Forecasting Weekly Dengue Cases in Brazilian Federative Units

Giovanni E. Zanardo[1], **Éfren L. Souza**[1,2], **Juan G. Colonna**[1], **Fabíola G. Nakamura**[1], **Eduardo F. Nakamura**[1]

[1] Institute of Computing – Federal University of Amazonas
[2] Institute of Engineering and Geosciences – Federal University of Western Pará

{gio.zanardo, juancolonna, fabiola, nakamura}@icomp.ufam.edu.br

efren.souza@ufopa.edu.br

Abstract. *Effective epidemic management requires proactive measures, making the accurate prediction of weekly dengue cases in specific regions essential for prevention and control strategies. In this study, we evaluated the effectiveness of classical statistical techniques and machine learning methods in forecasting the number of weekly dengue cases across 27 Brazilian federative units. For each unit, we explored a multivariate one-step prediction strategy, applying wavelet filtering to the features to enhance signal decomposition and improve predictive performance. We investigated the univariate LightGBM model trained on data from 26 cities (cross-learning) and validated it individually in each federative unit using the univariate leave-one-out (LOO) technique with one-step predictions. Additionally, we demonstrated the model's generalizability by training LightGBM on data from all Brazilian federative units and validating it in a different geographic location, San Juan. The LightGBM LOO model exhibited superior generalization compared to other shallow, deep, and foundation models, including TimeGPT-1 and MOIRAI.*

1. Introduction

Dengue cases in the Americas have risen sharply over the past four decades, from 1.5 million in the 1980s to 17.5 million between 2010 and 2019. Prior to 2023, the highest annual total was in 2019, with over 3.18 million cases, 28,208 severe cases, and 1,823 deaths—a fatality rate of 0.06% [World Health Organization 2023]. In Brazil, 7,866,769 cases were reported in the first 23 weeks of 2024, marking a 230% increase compared to the same period in 2023 [OPAS & OMS 2024]. The cumulative incidence reached 3,676 cases per 100,000 inhabitants, with 5,210 severe cases (0.07%) and 3,643 deaths. Although a vaccine is now available [Anumanthan et al. 2025], efficient allocation of public health resources remains essential. Therefore, anticipating dengue case trends through time series analysis is key to improving planning and response strategies.

Time series forecasting has benefited from deep learning models based on encoder-decoder architectures with self-attention mechanisms, such as Transformers [Vaswani et al. 2023]. These models fall within the universal forecasting paradigm and are considered foundation models [Moor et al. 2023], as they are pre-trained on large databases from various domains, including health, economics, and sales, to predict generic time series. In this context, pre-trained models like

TimeGPT-1 [Garza et al. 2023] and the Masked Encoder-based Universal Time Series Forecasting Transformer (MOIRAI) [Woo et al. 2024] leverage zero-shot learning [Rezaei & Shahidi 2020] to make predictions without requiring task-specific fine-tuning. Given the complex interaction between the number of dengue cases—represented as a time series—and environmental and socioeconomic covariates, including its higher incidence in low-income areas with inadequate sanitation and infrastructure [Bavia et al. 2020], these models emerge as promising alternatives for forecasting the weekly incidence of dengue.

Nonetheless, foundation models lack hyperparameter tuning and disregard the influence of other features of the dataset a priori, concentrating all their strengths on predicting only future values with based on past values of the same time series. Studies about time series forecasting [Makridakis et al. 2020, Elsayed et al. 2021, Makridakis et al. 2022] have shown that deep learning models are often not the best option, with models like LightGBM and XGBoost being superior. However, there is a lack of evaluation comparing these methods against deep learning models (without being pre-trained), and foundation models in the prediction of weekly dengue cases in Brazil.

Cross-learning refers to leveraging knowledge from different but related datasets or tasks to improve a model's forecasting performance [Semenoglou et al. 2021]. In time series forecasting, this often involves training a model on multiple related time series to capture shared patterns and improve generalization, making it akin to a transfer learning approach. In contrast, cross-validation is a statistical technique for evaluating model performance by splitting the data into training and validation sets. However, traditional cross-validation is not directly applicable to time series due to the sequential nature of the data, so a leave-one-out cross-validation scheme using Federative Units can be employed instead. The key distinction is that cross-learning enhances generalization by training across multiple datasets or tasks, while cross-validation systematically tests a model's predictive ability on unseen portions of the same dataset.

Thus, we propose and evaluate an ML regression pipeline to forecast weekly dengue cases across the 27 Brazilian federative units (FUs). We compare the predictive efficiency of various models, including classical statistical techniques, shallow regressors, and deep learning methods. Among the statistical models evaluated are Random Walk (RW), Holt's linear model (SEH), Simple Exponential Smoothing (SES), Holt-Winters (HW), ARIMA, and SARIMA. We also examine deep learning architectures such as CNNs, LSTMs, and their combinations, along with foundation models (TimeGPT-1 and MOIRAI) and shallow regressors, including Linear Regression (LR), LASSO, Ridge, SVR, Random Forest (RFR), Gradient Boosting (GBR), LightGBM, and XGBoost. The goal is to assess each model's performance in forecasting weekly dengue cases across Brazilian federative units. Additionally, we evaluate dataset reformatting strategies [Benidis et al. 2022] and several training-validation approaches, such as univariate time series combinations [Petropoulos & Svetunkov 2020], time-dependent cross-validation [Bergmeir et al. 2018], and *cross-learning* [Semenoglou et al. 2021].

Our main findings include:

- Demonstrating the effectiveness of *cross-learning* in forecasting weekly dengue cases across different localities, both within Brazil (covering all 27 federative units) and beyond (San Juan, Puerto Rico);

- We show that LightGBM has better performance compared to deep learning architectures and pre-trained foundation models like TimeGPT and MOIRAI for predicting weekly dengue cases in Brazilian federative units; and
- Our proposed ML pipeline achieves comparable performance to the multivariate forecasting model trained specifically for San Juan, as presented in [Zanardo et al. 2024], suggesting that our methodology with the univariate LightGBM generalizes better.

2. Related Work

The task of forecasting dengue cases has progressed notably, with various methods proposed to enhance predictive accuracy. Shaikh et al. [Shaikh et al. 2023] developed an early warning system for San Juan and Iquitos that not only forecasts dengue incidence but also suggests preventive actions. Their approach involved outlier removal, handling of missing data, and feature selection via the Neighbour Count-based Dragonfly Electric Fish Optimization (NC-DEFO) algorithm. The selected features trained an ensemble of ANN, CNN, and SVM models—collectively called the Optimized Ensemble Classifier (OEC)—which outperformed individual models across multiple metrics, including RMSE, MAE, SMAPE, and MASE.

Similarly, Panja et al. [Panja et al. 2023] proposed XEWNet, an ensemble of autoregressive neural networks using the Haar wavelet transform for preprocessing. Their model incorporated precipitation, identified via Granger causality as a key external factor, and tested different time windows for one-step forecasts using data from San Juan, Iquitos, and Ahmedabad. The best results for San Juan were achieved with a 26-week window, yielding RMSE and MAE of 7.69 and 5.66, respectively, highlighting the value of climatic variables in improving forecast performance.

More recently, Sebastianelli et al. [Sebastianelli et al. 2024] proposed a hybrid model combining CatBoost, SVR, and LSTM, with CNN-derived spatial features and Partial Least Squares (PLS) for dimensionality reduction, achieving strong generalization to Peruvian regions. In contrast, Chen & Moraga [Chen & Moraga 2025] developed a purely LSTM-based model using SHAP-selected lagged climate variables and spatial dependencies, applying a 7-year moving window to forecast dengue cases in Brazil, outperforming simpler LSTM variants and a Bayesian baseline.

While these studies have advanced the field of dengue forecasting, they differ from our proposed approach in key aspects. Unlike Shaikh et al. [Shaikh et al. 2023], who focused on classifier ensembles with heuristic feature selection, we adopt a regression-based pipeline tailored for time series forecasting. Additionally, our study differs from Panja et al. [Panja et al. 2023] by not incorporating external climatic variables like precipitation but instead exploring *cross-learning* strategies across different localities. In contrast to Sebastianelli et al. [Sebastianelli et al. 2024] and Chen & Moraga [Chen & Moraga 2025], who employed a deep learning approach incorporating spatial information, we emphasize the efficiency of shallow regressors over deep learning models in univariate forecasting.

3. Materials and Methods

3.1. Database

For the development of this work, two datasets were used: the San Juan dataset [US National Oceanic and Atmospheric Administration 2017] and the Info-

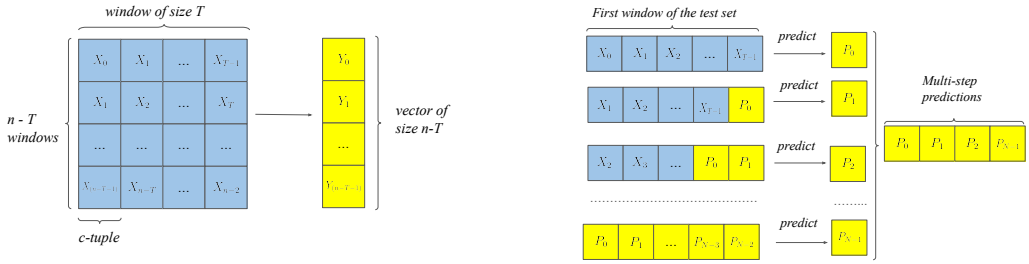
Dengue dataset [Codeco et al. 2018]. The main contents of each dataset used in our analyses include the number of weekly dengue cases, meteorological data, and demographic data.

InfoDengue is a Brazilian arbovirus alert system, using hybrid data generated by the integrated analysis of data extracted from social networks, along with climate and epidemiological data. All cities in Brazil are eligible to participate. The system currently includes 5,570 registered cities, covering all 27 Brazilian federative units. Data collection began in the first week of 2010. In our experiments, we used InfoDengue data from January 3, 2010, to May 19, 2024, focusing on the following features: data_iniSE (epidemiological week), casos (number of dengue cases), tempmin, tempmed, tempmax (average of daily minimum, mean, and maximum temperatures over the week, respectively), and umidmin, umidmed, umidmax (average of daily minimum, mean, and maximum relative humidity over the week, respectively).

The San Juan dataset comprises the number of dengue cases reported weekly in the city of San Juan (Puerto Rico) from 1990 to 2010. The dataset is in tabular format, consisting of 936 instances and 24 attributes, most of which represent time series.

3.2. Time Series Data Transformation Strategies

Shallow and deep machine learning models require dataset reformatting for supervised learning. The literature highlights two main strategies: *one-step* and *multi-step* prediction [Benidis et al. 2022]. In this work, we focus on predicting the weekly dengue cases.



(a) *One-step* scheme (b) *Multi-step* scheme

Figure 1. Illustration of one-step and multi-step processes.

As shown in Figure 1, column (a), the one-step process works as follows. Each row of the label matrix (highlighted in blue) contains T timesteps, corresponding to the chosen window size, which represents a temporal window of size T used to predict the next step. Each timestep X_j is a c-tuple, where c represents the number of selected features. The target column (highlighted in yellow) contains the predicted values, each associated with a corresponding row of the label matrix. The one-unit shift in each window explains the term *one-step*. If $c = 1$, then we have a univariate one-step approach. If $c > 1$, then we have a multivariate one-step approach.

For the multi-step process, illustrated in Figure 2, column (b), the model is first trained in a *one-step* manner using the designated training dataset. The prediction process then starts with the first window of the test set. The first predicted value, P_0, shifts the window one step left, replacing the last element X_0 with P_0. This updated window is used

to generate the next prediction, P_1. The process repeats iteratively until the prediction horizon, defined as N, is reached.

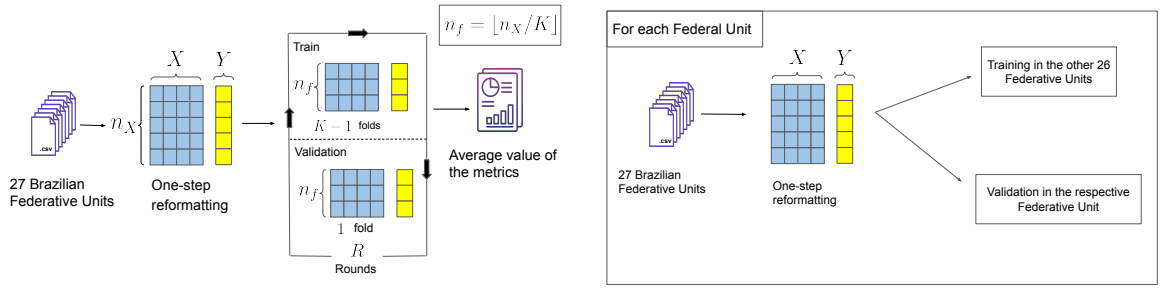
3.3. Wavelet Filtering

To analyze long-term trends in the time series while filtering out short-term fluctuations, we applied a low-pass filter. This approach ensures that rapid transitions in the data, which are captured by high-frequency components, do not interfere with the extraction of general patterns.

To implement the low-pass filter, we used wavelet decomposition, which divides the signal into high-frequency (detail) and low-frequency (approximation) components [Strang & Nguyen 1996]. The decomposition level is a hyperparameter that determines how many times the approximation component is further split into new approximation and detail components.

In our experiments, we tested all discrete wavelets available in the PyWavelets library [Lee et al. 2019], maintaining the decomposition level at one. We then reconstructed the signal using only the approximation coefficients, replacing the detail coefficients with zero values. This process ensures that the reconstructed time series maintains the same number of timesteps as the original but excludes high-frequency variations.

3.4. Randomized Cross-Learning and Leave-One-Out



(a) Randomized *cross-learning* approach (b) *Leave-one-out* (LOO) approach for Brazilian cities

Figure 2. Illustration of randomized cross-learning and leave-one-out.

Randomized cross-learning operates as follows. The process is inherently random because the training and validation sets differ in each round. It involves cross-learning, as the data comes from multiple datasets. Initially, we applied the one-step reformatting to all 27 datasets. Then, we concatenated the label matrices X (highlighted in blue) and the target columns Y (highlighted in yellow) to maintain the original relationship between each window and its corresponding target value. The resulting concatenated dataset has n_X rows.

Next, we split the dataset into K folds, each containing n_f instances. In every round r, we randomly select $K-1$ folds for training, leaving the remaining fold for validation. At the end of each round, we compute and store the performance metrics. Finally, each metric is reported as the average across all R rounds. This process is well illustrated in Figure 2, specifically in column (a).

For the leave-one-out approach, as shown in Figure 2, specifically in column (b), each dataset from the Brazilian FUs is reformatted using the one-step scheme. For each FU, the 26 datasets from the other FUs are used to train the model, while the dataset of the specific FU is reserved for validation. This strategy ensures that each model is tested in isolation with data not used during training, providing a robust evaluation of its performance across different regional contexts.

3.5. Model Evaluation

In the field of time series forecasting, the quality of a forecast model should not depend on the chosen metric, since robust methods generally perform well across a variety of metrics. Hence, when a method effectively models the data, the choice of a specific metric depends on the aspect the researcher wishes to highlight [Diamantis Koutsandreas & Assimakopoulos 2022]. To evaluate our models, we employed five metrics:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (1), \qquad \text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (2),$$

$$\text{MASE} = \frac{\frac{1}{h}\sum_{t=n+1}^{n+h}|y_t - \hat{y}_t|}{\frac{1}{n-1}\sum_{t=2}^{n}|y_t - y_{t-1}|} \quad (3), \quad \text{RMSSE} = \sqrt{\frac{\frac{1}{h}\sum_{t=n+1}^{n+h}(y_t - \hat{y}_t)^2}{\frac{1}{n-1}\sum_{t=2}^{n}(y_t - y_{t-1})^2}} \quad (4),$$

$$\text{BIAS} = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i) \quad (5),$$

where MAE stands for Mean Absolute Error, RMSE for Root Mean Squared Error, MASE for Mean Absolute Scaled Error, RMSSE for Root Mean Squared Scaled Error, and BIAS. In the definitions above, y_i represents the ground-truth values of the time series, $\hat{y}_i$ the predicted values, n the total number of observations, and h the forecast horizon. MAE and RMSE are common choices for evaluating forecast models.

However, because they are not scale-free, their interpretability can be compromised when comparing performance across multiple datasets with differently scaled time series. To address this issue, we include RMSSE [Makridakis et al. 2022] and MASE [Hyndman & Koehler 2006] in our analysis.

4. Preprocessing

For each Brazilian federative unit in the InfoDengue database, we filter characteristics using wavelets exclusively in the individual analysis of each dataset; conversely, for cross-learning approaches (using datasets from all capitals), we employ solely the weekly dengue case counts for training and validation.

In the individual analysis of each dataset, we use the multivariate *one-step* method, employing all relevant characteristics available in the dataset, as detailed previously. We

then partition the data into training and validation sets, reserving the last 100 instances exclusively for validation. Finally, we perform a comparative analysis of the models and the results obtained.

For the randomized *cross-learning* approach shown in column (a) of Figure 2, we train and validate the model using only the number of weekly dengue cases (univariate data). First, we reformatted all 27 datasets for supervised learning and concatenated the labels and targets into the matrices X and Y, respectively. Next, the number $n_f = \lfloor {}^{n_X}/_{K} \rfloor$ of elements in each *fold* was defined by taking the floor of the quotient between the number n_X of rows in matrix X and the number K of pre-defined *folds*.

After dividing into K *folds*, $K - 1$ *folds* were randomly selected to form the training set, and the remaining *fold* was used for validation. This process is termed a round. The number of rounds, R, is a hyperparameter. The final results for each metric are the average of the individual metrics calculated in each round.

The performance of the *cross-learning* approach was also analyzed through the use of *leave-one-out* (LOO). That is, for each federative unit, of the 27 datasets reformatted for supervised learning, we used 26 for model training and the *dataset* of the federative unit in question for validation, as shown in column (b) of Figure 2. For this strategy, we experimented with both *one-step* and *multi-step* schemes.

Finally, we applied the same *cross-learning* strategy using LOO, training the model with the InfoDengue database and validating it with the San Juan database, aiming to evaluate the model generalization potential.

5. Experimentation and Results

First, we performed a grid search using the K-fold strategy in LightGBM regressor to identify the optimal set of hyperparameters. Table 1 presents the hyperparameter space explored during the search. The best hyperparameters for the LightGBM *K-fold* and LightGBM LOO (*one-step* and *multi-step*) models were learning rate: 0.01; number of leaves: 127; maximum depth: 10; L1 regularization: 9; L2 regularization: 7; T: 30.

Table 1. Hyperparameter grids for experimented models.

Model	Hyperparameter grid
LightGBM	`learning rate`: [0.005, 0.01, 0.02, 0.05, 0.1] `number of leaves`: [31, 63, 127, 255] `maximum depth`: [5, 7, 10, 12, 15] `L1 regularization`: [0, 1, 3, 5, 7, 9, 11] `L2 regularization`: [0, 1, 3, 5, 7, 9] `window sizes` T: [5, 10, 15, 17, 22, 30]

For cross-learning K-fold validation, from the windows tested, the best result for the randomized approach was $T = 30$. Larger window sizes (i.e., $T > 30$) did not improve prediction performance and, in some cases, even led to worse results. In general, $R = 50$ was used for the number of rounds, and $K = 15$ *folds*. The number n_f of elements per *fold* varies according to the number of rows n_X in the label matrix X, which in turn depends on the window size T adopted in the experiment. For the best case presented in column (a) of Table 2, we have $T = 30$ and $n_f = \lfloor {}^{17710}/_{15} \rfloor = 1180$.

For the LightGBM LOO strategy, we trained the model using the best set of hyperparameters obtained by training on the 27 FUs of the InfoDengue database and validating it on the San Juan dataset. Figures 3(a)–3(b) illustrate the distributions of RMSE and RMSSE metrics across all 27 federative units, highlighting the overall superiority of the LightGBM LOO (one-step) model. Regarding the training and evaluation of the methods presented in these figures, the data processed by classical statistical models and the LightGBM LOO (*one-step* and *multi-step*) models were not filtered using *wavelets*.

Classical statistical models do not require dataset reformatting. In contrast, the shallow and deep machine learning models employed a multivariate *one-step* strategy, using the features described in the InfoDengue database section. Of the two pretrained models, MOIRAI is the only one that allows a windowing strategy, and this feature was explored. At the end of each table relating to a federative unit, the performance of LightGBM using *leave-one-out* (LOO) for both *one-step* and *multi-step* prediction is presented.

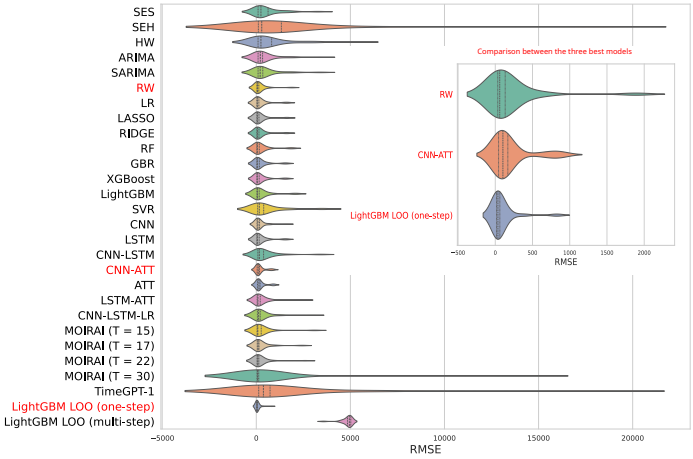
It is important to note that multivariate strategies and pre-trained models were trained separately for each of the 27 FUs. The LightGBM LOO models (one-step and multi-step) were trained on data from 26 FUs and validated on the remaining one, as explained in previous sections.

It is noted that the validation of the classical statistical models, and the shallow and deep machine learning models were done with the last hundred instances of the *dataset*. However, the validation of the LightGBM LOO (*one-step* and *multi-step*) was done with all instances of the dataset from the respective city (over 700 instances). The dataset from the city of Teresina (PI) contained only the total weekly cases column, so the approach in this case was exclusively univariate for both shallow and deep learning models.

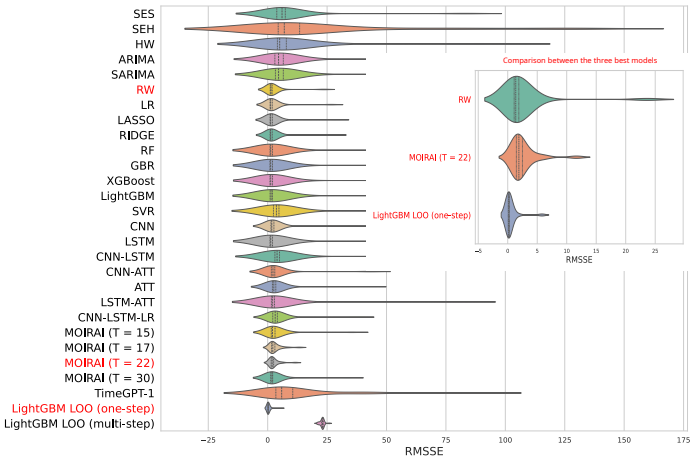
Analyzing Table 2, (a) presents the best result of the LightGBM one-step model using the randomized cross-learning approach shown in Figure 2(a); (b) shows the results obtained by training the LightGBM LOO (one-step) model on the 27 datasets from the Brazilian federative units using the best hyperparameter set found in the randomized cross-learning approach, and validating it on the full set of 936 instances from the San Juan dataset; (c) shows the results for the same LightGBM LOO (one-step) configuration as in (b), with the difference that validation was performed on the last 187 instances of the San Juan dataset. The RMSSE approaching zero indicates that LightGBM outperforms the random walk, demonstrating strong generalization.

Table 2. Performance of the LightGBM model is evaluated under three distinct scenarios. The optimal hyperparameter configuration identified in (a) was subsequently applied in scenarios (b) and (c) for comparative analysis.

	(a) Cross-Learning (K-fold)		(b) LOO one-step (all San Juan dataset)		(c) LOO one-step (last 187 instances)
Metric	**Train**	**Validation**	**Train**	**Validation**	**Validation**
MASE	0.64	0.75	0.67	0.16	0.12
RMSSE	**0.67**	**0.52**	**0.68**	**0.06**	**0.04**
MAE	27.95	31.87	28.73	6.80	4.98
RMSE	**141.66**	**104.97**	**140.78**	**11.49**	**7.55**
BIAS	-0.21	0.90	-0.24	1.89	2.24



(a) RMSE



(b) RMSSE

Figure 3. Numerical distributions of the RMSE (a) and RMSSE (b) performance metrics across all evaluated models.

6. Discussion

6.1. Infodengue database

Table 2 shows that the LightGBM *cross-learning* model with *K-fold* validation outperforms the random walk in terms of MASE and RMSSE. The higher MAE in testing compared to training is expected due to the larger training set (16,530 vs. 1,180 instances). The RMSE suggests robustness to outliers, while the BIAS metric indicates no significant tendency to underpredict or overpredict. Since the *folds* are randomly selected over 50 rounds, the model's performance remains stable despite variations in window order. This robustness motivated us to test LightGBM LOO across Brazilian federative units.

The LightGBM LOO (*one-step*) achieved the best RMSE across all Brazilian city datasets. Notably, it was validated on the full dataset, while other models used only the last 100 instances. Despite the larger validation set, it made fewer errors and showed greater robustness. The violin plot in Figure 3(a) confirms its superiority, with RMSE

values concentrated below 100 in each city.

The LightGBM LOO (*multi-step*), unlike the excellent performance of the LightGBM (*one-step*), was one of the worst models, which can be explained by the accumulation of previous errors generated at each prediction step. Therefore, due to the consecutive accumulation of errors, the model tends to perform worse across all metrics.

Classical statistical models made excessive errors, particularly SEH. This can be explained by the attempt to fit a linear model (SEH) to data with a non-linear nature. SES performed better than SEH and HW, showing performance similar to ARIMA and SARIMA. Note that, despite its simplicity, RW has metric values with distributions concentrated around lower values compared to machine learning models, with its distribution centered near the null value.

Shallow models yield similar results, with SVR performing the worst. Deep models also show similar performance but exhibit higher BIAS, suggesting poor generalization. CNN-ATT (CNN with an attention layer) stands out with the second smallest RMSE distribution and the best overall metrics among deep models. Notably, shallow models perform as well as or better than deep models without pre-training while requiring less computational effort. This raises questions about the necessity of deep learning for time series forecasting [Elsayed et al. 2021].

6.2. San Juan database

Finally, to assess the generalization potential of the best individual model across Brazilian federal units, we validated the LightGBM LOO (*one-step*) on the entire San Juan database and on the last 187 instances, with the goal of comparing it to models specifically trained on the San Juan database. The model outperformed all univariate models presented in the work [Zanardo et al. 2024].

However, the LightGBM LOO (*one-step*) was unable to surpass the models introduced in this work, such as LR (sym20) and CNN-LSTM-LR (sym20). This is possibly due to the specific characteristics of the San Juan *dataset*, on which these models were trained, as they are multivariate, whereas these characteristics are absent in the InfoDengue database.

7. Conclusions

This study investigates efficient methodologies for predicting weekly dengue cases in 27 Brazilian federative units, testing classical and machine learning models (shallow, deep, and ensemble), dataset reformatting strategies (one-step and multi-step), and preprocessing techniques like wavelets. The univariate LightGBM model with leave-one-out and one-step prediction outperformed deep models and foundational models like TimeGPT-1 and MOIRAI.

TimeGPT-1 lacked robustness, as shown by its RMSE distribution (Figure 3(a)). In contrast, MOIRAI performed better due to its *one-step* predictions and flexible window size, though its metrics worsened as the window increased (Figures 3(a) – 3(b)). Foundational models struggled due to training on diverse datasets, lack of hyperparameter tuning, and inadequate preprocessing in *zero-shot* settings. Studies like [Zeng et al. 2023] show that simple linear predictors often surpass pre-trained Transformers in forecasting, as time series data present domain-specific challenges that hinder zero-shot learning.

MASE and RMSSE were introduced because differences in dataset scales can skew performance evaluations when relying on RMSE and MAE. Interestingly, the LightGBM LOO (*one-step*) obtained the lowest values compared to other models, even though it was trained on a larger number of datasets. This highlights that, even when evaluating shallow, deep, and foundational models on a single dataset, these models, in general, did not surpass the performance of the RW.

Column (b) in Table 2 shows that the model does not suffer from *overfitting*, demonstrating good generalization capability. Therefore, although the LightGBM LOO *one-step* is not the best model in the specific context of San Juan, it demonstrates potential applicability beyond San Juan database, showing that the *cross-learning* strategy with a shallow model like LightGBM proved efficient in its univariate form, even without the use of specific *wavelet* filtering.

Despite its strong generalization capacity, the approach presents limitations, including the small validation set for San Juan, limited hyperparameter tuning in the InfoDengue dataset, and potential challenges related to high feature dimensionality. Future research may investigate PCA and alternative filtering methods (e.g., FFT, Kalman filters), incorporate seasonality, and refine multi-step forecasting strategies. Additional directions include the application of SHAP or residual analysis in univariate models, systematic comparisons between univariate and multivariate approaches, statistical significance testing, visualization of cumulative multi-step errors, and the integration of external variables such as climate and mobility data.

Acknowledgments

This work was carried out with support from the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES-PROEX) - Funding Code 001. This work was partially financed by the Fundação de Amparo à Pesquisa do Estado do Amazonas – FAPEAM – through the project POSGRAD 2024/2025. The work also counts with the support of the Centro de Inovação em Inteligência Artificial para a Saúde (CIIA-Saúde), Proc. 2020/09866-4 - FAPESP/MCTIC/CGI.

References

Anumanthan, G., Sahay, B., & Mergia, A. (2025). Current dengue virus vaccine developments and future directions. *Viruses*, 17:212.

Bavia, L., Melanda, F. N., de Arruda, T. B., Mosimann, A. L. P., Silveira, G. F., Aoki, M. N., Kuczera, D., Sarzi, M. L., Junior, W. L. C., Conchon-Costa, I., Pavanelli, W. R., dos Santos, C. N. D., Barreto, R. C., & Bordignon, J. (2020). Epidemiological study on dengue in southern brazil under the perspective of climate and poverty. *Scientific Reports*, 10:2127.

Benidis, K., Rangapuram, S. S., Flunkert, V., Wang, Y., Maddix, D., Turkmen, C., Gasthaus, J., Bohlke-Schneider, M., Salinas, D., Stella, L., Aubet, F.-X., Callot, L., & Januschowski, T. (2022). Deep learning for time series forecasting: Tutorial and literature survey. *ACM Computing Surveys*, 55(6):1–36.

Bergmeir, C., Hyndman, R. J., & Koo, B. (2018). A note on the validity of cross-validation for evaluating autoregressive time series prediction. *Computational Statistics & Data Analysis*, 120:70–83.

Chen, X. & Moraga, P. (2025). Forecasting dengue across brazil with lstm neural networks and shap-driven lagged climate and spatial effects. *BMC Public Health*, 25:973.

Codeco, C., Coelho, F., Cruz, O., Oliveira, S., Castro, T., & Bastos, L. (2018). Infodengue: A nowcasting system for the surveillance of arboviruses in brazil. *Revue d'Épidémiologie et de Santé Publique*, 66:S386. European Congress of Epidemiology "Crises, epidemiological transitions and the role of epidemiologists".

Diamantis Koutsandreas, Evangelos Spiliotis, F. P. & Assimakopoulos, V. (2022). On the selection of forecasting accuracy measures. *Journal of the Operational Research Society*, 73(5):937–954.

Elsayed, S., Thyssens, D., Rashed, A., Jomaa, H. S., & Schmidt-Thieme, L. (2021). Do we really need deep learning models for time series forecasting?

Garza, A., Challu, C., & Mergenthaler-Canseco, M. (2023). Timegpt-1.

Hyndman, R. J. & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International journal of forecasting*, 22(4):679–688.

Lee, G. R., Gommers, R., Waselewski, F., Wohlfahrt, K., & Leary, A. (2019). Pywavelets: A python package for wavelet analysis. *Journal of Open Source Software*, 4(36):1237.

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2020). The m4 competition: 100, 000 time series and 61 forecasting methods. *International Journal of Forecasting*, 36(1):54–74.

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2022). M5 accuracy competition: Results, findings, and conclusions. *International Journal of Forecasting*, 38(4):1346–1364.

Moor, M., Banerjee, O., Abad, Z. S. H., Krumholz, H. M., Leskovec, J., Topol, E. J., & Rajpurkar, P. (2023). Foundation models for generalist medical artificial intelligence. *Nature*, 616:259–266.

OPAS & OMS (2024). Atualização epidemiológica: Aumento de casos de dengue na região das américas. Acesso em: 13 jul. 2024.

Panja, M., Chakraborty, T., Nadim, S. S., Ghosh, I., Kumar, U., & Liu, N. (2023). An ensemble neural network approach to forecast dengue outbreak based on climatic condition. *Chaos, Solitons & Fractals*, 167:113124.

Petropoulos, F. & Svetunkov, I. (2020). A simple combination of univariate models. *International Journal of Forecasting*, 36(1):110–115. M4 Competition.

Rezaei, M. & Shahidi, M. (2020). Zero-shot learning and its applications from autonomous vehicles to covid-19 diagnosis: A review. *Intelligence-Based Medicine*, 3–4:100005.

Sebastianelli, A., Spiller, D., Carmo, R., Wheeler, J., Nowakowski, A., Jacobson, L. V., Kim, D., Barlevi, H., Cordero, Z. E. R., Colón-González, F. J., Lowe, R., Ullo, S. L., & Schneider, R. (2024). A reproducible ensemble machine learning approach to forecast dengue outbreaks. *Scientific Reports*, 14(1).

Semenoglou, A.-A., Spiliotis, E., Makridakis, S., & Assimakopoulos, V. (2021). Investigating the accuracy of cross-learning time series forecasting methods. *International Journal of Forecasting*, 37(3):1072–1084.

Shaikh, M. S. G., SureshKumar, D. B., & Narang, D. (2023). Development of optimized ensemble classifier for dengue fever prediction and recommendation system. *Biomedical Signal Processing and Control*, 85:104809.

Strang, G. & Nguyen, T. (1996). *Wavelets and filter banks*. Wellesley-Cambridge Press, Wellesley, MA, 2 edition.

US National Oceanic and Atmospheric Administration (2017). Dengue forecasting project website. Acessado em 21 de fevereiro de 2024.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2023). Attention is all you need.

Woo, G., Liu, C., Kumar, A., Xiong, C., Savarese, S., & Sahoo, D. (2024). Unified training of universal time series forecasting transformers. *arXiv preprint arXiv:2402.02592*.

World Health Organization (2023). Dengue - global situation. [Online; accessed 12-29-2023].

Zanardo, G., Éfren Souza, Nakamura, F., & Nakamura, E. (2024). Uma comparação entre métodos baseados em aprendizado de máquina para inferir número de casos semanais de dengue. In *Anais do LI Seminário Integrado de Software e Hardware*, pages 37–48, Porto Alegre, RS, Brasil. SBC.

Zeng, A., Chen, M., Zhang, L., & Xu, Q. (2023). Are transformers effective for time series forecasting? *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(9):11121–11128.