International Journal of Science, Strategic Management and Technology
Volume **02** Issue **03** March-**2026** | ISSN: **3108-1762 (Online)** | Impact Factor: 3.8
An International, Peer-Reviewed, Open Access Scholarly Journal Indexed in recognized academic databases

Machine Learning-Based Early Warning System for Predicting Weather-Driven Disease Outbreaks and Public Health Risk Assessment using Climate Parameters

[0]**Bhavikkumar Patel**, School of Computer Science Engineering & Technology, ITM SLS Baroda University, Vadodara, Gujarat, India

[0]bhavikkumar.patel@itmbu.ac.in

[1]**Prof. Raju Nakum** and [2]**Prof. Sunil Panchal**, School of Computer Science Engineering & Technology, ITM SLS Baroda University, Vadodara, Gujarat, India

[1]Raju.nakum@itmbu.ac.in [2]Sunil.panchal@itmbu.ac.in

https://doi.org/10.55041/ijsmt.v2i3.367

Cite this Article: Patel, B. (2026). Machine Learning-Based Early Warning System for Predicting Weather-Driven Disease Outbreaks and Public Health Risk Assessment using Climate Parameters. International Journal of Science, Strategic Management and Technology, *02*(03). https://doi.org/10.55041/ijsmt.v2i3.367

Abstract—Climate variability and shifting atmospheric patterns have emerged as critical determinants of vector-borne and waterborne disease incidence across tropical and subtropical regions. This paper presents a comprehensive machine learning- based early warning framework that integrates multi-source climate parameters—including temperature, relative humidity, rainfall intensity, wind velocity, and atmospheric pressure gradients—with epidemiological surveillance records to forecast disease outbreak risk at district and sub-district levels. The proposed system employs an ensemble architecture combining Random Forest, Gradient Boosting Machine (GBM), Long Short-Term Memory (LSTM) networks, and an Attention-Enhanced Convolutional Neural Network (ACNN) to capture both spatial heterogeneity and temporal dependencies inherent in climate- disease relationships. Training data spans seventeen years of merged meteorological observations sourced from the India Meteorological Department (IMD) and ISRO's MOSDAC satellite telemetry, cross-referenced with disease incidence records from the Integrated Disease Surveillance Programme (IDSP). Experimental results demonstrate an overall outbreak prediction accuracy of 93.7%, area under the receiver operating characteristic curve (AUC-ROC) of 0.96, and a mean absolute error of1.4 cases per 10,000 population for weekly incidence estimation. The framework incorporates an automated alert generation module that classifies health risk into four tiers—green, yellow, orange, and red—enabling district health officers to pre- position medical resources and initiate preventive interventions up to three weeks in advance of a projected epidemic threshold crossing. This work contributes a replicable, data-driven methodology applicable to resource-constrained public health systems in developing nations.

Keywords—machine learning; early warning system; disease outbreak prediction; climate parameters; LSTM; random forest; public health surveillance; ISRO MOSDAC; epidemiology; vector-borne diseases

I. INTRODUCTION

Over the past four decades, the relationship between meteorological fluctuations and the temporal clustering of infectious diseases has attracted sustained attention from both the epidemiological and computational communities. Diseases such as dengue fever, malaria, cholera, leptospirosis, and Japanese encephalitis are highly sensitive to ambient temperature, precipitation regimes, and relative humidity because these variables directly regulate vector breeding cycles, pathogen survival rates, and human exposure behaviors [1]. India alone recorded approximately 2.4 million dengue cases in a single year, with outbreak hotspots consistently correlating with monsoon onset and anomalous rainfall distribution, underscoring the urgency of anticipatory surveillance mechanisms [2].

Conventional disease surveillance systems in India and comparable low- and middle-income countries (LMICs) operate predominantly on passive reporting chains, whereby confirmed case counts reach national databases days or weeks after local transmission has already amplified. This temporal lag renders reactive containment strategies—mass drug administration, vector control campaigns, emergency hospitalization—far less cost-effective than if deployed proactively. An early warning system (EWS) capable of estimating outbreak probability three to four weeks ahead of epidemic threshold crossing would fundamentally alter public health resource allocation strategies [3].

The confluence of three enabling conditions has made data-driven early warning architecturally feasible at scale. First, satellite remote sensing operated by the Indian Space Research Organisation (ISRO) now provides continuous, high-resolution meteorological data products through the MOSDAC platform and the INSAT-3D/3DR series, offering gridded temperature, outgoing longwave radiation, sea surface temperature anomaly, and precipitation estimates at sub-daily cadences across the Indian subcontinent [4]. Second, the proliferation of supervised and deep learning algorithms—particularly ensemble tree methods and recurrent neural architectures—has demonstrated unprecedented capacity to model nonlinear, lagged, and spatially heterogeneous environment-disease interactions that classical statistical models fail to capture. Third, the digitization

of district-level disease incidence records through IDSP has created a longitudinal epidemiological corpus spanning over two decades, enabling model training at epidemiologically meaningful spatial resolutions [5].

This paper makes the following original contributions to the field: (i) a novel multi-model ensemble framework that fuses gradient-boosted trees with deep sequential learners optimized specifically for climate-lagged disease prediction; (ii) a dynamic risk tier classification engine that translates probabilistic outbreak forecasts into actionable public health alerts; (iii) systematic feature importance analysis across nine climate variables to elucidate which atmospheric drivers carry the highest predictive signal for each disease category; and (iv) a prospective validation study covering three Indian states over a 24-month holdout period that demonstrates operational readiness of the proposed system.

The remainder of this paper is organized as follows. Section II surveys related work in climate-driven disease modeling and machine learning-based epidemic forecasting. Section III describes the data sources, preprocessing pipeline, and feature engineering methodology. Section IV presents the architecture of the proposed ensemble early warning system. Section V details experimental setup and comparative evaluation. Section VI analyzes results and discusses limitations. Section VII concludes with directions for future research.

II. LITERATURE REVIEW

The domain of climate-sensitive disease modeling encompasses a rich body of interdisciplinary literature drawing from tropical medicine, atmospheric science, spatial statistics, and, more recently, computational intelligence. This section situates the present work within that body of knowledge.

A. Statistical and Epidemiological Approaches

Earlier quantitative approaches to linking climate signals with disease incidence relied heavily on generalized linear models (GLMs), distributed lag non-linear models (DLNMs), and autoregressive integrated moving average (ARIMA) frameworks. Bhatt et al. [6] mapped the global distribution of dengue fever using a boosted regression tree model conditioned on bioclimatic covariates, estimating 390 million dengue infections annually. Their work established that temperature and precipitation jointly determine the boundaries of dengue-permissive ecological niches, a finding later extended to the Indian context by researchers at the Indian Institute of Technology Bombay, who demonstrated that district-level dengue incidence follows a two-to-six week lag behind cumulative rainfall anomaly [7]. Similarly, researchers at the National Institute of Malaria Research (NIMR) and IIT Delhi applied negative binomial regression to IDSP malaria records, identifying minimum temperature, relative humidity, and normalized difference vegetation index (NDVI) as the three most parsimonious predictors of monthly case counts across endemic districts in Jharkhand and Odisha [8].

B. Machine Learning in Epidemic Surveillance

The adoption of machine learning methods has substantially expanded the predictive envelope of climate-disease models. Researchers at IIT Kharagpur demonstrated that a Random Forest classifier trained on lagged climate variables could correctly anticipate dengue outbreak weeks in six out of eight study districts in West Bengal with an F1-score exceeding 0.81, outperforming logistic regression baselines by a margin of 14 percentage points [9]. Gradient Boosted Decision Trees (GBDTs) have been applied to malaria forecasting in Rajasthan by a team at NIT Warangal, achieving mean absolute percentage error (MAPE) values below 12% on 30-day ahead forecasts—a performance level adequate for triggering vector control logistics [10].

Deep learning architectures have emerged as particularly powerful tools for modeling temporal dependencies in epidemiological data. LSTM networks, introduced by Hochreiter and Schmidhuber, are structurally suited to capturing the multi-week lag structures characteristic of climate-disease relationships [11]. A joint study by researchers from IIT Madras and the Tamil Nadu Health Department demonstrated that an LSTM trained on historical dengue case counts and meteorological records from the IMD's automatic weather stations achieved a 14-day forecast horizon accuracy of 87.3%, attributing the performance gain primarily to the network's ability to retain seasonal memory across monsoon cycles [12].

C. Satellite Remote Sensing for Health Applications

ISRO's contributions to health surveillance through remote sensing have been growing steadily. The MOSDAC (Meteorological and Oceanographic Satellite Data Archival Centre) platform maintained by the Space Applications Centre (SAC), Ahmedabad, provides operationally validated gridded meteorological products from INSAT-3D and INSAT-3DR at 4 km spatial resolution [4]. Researchers at SAC demonstrated the utility of INSAT-derived land surface temperature and outgoing

longwave radiation data for identifying thermally anomalous zones that correlate with vector breeding acceleration [13]. The BAPS (Bochasanwasi Akshar Purushottam Sanstha) social welfare network has also collaborated with state health departments in Gujarat to implement community-level dengue awareness campaigns triggered by satellite-derived rainfall alerts, illustrating a practical application of space-based data in grassroots public health [14].

D. Risk Stratification and Alert Systems

Beyond point forecasts of case counts, operationally useful EWS architectures must translate probabilistic outputs into tiered risk classifications amenable to non-technical decision-makers. The WHO's Health Emergency and Disaster Risk Management framework recommends traffic-light alert schemes where each tier triggers a predefined set of preparedness actions. A study at NIT Trichy operationalized such a framework for cholera along the Cauvery river basin, demonstrating that four-tier risk classification with weekly update cycles reduced average emergency response latency by 11 days compared to passive surveillance baselines [15]. The present work extends this concept by integrating multi-disease, multi-parameter climate inputs within a unified probabilistic risk engine.

International Journal of Science, Strategic Management and Technology
Volume 02 Issue 03 March-2026 | ISSN: 3108-1762 (Online) | Impact Factor: 3.8
An International, Peer-Reviewed, Open Access Scholarly Journal Indexed in recognized academic databases

III. DATA SOURCES AND PREPROCESSING

A. Epidemiological Data

Weekly disease incidence records for dengue, malaria, leptospirosis, and acute diarrheal disease (ADD) were acquired from the Integrated Disease Surveillance Programme (IDSP) portal covering 112 districts across three Indian states—Gujarat, Rajasthan, and Odisha—for the period January 2006 to December 2022. The dataset comprised approximately 196,000 district- week observations after excluding records with greater than 15% missingness in any single quarter. Outbreak events were defined operationally as weeks in which reported incidence exceeded the 95th percentile of historical case distribution for that district and calendar week, following the approach standardized by the National Centre for Disease Control [2].

B. Meteorological and Satellite Data

Climate variables were assembled from three complementary sources. First, ground-based meteorological observations including maximum and minimum temperature, relative humidity, wind speed, and pan evaporation were obtained from 87 India Meteorological Department (IMD) stations within the study geography at daily temporal resolution. Second, gridded precipitation data at 0.25° spatial resolution were sourced from IMD's APHRODITE-merged gridded rainfall product. Third, satellite-derived land surface temperature, outgoing longwave radiation, and aerosol optical depth fields were extracted from ISRO's MOSDAC archive, specifically the INSAT-3D L2B product suite, at 6-hourly intervals with subsequent daily aggregation [4]. An additional layer of vegetation condition, represented by 16-day composite NDVI derived from ISRO Resourcesat-2 LISS-III imagery, was appended as an indirect proxy for standing water and larval habitat availability [13].

C. Feature Engineering

A total of 47 predictor features were constructed from nine base climate variables. Lagged features were generated at 1-, 2-, 3-, 4-, 6-, and 8-week delays for each meteorological variable, reflecting the incubation and extrinsic incubation periods relevant to the study diseases. Rolling statistical summaries (mean, standard deviation, minimum, maximum) over 4-week and 8-week windows were computed to capture accumulation effects. A monsoon phase indicator—encoding pre-monsoon, active monsoon, withdrawal, and post-monsoon periods—was included as a categorical covariate. Additionally, three interaction terms representing temperature-humidity coupling, rainfall-NDVI co-evolution, and sea surface temperature anomaly- precipitation linkage were derived based on prior epidemiological literature [7][8].

D. Data Integration and Quality Control

Temporal alignment across datasets with differing native cadences was achieved through cubic spline interpolation for sub-weekly to weekly upscaling and district-level spatial interpolation using inverse distance weighting (IDW) for point observations to areal district averages. Missing climate data arising from instrument outages were imputed using a k-nearest- neighbor algorithm (k=5) conditioned on climatologically similar stations. The final integrated dataset was split chronologically: years 2006–2019 formed the training corpus (approximately 75%), 2020 constituted a validation set for hyperparameter optimization (approximately 7%), and 2021–2022 served as an independent prospective test set (approximately 18%). Stratified sampling was not applied in order to preserve the temporal autocorrelation structure essential for sequential models.

IV. PROPOSED METHODOLOGY

A. Overall System Architecture

The proposed Early Warning System (EWS) adopts a layered ensemble architecture consisting of four functional stages: (1) data ingestion and preprocessing pipeline, (2) base learner prediction generation, (3) meta-learner ensemble fusion, and

(4) risk tier classification and alert dissemination. The system is designed to execute weekly inference cycles, ingesting the latest seven days of meteorological observations and satellite data updates to produce district-level outbreak probability estimates and corresponding risk tier assignments for the subsequent one-to-three week horizon.

B. Base Learner Models

Four base learner models form the predictive core of the ensemble. The first model is a Random Forest (RF) classifier comprising 500 decision trees trained with a maximum depth of 25, utilizing Gini impurity as the splitting criterion and bootstrapped subsampling with replacement. Random Forest was selected for its intrinsic resistance to overfitting, interpretable feature importance scores, and robust performance on tabular climate data with mixed numeric and categorical features [9].

The second base learner is a Gradient Boosting Machine (GBM) implemented via XGBoost with 800 estimators, a learning rate of 0.05, and a maximum tree depth of 6. Regularization parameters L1 (α=0.1) and L2 (λ=1.0) were tuned via 5-fold time- series cross-validation to prevent overfitting to seasonal patterns present in the training corpus [10].

The third model is a Long Short-Term Memory (LSTM) network configured with two stacked LSTM layers of 128 and 64 units respectively, followed by two fully connected layers with ReLU activation and a sigmoid output unit for binary outbreak classification. The network accepts input sequences of length 8 weeks, enabling it to directly model the delayed relationship between meteorological antecedents and disease case counts. Dropout regularization at 0.3 was applied between LSTM layers to mitigate gradient co-adaptation. The Adam optimizer with a learning rate of 0.001 and batch size of 64 was used for training over 150 epochs [11][12].

The fourth base learner is an Attention-Enhanced Convolutional Neural Network (ACNN) that applies 1D convolutional filters (kernel sizes 3 and 5) across the temporal feature dimension to extract local patterns, followed by a self-attention mechanism that weights the relative contribution of different time steps before global average pooling and classification. This architecture captures both local temporal motifs (e.g., sudden temperature spikes preceding outbreak) and global seasonal context simultaneously.

C. Meta-Learner Ensemble Fusion

The out-of-fold probability predictions from all four base learners were concatenated into a meta-feature matrix and supplied as input to a logistic regression meta-learner. This stacking approach was preferred over simple averaging because it allows the meta-learner to dynamically weight base models according to their demonstrated accuracy on the specific disease and geographic context, rather than assigning uniform weights. The meta-learner was trained using 5-fold temporal cross- validation on the training set only, with base learner predictions obtained on held-out folds to prevent target leakage [16].

D. Risk Tier Classification

The ensemble's final calibrated probability output P_outbreak was mapped to a four-tier risk classification as follows: Green ($P < 0.25$, routine surveillance, no specific action required), Yellow ($0.25 \leq P < 0.55$, heightened monitoring and pre- positioning of rapid diagnostic tests), Orange ($0.55 \leq P < 0.80$, activation of vector control operations and public advisory issuance), and Red ($P \geq 0.80$, emergency response activation, hospital surge capacity triggers). These thresholds were calibrated jointly with district health officers from Gujarat and Rajasthan through a structured expert elicitation process to ensure operational alignment with government preparedness protocols [3][15].

International Journal of Science, Strategic Management and Technology
Volume **02** Issue **03** March-**2026** | ISSN: **3108-1762 (Online)** | Impact Factor: 3.8
An International, Peer-Reviewed, Open Access Scholarly Journal Indexed in recognized academic databases

TABLE I. Performance Comparison of Base Learner Models and Ensemble System

Model	Accuracy (%)	AUC-ROC	Precision	Recall	F1-Score
Random Forest	88.4	0.912	0.871	0.893	0.882
GBM (XGBoost)	90.1	0.928	0.889	0.913	0.901
LSTM Network	91.6	0.941	0.907	0.924	0.915
ACNN	90.8	0.934	0.896	0.918	0.907
Proposed Ensemble	93.7	0.960	0.931	0.943	0.937

V. EXPERIMENTAL SETUP AND EVALUATION

A. Implementation Environment

All models were implemented in Python 3.10 using scikit-learn 1.2, XGBoost 1.7, and TensorFlow 2.12 with Keras API. Training of deep learning components was performed on an NVIDIA A100 80 GB GPU cluster provisioned through the High Performance Computing facility at IIT Gandhinagar. Geospatial preprocessing, including satellite data extraction and IDW interpolation, was implemented in GDAL 3.6 and GeoPandas 0.13. The EWS backend was containerized using Docker and deployed as a REST API using FastAPI, enabling integration with district health management information systems.

B. Evaluation Metrics

Model performance was evaluated using binary classification metrics—accuracy, precision, recall, F1-score, and AUC-ROC—for outbreak event detection. For continuous weekly incidence estimation, mean absolute error (MAE), root mean squared error (RMSE), and symmetric mean absolute percentage error (sMAPE) were computed. Statistical significance of performance differences between the ensemble and the best individual base learner (LSTM) was assessed using the Diebold- Mariano test with a 5% significance threshold. All evaluation metrics were computed exclusively on the 2021–2022 prospective holdout set to ensure unbiased performance estimation.

C. Comparative Baselines

The proposed ensemble was benchmarked against three reference systems: (i) a seasonal ARIMA model conditioned on historical case counts without climate covariates; (ii) a multivariable logistic regression model with the same 47 climate features as input; and (iii) a single LSTM model—the strongest individual base learner—to isolate the marginal benefit of ensemble fusion. Additionally, the system was compared against the passive IDSP surveillance baseline, in which the effective detection lag approximates the time between epidemic threshold crossing and national reporting system notification.

TABLE II. Comparative Evaluation Against Baseline Methods on 2021–2022 Prospective Test Set

Method	Accuracy (%)	AUC-ROC	MAE (cases/10k)	Lead Time (weeks)
Seasonal ARIMA	74.2	0.761	4.8	1
Logistic Regression + Climate	81.9	0.845	3.6	2
Single LSTM	91.6	0.941	2.1	3
Proposed Ensemble EWS	93.7	0.960	1.4	3
Passive IDSP Surveillance	—	—	—	–1 to 0

VI. RESULTS AND DISCUSSION

A. Overall Predictive Performance

The proposed ensemble EWS achieved an overall outbreak classification accuracy of 93.7% and an AUC-ROC of 0.960 on the prospective 2021–2022 test dataset, representing statistically significant improvements over all baseline methods ($p < 0.01$, Diebold-Mariano test). The ensemble outperformed the best individual base learner—the LSTM—by 2.1 percentage points in accuracy and 0.019 AUC units, confirming that the diversity of learning signals captured by gradient-boosted trees and convolutional attention mechanisms complements the temporal memory of the recurrent component in ways that improve ensemble generalization [16].

The mean absolute error of 1.4 cases per 10,000 population per week for continuous incidence estimation represents a 33% reduction relative to the single LSTM model (2.1 cases/10k) and a 71% reduction relative to the ARIMA baseline (4.8 cases/10k). These gains are epidemiologically consequential: at a district population of 500,000 persons, the difference between the ARIMA and ensemble MAE corresponds to an estimation error differential of approximately 17 cases per week—sufficient to determine whether emergency hospital bed pre-positioning is warranted.

B. Feature Importance Analysis

Aggregated Shapley Additive Explanation (SHAP) values computed across the Random Forest and GBM components of the ensemble revealed that two-week lagged cumulative rainfall, four-week lagged minimum temperature, and current-week relative humidity collectively accounted for 61.3% of the total feature importance mass for dengue outbreak prediction. For malaria, minimum temperature at a four-week lag and NDVI at a two-week lag emerged as the dominant drivers, consistent with the biological sensitivity of Anopheles mosquito breeding to nighttime temperature thresholds and vegetation-associated larval habitat. These findings align with and extend the observations reported by IIT Bombay for dengue [7] and IIT Delhi for malaria [8], providing cross-model validation of the climate-disease linkage structure.

C. Risk Tier Classification Validation

Prospective validation of the four-tier risk alert system across the 2021–2022 holdout period demonstrated that Red-tier alerts correctly preceded actual outbreak escalation (defined as incidence exceeding the 95th historical percentile) in 89.4% of instances, with a mean advance notification of 18.3 days (approximately 2.6 weeks) before local reporting systems flagged the anomaly. False positive Red-tier alerts accounted for 10.6% of all Red-tier alert instances, a rate deemed operationally acceptable by health officials in a co-design workshop conducted with Gujarat state health authorities, given the asymmetric cost structure in which the consequence of missing an outbreak substantially exceeds the cost of an unnecessary preparedness activation [3].

D. Limitations

Several limitations warrant acknowledgment. First, the model's training geography is confined to three Indian states; transferability to different ecological and healthcare system contexts—particularly those with structurally different disease surveillance infrastructures—requires prospective validation before deployment. Second, the system currently treats each district as an independent spatial unit, neglecting cross-district transmission pathways that may matter for spatially contiguous outbreak spread, a limitation that future work incorporating spatial autoregressive components could address. Third, the satellite data products utilized in training have a native latency of approximately 24–48 hours relative to real-time, which constrains the operational refresh frequency of the alert system.

VII. CONCLUSION

This paper presented a machine learning-based early warning system for predicting weather-driven disease outbreaks that integrates satellite-derived and ground-based climate parameters with longitudinal epidemiological surveillance records. The proposed ensemble architecture—combining Random Forest, Gradient Boosting Machine, LSTM, and Attention-Enhanced CNN models under a stacking meta-learner—achieved an outbreak prediction accuracy of 93.7% and an AUC-ROC of 0.960 on a rigorous two-year prospective holdout dataset, outperforming all evaluated baseline methods. The four-tier automated risk classification engine demonstrated a mean advance warning lead of 18.3 days relative to passive surveillance systems, offering district health authorities a practically significant window for preparedness action.

The framework's integration of ISRO MOSDAC satellite telemetry with IMD ground observations as dual climate sensing streams represents a methodological contribution that is directly replicable in other ISRO-served geographies across South and Southeast Asia. Future extensions will incorporate mobility data from anonymized telecommunications records to model human movement-driven transmission amplification, spatial graph neural networks to capture inter-district disease propagation dynamics, and near-real-time ensemble model updating through online learning protocols to adapt to emerging climate-disease relationship shifts induced by long-term climate change.

The authors anticipate that this system, once fully operationalized within the national health information architecture of India, could contribute meaningfully to the country's epidemic preparedness infrastructure, reducing the human and economic burden of preventable disease outbreaks through timely, evidence-based, climate-informed public health action.

ACKNOWLEDGMENT

The authors sincerely thank the India Meteorological Department (IMD), the Integrated Disease Surveillance Programme (IDSP), and ISRO's Meteorological and Oceanographic Satellite Data Archival Centre (MOSDAC), Space Applications Centre, Ahmedabad, for providing the climate and epidemiological datasets that made this research possible. The authors also acknowledge the computational support extended by the HPC facility at IIT Gandhinagar. This work received no specific grant funding from public, commercial, or not-for-profit funding agencies.

REFERENCES

[1] S. S. Morse, "Factors in the emergence of infectious diseases," Emerg. Infect. Dis., vol. 1, no. 1, pp. 7-15, Jan. 1995.

[2] National Centre for Disease Control (NCDC), Ministry of Health & Family Welfare, Govt. of India, "Dengue Fever Annual Report 2022," NCDC, New Delhi, India, Tech. Rep., 2022.

[3] World Health Organization (WHO), "Early Warning, Alert and Response System (EWARS): Technical Framework for Epidemic-Prone Disease Surveillance," WHO Press, Geneva, Switzerland, Tech. Doc. WHO/HSE/GCR/LYO/2014.4, 2014.

[4] Space Applications Centre (SAC), ISRO, "INSAT-3D/3DR Meteorological Data Products: Level-2B Algorithms and Validation Report," SAC/EPSA/AOSG/TR/01/2019, Ahmedabad, India, 2019.

[5] Integrated Disease Surveillance Programme (IDSP), NCDC, "IDSP Operational Guidelines and Reporting Protocol, 5th Edition," Ministry of Health & Family Welfare, New Delhi, India, 2020.

[6] S. Bhatt et al., "The global distribution and burden of dengue," Nature, vol. 496, no. 7446, pp. 504-507, Apr. 2013.

[7] A. K. Sharma, R. Bhatia, and P. Chaudhary, "Rainfall-lagged dengue incidence modeling for Indian metropolitan districts using climate- adjusted negative binomial regression," J. Epidemiol. Community Health, vol. 74, no. 9, pp. 781-789, Sep. 2020. [IIT Bombay]

[8] V. Pandey, S. Kumar, and N. Das, "Meteorological determinants of malaria incidence in tribal districts of Jharkhand and

[9] Odisha: A multi- year panel data analysis," Malar. J., vol. 20, no. 1, p. 314, Aug. 2021. [IIT Delhi & NIMR]

[10] D. Mukherjee, S. Bose, and T. Roy, "Random forest classification for anticipatory dengue outbreak detection using climate-lagged predictors in West Bengal, India," Comput. Biol. Med., vol. 151, p. 106235, Dec. 2022. [IIT Kharagpur]

[11] P. R. Naidu and K. Srinivasan, "Gradient boosted decision trees for 30-day malaria incidence forecasting in semi-arid Rajasthan: A cross-validated climate covariate approach," Expert Syst. Appl., vol. 211, p. 118658, Jan. 2023. [NIT Warangal]

[12] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Comput., vol. 9, no. 8, pp. 1735-1780, Nov. 1997.

[13] K. Rajan, S. Anand, and M. Krishnaswamy, "Deep recurrent learning for dengue forecasting under monsoon variability: A statewide LSTM study in Tamil Nadu," IEEE J. Biomed. Health Inform., vol. 27, no. 4, pp. 1823-1832, Apr. 2023. [IIT Madras]

[14] B. S. Gohil, P. C. Joshi, and H. Shah, "Application of INSAT-3D land surface temperature and outgoing longwave radiation products for vector-borne disease risk zone identification in India," J. Indian Soc. Remote Sens., vol. 50, no. 11, pp. 2147-2159, Nov. 2022. [Space Applications Centre, ISRO, Ahmedabad]

[15] BAPS Swaminarayan Sanstha Public Health Initiative, "Community-Integrated Dengue Prevention Programme in Gujarat Using Satellite-Derived Rainfall Alerts: Impact Assessment 2021," BAPS Charitable Trust, Ahmedabad, India, Internal Programme Report, 2021.

[16] A. Sundaram, G. Krishnan, and R. Natarajan, "Four-tier risk stratification for cholera early warning along the Cauvery river basin using machine learning and meteorological triggers," Int. J. Environ. Res. Public Health, vol. 20, no. 3, p. 2541, Feb. 2023. [NIT Tiruchirappalli]

[17] D. H. Wolpert, "Stacked generalization," Neural Netw., vol. 5, no. 2, pp. 241-259, 1992.

[18] P. Mehendale, A. Mishra, and S. Joshi, "Satellite-based NDVI as a proxy for Aedes mosquito larval habitat quantification in urban Gujarat: Integration with ISRO Resourcesat-2 imagery," Remote Sens. Appl.: Soc. Environ., vol. 31, p. 100994, Aug. 2023. [ISRO - National Remote Sensing Centre, Hyderabad]

[19] G. Verma, A. Sinha, and P. Tiwari, "Ensemble stacking architectures for climate-sensitive disease forecasting: A systematic comparison across Indian climate zones," Artif. Intell. Med., vol. 145, p. 102676, Nov. 2023. [IIT Roorkee]

[20] S. N. Patel and R. Gupta, "Convolutional attention networks for spatiotemporal disease outbreak prediction using multi-sensor satellite data," IEEE Trans. Geosci. Remote Sens., vol. 62, p. 4700213, 2024. [SVNIT Surat]

[21] A. R. Desai, M. Nair, and T. Subramanian, "Operationalizing climate-informed disease early warning systems in resource-constrained health settings: Lessons from a multi-state pilot in India," Lancet Reg. Health Southeast Asia, vol. 22, p. 100337, Mar. 2024. [IIT Gandhinagar & ISRO Space Applications Centre]