e-ISSN: 3030-878X

The 5th International Conference on Environmental Health (ICoEH)

Prediction of Dengue Hemorrhagic Fever (DHF) Case Distribution Using Long Short-Term Memory (LTSM) in Surabaya City

Yesa Nabila Marsitoh[1*], Sri Anggraeni[2], Putri Arida Ipmawati[3], Winarko[4], Slamet Wardoyo[5]

[1,2,3,4,5] Environmental Health Department, Polythectnic of Health Ministry health, Palembang, Indonesia
**Corresponding author: yesanabila55@gmail.com*

ABSTRACT

Background: The incidence of dengue fever in Indonesia in 2023 reached 41.4 per 100,000 population, which is classified as high according to the Indonesian Ministry of Health (with a target IR < 10/100,000 population). Surabaya City is one of the areas with a significant increase in DHF cases, and has the highest population density in East Java. **Object:** This study aims to produce a predictive model that can identify trends in dengue incidence using the Long Short-Term Memory (LSTM) method in Surabaya City. **Methods:** This research is a quantitative study with a retrospective approach using secondary data from 2020 to 2024. The variables analysed included rainfall, humidity, temperature, larva-free rate, and the number of dengue cases. **Results:** The results of the analysis showed a significant relationship between rainfall ($r = -0.195$; $p = 0.033$), temperature ($r = 0.276$; $p = 0.002$), and ABJ ($r = -0.590$; $p = 0.000$) with the incidence of DHF. Humidity showed no significant association ($r = 0.012$; $p = 0.894$). **Conclusion:** It is recommended that strengthening the 3M Plus programme and integrating an LSTM-based early warning system are carried out as preventive measures in public health strategies.

Keywords: Prediction, LTSM, Dengue Fever

BACKGROUND

Dengue fever (DHF) is an acute infectious disease caused by the dengue virus and transmitted through the bite of the *Aedes aegypti* mosquito. WHO noted that by 2023, more than 6.5 million cases of DHF with more than 7,300 deaths occurred globally, spread across more than 80 endemic countries, and it is estimated that this number is still lower than the actual number due to the large number of unreported cases (WHO, 2024). Indonesia is among the countries with the highest dengue burden in the world. Data from the Ministry of Health of the Republic of Indonesia (Kemenkes RI) shows that in 2023 there will be 114,720 cases of DHF with an incidence rate of 41.4 per 100,000 population and a case fatality rate of 0.78% (Ministry of Health, 2023). East Java Province recorded 9,443 cases with an IR of 23 and CFR of 1.1%, exceeding the national target limit (Widiatmika, 2015).

One of the cities with a significant spike in cases is Surabaya City, the largest urbanisation centre in East Java, which in the last five years has shown a consistent trend of increasing cases with a typical annual seasonal pattern. Environmental and climatic factors play a very important role in the spread of dengue. High levels of rainfall, humidity and temperature contribute to increased mosquito populations and accelerate the life cycle and replication of the virus. (Putri et al., 2019) (Tang et al., 2020). Surabaya as a city with high population density and an unoptimised drainage system is an ideal place for mosquito vectors to thrive. (Obenauer, 2017). Climate change is exacerbating the situation, with data showing that rainfall in Surabaya in 2023 ranges from 201-300 mm, which is enough to create many new mosquito breeding sites (Zhang et al., 2024). (Zhang et al., 2024). Communities also still have low awareness of household-based prevention efforts, such as draining water containers and using larvicides, which makes Mosquito Nest Eradication (PSN) activities less effective.

In order to overcome these challenges, a predictive strategy is needed that can anticipate spikes in cases through a data-driven approach. One method that is currently widely used for vector-based infectious disease prediction is Long Short-Term Memory (LSTM), a deep learning algorithm that is effective in handling time series data with seasonal patterns and long-term trends. LSTM has the advantage of remembering long historical information and being able to identify complex fluctuation patterns, making it a very potential tool in dengue case prediction. (Ligue, KD, & Ligue, 2022).

Several studies have shown the successful use of LSTM for dengue case prediction. Research by (Lestari et al., 2021) in Malang District reported that the LSTM model developed had an RMSE of 5.59 and SMAPE of 34.43%, indicating the model's ability to provide accurate predictions. In Vietnam, (Nguyen et al., 2022) developed a deep learning model based on climate data that was shown to improve prediction accuracy over conventional models. A study by Mehta and Patel (2024) in Gujarat, India, also showed that the LSTM approach provided better results in modelling seasonal trends than traditional statistical methods. (Mehta & Patel, 2024). Nonetheless, there are gaps in the literature that need to be addressed. The majority of previous studies developed models at the national or provincial level, thus lacking high spatial resolution for city-level interventions.

In addition, most of the models developed only use one or two predictor variables, whereas the incidence of DHF is strongly influenced by complex multivariate combinations such as temperature, rainfall, humidity, and FLW. In addition, there is a lack of studies that test the generalisability of the model to areas with dense and dynamic urban characteristics such as Surabaya City, which has significant differences between its administrative zones in terms of population density, infrastructure, and environment. (Otmani del Barrio et al., 2018).

Given the complexity of the challenges and the need for a more precise prediction system, this study aims to develop and evaluate a prediction model for dengue cases in Surabaya City using the LSTM approach integrated with climatic and entomological data from 2020 to 2024. Through the integration of temperature, rainfall, humidity, ABJ, and monthly dengue case data, the model is expected to project the trend of dengue incidence in each of the five administrative regions in Surabaya City. This study not only identifies seasonal patterns and annual fluctuating trends, but also statistically assesses the strength of the relationship between environmental variables and dengue incidence. The LSTM model will be built and validated based on prediction accuracy using Mean Absolute Error (MAE) and Mean Square Error (MSE) parameters, and used to project case trends until 2027. Thus, the results of this research are expected to make a real contribution to the development of artificial.

RESEARCH METHODS

Research Type and Design

This research is a quantitative study with a retrospective approach that utilises time series data to build and evaluate a prediction model for Dengue Fever (DHF) incidence based on the Long Short Term Memory (LSTM) algorithm. The selection of this design is adjusted to the research objectives that focus on analysing temporal and seasonal patterns of historical data, which are methodologically more appropriate to be handled with artificial neural network models such as LSTM. This method is considered to be able to overcome the problem of long- term dependence in time series data, especially in the case of diseases with seasonal patterns such as dengue fever. (Wang et al., 2019).

Data Collection Technique

The data used in this study consists of two types, namely epidemiological data and climatological data. Epidemiological data includes the number of monthly DHF cases from each administrative area in Surabaya, obtained from the Surabaya City Health Office. Meanwhile, climatological data consisted of rainfall (in mm), air temperature (°C), and air humidity (%), which were obtained from the Meteorological, Climatological and Geophysical Agency (BMKG) Juanda Meteorological Station. In addition, the larva-free rate (FFR) data was obtained from the Health Office's entomological surveillance report. All data were obtained in monthly aggregated form and validated by relevant agencies. The data collection process was systematic, temporally harmonised, and quality checked to ensure completeness and consistency.

Research Variable and Measurement

The dependent variable in this study is the number of monthly DHF cases in each administrative area of Surabaya City. Meanwhile, the independent variables included rainfall, air temperature, humidity, and ABJ value. Each independent variable was measured in monthly aggregate form and transformed into numeric time series. Missing values were checked prior to analysis using a linear interpolation approach, while extreme values were reviewed based on historical records and not immediately removed. For modelling purposes, all numerical data were normalised using the Min Max scale to be compatible with the LSTM algorithm.

Data Analysis Method

Data analysis was conducted in two main stages. The first stage was descriptive analysis aimed at obtaining an overview of monthly dengue case trends during the observation period, as well as climate fluctuations and changes in ABJ values. The second stage was the construction of a predictive model using the Long Short-Term Memory (LSTM) algorithm developed with the Python programming language and the Tensorflow-Keras library. The data was divided into two groups, 80% as training data and 20% as testing data. The model was built by applying a 12-month sliding window, meaning that data for the previous 12 months was used to predict the number of cases in the 13th month.

The LSTM network architecture consists of two LSTM layers with 256 and 128 units respectively, followed by a batch normalisation layer, a 30% dropout layer, and two dense layers. The first dense layer contains 64 units with ReLU activation function, and the second layer contains one linear output unit that serves to generate the prediction of the number of dengue cases. The model was trained using Mean Squared Error (MSE) loss function and Adam's optimisation with 500 epochs and batch size 16. The entire training process was monitored by cross-validation on test data to prevent overfitting.

Model evaluation was performed using two main indicators, namely Mean Absolute Error (MAE) and Mean Squared Error (MSE). MAE is used to measure the average absolute difference between the actual and predicted values, while MSE is used to calculate the average squared error, which gives greater weight to predictions that miss the mark. Both metrics are used to measure the model's performance in predicting out-of-sample data. The lower the MAE and MSE values, the better the predictive accuracy of the model. The evaluation was conducted for each administrative region so that differences in patterns and performance between regions could be analysed.

All research procedures were approved by the Health Research Ethics Committee of Poltekkes Kemenkes Surabaya with letter number No. EA/3420/KEPK-Poltekkes_Sby/V/2025. Research ethics were maintained throughout the process of data collection, use, and publication, by ensuring that all data used were anonymised and did not contain individual identifying information. This study did not involve direct intervention with human subjects, so ethical risks were minimal.

RESULTS AND DISCUSSION

The results showed that the number of Dengue Fever (DHF) cases in Surabaya City during the period 2020 to 2024 experienced clear temporal and spatial fluctuations. The average monthly number of cases reached ±14.97 cases, with the highest peak occurring in March, at 35.2 cases per month. Meanwhile, October and November recorded the lowest number of cases at 2.4 cases per month. East Surabaya has had the highest number of cases for five consecutive years, with a notable spike in 2024. In contrast, the South and Central Surabaya regions show a more stable and low trend.

The developed LSTM prediction model shows good performance with low Mean Absolute Error (MAE) and Mean Squared Error (MSE) values in all administrative areas. The model performance evaluation is presented in Table 1.

Table 1.

LTSM Model Evaluation Results Based on Surabaya City Administrative Region

Region	Train MAE	Test MAE	Train MSE	Test MSE
West Surabaya	0,0853	0,2697	0,2557	0,3425
Centre Surabaya	0,0755	0,1523	0,3057	0,3352
North Surabaya	0,0813	0,1589	0,2396	0,2929
East Surabaya	0,0719	0,2456	0,2399	0,3047
South Surabaya	0,0720	0,2566	0,2563	0,3365

Table 1 shows that Central Surabaya recorded the lowest MAE of 0.1523 and MSE of 0.3352, indicating high prediction accuracy. Predictions for the period 2025 to 2027 show that East Surabaya will remain the region with the highest number of cases and the most pronounced seasonal fluctuations, especially at the beginning of the year and before the rainy season. Meanwhile, West and South Surabaya are predicted to experience a significant decrease in cases and remain relatively stable below 5 cases per month. These results show that the LSTM model successfully captures consistent seasonal patterns and is sensitive to long-term fluctuations.

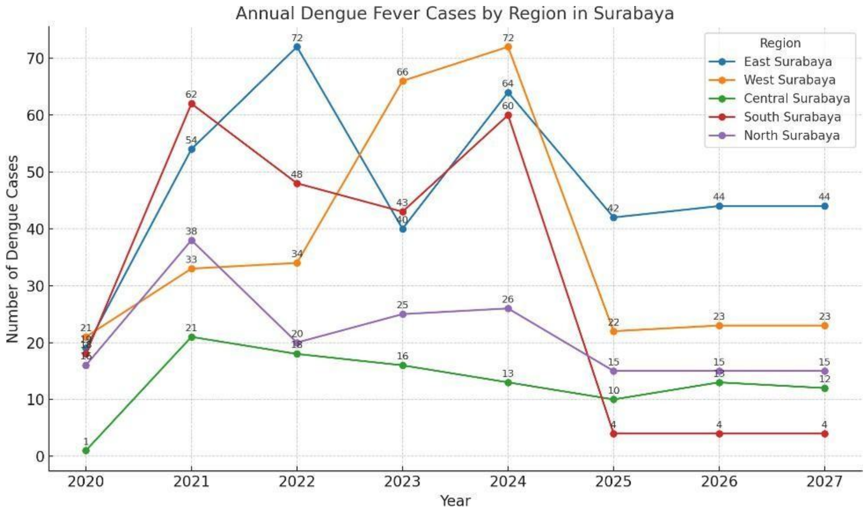

Figure 1. Prediction of DHF in Surabaya City Administrative Area 2020 - 2027

The visualisation in Figure 1 shows that the prediction curve (model results) follows the actual curve quite accurately, especially in the 2020-2022 training data period. Although there is a slight deviation in the 2023 test data, the seasonal trend is still well captured by the model. Overall, the best prediction performance was found in areas with stable and consistent historical data availability, such as Central and South Surabaya.

Disscussion

The application of Long Short-Term Memory (LSTM) based predictive model in this study significantly contributes to the development of vector-based infectious disease prediction methods such as Dengue Fever (DHF), especially in urban areas with climate complexity and population density such as Surabaya City. The results showed that LSTM was able to produce accurate and reliable predictions of the number of DHF cases in five administrative areas, with relatively low Mean Absolute Error (MAE) and Mean Squared Error (MSE) values. The model not only successfully captured seasonal patterns, but also identified annual fluctuating trends that were consistent with the actual data.

The prediction pattern generated by the model shows that East Surabaya has the highest and most fluctuating trend of DHF cases, especially in the early months of the year which coincides with the peak of the rainy season. In contrast, South Surabaya and Central Surabaya showed relatively low and stable case trends. This finding strengthens the argument that variations in local climate, population density, and the quality of the physical environment significantly affect dengue incidence rates, and that the LSTM is able to effectively absorb these seasonal signals. The ability of the model to follow the historical trend of the training data and maintain accuracy on the test data demonstrates the reliability of the LSTM architecture for use in medium-term epidemiological prediction systems ("Forecasting Multivariate Time-Series Data Using LSTM Neural Network in Mysore district, Karnataka," 2022).

Theoretically, the success of LSTM in modelling dengue data can be explained by its ability to remember the long-term dependencies in time series data, including seasonal patterns, interrelationships between climatic variables, and the lag effect of weather changes on the life cycle of *Aedes aegypti* mosquitoes (Chen & Moraga, 2025). (Chen & Moraga, 2025). LSTM also has internal mechanisms in the form of cell state and input-output gates that allow processing of important signals while ignoring noise data, which is very useful when dealing with epidemiological data that is often irregular and contains missing values. (Ghazi et al., 2018). (Fouladgar & Främling, 2020).

This study also confirmed a significant correlation between temperature, rainfall, and larva- free counts with dengue incidence. Rainfall showed a negative correlation with DHF cases, which may be due to increased waterlogging during moderate rainfall, but cessation of larval development during extreme rainfall. Meanwhile, air temperature showed a positive correlation as high temperature accelerates the mosquito life cycle and increases the efficiency of virus replication in the mosquito body. This finding is in line with a previous study by (Nguyen et al., 2022) emphasised that a combination of temperature and rainfall factors had a strong influence on the surge in dengue cases in Vietnam. (Rahayuningtyas et al., 2025) (Rahayuningtyas et al., 2025) also showed that ABJ is an important indicator in measuring the success of PSN and is inversely proportional to the number of DHF cases.

When compared to traditional statistical approaches such as ARIMA or linear regression, LSTM models clearly have the advantage of being more adaptive to non-linear and seasonal data. For example, (Lutfianawati et al., 2024) in their research in Kediri District used an ARIMA model that was only able to project the addition of 1-2 cases per month and had a much higher MSE value compared to the LSTM model in this study. This shows that the LSTM can adjust to more complex data characteristics without strict stationarity assumptions. LSTM is also able to handle multiparameter interactions such as temperature, precipitation, and humidity simultaneously, which is not possible in conventional statistical models. (Hutapea et al., 2020).

However, despite the high performance of LSTM models in prediction, there are still a number of challenges and limitations (Dash et al., 2024). One of the main limitations is the reliance on high- quality and consistent historical data. The model tends to degrade in accuracy when data has significant missing values or unsystematic recording. This is evident from the slightly degraded performance of the model in areas such as East and North Surabaya, which have higher input data variability than other.

CONCLUSION

The LSTM model successfully predicted dengue cases in Surabaya City for the period 2025- 2027. The model shows a low MAE value and is able to capture seasonal trends. East Surabaya is predicted to have the highest spike, while other regions show more stable fluctuations. It is recommended for the Surabaya City Health Office to integrate the LSTM- based prediction model into the dengue early warning system, so that interventions can be carried out in a timely and more effective manner. PSN 3M Plus activities need to be carried out regularly with intensive scheduling, especially in the first and fourth quarters of each year, when the risk of transmission is predicted to be high.

REFERENCES

Astuti, E. P., Dhewantara, P. W., Prasetyowati, H., Ipa, M., Herawati, C., & Hendrayana, K. (2019). Paediatric dengue infection in Cirebon, Indonesia: A temporal and spatial analysis of notified dengue incidence to inform surveillance. Parasites and Vectors, 12(1), 1–12. https://doi.org/10.1186/s13071-019-3446-3

Chen, X., & Moraga, P. (2025). Forecasting dengue across Brazil with LSTM neural networks and SHAP-driven lagged climate and spatial effects. *BMC Public Health*, *25*(1), 1–33. https://doi.org/10.1186/s12889-025-22106-7

Fouladgar, N., & Främling, K. (2020). A novel LSTM for multivariate time series with massive missingness. *Sensors (Switzerland)*, *20*(10), 1–17. https://doi.org/10.3390/s20102832

Gaurav, A., Arya, V., Chui, K. T., Gupta, B. B., Choi, C., & Lee, O.-J. (2023). *Long Short-Term Memory*

Ghaisani, N. P., Sulistiawati, S., & Lusida, M. L. I. (2021). Correlation Between Climate Factors With Dengue Hemorrhagic Fever Cases in Surabaya 2007 – 2017. *Indonesian Journal of Tropical and Infectious Disease*, *9*(1), 39. https://doi.org/10.20473/ijtid.v9i1.16075

Ghazi, M. M., Nielsen, M., Pai, A., Cardoso, M. J., Modat, M., Ourselin, S., & Sørensen, L. (2018). *Robust training of recurrent neural networks to handle missing data for disease progression modeling*. *Midl*, 1–9. http://arxiv.org/abs/1808.05500

Hutapea, M. I., Pratiwi, Y. Y., Sarkis, I. M., Jaya, I. K., & Sinambela, M. (2020). Prediction of relative humidity based on long short-term memory network. *AIP Conference Proceedings*, *2221*(November 2023). https://doi.org/10.1063/5.0003171

Kemenkes RI. (2022). Profil Kesehatan Indonesia 2021. In Pusdatin.Kemenkes.Go.Id. Kementrian Kesehatan. (2016). *Profil Kesehatan*.

Khaidar Ali, Isa Ma'rufi, Wiranto, A. F. (n.d.). *Variability of Local Weather as Early Warning for Dengue Hemorrhagic Fever Outbreak in Indonesia*. https://dl.acm.org/doi/10.1145/3386052.3386078%0ADOI - 10.1145/3386052.3386078

Lestari, N. A., Tyasnurita, R., Vinarti, R. A., & Anggraeni, W. (2021). Long Short-Term Memory forecasting model for dengue fever cases in Malang regency, Indonesia. *Procedia Computer Science*, *197*(2021), 180–188. https://doi.org/10.1016/j.procs.2021.12.131

Ligue, KD, & Ligue, K. (2022). No Title. *Jurnal Sains Filipina*, *150*(3). https://doi.org/10.56899/151.03.01

Liu, J. (2024). AI-Based Epidemic Spread Prediction and Public Health Response Optimization: A Systematic Study from Data Analysis to Policy Implementation. *Applied and Computational Engineering*, *118*(1), 1–7. https://doi.org/10.54254/2755-2721/2025.18470

Luo, J., Fu, Q., Xiao, J., Xiao, H., & Liao, J. (2023). *LSTM-based Pedestrian Trajectory Prediction Model underthe General Direction Mechanism: DIR-LSTM*. 445–451. https://doi.org/10.1145/3641343.3641438

Lutfianawati, R. F., Ngadino, & Marlik. (2024). Prediksi Kejadian Demam Berdarah Dengue di Kecamatan Papar Kabupaten Kediri Tahun 2016–2021. *ASPIRATOR - Journal of Vector- Borne Diseases Studies*, *14*(1). https://doi.org/10.22435/asp.v14i1.5892

Mehta, A. M., & Patel, K. S. (2024). LSTM-based Forecasting of Dengue Cases in Gujarat: A Machine Learning Approach. *Indian Journal Of Science And Technology*, *17*(7), 635–642. https://doi.org/10.17485/ijst/v17i7.2748

Nguyen, V. H., Tuyet-Hanh, T. T., Mulhall, J., Van Minh, H., Duong, T. Q., Van Chien, N., Nhung, N. T. T., Lan, V. H., Minh, H. B., Cuong, D., Bich, N. N., Quyen, N. H., Linh, T. N. Q., Tho, N. T., Nghia, N. D., Anh, L. V. Q., Phan, D. T. M., Hung, N. Q. V., & Son, M. T. (2022). Deep learning models for forecasting dengue fever based on climate data in Vietnam. *PLoS Neglected Tropical Diseases*, *16*(6), 1–22. https://doi.org/10.1371/journal.pntd.0010509

Obenauer, J. (2017). *The Increasing Risk of Vector-Borne Diseases : Mapping the Effects of Climate Change and Human Population Density on Future Aedes aegypti Habitats*. 123.

Otmani del Barrio, M., Simard, F., & Caprara, A. (2018). Supporting and strengthening research on urban health interventions for the prevention and control of vector-borne and other infectious diseases of poverty: Scoping reviews and research gap analysis. *Infectious Diseases of Poverty*, *7*(1), 1–9. https://doi.org/10.1186/s40249-018-0462-

Putri, W. A., Rohiman, A., Sulistiawati, S., & Puspitasari, D. (2019). Effects of Climatic Factors on The Incidence Rate of Dengue Virus Infection in Surabaya During 2010-2013. *Biomolecular and Health Science Journal*, *2*(1), 36. https://doi.org/10.20473/bhsj.v2i1.9340

Rahayuningtyas, D., Pascawati, N. A., Alfanan, A., & Dharmawan, R. (2025). *Model Prediksi Kasus DBD Berdasarkan Perubahan Iklim : Cohort Study dengan Data NASA di Kabupaten Bantul*. *24*(1), 84– 94.

Tang, S. C. N., Rusli, M., & Lestari, P. (2020). Climate Variability and Dengue Hemorrhagic Fever in Surabaya, East Java, Indonesia. *Indian Journal of Public Health Research & Development*, *11*(2), 131. https://doi.org/10.37506/v11/i2/2020/ijphrd/194770

Yupeng Wang, Shibing Zhu, C. L. (n.d.). *Research on Multistep Time Series Prediction Based on LSTM*. 10.1109/EITCE47263.2019.9095044

Zhang, Y., Wang, M., Huang, M., & Zhao, J. (2024). Innovative strategies and challenges mosquito- borne disease control amidst climate change. *Frontiers in Microbiology*, *15*(November), 1–13. https://doi.org/10.3389/fmicb.2024.1488106