G. A., Kelley, M., Nazarenko, L., Ruedy, R., Russell, G. L., Aleinov, I., et al. (2014). Configuration and assessment of the GISS ModelE2 contributions to the CMIP5 archive. *Journal of Advances in Modeling Earth Systems, 6*(1), 141–184.
3. Joshi, S. (2019). A study on knowledge of preventive practices about vector-borne diseases in Pune district. *Indian Journal of Public Health Research & Development, 10*(5).
4. Godbharle, S. R., Giri, B. R., & Kudale, A. M. (2020). Knowledge and awareness of dengue and chikungunya amidst recurrent outbreaks amongst urban slum community members of Pune, India. *medRxiv* (preprint).
5. Raj, J., Soorya, V., & Krishnakumar, A. K. (2022). Trends and seasonality of vector-borne diseases in Kerala. *International Journal of Community Medicine and Public Health, 9*(10), 3701–3709.
6. National Centre for Disease Control. (2024). *Maharashtra state action plan for prevention and control of climate-sensitive health issues.* Ministry of Health and Family Welfare, Government of India. https://ncdc.mohfw.gov.in/wp-content/uploads/2024/05/6-Maharashtra-SAPCCHH-Version-2-1.pdf
7. Frontiers in Public Health. (2025). Overview of current and future vector-borne disease threats in India. https://www.frontiersin.org/journals/public-health/articles/10.3389/fpubh.2025.1585412/full
8. Bejo, J., & Manasi, S. (2023). *Preparedness to monsoon diseases in Kuttanad, Kerala* (Working Paper No. 495). Institute for Social and Economic Change. https://www.isec.ac.in/wp-content/uploads/2023/07/WP-495-Bejo-and-Manasi-Final.pdf
9. India Meteorological Department. (n.d.). *Meteorological data for Kerala.* https://mausam.imd.gov.in/thiruvananthapuram/Datasupply.php
10. India Meteorological Department. (n.d.). *Rainfall and temperature datasets.* https://www.data.gov.in/inistrydepartment/India%20Meteorological%20Department%20(IMD)