

Predicting Dengue Fever Outbreaks through Support Vector Machine: The Role of Climate Variables and Time-Lagged Data

Hetty Meileni ^{1,2}, Ermatita ^{3,*}, Abdiansah ⁴, and Nyayu Latifah Husni ⁵

¹ Doctoral Program in Engineering Science, Faculty of Engineering, Universitas Sriwijaya, Palembang, Indonesia

² Management of Informatics Department, Politeknik Negeri Sriwijaya, Palembang, Indonesia

³ Faculty of Computer Science, Universitas Sriwijaya, Palembang, Indonesia

⁴ Artificial Intelligence Research Development (AIRD), Universitas Sriwijaya, Palembang, Indonesia

⁵ Department of Electrical Engineering, Politeknik Negeri Sriwijaya, Palembang, Indonesia
Email: meileni@polsri.ac.id (H.M.); ermatita@unsri.ac.id (E.); abdiansah@unsri.ac.id (A.);
nyayu_latifah@polsri.ac.id (N.L.H.)

*Corresponding author

Abstract—Dengue fever outbreaks (DENGUE FEVER) represent a significant public health concern in tropical regions, including South Sumatra Province, where rapid climate changes and complex environmental factors contribute to the unpredictability of outbreaks. This study develops a hybrid predictive model utilizing Support Vector Machine (SVM) and Support Vector Regression (SVR) to identify high-risk areas and forecast the future trends of dengue cases. SVM, known for its classification ability, is employed to precisely classify endemic and non-endemic areas, while SVR, with its capacity to model temporal dynamics, is used to predict the number of future cases based on climate variables, such as temperature, rainfall, and humidity, including time-lagged data to capture delayed environmental effects. The model's performance was evaluated using real-world data, revealing that integrating SVM and SVR significantly improves both spatial and temporal predictions of DENGUE FEVER outbreaks. SVM's classification output helps identify areas prone to outbreaks, while SVR provides a detailed forecast of potential case numbers. The model demonstrated high accuracy in mapping endemic zones and predicting case trends, thus addressing both the spatial and temporal aspects of DENGUE FEVER epidemiology. The strength of this approach lies in its ability to process high-dimensional and time-lagged data, providing insights into the delayed effects of environmental factors on disease transmission. The predictive model is valuable for identifying risk areas and assisting health authorities in resource allocation and intervention planning. This study contributes to developing more reliable early warning systems for DENGUE FEVER and lays the groundwork for applying this hybrid machine learning method to other infectious diseases. The results offer significant implications for enhancing preventive measures and public health management in tropical regions.

Keywords—dengue fever, Support Vector Machine (SVM), climate variables, time-lagged data, disease prediction

Manuscript received December 9, 2024; revised February 1, 2025; accepted February 26, 2025; published June 19, 2025.

I. INTRODUCTION

Dengue fever, caused by the dengue virus transmitted by *Aedes* mosquitoes, is a significant global public health issue [1, 2]. The disease presents with symptoms ranging from mild fever to severe complications like Dengue Hemorrhagic Fever (DHF) and Dengue Shock Syndrome (DSS) [3]. Common symptoms include high fever, muscle and joint pain, rash, and hemorrhagic episodes. Early and accurate diagnosis is crucial for reducing mortality [4]. While dengue infections are typically self-limiting, they burden tropical and subtropical regions [4]. The lack of specific treatment and simple diagnostic methods for early-stage illness contributes to the challenge of managing dengue [5]. Developing and implementing long-term control strategies to reduce the global burden of dengue remains a priority for public health specialists.

Dengue fever poses significant health, economic, and social challenges globally. Climate factors play a crucial role in dengue transmission, with temperature, rainfall, and humidity influencing outbreak patterns [6]. Predictive models using climate indicators have shown promise in forecasting dengue outbreaks, as demonstrated in French Guiana [7]. Climate change projections suggest an increased risk of dengue in the future, necessitating improved surveillance and control strategies [8]. In Saudi Arabia, unique factors such as large numbers of migrant workers and religious pilgrims contribute to dengue transmission, climate, and urbanization. Research on climate-dengue relationships is essential for developing effective early warning systems and enhancing public health preparedness [9]. However, the impact of climate on dengue transmission can vary across regions, highlighting the need for location-specific studies and interventions [8, 9].

Recent studies have explored various machine learning approaches for predicting dengue outbreaks. Mills *et al.* [10] highlighted the potential of data-driven

models using climatic and epidemiological data to forecast dengue incidence. Hoyos *et al.* [11] conducted a systematic review, identifying logistic regression as the most common technique for dengue diagnosis and linear regression for spatial analysis. Siriyasatien *et al.* [12] provided a comprehensive overview of dengue prediction methods, emphasizing the importance of data sources, preparation techniques, and model evaluation. Ismail *et al.* [13] demonstrated the effectiveness of Support Vector Machine (SVM) with linear kernel in predicting dengue outbreaks in Selangor, Malaysia, achieving 70% accuracy. They found that the week of the year was the most crucial predictor. These studies collectively underscore the growing significance of machine learning in dengue outbreak prediction and the need for further research to enhance model performance and applicability.

Furthermore, this study develops a predictive model for mapping dengue fever endemic areas by integrating Support Vector Machine and Geospatial Artificial Intelligence (GeoAI) using various parameters. In this model, SVM serves as the primary classification method to predict regions at high risk for dengue spread, while GeoAI strengthens spatial analysis to provide a deeper understanding of the disease's geographic distribution.

The novelty of this research lies in the integrative approach between SVM and GeoAI in predicting dengue-endemic areas. According to the literature review and the author's knowledge, no studies have combined these two methods simultaneously to map dengue risk. The integration of SVM and GeoAI improves prediction accuracy and enriches the analysis by considering spatial dimensions in more detail. This approach offers a significant methodological innovation in the field of epidemiology, particularly in predictive modeling of infectious diseases based on spatiotemporal data. Hence, the aim of this study is to develop a predictive model for dengue-endemic areas by integrating Support Vector Machine and Geospatial Artificial Intelligence to achieve high prediction accuracy. Additionally, the study will assess the accuracy of the predictions generated using these methods, providing insights into their effectiveness for mapping and forecasting dengue outbreaks.

II. LITERATURE REVIEW

A. Deeper Exploration of the Findings' Implications

The findings of the study underscore the transformative potential of predictive models, such as those combining SVM and SVR in shaping proactive public health strategies. By accurately forecasting dengue fever outbreaks, these models enable policymakers to anticipate high-risk periods and geographic locations, ensuring that resources such as medical supplies, vector control measures, and public awareness campaigns are allocated effectively [14]. Climate variables and time-lagged data provide essential insights into the patterns driving outbreaks, making the hybrid model particularly adept at responding to regional environmental factors like rainfall and temperature fluctuations. This integration of data-

driven insights into public health planning can mitigate the human and economic toll of dengue fever outbreaks, particularly in vulnerable areas like South Sumatra Province [15]. Beyond resource allocation, the study highlights a crucial role for predictive models in guiding intervention strategies tailored to local contexts. For example, early warnings based on model predictions can prompt pre-emptive actions such as targeted pesticide applications, community clean-up drives, or the deployment of mobile health clinics. Furthermore, these models can inform public health education initiatives by identifying periods when heightened awareness is most critical, fostering community participation in prevention efforts [16, 17]. By aligning intervention strategies with real-time and forecasted outbreak data, policymakers can shift from reactive to preventive health management, more broadly establishing a robust framework for combating vector-borne diseases. This approach represents a significant advancement in leveraging technology and data analytics for public health resilience.

B. Dengue Fever Transmission Dynamics

Dengue fever is primarily transmitted by *Aedes aegypti* and *Aedes albopictus* mosquitoes [18]. The virus spreads through biological transmission, where infected mosquitoes bite humans after an extrinsic incubation period, and mechanical transmission, where mosquitoes transfer the virus directly between humans without incubation [19]. Mosquito immune systems, including antimicrobial peptides and RNA interference, play a crucial role in controlling dengue virus replication [20]. Temperature significantly affects dengue transmission by influencing mosquito survival, virus proliferation, and vector competence. Environmental factors like breeding sites and temperature are critical for mosquito survival and virus persistence [21]. Mathematical models incorporating these factors can help predict outbreaks and evaluate control measures, such as fumigation. Understanding these complex interactions is essential for developing effective strategies to combat dengue transmission.

Complex interactions between climate and socio-economic factors influence dengue fever transmission. Temperature is dominant in long-term dengue transmission, while socio-economic factors have greater explanatory power during outbreaks [22]. Key socio-economic determinants include population density, socioeconomic stratum, and proximity to tire shops and plant nurseries [23]. Unemployment rates and the number of people per premise also contribute to spatial heterogeneity in dengue cases [24]. While climate factors may act as limiting factors globally, local processes can dominate at smaller scales [25]. Rainfall's role in spatial distribution during epidemics appears less significant. Future climate projections suggest that a 3°C temperature increase could double mean incidence rates during epidemics by the end of the 21st century [24]. Understanding these multifaceted dynamics is crucial for developing effective dengue prevention and control strategies.

C. Predictive Modeling in Epidemiology

Predictive modeling for dengue outbreaks has evolved from traditional methods to advanced machine learning techniques. While regression and time series models have been used historically, machine learning algorithms have shown superior performance in forecasting dengue cases and outbreaks [26]. These models integrate various data sources, including dengue surveillance, meteorological variables, and socio-economic factors, to improve prediction accuracy [27]. Studies have identified humidity and maximum temperature as significant climate factors influencing dengue incidence [28]. Random Forest Regression and Facebook Prophet models have effectively predicted dengue hotspots. Machine learning models have shown advantages in near-term predictions of weekly case counts when surveillance data is available and in forecasting outbreaks 12 weeks without surveillance data [26, 28]. These advancements in predictive modeling contribute to the development of early warning systems, enabling proactive measures to manage dengue outbreaks effectively.

D. Support Vector Machine in Disease Prediction

Support Vector Machines are powerful supervised learning algorithms used for classification and regression tasks [29]. The performance of SVMs is significantly influenced by kernel function selection, with the Gaussian Radial Basis Function (RBF) often being the default choice [30]. Recent advancements include kernel parameter optimization using Sliding Mode Control for improved accuracy and speed [29], and a novel distance-based kernel for binary-type features that outperforms existing approaches [31]. New SVM models utilizing piecewise linear functions and mixed-integer linear programming techniques have also been developed, offering effective data separation and inbuilt feature selection [32]. These innovations demonstrate the ongoing evolution of SVMs, enhancing their applicability and performance across various domains and dataset types.

SVM has shown promising results in predicting dengue outbreaks. Studies have demonstrated its effectiveness in forecasting dengue incidence using climate variables, mosquito infection rates, and socioeconomic data [33]. SVM with linear and radial basis function kernels has achieved high accuracy in predicting dengue morbidity rates, outperforming traditional models [34]. Integrating *Aedes aegypti* mosquito infection rates significantly improved prediction accuracy compared to climate factors. Researchers have also explored using other machine-learning techniques and generalized additive models for dengue prediction, incorporating spatial dependence and socioeconomic factors [35]. These studies highlight the potential of data-driven approaches in enhancing dengue outbreak forecasting, which can aid in timely implementation of preventive measures [36].

E. Role of Climate Variables in Dengue Prediction

Climate variables significantly influence dengue fever transmission, as demonstrated by several studies. Temperature, rainfall, and humidity play crucial roles in the mosquito lifecycle and virus development [37, 38].

These factors vary across seasons, with the highest dengue incidence typically occurring during post-monsoon periods [39, 40]. Research in Taiwan, Cambodia, and India has shown that incorporating climate data into predictive models can improve dengue forecasting accuracy [37, 38]. Notably, a time lag of 2 months between climatic factors and dengue occurrence was found to provide the best correlation in some studies [39]. These climate-based models can serve as early warning systems, allowing public health authorities to implement timely preventive measures. However, the relationship between climate and dengue transmission may vary by locality, suggesting that localized approaches may be more effective [38].

F. Time-Lagged Data in Epidemiological Modeling

Time-lagged data analysis has proven effective in predicting dengue outbreaks by incorporating historical meteorological and disease surveillance information. Studies have shown that combining past weather data and dengue case counts can significantly improve outbreak forecasting [41]. Rainfall, temperature, and humidity are key predictors, with lag times ranging from weeks to months. For instance, the highest relative risk of dengue incidence was observed 10 weeks after rainfall events in Brazil [42]. In Bangladesh, humidity levels 6–8 months prior to outbreaks were strong predictors of case numbers [43]. Similarly, in Malaysia, maximum temperature, relative humidity, and rainfall at 4–6-month lags were significant predictors [44]. These time-lagged models enable early warning systems, allowing health authorities to implement preventive measures before outbreaks reach critical levels. However, model accuracy may vary across geographic locations, emphasizing the need for location-specific studies.

G. Integration of SVM with Climate Variables and Time-Lagged Data

Recent studies have demonstrated the effectiveness of integrating machine learning models with climate variables and time-lagged data for predicting dengue outbreaks. SVM and Bayes Network models have shown promising results in forecasting dengue incidence [45]. These models incorporate meteorological factors, including temperature, rainfall, and humidity, alongside historical disease surveillance data to improve prediction accuracy [41]. Long-term predictors, such as the number of rainy days and average humidity up to six months prior to an outbreak, have been identified as significant factors [43]. Integrating these variables allows for capturing complex, non-linear relationships and delayed effects on disease transmission. While data quality and alignment challenges exist, these integrated approaches have demonstrated superior performance compared to traditional statistical methods, offering valuable tools for early warning systems and public health interventions.

H. Descriptions of SVM and SVR

Support Vector Machine (SVM) and Support Vector Regression (SVR) are machine learning algorithms that rely on the tuning of key parameters to achieve optimal predictive performance. SVM, which is used for

classification tasks, requires parameter tuning for C (regularization parameter), γ (degree of influence of individual data points), and the kernel type (e.g., linear, polynomial, or radial basis function/RBF). The C parameter controls the trade-off between minimizing training data errors and the model's ability to generalize to new data. A high C value prioritizes reducing errors on training data but may lead to overfitting, whereas a lower C value results in a simpler model with greater tolerance for errors. The γ parameter determines how far the influence of a single data point extends; a high γ focuses on nearby data points, while a low γ considers more distant data.

Meanwhile, SVR, which is used for regression tasks, requires optimization of parameters such as C , ϵ , and the kernel. The ϵ parameter defines the tolerance margin for prediction errors, allowing the model to ignore small deviations from the actual values, which in turn helps reduce the risk of overfitting to noise. The kernel function enables the transformation of input data into a higher-dimensional space to handle non-linear relationships. For example, the Radial Basis Function (RBF) kernel is highly effective in capturing complex patterns by creating flexible decision boundaries.

In this study, parameter optimization is performed using the grid search method combined with cross-validation techniques. The grid search method systematically explores combinations of parameter values (such as C , γ , and ϵ) to identify the best configuration based on relevant performance metrics, such as accuracy for SVM and Mean Absolute Error (MAE) for SVR. The use of cross-validation ensures that the selected parameters generalize well across different data subsets, minimizing the risk of overfitting. Through careful parameter tuning, the resulting model achieves high prediction accuracy and reliable performance, effectively capturing the spatial and temporal dynamics of dengue fever outbreaks.

III. MATERIALS AND METHODS

The use of predictive models with regression is an essential step in this research, particularly for predicting the intensity and distribution of dengue fever cases based on environmental data. Support Vector Regression, a robust regression method, is well-suited for modeling the non-linear relationship between environmental variables and dengue occurrences. SVR can handle the complexity of the data and provide more accurate predictions compared to traditional regression methods. In this study, SVR can be used to forecast the number of future dengue fever cases, considering environmental factors that have been shown to have a strong correlation with disease spread.

Data collection is a crucial initial step, as the gathered data will be used to build and test a predictive model for identifying dengue fever endemic areas, integrating SVM, SVR and GeoAI. The data collection focuses on two main categories: epidemiological and climate data. These two

types of data are then integrated to produce a more comprehensive and accurate predictive model.

Epidemiological data is obtained from official sources such as reports from the Health Department of South Sumatra Province, Indonesia, the Health Departments of each city and district, the South Sumatra Provincial Statistics Office (BPS), and the BPS offices of each city and district, as well as hospital records documenting dengue fever case incidences. This data includes detailed information on the number of reported cases and the geographical distribution of those cases. The data is collected for the last seven years (2017–2023) to ensure sufficient temporal variation for analysis.

Environmental and climate data are collected from reliable sources, such as the Meteorological, Climatological, and Geophysical Agency (BMKG), the South Sumatra Provincial Statistics Office (BPS), and the BPS offices of each city and district. The variables collected include air temperature, rainfall, and relative humidity. This data is gathered with monthly and annual temporal resolutions to align with the epidemiological data period, and spatial resolution allows for analysis at the sub-district or village level.

Once the epidemiological, environmental, and climate data have been collected, the next step is to integrate these two types of data into a single database that can be used for model training. This integration process involves temporal alignment between the epidemiological and environmental data, so that each point of epidemiological data can be linked with the relevant environmental conditions at the time of the incident. Additionally, spatial data is processed to ensure compatibility between the geographical resolution of the epidemiological and environmental data, allowing for accurate spatial analysis.

The integrated data is then normalized to eliminate scale differences between the various variables and further processed to identify outliers or inconsistent data. This is important to ensure that the predictive model developed is not biased by unrepresentative data. With the integrated epidemiological and environmental data, the next step is model development, where this data will be used to train predictive models based on SVM and SVR with the support of GeoAI. The model development in this study involves the application of Support Vector Machine and Support Vector Regression integrated with Geospatial Artificial Intelligence. The primary goal of this model development is to predict dengue fever endemic areas and estimate numerical variables related to disease spread, such as case numbers or severity levels.

SVM is used in this study to classify areas as endemic or non-endemic. The SVM model is trained using a dataset with relevant features such as epidemiological, environmental, and geospatial data. The SVM model will learn from patterns in historical data to differentiate between areas with high dengue fever risk (endemic) and those with low or no risk (non-endemic). Thus, SVM helps identify regions that require special attention in dengue fever prevention and control efforts.

IV. RESULTS AND DISCUSSION

A. Dengue Fever Visualization

The visualization of dengue fever case trends in South Sumatra Province from 2017 to 2023 is presented in Fig. 1. This graph shows the fluctuation in dengue fever cases in each district/city, aiming to identify temporal patterns and variations in case rates across different regions. This visualization provides an initial overview of the areas with the highest case rates.

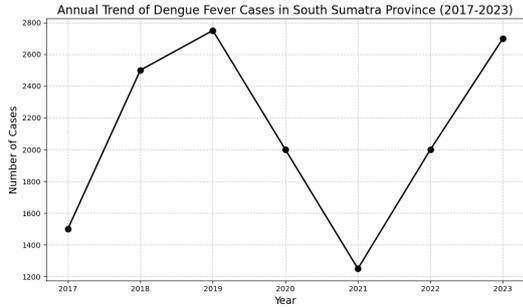


Fig. 1. Dengue fever cases in South Sumatra Province (2017–2023).

A graph showing the dengue fever cases in South Sumatra Province from 2017 to 2023 is presented in Fig. 1. There was an increase in cases from 1452 in 2017 to 2814 in 2019. A significant decrease occurred in 2020, with cases dropping to 2359. The lowest point in this range was

reached in 2021, with 1135 cases. However, the cases began to rise again, increasing from 2854 in 2022 to 2754 in 2023. The chart in Fig. 2 illustrates the distribution of total dengue fever cases across the districts of South Sumatra Province from 2017 to 2023. It shows that Palembang had the highest number of cases, with 4691 cases, far surpassing other districts. Lahat, with 2607 cases, ranks second, indicating a significant potential risk for disease spread in the area. Oku Timur recorded the lowest number of cases, with 251 cases. Banyuasin and Musi Banyuasin also reported relatively high case numbers, each with 2085 and 2607 cases, respectively.

The comparison between the number of dengue fever cases and dengue fever-related deaths from 2017 to 2023 is presented in Fig. 3. It can be seen that the number of dengue fever cases fluctuated significantly year to year. 2017 there were 3028 cases, which steadily increased, peaking at 3900 cases in 2019. In 2020, the number of cases decreased to 3259, reaching its lowest point in 2021 with 2407 cases. Cases began to rise again in the next two years, reaching 4271 cases in 2023, the highest peak within the observed period. The number of deaths related to dengue fever showed a more stable and relatively low trend. In 2017, there were 72 deaths, which increased to 128 deaths in 2018, but then decreased in the following years. In 2020, the death toll sharply dropped to 34, and in 2021, only 21 deaths were recorded, the lowest throughout the period. Subsequently, the number of deaths increased to 116 in 2022 and slightly decreased to 70 deaths in 2023.

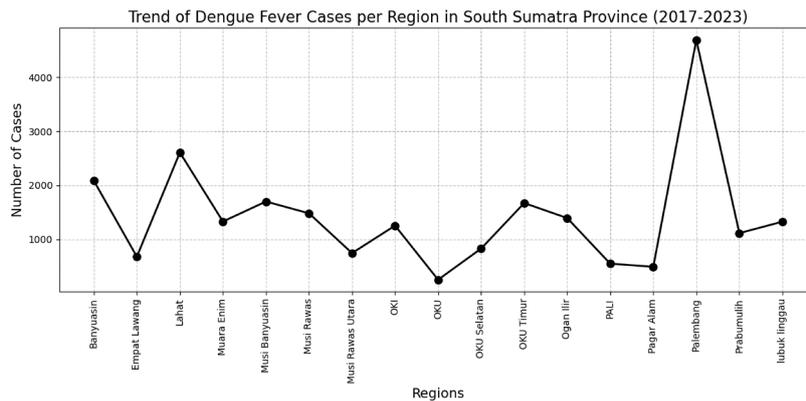


Fig. 2. Total dengue fever cases per district (2017–2023).

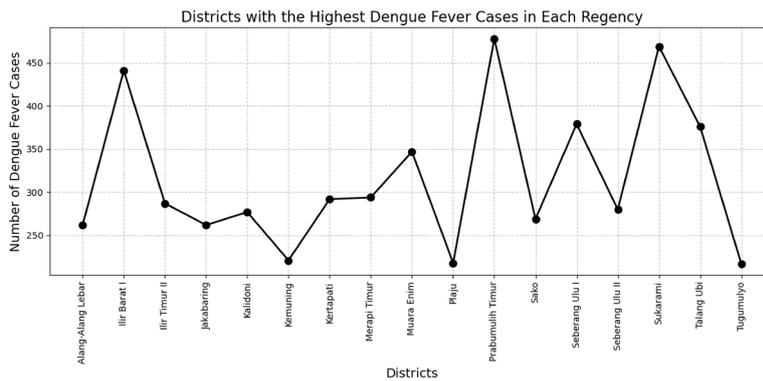


Fig. 3. The highest dengue fever cases by district.

B. Support Vector Machine Modeling

This study uses SVM to predict the endemic status of Dengue Hemorrhagic Fever (DHF) based on several climate variables and population density. The dataset is structured considering several key features, including humidity, rainfall, temperature, population density, DHF cases, and DHF deaths, with the target or label variable being “endemic”. In the model training phase, the data is split into 80% for training and 20% for testing. The training data is then normalized using Min-Max Scaler to ensure that all data is within the same range, allowing the model to function more optimally. Parameter optimization is carried out using Grid Search to find the best combination of parameters. The parameters tested are as follows:

1. C: [0.1, 1, 10, 100], which represents the regularization strength of the model.
2. Gamma: [‘scale’, ‘auto’, 0.1, 0.01, 0.001], which controls the model complexity, especially for non-linear kernels.
3. Kernel: [‘rbf’, ‘linear’, ‘poly’, ‘sigmoid’], which determines the type of kernel to separate the data.

The optimization results in the best combination of parameters with C = 100, gamma = ‘scale’, and kernel = linear. This optimal model is selected based on the highest accuracy on the training data and is saved in a pickle format as svm_endemi_model.pkl. The best model is then applied to the testing data to evaluate its performance in classifying endemic status. Fig. 4 below shows the confusion matrix generated from the model’s predictions on the test data. In the confusion matrix shown in Fig. 4, the interpretation is as follows:

1. True Negative (TN): 58 data points from the non-endemic class were correctly classified as non-endemic.
2. False Positive (FP): 2 data points from the non-endemic class were incorrectly classified as endemic.
3. False Negative (FN): 1 data point from the endemic class was incorrectly classified as non-endemic.
4. True Positive (TP): 281 data points from the endemic class were correctly classified as endemic.

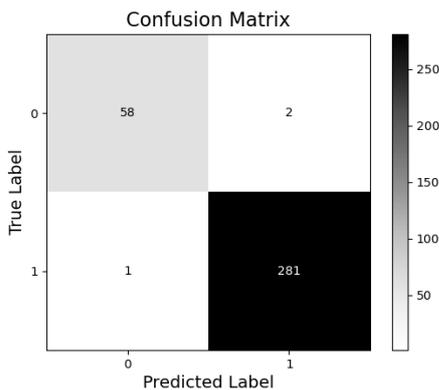


Fig. 4. Confusion matrix.

C. Correlation Data Correlation in Banyuasin III Subdistrict

A correlation analysis of the variables studied in Banyuasin III District, Palembang City is presented in Fig. 5. This analysis highlights the strength of the linear relationship between key variables, such as humidity, rainfall, temperature, population density, and the number of dengue fever cases. The correlation coefficients are represented as positive values, where higher values indicate a stronger relationship between two variables. For instance, a high correlation suggests that changes in one variable are significantly associated with parallel changes in another. This visualization aims to identify the most influential environmental and demographic factors affecting the dynamics of dengue fever transmission in the region.

1) The correlation of variables in Banyuasin III district

Fig. 5 illustrates the relationships between environmental, social, and epidemiological variables influencing the spread of dengue fever in the area. Correlation is measured using correlation coefficients, which indicate the strength and direction of the linear relationship between each pair of variables. The analysis results reveal that temperature has a very strong correlation (0.92) with the number of dengue cases, indicating that rising temperatures significantly impact Aedes mosquito activity, including faster virus replication. Additionally, rainfall shows a moderate positive correlation (0.65), reflecting its role in creating ideal breeding habitats for mosquitoes, particularly through stagnant water accumulation. Population density also exhibits a moderate positive correlation (0.51), suggesting that densely populated areas tend to have a higher risk of transmission due to increased human-vector interactions. On the other hand, humidity demonstrates a weak positive correlation (0.23), indicating that while humidity supports mosquito longevity, its impact on case transmission is less significant compared to other variables.

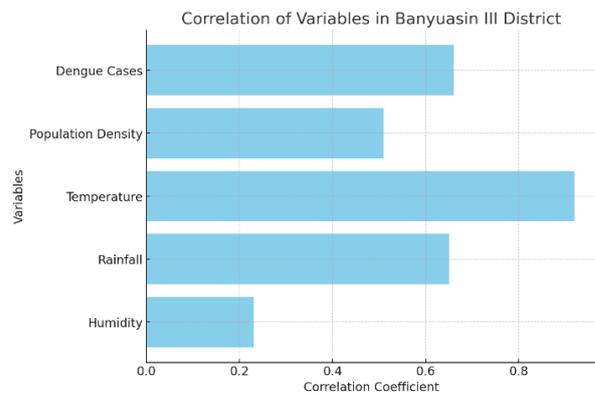


Fig. 5. Correlation matrix of Banyuasin III district.

Based on these findings, temperature, rainfall, and population density are identified as the primary factors that should be considered in dengue fever mitigation and control strategies in Banyuasin III District.

2) Correlation in Tebing Tinggi subdistrict

The relationship between variables affecting the spread of Dengue Hemorrhagic Fever (DHF) cases in Tebing Tinggi, Empat Lawang Regency, was analyzed using a Pearson correlation matrix. The results of this correlation analysis are presented in Fig. 6 to show the strength of the linear relationship between environmental variables (humidity, rainfall, and temperature), social factors (population density), the number of DHF cases, and the mortality rate due to DHF. The analysis reveals significant correlations between several variables. The analysis results show that humidity has the strongest correlation (0.93), indicating that high humidity conditions significantly support the activity of *Aedes* mosquitoes, the primary vector of dengue fever. Population density also has a high correlation (0.77), suggesting that densely populated areas increase the risk of disease transmission due to more frequent interactions between humans and mosquitoes. Temperature exhibits a moderate correlation (0.70), highlighting its influence on the mosquito life cycle and virus replication capacity. Rainfall has a lower correlation (0.50), which may be related to its role in providing breeding habitats for mosquitoes.

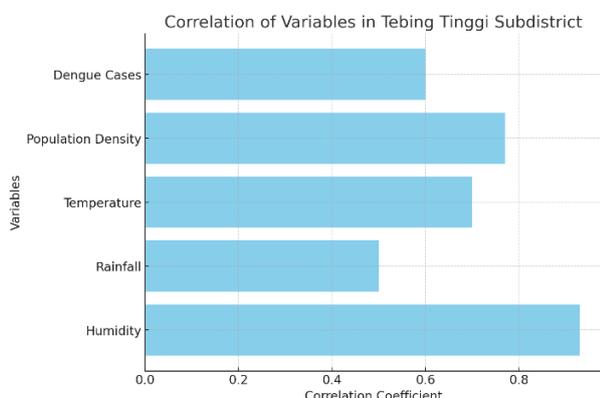


Fig. 6. Correlation in Tebing Tinggi subdistrict.

These findings emphasize that humidity, population density, and temperature are the key factors that need to be considered in efforts to mitigate the risk of dengue fever in Tebing Tinggi District.

3) Correlation in Merapi Timur subdistrict

The linear relationships between variables affecting the spread of dengue fever cases in the Merapi Timur area, Lahat Regency, was analyzed using the Pearson correlation matrix. This matrix illustrates the strength of correlations between environmental variables (humidity, rainfall, and temperature), social factors (population density), the number of dengue fever cases, and the mortality rate due to dengue fever. Based on the correlation matrix presented in Fig. 7, humidity has a significant positive correlation (0.76), indicating that high humidity conditions support increased mosquito activity as the primary vector for dengue fever transmission. Rainfall also shows a moderate positive correlation (0.67),

reflecting its role in creating stagnant water, which serves as a breeding ground for mosquitoes. Population density exhibits a moderate positive correlation (0.47), suggesting that increased human-mosquito interactions in densely populated areas contribute to the risk of transmission. In contrast, temperature has a significant negative correlation (-0.76), indicating that higher temperatures in this region may lead to a decline in mosquito populations or affect their life cycle. These findings highlight that humidity, rainfall, and temperature play crucial roles in the dynamics of dengue fever transmission in Merapi Timur District.

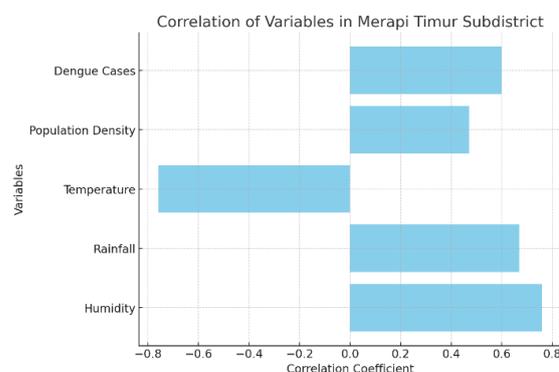
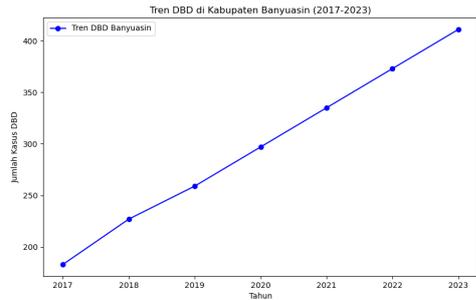


Fig. 7. Correlation in Merapi Timur subdistrict.

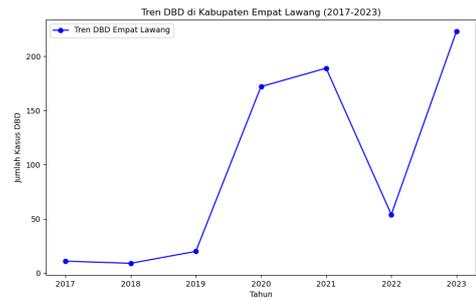
D. Annual Data Correlation

Fig. 8 presents the normalized annual trend data for each analyzed variable, including humidity, rainfall, temperature, population density, number of dengue cases, and dengue mortality rates during the period from 2017 to 2023. This graph provides a comparative overview of the relative changes in the values of each variable over time. The visualization aims to understand the temporal dynamics of these variables and the potential relationship between environmental and demographic factors with the occurrence and impact of dengue fever. The graph from Fig. 9 shows normalized data for the variables of humidity, rainfall, temperature, population density, dengue cases, and dengue-related deaths from 2017 to 2023. These variables exhibit fluctuating patterns from year to year that influence the spread of dengue. In certain years, such as 2020 and 2022, temperature, humidity, and rainfall were at high levels, coinciding with an increase in dengue cases, suggesting a potential positive correlation between these environmental factors and dengue incidence.

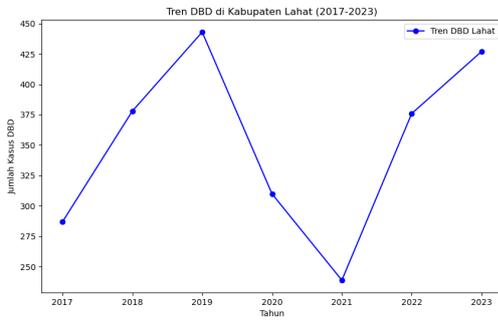
Meanwhile, population density shows a more stable trend with gradual increases, although in certain years, such as 2021, density decreased, and both cases and deaths from dengue were at low points. A drastic decline in some variables, such as rainfall and temperature, in 2019 and 2021 coincided with lower dengue mortality rates. Overall, the graph highlights how environmental factors like temperature, humidity, and rainfall play a crucial role in determining the patterns of dengue cases and fatalities each year.



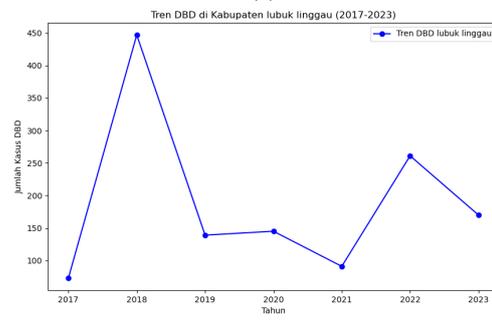
(a)



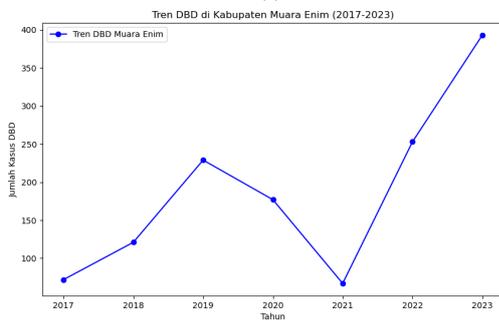
(b)



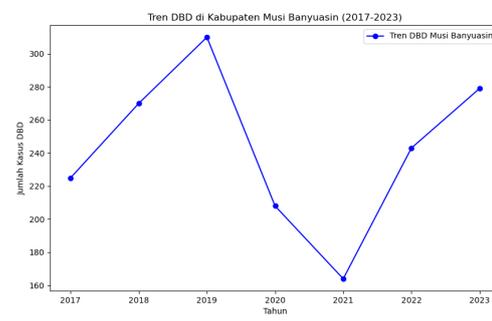
(c)



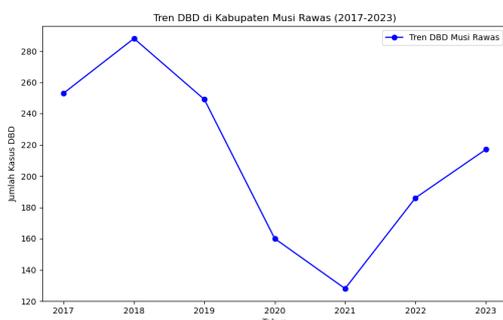
(d)



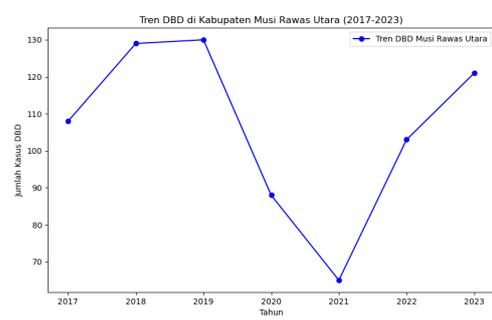
(e)



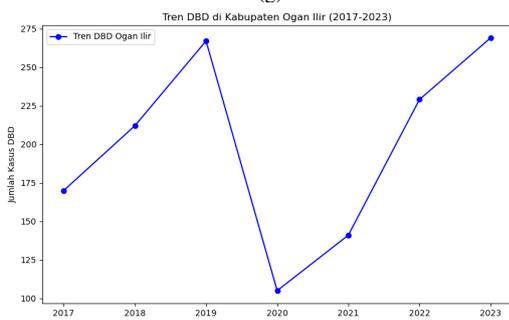
(f)



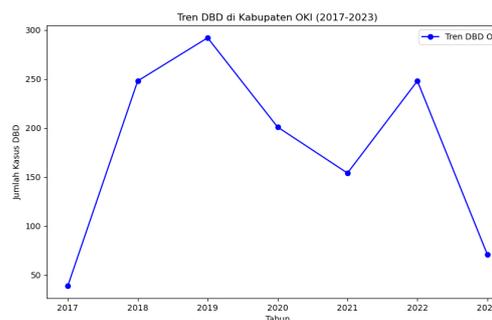
(g)



(h)



(i)



(j)

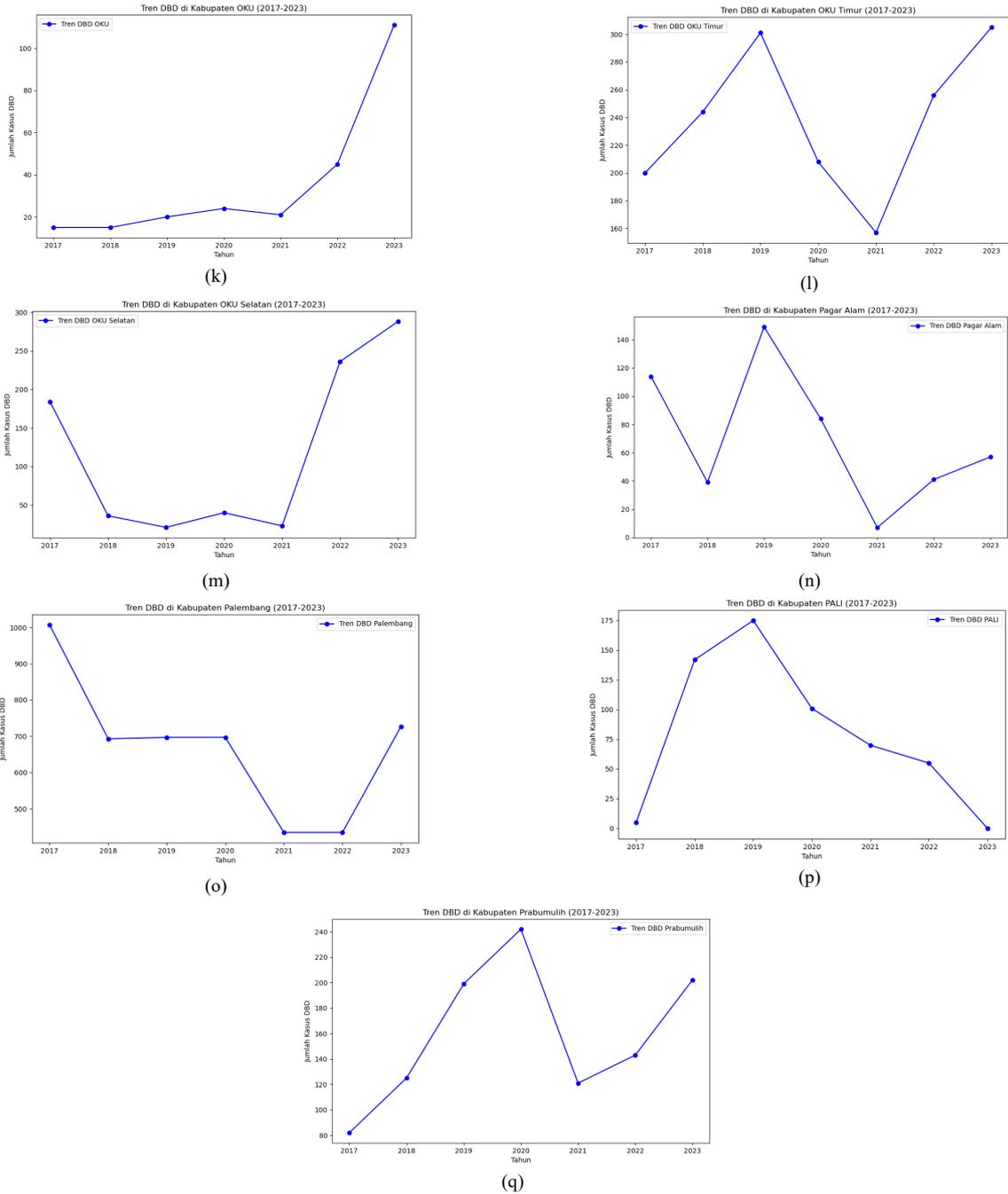


Fig. 8. Dengue fever cases in South Sumatra Province 2017–2023. (a) Banyuasin; (b) Empat Lawang; (c) Lahat; (d) Lubuk Linggau; (e) Muara Enim; (f) Musi Banyuasin; (g) Musi Rawas; (h) Musi Rawas Utara; (i) Ogan Ilir; (j) Oki; (k) Oku; (l) Oku Timur; (m) Oku Selatan; (n) Pagar Alam; (o) Palembang; (p) Pali; (q) Prabumulih

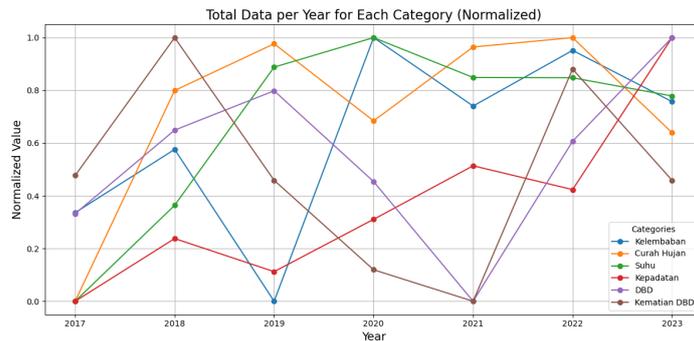


Fig. 9. Normalized data for the variables of humidity, rainfall, temperature, population density, dengue cases, and dengue-related deaths.

E. Support Vector Regression (SVR) Modeling

This section explains the application of the SVR method in predicting the endemic areas of Dengue fever based on climate and population density variables in South Sumatra. This modeling aims to generate accurate predictions regarding the potential spread of dengue fever by utilizing SVR’s ability to handle non-linear data and complex variables.

F. SVR Model for Banyuasin III District

Fig. 10 shows a comparison between the actual values and the predicted cases of Dengue fever in the Banyuasin III area using the SVR method with four types of kernels: linear, sigmoid, polynomial, and Radial Basis Function (RBF). With the linear kernel, the prediction results are close to the actual values, with the predicted line consistently following the actual line pattern. This

indicates that the linear kernel performs well, especially for data with linear relationships. In contrast, the sigmoid kernel shows greater deviations between the actual and predicted values. The prediction pattern of this kernel does not closely reflect the actual relationship, suggesting that the sigmoid kernel is less suitable.

The polynomial kernel produces reasonably good predictions, following the actual data pattern, particularly in capturing potential non-linear relationships. However, some deviations indicate slightly lower performance compared to the linear kernel. The RBF kernel demonstrates the ability to capture more complex data patterns, with predictions close to the actual values. However, there are some discrepancies at certain points, indicating that the parameters of this kernel can still be further optimized. Overall, the linear kernel provides the most accurate predictions compared to the other kernels.

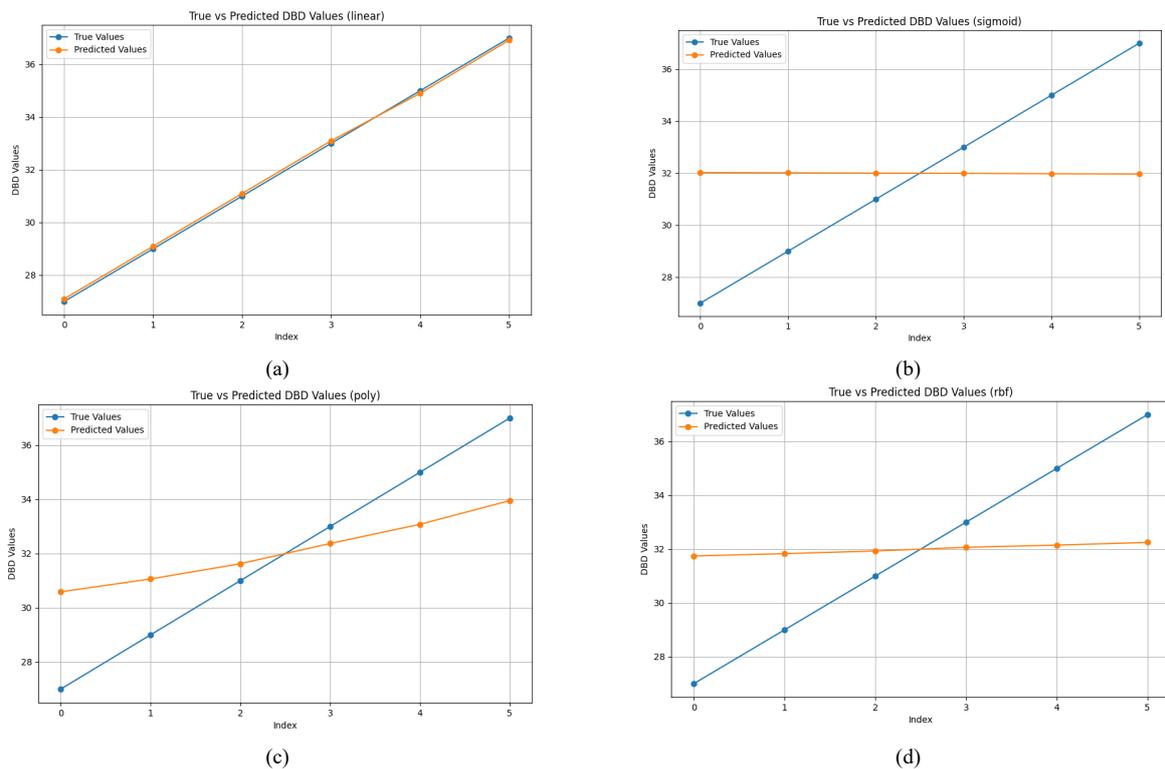


Fig. 10. Prediction results for Banyuasin III district. (a) Linear; (b) Sigmoid; (c) Polynomial; (d) RBF.

G. SVR in Tebing Tinggi District

Fig. 11 shows a comparison between the actual values and the predicted cases of Dengue fever in Tebing Tinggi District using the SVR method with four types of kernels: linear, sigmoid, polynomial, and Radial Basis Function (RBF). For the linear kernel (a), the prediction results show a pattern that closely follows the actual values, but there are significant deviations at some data points, particularly at the initial index. This suggests that the linear kernel is less optimal, even though the overall prediction trend aligns with the actual values. With the sigmoid kernel (b), the prediction line appears relatively flat

compared to the actual values, with consistent deviations at nearly all points. This pattern indicates that the sigmoid kernel struggles to capture the variation in the actual data, leading to inaccurate predictions. The polynomial kernel (c) yields slightly better results than the sigmoid, with a prediction pattern closer to the actual values. The RBF kernel (d) performs reasonably well, with more stable prediction results compared to the sigmoid and polynomial kernels. However, the prediction line still shows deviations from the actual values at some points. Based on the graphical analysis, it can be concluded that the linear and RBF kernels perform better than the sigmoid and polynomial kernels.

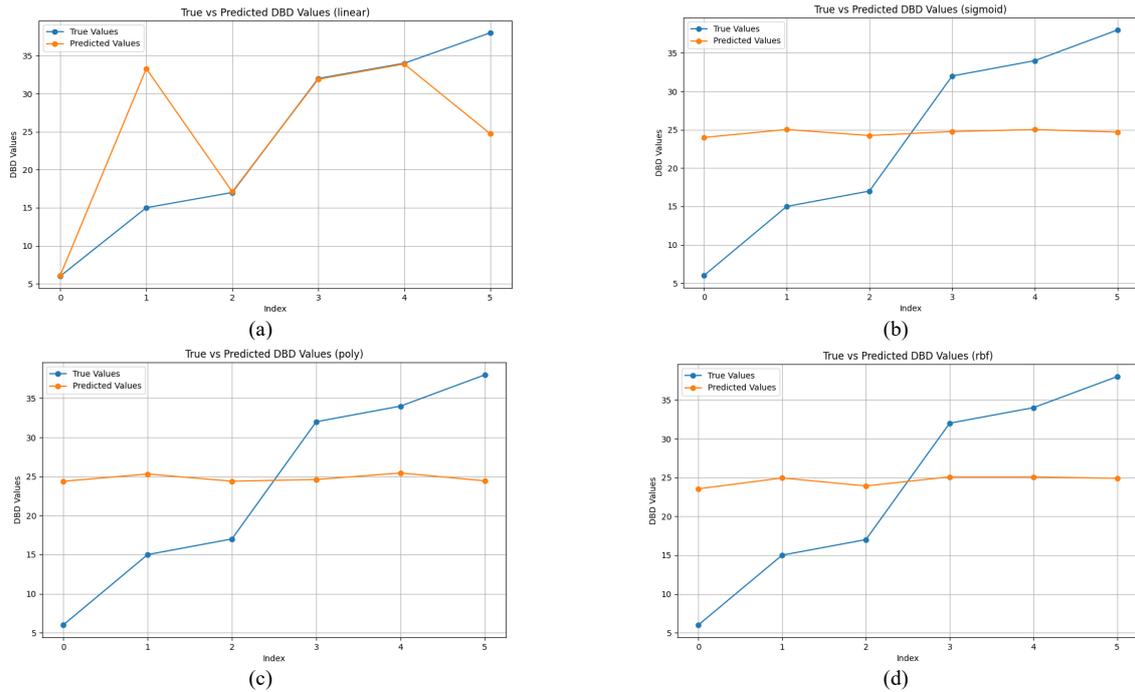


Fig. 11. Prediction results for Tebing Tinggi district. (a) Linear; (b) Sigmoid; (c) Polynomial; (d) RBF.

H. SVR in Merapi Timur District

Fig. 12 illustrates a comparison between the actual values and the predicted cases of Dengue fever in Merapi Timur District using the SVR method with four types of kernels: linear, sigmoid, polynomial, and Radial Basis Function (RBF). For the linear kernel (a), the pattern is relatively close to the actual values, but there are significant deviations at some points, particularly at the initial and final indices. The sigmoid kernel (b) shows larger deviations compared to the linear kernel. The prediction pattern tends to flatten at some indices, making

it less capable of representing the fluctuations of the actual values. This indicates that the sigmoid kernel is unsuitable because it fails to capture the non-linear relationships. The polynomial kernel (c) shows a better prediction pattern than the sigmoid kernel, with the predicted line being closer to the actual values at most indices. However, there are still some discrepancies at certain points. The RBF kernel (d) produces more stable predictions and closely matches the actual values at most indices. This kernel performs better in capturing non-linear relationships in the data, with smaller deviations compared to the other kernels.

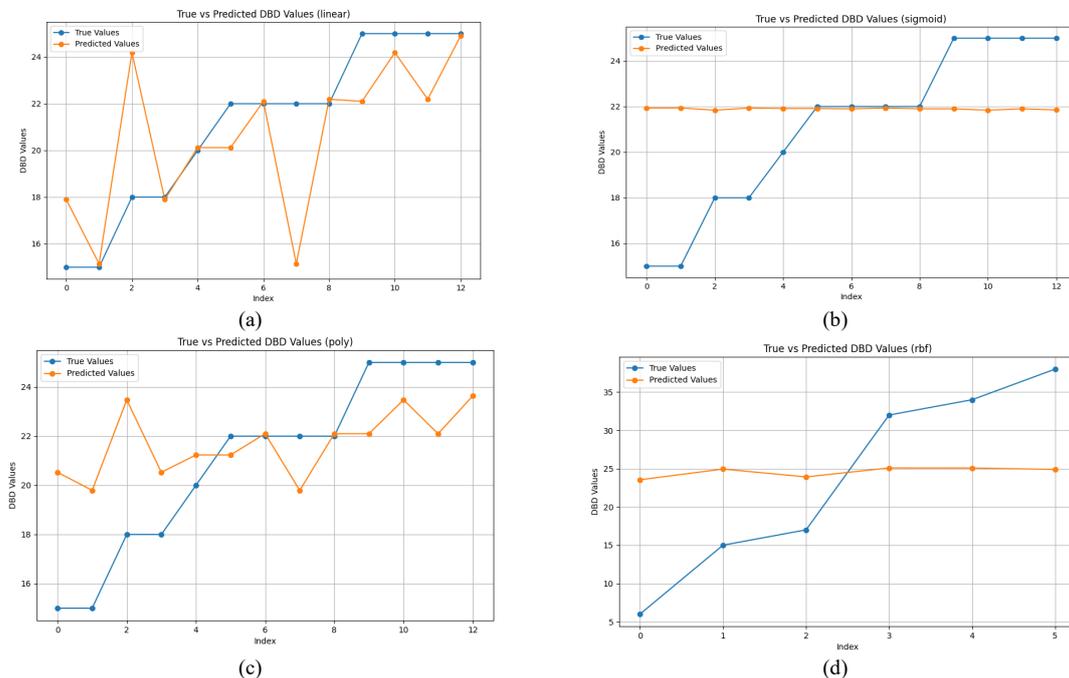


Fig. 12. Prediction results for Merapi Timur district. (a) Linear; (b) Sigmoid; (c) Polynomial; (d) RBF.

I. Further Discussion

This study highlights the significant role of climate variables—temperature, rainfall, and humidity—in influencing dengue fever outbreaks in South Sumatra Province. Temperature has a very strong positive correlation (0.92) with dengue cases, as rising temperatures accelerate *Aedes* mosquito activity and the virus replication cycle. However, extreme temperatures ($>35^{\circ}\text{C}$) are known to reduce mosquito survival rates. Rainfall exhibits a moderate positive correlation (0.65) since it creates optimal breeding habitats for mosquitoes, particularly within a rainfall range of 50–150 mm per month. Conversely, excessive rainfall can reduce mosquito habitats by washing away stagnant water. Humidity also contributes to dengue transmission by extending mosquito lifespan, although its correlation (0.23) is weaker compared to temperature and rainfall. The interaction between these climatic variables amplifies their individual effects, particularly in areas with high temperatures combined with high humidity or during wet months when temperatures exceed 27°C , significantly increasing the incidence of dengue fever cases.

The findings of this study have strategic implications for public health management. By integrating predictive models with real-time climate data, an early warning system can be developed to identify high-risk periods, allowing for targeted vector control measures and public awareness campaigns. The use of spatial classification through the SVM approach enables a more efficient allocation of resources, ensuring that high-risk areas receive priority for interventions such as fogging and larvicide distribution. Additionally, community education programs can focus on raising awareness about the impact of climate conditions, encouraging local efforts to eliminate mosquito breeding sites during the rainy season. This study also emphasizes the importance of implementing locally tailored strategies that align with specific climate patterns at the district/city level, ensuring timely and effective interventions to reduce dengue fever outbreaks.

J. Parameter Optimization and the Impact on Model Performance

The parameter optimization process for SVM and SVR models plays a critical role in balancing bias and variance, directly impacting predictive performance. For SVM, key parameters include the regularization parameter, C , which controls the trade-off between achieving a low training error and avoiding overfitting, and γ , which determines the influence range of a single training point in the model. The choice of kernel function—whether linear, RBF, or another type—is crucial in defining the decision boundary in the feature space. For SVR, parameter optimization revolves around ϵ , which defines the margin of tolerance for prediction errors, C , and the kernel to capture non-linear patterns. The interplay between these parameters allows the models to generalize well across datasets, ensuring high reliability in prediction.

Advanced optimization techniques such as grid search and cross-validation are utilized to identify optimal

parameter values. Grid Search systematically explores a predefined range of parameter combinations, while cross-validation ensures the selected parameters perform consistently across different data splits. For instance, the regularization parameter C might be adjusted in a range of 0.1–100 to balance underfitting and overfitting. Similarly, the ϵ value is fine-tuned for SVR to ensure the model tolerates small errors without compromising overall accuracy. The results of these optimization processes are evaluated using performance metrics such as accuracy for SVM and Mean Absolute Error (MAE) or Root Mean Square Error (RMSE) for SVR. These metrics clearly indicate the model's predictive capability and ability to adapt to real-world complexities.

The effectiveness of parameter optimization is demonstrated in the enhanced accuracy and reliability of the models, particularly in applications like predicting endemic regions or forecasting disease cases. By selecting appropriate parameters (e.g., $C = 100$ and a linear kernel for SVM), the models can effectively capture patterns in complex datasets while avoiding pitfalls like overfitting or underfitting. The ability to accurately map endemic regions or predict health trends is crucial for public health interventions, as it enables better resource allocation and strategic planning. Ultimately, the thorough optimization of these parameters ensures that the models deliver high predictive accuracy and adapt to diverse and evolving data conditions, making them indispensable tools in data-driven decision-making for public health.

K. Data Collection, Potential Biases and Accuracy of the Model.

In this study, data processing and preprocessing played a pivotal role in ensuring the reliability and accuracy of the predictive models. A crucial first step was data normalization, which adjusted the scales of different variables to a common range. Variables such as epidemiological data (e.g., case counts) and environmental data (e.g., rainfall, temperature, and humidity) often exist on vastly different scales. Without normalization, variables with larger numerical ranges could disproportionately influence the training process, potentially skewing the model's outcomes. By aligning the scales, normalization ensured that all variables were treated with equal importance, enhancing the model's ability to identify patterns and relationships between predictors and outcomes effectively.

Outlier detection and removal were another critical aspect of pre-processing [16, 17]. Outliers representing extreme or unrepresentative data points can severely distort statistical analyses and model training. For example, an unusually high or low dengue case count caused by data entry errors or exceptional circumstances might not align with typical trends in the dataset. By identifying and excluding such outliers, the study reduced the risk of models overfitting to aberrant data points, improving their generalizability. Advanced statistical and machine learning techniques, such as the Interquartile Range (IQR) method or clustering-based approaches, may have been employed to identify these anomalies. Removing these inconsistencies allowed the models to

focus on the core patterns and trends that are more representative of real-world conditions [15].

Cross-validation was implemented as an essential strategy to evaluate the robustness and reliability of the models. By dividing the dataset into multiple training and testing subsets, this technique ensured that the models were not overly tailored to a single dataset configuration [46]. For example, k -fold cross-validation systematically split the dataset into k partitions, training the model on $k-1$ subsets and testing it on the remaining one. This process was repeated k times, providing a comprehensive measure of the model's performance across different data splits. Such an approach minimizes the risk of bias and overfitting, as it [47, 48]. This step was particularly important in this study, as it helped confirm the model's capability to generalize its predictions to real-world scenarios, such as mapping dengue endemic regions and forecasting case numbers.

Together, these data processing and preprocessing steps ensured that the predictive models developed in this study were both accurate and reliable. Normalization provided a balanced foundation for training by harmonizing variable scales, while outlier detection safeguarded against misleading patterns from extreme data points. Cross-validation further bolstered the study's methodological rigor by providing a consistent and unbiased framework for performance evaluation. These measures collectively enhanced the study's ability to map dengue endemic regions and predict future cases, offering practical insights into combating the spread of dengue in resource-limited settings. The thoroughness of these preprocessing steps underscores their importance in creating robust predictive models capable of addressing real-world public health challenges [48, 49].

L. Comparison with Other Machine Learning Models

Kernel functions in SVM and SVR play a crucial role in determining the effectiveness of these models in capturing data relationships. The linear kernel is computationally efficient and works well for datasets with clear linear separations. However, its simplicity limits its application to more complex datasets [50]. In contrast, the polynomial kernel introduces a higher level of flexibility through its ability to capture non-linear relationships using adjustable polynomial degrees. While powerful, the polynomial kernel requires careful tuning to avoid overfitting, especially with noisy or large datasets [51, 52].

The RBF kernel, or Radial Basis Function kernel, is widely recognized for its ability to model complex, non-linear relationships. Its strength lies in its capacity to handle datasets influenced by numerous features or environmental variables, such as rainfall and humidity, which are often critical in epidemiological studies. This kernel's versatility and robust performance make it a preferred choice in domains requiring precise pattern recognition. Although the sigmoid kernel can also model non-linear relationships, it is less commonly used due to stability concerns and suboptimal performance in complex tasks [52, 53].

SVMs with kernel functions offer a unique blend of capabilities that distinguish them from other machine

learning models, such as Decision Trees, Random Forests, and Neural Networks. Decision Trees and their ensemble counterpart, Random Forests, excel at modeling non-linear relationships and provide highly interpretable results [54]. Their ability to seamlessly manage categorical and continuous variables makes them versatile for many practical applications. However, these models may struggle with high-dimensional datasets due to overfitting or computational inefficiencies, particularly when the number of features significantly exceeds the number of observations. SVMs, on the other hand, shine in such high-dimensional settings, especially when paired with kernel functions like the RBF kernel [55, 56]. These functions enable SVMs to transform data into higher-dimensional spaces, allowing them to construct optimal separating hyperplanes that effectively classify complex patterns. Neural Networks, particularly deep learning architectures, surpass SVMs in performance for very large and unstructured datasets, such as image and text data, but at the cost of higher computational demands and greater complexity in tuning. This makes SVMs preferred or smaller, structured datasets where precision is paramount [50].

Kernel-based SVMs, particularly those utilizing the RBF kernel, achieve a balance between computational efficiency and predictive accuracy that is challenging for other models to replicate. The RBF kernel excels in modelling non-linear relationships in data by creating flexible decision boundaries, making it highly effective for tasks involving complex patterns or interactions among features [54, 55]. In epidemiological studies, for example, SVMs with RBF kernels have demonstrated significant success in predicting disease outbreaks, such as dengue fever, by leveraging environmental variables like rainfall and humidity. This adaptability also extends to other domains, such as biomedical applications, where precise pattern recognition in structured datasets is critical. Financial prediction tasks, such as stock price movements or credit risk assessments, similarly benefit from the precision and robustness of kernel-based SVMs [56]. While Neural Networks may offer superior performance in these domains under ideal conditions, their reliance on large-scale data and high computational costs often make SVMs a more pragmatic option in resource-constrained environments or when data availability is limited [50].

Overall, SVMs with RBF kernels provide a compelling alternative for applications requiring a combination of precision, interpretability, and manageable computational requirements. Their ability to handle high-dimensional data while avoiding overfitting—thanks to robust regularization techniques—makes them well-suited for small-scale or medium-scale datasets where other models may falter. These models efficiently identify decision boundaries by focusing on support vectors, even in scenarios with overlapping classes or complex distributions. Furthermore, kernel-based SVMs provide high customizability through kernel selection and hyperparameter tuning, allowing them to adapt to a wide range of tasks. This versatility ensures their continued relevance in fields as diverse as epidemiology, finance,

and biomedical research. In an era increasingly dominated by deep learning, SVMs with kernel functions remain a reliable and powerful choice for structured datasets, bridging the gap between traditional statistical models and advanced machine learning architectures.

V. CONCLUSION

This study develops a predictive model to identify areas at risk of dengue fever outbreaks in South Sumatra Province by utilizing SVM and SVR. This approach combines the classification ability of SVM to map endemic areas with high precision and the capability of SVR to predict the quantitative trends of case numbers based on climate variables such as temperature, rainfall, and humidity, as well as time-lagged data. The analysis results show that this integrative approach significantly improves the accuracy and effectiveness of predictions. SVM successfully classifies areas as endemic or non-endemic by considering non-linear patterns between variables, while SVR captures the temporal dynamics of case spread, providing more detailed numerical predictions of potential future cases. The combination of both creates a more comprehensive model, allowing for simultaneous spatial and temporal analysis.

The strength of this model lies in its ability to accommodate the complexity of high-dimensional data and utilize time-lagged data to understand the delayed effects of environmental variables on dengue fever outbreaks. With this model, health authorities can use the predictive results to identify high-risk areas, determine the required resources, and design more timely and effective interventions. This research contributes academically by offering a hybrid machine learning-based approach that can be adapted for studies on other infectious disease epidemiology. Practically, this model also serves as a foundation for developing more accurate early warning systems, strengthening preventive efforts, and supporting sustainable public health in tropical regions.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Conceptualization, Hetty Meileni and Ermatita. Methodology, Hetty Meileni, Ermatita and Abdiansah. Investigation and Validation, Ermatita and Nyayu Latifah Husni. Writing-Review and Editing, Hetty Meileni, Ermatita, Nyayu Latifah Husni. All authors had approved the final version.

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to the following institutions for their support and contributions to this research: the Doctoral Program in Engineering Science, Faculty of Engineering, Universitas Sriwijaya, Palembang, Indonesia; the Faculty of Computer Science, Universitas Sriwijaya, Palembang, Indonesia; the Artificial Intelligence Research Development (AIRD),

Universitas Sriwijaya, Palembang, Indonesia; and the Department of Electrical Engineering, Politeknik Negeri Sriwijaya, Palembang, Indonesia. Their resources, guidance, and encouragement were invaluable throughout this study.

REFERENCES

- [1] N. Singh, A. P. Singh, and A. P. Singh, "Dengue virus: A review," *Indian Journal of Pharmacy and Pharmacology*, vol. 8, no. 4, pp. 243–247, 2022.
- [2] O. J. Brady and S. I. Hay, "The global expansion of dengue: How *Aedes aegypti* mosquitoes enabled the first pandemic arbovirus," *Annu. Rev. Entomol.*, vol. 65, no. 1, pp. 191–208, Jan. 2020.
- [3] Z. Thant *et al.*, "Dengue virus infections: Global scenario, classification, and clinical manifestations—A review," *Research Journal of Pharmacy and Technology*, vol. 9, no. 1, pp. 83–90, Mar. 2016.
- [4] S. Hasan *et al.*, "Dengue virus: A global human threat: Review of literature," *Journal of International Society of Preventive and Community Dentistry*, vol. 6, no. 1, pp. 1–6, 2016.
- [5] M. Hoque, "Dengue: Essence of its causes, symptoms and preventive measures," *International Journal of Research Publication and Reviews*, vol. 4, pp. 726–730, 2023.
- [6] Y. L. Hii *et al.*, "Research on climate and dengue in Malaysia: A systematic review," *Current Environmental Health Reports*, vol. 3, pp. 81–90, Mar. 2016.
- [7] A. Adde *et al.*, "Predicting dengue fever outbreaks in French Guiana using climate indicators," *PLoS Negl. Trop. Dis.*, vol. 10, no. 4, e0004681, April 2016.
- [8] K. L. Ebi and J. Nealon, "Dengue in a changing climate," *Environ. Res.*, vol. 151, pp. 115–123, Nov. 2016.
- [9] K. K. Altassan *et al.*, "Dengue fever in Saudi Arabia: A review of environmental and population factors impacting emergence and spread," *Travel Medicine and Infectious Disease*, vol. 30, pp. 46–53, 2019.
- [10] C. Mills and C. A. Donnelly, "Climate-based modelling and forecasting of dengue in three endemic departments of Peru," *PLoS Negl. Trop. Dis.*, vol. 18, no. 12, e0012596, Dec. 2024.
- [11] W. Hoyos, J. Aguilar, and M. Toro, "Dengue models based on machine learning techniques: A systematic literature review," *Artif. Intell. Med.*, vol. 119, 102157, Sep. 2021.
- [12] P. Siryasatien *et al.*, "Dengue epidemics prediction: A survey of the state-of-the-art based on data science processes," *IEEE Access*, vol. 6, pp. 53757–53795, Sep. 2018.
- [13] S. Ismail *et al.*, "The practicality of Malaysia dengue outbreak forecasting model as an early warning system," *Infect. Dis. Model.*, vol. 7, no. 3, pp. 510–525, Sep. 2022.
- [14] T. L. Drake *et al.*, "Geographic resource allocation based on cost effectiveness: An application to malaria policy," *Appl. Health Econ. Health Policy*, vol. 15, no. 3, pp. 299–306, June 2017.
- [15] H. Meileni *et al.*, "Advancements and challenges in geospatial artificial intelligence, evaluating support vector machines models for dengue fever prediction: A structured literature review," *International Journal of Advanced Computer Science and Applications*, vol. 15, no. 9, pp. 638–645, 2024.
- [16] M. D. Eastin *et al.*, "Intra- and interseasonal autoregressive prediction of dengue outbreaks using local weather and regional climate for a tropical environment in Colombia," *The American Society of Tropical Medicine and Hygiene*, vol. 91, no. 3, pp. 598–610, Sep. 2014.
- [17] B. Mazhar *et al.*, "Development of data-driven machine learning models and their potential role in predicting dengue outbreak," *J. Vector Borne Dis.*, vol. 61, no. 4, pp. 503–514, Oct. 2024.
- [18] J. M. Cruz *et al.*, "A new stochastic model for the aedes aegypti life cycle and the dengue virus transmission," *Applied Sciences*, vol. 13, no. 10, 6214, May 2023.
- [19] L. Laura *et al.*, "Biological and mechanical transmission models of dengue fever," *Communication in Biomathematical Sciences*, vol. 2, no. 1, pp. 12–22, 2019.
- [20] D. Mukherjee *et al.*, "The mosquito immune system and the life of dengue virus: What we know and do not know," *Pathogen*, vol. 8, no. 2, 77, June 2019.

- [21] Z. Liu *et al.*, “Corrigendum: The effect of temperature on dengue virus transmission by Aedes mosquitoes,” *Frontiers in Cellular and Infection Microbiology*, vol. 13, 1242173, Oct. 2023.
- [22] C. Li *et al.*, “Interaction of climate and socio-ecological environment drives the dengue outbreak in epidemic region of China,” *PLoS Negl. Trop. Dis.*, vol. 15, no. 10, e000976, Oct. 2021.
- [23] E. Delmelle *et al.*, “A spatial model of socioeconomic and environmental determinants of dengue fever in Cali, Colombia,” *Acta Trop.*, vol. 164, pp. 169–176, Dec. 2016.
- [24] M. Teurlai *et al.*, “Socio-economic and climate factors associated with dengue fever spatial heterogeneity: A Worked example in New Caledonia,” *PLoS Negl. Trop. Dis.*, vol. 9, no. 12, e0004211, Dec. 2015.
- [25] R. Lowe, “Understanding the relative importance of global dengue risk factors,” *Trans. R. Soc. Trop. Med. Hyg.*, vol. 109, no. 10, pp. 607–608, Aug. 2015.
- [26] C. M. Benedum *et al.*, “Weekly dengue forecasts in Iquitos, Peru; San Juan, Puerto Rico; and Singapore,” *PLoS Negl. Trop. Dis.*, vol. 14, no. 10, e0008710, Oct. 2020.
- [27] R. Jain *et al.*, “Prediction of dengue outbreaks based on disease surveillance, meteorological and socio-economic data,” *BMC Infect. Dis.*, vol. 19, no. 1, pp. 1–16, Mar. 2019.
- [28] S. Patil and S. Pandya, “Forecasting dengue hotspots associated with variation in meteorological parameters using regression and time series models,” *Front Public Health*, vol. 9, 798034, Nov. 2021.
- [29] M. Yalsavar *et al.*, “Kernel parameter optimization for support vector machine based on sliding mode control,” *IEEE Access*, vol. 10, pp. 17003–17017, Feb. 2022.
- [30] I. S. Al-Mejibli, J. K. Alwan, and D. H. Abd, “The effect of gamma value on support vector machine performance with different kernels,” *Int. J. Electr. Comput. Eng. (IJECE)*, vol. 10, no. 5, pp. 5497–5506, Oct. 2020.
- [31] N. Amaya-Tejera *et al.*, “A distance-based kernel for classification via Support Vector Machines,” *Front Artif. Intell.*, vol. 7, 1287875, Feb. 2024.
- [32] J. A. Warwicker and S. Rebennack, “Support vector machines within a bivariate mixed-integer linear programming framework,” *Expert Syst. Appl.*, vol. 245, 122998, July 2024.
- [33] N. A. M. Salim *et al.*, “Prediction of dengue outbreak in Selangor Malaysia using machine learning techniques,” *Sci. Rep.*, vol. 11, no. 1, 939, Jan. 2021.
- [34] S. Khan *et al.*, “Analysis of dengue infection based on Raman spectroscopy and Support Vector Machine (SVM),” *Biomed. Opt. Express*, vol. 7, no. 6, pp. 2249–2256, June 2016.
- [35] O. S. Baquero, L. M. R. Santana, and F. Chiaravalloti-Neto, “Dengue forecasting in São Paulo city with generalized additive models, artificial neural networks and seasonal autoregressive integrated moving average models,” *PLoS One*, vol. 13, no. 4, e0195065, Apr. 2018.
- [36] I. Sanchez-Gendriz *et al.*, “Data-driven computational intelligence applied to dengue outbreak forecasting: A case study at the scale of the city of Natal, RN-Brazil,” *Sci. Rep.*, vol. 12, no. 1, 6550, Apr. 2022.
- [37] Y. H. Lai, “The climatic factors affecting dengue fever outbreaks in southern Taiwan: An application of symbolic data analysis,” *Biomed Eng Online*, vol. 17, pp. 1–14, Nov. 2018.
- [38] Y. Choi *et al.*, “Effects of weather factors on dengue fever incidence and implications for interventions in Cambodia,” *BMC Public Health*, vol. 16, pp. 1–7, Mar. 2016.
- [39] V. G. Ramachandran *et al.*, “Empirical model for estimating dengue incidence using temperature, rainfall, and relative humidity: A 19-year retrospective analysis in East Delhi,” *Epidemiol Health*, vol. 38, e2016052, Nov. 2016.
- [40] M. N. Karim *et al.*, “Climatic factors influencing dengue cases in Dhaka city: A model for dengue prediction,” *Indian Journal of Medical Research*, vol. 136, no. 1, pp. 32–39, July 2012.
- [41] A. L. Ramadona *et al.*, “Prediction of dengue outbreaks based on disease surveillance and meteorological data,” *PLoS One*, vol. 11, no. 3, e0152688, Mar. 2016.
- [42] M. A. Horta *et al.*, “Temporal relationship between environmental factors and the occurrence of dengue fever,” *Int. J. Environ. Health Res.*, vol. 24, no. 5, pp. 471–481, Jan. 2014.
- [43] O. T. Muurlink *et al.*, “Long-term predictors of dengue outbreaks in Bangladesh: A data mining approach,” *Infect. Dis. Model.*, vol. 3, pp. 322–330, 2018.
- [44] V. J. Jayaraj *et al.*, “Developing a dengue prediction model based on climate in Tawau, Malaysia,” *Acta Trop.*, vol. 197, 105055, Sep. 2019.
- [45] F. Y. Nejad and K. D. Varathan, “Identification of significant climatic risk factors and machine learning models in dengue outbreak prediction,” *BMC Med. Inform. Decis. Mak.*, vol. 21, no. 1, 141, Dec. 2021.
- [46] A. Sebastianelli *et al.*, “A reproducible ensemble machine learning approach to forecast dengue outbreaks,” *Sci. Rep.*, vol. 14, no. 1, 3807, Feb. 2024.
- [47] Neelam and B. Gaur, “Data preprocessing: A step-by-step guide for clean and usable data,” *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, vol. 10, no. 2, pp. 1148–1153, Sep. 2019.
- [48] E. Sylvestre *et al.*, “Data-driven methods for dengue prediction and surveillance using real-world and big data: A systematic review,” *PLoS Negl. Trop. Dis.*, vol. 16, no. 1, e0010056, Jan. 2022.
- [49] K. Kesorn *et al.*, “Morbidity rate prediction of Dengue Hemorrhagic Fever (DHF) using the support vector machine and the aedes aegypti infection rate in similar climates and geographical areas,” *PLoS One*, vol. 10, no. 5, e0125049, May 2015.
- [50] Y. Peng and M. H. Nagata, “An empirical overview of nonlinearity and overfitting in machine learning using COVID-19 data,” *Chaos Solitons Fractals*, vol. 139, 110055, Oct. 2020.
- [51] D. Nurwaha, “Comparison of kernel functions of support vector machines: A case study for the solar cell output power prediction,” *International Journal of Energy Applications and Technologies*, vol. 7, no. 1, pp. 1–6, Apr. 2020.
- [52] I. K. Nti *et al.*, “An empirical assessment of different kernel functions on the performance of support vector machines,” *Bulletin of Electrical Engineering and Informatics*, vol. 10, no. 6, pp. 3403–3411, Dec. 2021.
- [53] A. O. Kuyoro, S. Alimi, and O. Awodele, “Comparative analysis of the performance of various support vector machine kernels,” in *Proc. 2022 5th Information Technology for Education and Development (ITED)*, 2022, pp. 1–7.
- [54] M. Somvanshi *et al.*, “A review of machine learning techniques using decision tree and support vector machine,” in *Proc. 2016 International Conf. on Computing Communication Control and automation (ICCUBEA)*, 2016, pp. 1–7.
- [55] E. Y. Boateng, J. Otoo, and D. A. Abaye, “Basic tenets of classification algorithms k-nearest-neighbor, support vector machine, random forest and neural network: A review,” *Journal of Data Analysis and Information Processing*, vol. 8, no. 4, pp. 341–357, Nov. 2020.
- [56] X. Liu, “Comparison of different machine learning models: Linear model, forest and SVM,” *Applied and Computational Engineering*, vol. 51, no. 1, pp. 225–230, Mar. 2024.

Copyright © 2025 by the authors. This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited ([CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)).