Developing an Early Detection Model for improving Vector-Borne Disease Surveillance

Abhishek Raisinghani[1], Kakoli Saha[2]
[1]Greenko School of Sustainability, Sustainable Engineering, IIT Hyderabad, Telangana - 502284, India
gs24mtech11018@iith.ac.in
[2] Department of Planning, School of planning and Architecture, Bhopal, Bhopal, 462030, India
kakolisaha@spabhopal.ac.in

KEY WORDS: Remote Sensing, GIS, Machine Learning, Dengue, Vector-Borne Diseases, Health Surveillance

Abstract

Given the increasing incidence of vector-borne diseases, there is a need for an effective predictive model, to support timely public health responses in urban areas. However, most of the study have been limited to district level and few climatic variables, which may not be sufficient for localized mitigation efforts. To bridge this gap an Early Detection model for dengue fever is developed, analysing key spatial-temporal variables influencing local transmission. The model integrates meteorological variables such as rainfall, temperature, humidity along with physical factors such as NDVI, land cover and population distribution. Dengue cases data was obtained form District Medical Office, Bhopal, while other independent variables were generated through Geographic Information System (GIS) operations on Earth Observation (EO) datasets such as Landsat, MODIS, etc. Multiple Linear Regression (MLR), Generalized Linear Model (GLM), and Sequential Regression Model (SRM) were employed to capture temporal and spatial dependencies. SRM emerged as the most effective model (Adjusted Pseudo-R^2 = .407, RMSE Test = 1.810), outperforming MLR (Adjusted -R^2 = . 256, RMSE Test = 1.701) and GLM (Adjusted Pseudo-R^2 = .555, RMSE Test = 2.623), to identify high-risk areas. Humidity, NDVI, LULC water and forest significantly influenced dengue cases, as these factors favours mosquito breeding. The study highlights the effectiveness of GIS and Machine Learning (ML) in strengthening the disease surveillance and control. Applying this approach in Indian cities such as Bhopal, Madhya Pradesh, demonstrated its potential to facilitated timely, targeted, tailored and resource efficient interventions.

1. Introduction

Dengue, a disease transmitted by Aedes mosquitoes, impacts over 3.9 billion people globally, causing severe public health and economic burden (WHO, 2019). In India alone, annual incidences of dengue rose form 12.5 thousand cases in 2008 to 233.2 thousand cases in 2022 (The National Center for Vector Borne Diseases Control, 2023). WHO, (2019) advocates early detection as one such critical strategy. At present disease prevention mostly rely on vector control methods, therefore early detection is needed to enable public health authorities to respond swiftly to prevent outbreaks.

Several literatures have linked the dynamics of disease transmission to meteorological and physical factors. Rainfall provides breeding habitats for vectors (Kakarla et al., 2019; Mutheneni et al., 2017). Higher temperatures promote longer lifespan for mosquitoes and reduce incubation period. Several of literature referred to mean, maximum and minimum temperatures (Choi et al., 2016; Hossain et al., 2022; Mutheneni et al., 2017; Sarma et al., 2022; Singh et. al., 2022), while Pakhare et al., (2016) used daily diurnal temperature variation. Pakhare et al., (2016) also used humidity, as it aided in vector’s ability to fly, expanding disease capacity to spread. Climate change has further worsened dengue transmission due to expansion of suitable habitats for disease vectors, exposing millions to previously unrecognized threat (Houtman et al., 2022; Lowe et al., 2021; Mordecai et al., 2017). Harsha et al., (2023) stated that built-up areas have high dengue risk due to water collection potential. Scholars have (Harsha et al., 2023 and Sarma et al., 2022) identified Normalized Difference Vegetation Index (NDVI) as crucial variable, as overlooked areas with dense vegetation can act as habitat for vector breeding. Harsha et al., (2023) integrated Topographic Wetness Index (TWI) in there model, to highlight locations at risk of flooding. Sarma et al., (2022) used population distribution as an indicator as high population have higher probability of interaction with vector.

Few methods have been implemented to act as an early warning system to aid in timely decision-making. Dengue forecasting Model Satellite-based System (D-MOSS) developed by HR Wallingford (Kaiser, 2019), launched in Vietnam in 2019, followed by Malaysia and Sri Lanka in 2020 (European Space Agency, 2020), projecting dengue at district level up to seven month ahead of time. Another is Early Warning and Response System (EWARS), in Mexico in 2012, projecting dengue at district level up to 3 months in advance (Cardenas et al., 2022). Both these models relied on temperature, rainfall, and humidity factors only. Scholars have studied the correlation between dengue cases and diverse meteorological and physical parameters. But most of the predictive models have been limited to district level analysis and few climatic variables, which may not address the needs for localized mitigation efforts. In this research, an attempt has been made to create a predictive model at ward level (lowest administrative unit in Indian cities), incorporating both meteorological and physical variables. This will help identify the potential areas with high dengue risk and allow for targeted interventions saving time and resources.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-G-2025-689-2025

2. Study Area: Bhopal Municipal Corporation

The study is undertaken in the Bhopal Municipal Corporation (BMC), positioned at coordinates 23.25˚N, 77.40˚E. Bhopal is a Tier 2 city, with an area of 463 km^2, population around 23 lakh, density of 5,039 persons per km^2, and divided into 85 wards (Directorate of Town and Country Planning MP, 2020). Bhopal was selected for this study due to its climatic and topographical conditions that are suitable for dengue vector breeding. The city's rolling terrain and high annual precipitation (1260 mm) create favorable conditions for standing water, a critical factor for mosquito breeding. Additionally, its average temperature (25°C) falls within the optimal range for Aedes aegypti breeding (Pakhare et al., 2016), while high relative humidity (87% during monsoon) further supports vector activity. Bhopal's green cover (18.1%) also provides suitable habitats for mosquito breeding (Directorate of Town and Country Planning MP, 2020). These factors, combined with comprehensive data availability, make Bhopal an ideal location for developing the proposed Dengue Early Detection (DED) model.

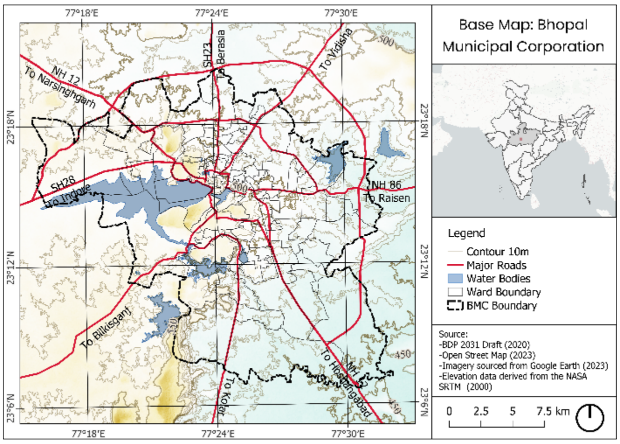

Figure 1. Location of the BMC in Madhya Pradesh, India

In 2014, there was a dengue outbreak in Bhopal with 300 reported cases. Since then, the city has experienced a significant rise in dengue incidences. By 2019, the number of cases surged to 1,705 (DMO Bhopal, 2023), a 14.9-fold increase compared to 2014, far exceeding the 2.3-fold rise observed at the national level (Sarma et al., 2022). This sharp increase highlights the need for improved vector-borne disease management strategies. While there was a notable decline in reported cases in 2020, likely due to the impact of COVID-19 and underreporting, a subsequent rise in cases further emphasizes the ongoing challenge. Considering this the time-period of this study focuses on time duration between 2014 and 2019.

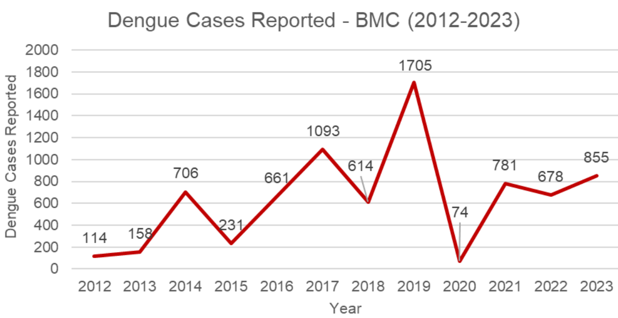

Figure 2. Reported Dengue Cases – BMC (2012-2023)

3. Methodology

3.1 Data Collection and Preparation

For predictive modelling a range of data acquired were Ward wise monthly dengue cases, Monthly precipitation, Land Surface Temperature (LST), Monthly specific humidity, NDVI, Topographic Wetness Index (TWI) and population distribution for year 2010. These datasets were generated through various GIS operations on EO data. Table 1 provides descriptions of the EO data used in this research. The LULC was created using the Landsat 7 and Landsat 8 Level 2, Collection 2, Tier 1 Surface Reflectance data (Vermote et al., 2016) tri-annually, pre-monsoon (February – May), monsoon (June – September), and post-monsoon (October – January). Random Forest supervised classification (Khushaktov, 2023) on multi-spectral band composite (Blue, Green, Red, NIR, SWIR 1 and SWIR 2) was conducted and classified into five broad categories (Water Body, Built-up, Barren, Forest, and Agriculture). The data was trained and tested using 100 datapoint (sample size calculated using Stratified Random Sampling) collected from google earth pro historical imagery feature, with average accuracy and kappa coefficient of 84.5% and 77.9% respectively. Geospatial processes for LST, Humidity, NDVI, LULC were conducted in Google Earth Engine, due to its ability to analyze large data sets simultaneously. Population data of 2010 form GHSL (European Commission,2010) was used to estimate subsequent months data at a linear growth rate of 29.72% as calculated by Directorate of Town and Country Planning MP (2020) in draft-Bhopal Development Plan-2031. All the monthly raster data were resampled to 30 meters resolution and then aggregated at ward level using zonal statistics techniques in QGIS for further analysis.

Table1: Description of the data used

Type of data	Period	Source
Monthly 'Dengue Fever' cases of Bhopal	2012-2019	District Medical Office, Bhopal
Monthly precipitation	2011-2019	Center for Hydrometeorology and Remote Sensing (CHRS)
Land Surface Temperature (LST)	2011-2013 2013-2019 for months with high cloud cover	Landsat 7 Landsat 8 MODIS
LULC	2011-2013 2013-2019	Landsat 7 Landsat 8
Monthly specific humidity	2011-2019	FLDAS
NDVI	2011-2013 2013-2019 for months with high cloud cover	Landsat 7 Landsat 8 MODIS
Topographic Wetness Index	2000	SRTM DEM of 1 Arc-Second
Population distribution	2010	GHSL

3.2 Studying Relationship between Dengue Cases and Influencing Variables

https://doi.org/10.5194/isprs-annals-X-G-2025-689-2025 | © Author(s) 2025. CC BY 4.0 License.

3.2.1 Temporal Relationship

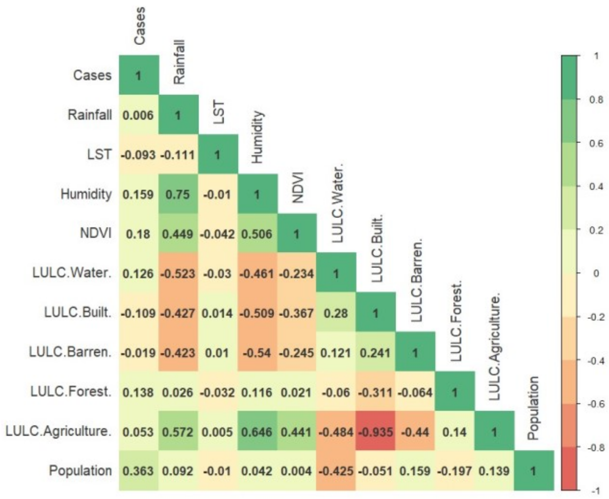

Figure 3. Time Series- Correlation Matrix

To account for inherent biases, temporal relationships between variables were analyzed prior to model preparation. A correlation matrix was used to assess the relationships between the dependent variable and independent variables, as well as interrelationships among the independent variables as seen in Figure 3. Initial results showed low correlations between dengue cases and independent variables (-.093 < r < .363), likely due to non-linear relationships, outliers, or spatial-temporal lag effects. Multicollinearity among independent variables was low. A temporal lag analysis, adjusting for monthly lags up to six months, revealed that correlations improved after accounting for lag as shown in Figure 4.

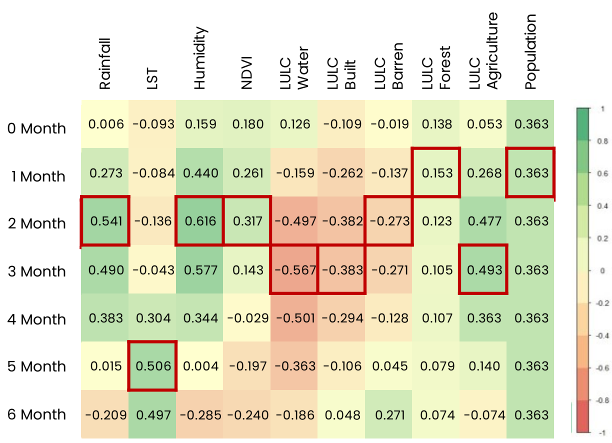

Figure 4. Monthly Lag Matrix

As seen in Figure 4 and Figure 5, rainfall (r = .541), humidity (r = .616), NDVI (r = .317), and LULC barren (r = -.273) showed the highest correlation with a two-month lag, LULC water (r = -.567), LULC built (r = -.383), and LULC agriculture (r = .493) with a three-month lag, and LST (r = .506) with a five-month lag. While moderate multicollinearity emerged in the lag-adjusted dataset, the improved correlations indicated that the model could capture complex relationships, enhancing predictive accuracy. Key variables like rainfall, LST, humidity, and LULC agriculture were identified as important predictors for dengue case variation over time.

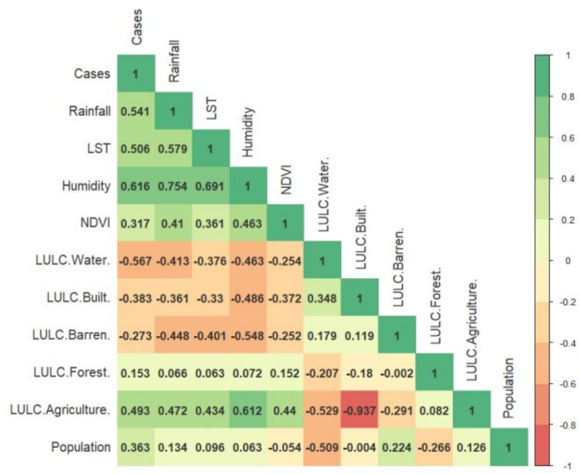

Figure 5. Time Series - Lag Adjusted - Correlation Matrix

3.2.2 Spatial Relationship

Ward wise total dengue cases during the study period were correlated with the average values of the independent variables using a correlation matrix. Results showed mostly low correlations between dengue cases and independent variables, except for humidity (r = .481) and population (r = .484), which had moderate correlations as seen in Figure 6. The low spatial correlation may be attributed to factors such as non-linear relationships, outliers, or spatial-temporal lags due to the seasonal and dynamic nature of the variables.

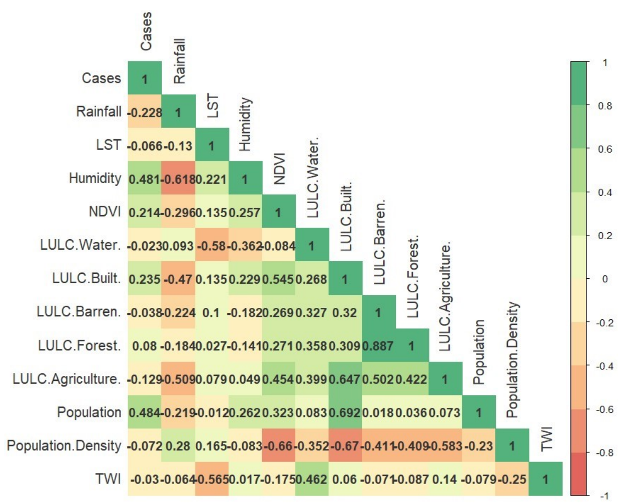

Figure 6. Spatial - Correlation Matrix

The analysis revealed that lag-adjustment improved correlations, capturing complex relationships between predictors like rainfall, humidity, and land use. Moderate intercorrelations were observed after lag adjustment, while low spatial correlations suggested localized effects. These findings emphasize the need for incorporating lagged variables into predictive models.

4. Dengue Early Detection (DED) Model

In the literature reviewed, scholars used various techniques to study relationship between variables. Choi et al., (2016); Singh

& Chaturvedi, (2022) used GLM with negative binomial regression, Hossain et al., (2022) utilised Quasi-Poisson Regression, Pakhare et al., (2016) employed MLR, Sarma et al., (2022) used distributed lag non-linear model combined with quasi-Poisson regression, and Harsha et al., (2023) used AHP and F-AHP. Considering the scope and limitations of the study, two methods were tested MLR and GLM with Quasi-Poisson Regression. MLR was conducted to assess a basic understanding of influence variables had on dengue cases, and GLM since the dependent variables satisfied the assumption of overdispersion i.e., variance (4.28) greater than mean (0.60). The independent variables were standardized using the z-score method (Zach, 2021) for better relative interpretation of influencing factors. The dependent variable (dengue cases) was left unaltered to ensure compatibility with the GLM with Quasi-Poisson Regression (Log Link), which cannot process negative values.

The models were trained for 4 years of data (66.7%) and tested for 2 years of data (33.3%), in temporal order as randomly selecting training data would violate temporal order and likely lead to poor model performance. The model was prepared in R-Studio using ‘lm’ and ‘glm’ function for MLR and GLM respectively. To assess and compare the strength of models to select the best fit model, adjusted coefficient of determination (R^2) for (MLR) Adjusted Pseudo[1] R^2 for (GLM) were used since GLM models does not have normal R^2. Adjusted R^2 measures how well the model explains variations in dengue cases while accounting for number of predictors. To check the magnitude of the errors, Root Mean Squared Error (RMSE) was utilised, showing how far the model’s value are from the actual values. Also, Pearson Correlation was calculated between predicted and actual values both temporally and spatially to measure how well the model capture the temporal and spatial variation of dengue cases.

4.1 Model Preparation

4.1.1 Multiple Linear Regression Model

Table 2: MLR Results

	Coeff.	P-Value	Sig. Level
Intercept	34.281	2.05E-10	< 0.001
Rainfall	0.015	0.741	<1
LST	0.166	3.02E-04	< 0.001
Humidity	0.836	2.00E-16	< 0.001
NDVI	0.389	<2.00E-16	< 0.001
LULC(Water)	0.133	0.615	<1
LULC (Built)	-5.084	7.35E-05	< 0.001
LULC(Barren)	-0.282	0.473	<1
LULC(Forest)	2.431	0.007	< 0.01
LULC(Agriculture)	-0.774	0.006	< 0.01
Population	4.268	<2.00E-16	< 0.001
TWI	0.167	4.62E-05	< 0.001

First MLR model was developed using all variables. Most variables demonstrated high significance. Population and LULC (Forest) were the strongest positive predictors of dengue incidence, while LULC (Built) exhibited the most substantial negative influence as seen in Table 2. This model yielded a low Adjusted R^2 ($R^2 = .26$), with RMSE values for the training and test datasets at 2.24 and 1.70, respectively. The model had good Temporal Pearson Correlation (R = .79) and Spatial Pearson Correlation (R = .55).

4.1.2 Generalized Linear Model

Like the MLR results, Population and LULC (Forest) had the most positive influence on dengue cases, while LULC (Built) had the most negative effect. However, in the GLM model, the influence of humidity also increased notably as shown in Table 3. The model achieved an Adjusted Pseudo R^2 ($R^2 = .55$), with RMSE values of 3.40 for training and 2.62 for testing. In this case, Temporal Pearson Correlation (R = .86) improved but Spatial Pearson Correlation (R = .52) reduced slightly.

Table 3. GLM Results

	Coeff.	P-Value	Sig. Level
Intercept	26.051	6.02E-15	< 0.001
Rainfall	-0.208	<2.00E-16	< 0.001
LST	0.005	0.890	<1
Humidity	1.768	<2.00E-16	< 0.001
NDVI	0.160	<2.00E-16	< 0.001
LULC(Water)	-0.145	0.535	<1
LULC (Built)	-3.856	1.69E-04	< 0.001
LULC(Barren)	-0.611	0.166	<1
LULC(Forest)	2.456	2.61E-06	< 0.001
LULC(Agriculture)	-0.965	7.84E-05	< 0.001
Population	3.662	<2.00E-16	< 0.001
TWI	0.152	3.36E-07	< 0.001

4.1.3 Sequential Regression Model with GLM

From the previous results it was seen that GLM showed better results than MLR but was not able to capture both spatial and temporal variations very well. To better capture both aspects sequential regression method (von Hippel, 2007) was adopted. Were first a dummy variable will be predicted which will capture temporal aspect of the data and then that dummy variable will be used as predictor in the next model to capture spatial aspects. Which variables will better capture temporal variation and which spatial was based on the time series and spatial correlation matrix discussed previously in part 3.2.

Table 4 : SRM Model GLM Results (Temporal)

	Coeff.	P-Value	Sig. Level
Intercept	-6.681	2.03E-06	< 0.001
Rainfall	-0.207	< 2E-16	< 0.001
LST	-0.010	0.791	<1
Humidity	1.823	<2E-16	< 0.001
NDVI	0.177	< 2E-16	< 0.001
LULC(Water)	0.551	0.008	< 0.01
LULC(Barren)	-0.780	0.082	<0.05
LULC(Forest)	2.042	1.16E-04	< 0.001
LULC(Agriculture)	-1.199	2.22E-08	< 0.001

First a temporal model was trained using Rainfall, LST, Humidity, NDVI, LULC(Water), LULC(Barren), LULC(Forest), LULC(Agriculture) as predictors in MLR. The results showed that Rainfall, Humidity NDVI, LULC (Forest) and LULC (Agriculture) were most significant variables, and

[1] Adjusted Pseudo coefficient of determination in GLR, approximates R^2 specifically adapted for GLMs (Smith & McKenna, 2013)

https://doi.org/10.5194/isprs-annals-X-G-2025-689-2025 |

humidity, forest, and agriculture most influenced the dengue cases as seen in Table 4.

This model was used to predict dengue cases (CasesP) which was then used as predictor variables in next step of the model. Next a GLM model with Quasi-Poisson Regression was modelled with CasesP, LULC (Built), Population and TWI, to capture spatial variation. Here the results showed that all the variables were most significant, but variables population and built-up area most influenced the dengue cases as seen in Table 5. Final Adjusted Pseudo R^2 ($R^2 = .41$) and RMSE Train and Test were 2.77 and 1.81 respectively. In this case Temporal Pearson Correlation (R = .83) was between previous two models but Spatial Pearson Correlation (R = .61) improved significantly.

Table 5 : SRM Model GLM Results (Spatial)

	Coeff.	P-Value	Sig. Level
Intercept	25.639	5.77E-12	< 0.001
LULC(Built)	-8.581	<2E-16	< 0.001
Population	4.898	<2E-16	< 0.001
TWI	0.099	2.87E-04	< 0.001
CasesP	0.492	<2e-16	< 0.001

4.2 Model Selection and Result

After evaluating all three models, MLR, GLM, and SRM - the best fit was determined based on R^2 and RMSE values. As shown in Table 6, the GLM model had the highest Adjusted Pseudo R^2 ($R^2 = .55$), followed by the Sequential model and then the MLR model. However, based on RMSE values, the GLM had the highest error, MLR had the lowest error for both training and testing datasets, with the SRM model performing second best.

Table 6: Model Results

	MLR Model	GLM Model	SRM Model
Adjusted R^2	0.256	0.555	0.407
RMSE Train	2.234	3.397	2.769
RMSE Test	1.701	2.623	1.810

Pearson correlation tests further showed that the GLM and SRM model captured temporal variation very well, but SRM outperformed the MLR and GLM models in capturing spatial variation as seen in Table 7. Overall SRM performed better in capturing both temporal and spatial variation.

Table 7: Pearson R Test

Pearson R	MLR Model	GLM Model	SRM Model
Temporal	0.79	0.86	0.83
Spatial	0.55	0.52	0.61

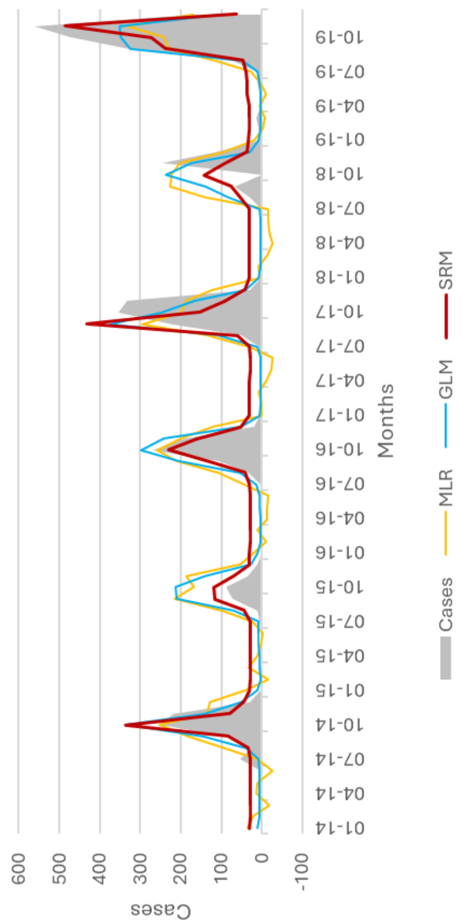

Figure 6: Time Series - Model Results

Visually, as seen in Figure 6. all models captured the seasonal dynamics of dengue cases, but the SRM closely followed the actual trend, while other models underestimated or overestimated peak cases. Also as seen in and Figure 7, SRM estimated the clustering in the south-east direction of the city and relatively lower cases in norther part of city, like observed cases better than other models.

Figure 7: Total Ward Wise Model Results (2014-2019)

The results in Table 4 and Table 5 also show a significant positive relationship between dengue cases and humidity, NDVI, LULC water and LULC forest as these factors favor mosquito breeding and survival. Areas with such characteristics can promote mosquito density raising transmission risk. Also, the population had a highly positive spatial relationship as more human-mosquito interaction facilitated more disease transmission. LULC Barren had a high negative relationship, due to a lack of vegetation and population, reducing breeding sites as well as human-mosquito interaction.

Surprisingly, rainfall, LULC agriculture and built showed a negative relationship with dengue cases. Rainfall can increase mosquito populations, but excessive rain may flush out mosquito larvae. Irrigated agricultural land can increase mosquito breeding but the possible use of pesticides can have negative effects. Generally associated with higher dengue risk due to human-mosquito interactions, factors such as limited vegetation or better infrastructure can limit mosquito growth. Overall, the results highlight the complex interaction of environmental factors with mosquito breeding and dengue cases.

5. Conclusion and Wider Implications

This study has shown the potential use of GIS and ML in strengthening 'Dengue Fever' surveillance. By developing a Dengue Early Detection Model through analysis of spatial-temporal meteorological and physical factors such as rainfall, temperature, humidity, land cover, vegetation index and population distribution. This study attempts to predict dengue cases by including both temporal and spatial dependences of dengue vector, which many prior studies have not addressed. The research also highlights the importance of considering exposure-response lag effect which further enhances prediction accuracy.

Three models—MLR, GLM, and an SRM—were developed to predict dengue cases. The SRM, which incorporated both temporal and spatial aspects, emerged as the best-performing model with an Adjusted Pseudo-R^2 ($R^2 = .407$) and relatively low RMSE values for training and testing datasets. This model not only captures seasonal dengue dynamics but also is able identify high-risk areas at ward level. While the forecasting period is shorter compared to D-Moss, EWARS models, and other studies, this model significantly improved spatial resolution, from district to ward level.

Beyond Bhopal, these results have broader implications for urban planning and public health in other dengue-endemic regions with similar characteristics. But the methodology can be used to create similar models for regions with different climatic or socio-economic characteristics. By integrating climate change and land use prediction models, it can be employed to forecast future dengue scenarios. Also, by incorporating local factors such as mosquito breeding sites identified by the District Malaria Office (DMO), poorly maintained green spaces, open drains, and waste disposal areas, the model's accuracy and relevance can be further enhanced. Additionally, integrating real-time environmental data and dengue case reports would improve the model's timeliness, allowing for quicker responses to outbreaks. This framework is also adaptable to other vector-borne diseases such as Zika and Chikungunya, making it a versatile tool for public health authorities. By enabling localized, data-driven interventions, this model can help save time and resources, empowering local authorities to implement more effective, targeted strategies for disease control and prevention.

References

Cardenas, R., Hussain-Alkhateeb, L., Benitez-Valladares, D., Sánchez-Tejeda, G., & Kroeger, A. (2022). The Early Warning and Response System (EWARS-TDR) for dengue outbreaks: Can it also be applied to chikungunya and Zika outbreak warning? *BMC Infectious Diseases*, *22*, 235. https://doi.org/10.1186/s12879-022-07197-6

Choi, Y., Tang, C. S., McIver, L., Hashizume, M., Chan, V., Abeyasinghe, R. R., Iddings, S., & Huy, R. (2016). Effects of weather factors on dengue fever incidence and implications for interventions in Cambodia. *BMC Public Health*, *16*(1), 241. https://doi.org/10.1186/s12889-016-2923-2

Directorate of Town and Country Planning MP. (2020). *Bhopal Development Plan-2031, (Draft)*. https://mptownplan.gov.in/LU-panel/Bhopal/Amrut/ENGLISH/VOL1.pdf

European Commission. (n.d.). *Global Human Settlement—GHSL Data overview*. Retrieved February 8, 2024, from https://ghsl.jrc.ec.europa.eu/data.php

European Space Agency. (2020, October 12). *D-MOSS | SDG*. https://sdg.esa.int/activity/d-moss-4821

Harsha, G., Anish, T. S., Rajaneesh, A., Prasad, M. K., Mathew, R., Mammen, P. C., Ajin, R. S., & Kuriakose, S. L. (2023). Dengue risk zone mapping of Thiruvananthapuram district, India: A comparison of the AHP and F-AHP methods. *GeoJournal*, *88*(3), 2449–2470. https://doi.org/10.1007/s10708-022-10757-7

Hossain, M. P., Zhou, W., Ren, C., Marshall, J., & Yuan, H.-Y. (2022). Prediction of dengue annual incidence using seasonal climate variability in Bangladesh between 2000 and 2018. *PLOS Global Public Health*, *2*(5), e0000047. https://doi.org/10.1371/journal.pgph.0000047

Houtman, J., Shultz, L., Glassman, R., Gilmour, J., Rivera, J. M., & Scarpino, S. (2022, November 10). The Increasing Burden of Dengue Fever in a Changing Climate. *The Rockefeller Foundation*. https://www.rockefellerfoundation.org/insights/perspective/the-increasing-burden-of-dengue-fever-in-a-changing-climate/

Kaiser, A. (2019, June 27). *Dengue MOdel Forecasting Satellite-based System (D-MOSS) (IPP funded) – Space for Development*. https://www.spacefordevelopment.org/catalogue/dengue-mosquito-simulation-from-satellites-d-moss-hr-wallingford/

Kakarla, S. G., Caminade, C., Mutheneni, S. R., Morse, A. P., Upadhyayula, S. M., Kadiri, M. R., & Kumaraswamy, S. (2019). Lag effect of climatic variables on dengue burden in India. *Epidemiology & Infection*, *147*, e170. https://doi.org/10.1017/S0950268819000608

Khushaktov, M. F. (2023, August 26). Introduction Random Forest Classification By Example. *Medium*. https://medium.com/@mrmaster907/introduction-random-forest-classification-by-example-6983d95c7b91

Lowe, R., Lee, S. A., O'Reilly, K. M., Brady, O. J., Bastos, L., Carrasco-Escobar, G., de Castro Catão, R., Colón-González, F. J., Barcellos, C., Carvalho, M. S., Blangiardo, M., Rue, H., & Gasparrini, A. (2021). Combined effects of hydrometeorological hazards and urbanisation on dengue risk in Brazil: A spatiotemporal modelling study. *The Lancet Planetary Health*, *5*(4), e209–e219. https://doi.org/10.1016/S2542-5196(20)30292-8

Mordecai, E. A., Cohen, J. M., Evans, M. V., Gudapati, P., Johnson, L. R., Lippi, C. A., Miazgowicz, K., Murdock, C. C., Rohr, J. R., Ryan, S. J., Savage, V., Shocket, M. S., Ibarra, A. S., Thomas, M. B., & Weikel, D. P. (2017). Detecting the impact of temperature on transmission of Zika, dengue, and chikungunya using mechanistic models. *PLOS Neglected Tropical Diseases*, *11*(4), e0005568. https://doi.org/10.1371/journal.pntd.0005568

Mutheneni, S. R., Morse, A. P., Caminade, C., & Upadhyayula, S. M. (2017). Dengue burden in India: Recent trends and importance of climatic parameters. *Emerging Microbes & Infections*, *6*(1), 1–10. https://doi.org/10.1038/emi.2017.57

Pakhare, A., Sabde, Y., Joshi, A., Jain, R., Kokane, A., & Joshi, R. (2016). A study of spatial and meteorological determinants of dengue outbreak in Bhopal City in 2014. *Journal of Vector Borne Diseases*, *53*(3), 225–233.

Sarma, D. K., Kumar, M., Balabaskaran Nina, P., Balasubramani, K., Pramanik, M., Kutum, R., Shubham, S., Das, D., Kumawat, M., Verma, V., Dhurve, J., George, S. L., Balasundreshwaran, A., Prakash, A., & Tiwari, R. R. (2022). An assessment of remotely sensed environmental variables on Dengue epidemiology in Central India. *PLoS Neglected Tropical Diseases*, *16*(10), e0010859. https://doi.org/10.1371/journal.pntd.0010859

Singh, P. S., & Chaturvedi, H. K. (2022). A retrospective study of environmental predictors of dengue in Delhi from 2015 to 2018 using the generalized linear model. *Scientific Reports*, *12*(1), 8109. https://doi.org/10.1038/s41598-022-12164-x

Smith, T. J., & McKenna, C. M. (2013). *A Comparison of Logistic Regression Pseudo R 2 Indices*. https://www.semanticscholar.org/paper/A-Comparison-of-Logistic-Regression-Pseudo-R-2-Smith-McKenna/6e21d24f52645f9e871f6896a6bf13267d2ad7f0

The National Center for Vector Borne Diseases Control. (2023, September 17). *DENGUE SITUATION IN INDIA*. https://ncvbdc.mohfw.gov.in/index4.php?lang=1&level=0&linkid=431&lid=3715

Vermote, E., Justice, C., Claverie, M., & Franch, B. (2016). Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product. *Remote Sensing of Environment*, *185*, 46–56. https://doi.org/10.1016/j.rse.2016.04.008

von Hippel, P. T. (2007). Regression with Missing Ys: An Improved Strategy for Analyzing Multiply Imputed Data. *Sociological Methodology*, *37*(1), 83–117. https://doi.org/10.1111/j.1467-9531.2007.00180.x

WHO. (2019, October 24). *Dengue and severe dengue*. World Health Organization. https://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue

Zach. (2021, August 12). Z-Score Normalization: Definition & Examples. *Statology*. https://www.statology.org/z-score-normalization/