

Evaluating the Statistical Significance of Grid Search Optimization in Support Vector Regression for Dengue Forecasting Using Nested Time-Series Cross-Validation

Ovien Y. Caesarizky^{1, 2}, Ahmad Ilham^{1, 2, ✉}, Laelatul Khikmah³, Faiq Madani⁴

1. Department of Informatics, Universitas Muhammadiyah Semarang, Semarang, IDN

2. Intelligent Data Science Research Group, Universitas Muhammadiyah Semarang, Semarang, IDN

3. Department of Statistics, Institut Teknologi Statistika dan Bisnis Muhammadiyah Semarang, Semarang, IDN

4. Department of Artificial Intelligence, Universitas Muhammadiyah Semarang, Semarang, IDN

Received: May 09, 2026 | Review began: June 10, 2026 | Review ended: August 02, 2026 | Published: August 08, 2026

© Copyright 2026

This is an open access article distributed under the terms of the Creative Commons Attribution License CC-BY 4.0., which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract

Background

Support Vector Regression (SVR) is widely employed for epidemic forecasting because kernel methods effectively model nonlinear, multivariate, and lagged climate-disease relationships. While practitioners routinely apply exhaustive grid-search hyperparameter optimization, whether such computationally expensive tuning produces statistically significant predictive enhancements under strict, leakage-free time-series validation remains an open empirical question.

Objective

To rigorously evaluate the statistical significance and practical utility of grid-search hyperparameter optimization for SVR-based weekly dengue case forecasting.

Methods

Utilizing the publicly available DengAI benchmark for San Juan, Puerto Rico, and Iquitos, Peru, from 1990 to 2013 (N = 1,872 city-weeks), this study constructs a strictly leakage-free pipeline. This pipeline incorporates median imputation, training-exclusive Z-score outlier mitigation, standardization, and Principal Component Analysis. We compared the default SVR configuration against a grid-tuned SVR optimized over the regularization constant, epsilon-insensitive tube, and kernel coefficient. Three validation protocols were applied: a chronological 80:20 hold-out, rolling-origin time-series cross-validation, and nested time-series cross-validation utilizing five outer folds. Statistical inference relied on paired Wilcoxon signed-rank tests applied to outer-fold errors, complemented by Cohen's d_z effect sizes, percentile-bootstrap 95% confidence intervals, and Holm-Bonferroni multiplicity correction. Methodological robustness was examined by systematically varying the number of retained principal components.

Results

On the chronological hold-out evaluation, the grid search reduced the Mean Absolute Error (MAE) by a marginal 0.30% in Iquitos and 0.59% in San Juan. Under nested cross-validation, the mean difference in MAE between the tuned and default models demonstrated a 0.024-case reduction in Iquitos ($p = 1.0000$) and a 0.494-case increase in San Juan ($p = 0.1875$). Neither result achieved statistical significance following multiplicity correction. Furthermore, the selected hyperparameters exhibited severe instability across the outer validation folds, particularly for San Juan. The optimization process also incurred an additional 11.48 seconds and 14.80 seconds of computational wall-clock overhead per city, respectively, without delivering a corresponding accuracy benefit.

How to cite this article:

Caesarizky O Y, Ilham A, Khikmah L, et al. (August 08, 2026) Evaluating the Statistical Significance of Grid Search Optimization in Support Vector Regression for Dengue Forecasting Using Nested Time-Series Cross-Validation. Cureus J Comput Sci 3 : es44389-026-00212-w. DOI https://doi.org/10.7759/s44389-026-00212-w

Conclusion

Within the context of the DengAI benchmark and a robust leakage-free preprocessing pipeline, the default SVR configuration is statistically indistinguishable from its grid-tuned counterpart. These findings suggest that hyperparameter optimization in epidemic forecasting should not be executed automatically. Instead, it must be strategically guided by a pre-registered Smallest Effect Size of Interest and an explicit computational stopping rule. The comprehensive methodological framework presented in this study offers a potentially transferable framework for evaluating optimization utility in other low-fold time-series forecasting domains.

Categories: AI applications, Evaluation Methods, Machine Learning (ML)

Keywords: dengue forecasting, support vector regression, grid search, nested time-series cross-validation, statistical significance, epidemic early-warning systems

Introduction

Dengue fever continues to impose an escalating health and economic burden across tropical and subtropical regions. The World Health Organization reported that 2023 was a historic high, with more than 6.5 million cases and over 7,300 deaths recorded across more than 80 countries [1]. Indonesia exemplifies this trend: by epidemiological week 43 of 2024, national surveillance had documented 210,644 cases [1], reflecting both climatic shifts and limited entomological control capacity [2,3]. In this context, accurate weekly case forecasts are not merely of academic interest; they are operationally critical for resource allocation, vector-control planning, and early-warning systems.

The transmission dynamics of dengue are tightly coupled to environmental drivers, including ambient temperature, relative humidity, precipitation, and vegetation indices, whose effects on the mosquito life cycle and viral replication are nonlinear and frequently lagged by several weeks [3,4]. Classical linear time-series models such as AutoRegressive Integrated Moving Average and its seasonal variants remain a common baseline. For instance, Waseem et al. [5] compared Seasonal AutoRegressive Integrated Moving Average (SARIMA) and Seasonal AutoRegressive Integrated Moving Average with exogenous Variables (SARIMAX) on dengue surveillance data and reported Mean Absolute Error (MAE) values of 27.39 and 25.52, respectively, with SARIMAX benefiting from exogenous climatic regressors. Although such linear and seasonal models can be effective when the underlying dynamics are mild, they often fail to represent the multivariate nonlinearities that characterize epidemic processes, motivating the increasing adoption of machine-learning approaches for epidemic forecasting [4].

Support Vector Regression (SVR) is particularly attractive in this setting: by mapping inputs into a reproducing kernel Hilbert space, SVR can capture complex relationships while controlling model capacity through margin-based regularization [6]. However, SVR is known to be sensitive to its hyperparameters, specifically the regularization constant C , the ϵ -insensitive loss width, and the kernel coefficient γ , and its predictive performance can vary substantially across configurations [7]. The conventional response to this sensitivity is an exhaustive grid search over a predefined hyperparameter range [8,9]. The literature documenting grid search reports impressive gains in some application domains. In the COVID-19 setting, Safira et al. [10] applied SVR with grid search and time-series cross-validation, achieving a MAPE of 0.4946% and an R^2 of 85.65%, outperforming untuned baselines. Beyond epidemiology, optimized SVR has demonstrated remarkable predictive capacity across diverse forecasting domains. For instance, Seyam et al. [11] integrated SVR into a high-performance ensemble stacking framework for food demand forecasting to systematically reduce waste. Such results underscore the efficacy of SVR when coupled with rigorous optimization, a trend similarly observed in financial markets for stock price prediction [12], commodity pricing analysis [13], and complex thermal engineering applications such as radiant heat transfer prediction [14].

Yet, a more cautious reading reveals that such computational gains are inherently dataset-dependent. The relative benefit of hyperparameter tuning strongly correlates with intrinsic data properties, including variance, background noise levels, and feature redundancy [9]. In health informatics applications, the overall efficacy of machine learning is frequently driven by robust dimensionality reduction and optimal feature representation. For instance, Agatra et al. [15]

How to cite this article:

Caesarizky O Y, Ilham A, Khikmah L, et al. (August 08, 2026) Evaluating the Statistical Significance of Grid Search Optimization in Support Vector Regression for Dengue Forecasting Using Nested Time-Series Cross-Validation. *Cureus J Comput Sci* 3 : es44389-026-00212-w. DOI <https://doi.org/10.7759/s44389-026-00212-w>

demonstrated that addressing feature redundancy through transformations like Principal Component Analysis (PCA) significantly improves model stability and predictive accuracy, emphasizing that proper data geometry often dictates algorithmic success. For dengue forecasting specifically, traditional regression approaches have also been rigorously benchmarked on the DengAI dataset, where Poisson regression achieved an MAE of 25.6 [16]. A harmonized synthesis of these approaches is provided in Table 1 to facilitate a direct methodological comparison.

Ref.	Method	Dataset	Validation	Metric	Reported result
[5]	SARIMA, SARIMAX	DengAI (dengue)	—	MAE	SARIMAX 25.52; SARIMA 27.39
[10]	SVR + Grid Search (poly.)	COVID-19 (Indonesia)	Time-series CV	MAPE; R ² ; MSE	0.4946%; 85.65%; 161,606.1
[12]	SVR + Grid Search	Stocks (INDF, MYOR)	—	MAPE; R ²	5.57%; 2.95%
[16]	Poisson Regression	DengAI (dengue)	—	MAE	25.6
[13]	SVR + Grid Search	Gold prices (Indonesia)	—	MAPE	1.28%
[14]	SVR (untuned)	Industrial gas usage	—	MAPE	7.7513%

TABLE 1: Selected literature on machine learning forecasting with and without hyperparameter tuning

CV, Cross-Validation; MAE, Mean Absolute Error; MAPE, Mean Absolute Percentage Error; MSE, Mean Squared Error; SARIMA, Seasonal AutoRegressive Integrated Moving Average; SARIMAX, Seasonal AutoRegressive Integrated Moving Average with eXogenous Variables; SVR, Support Vector Regression

Two methodological gaps are visible in the synthesis above. First, although tuned SVR frequently appears to improve accuracy in non-dengue domains [10-14], dengue-focused results are mixed across model families [5,16,17], and very few studies test whether tuning effects survive temporally consistent validation on epidemic series. Many published evaluations report a single point estimate from a single random or chronological split, without quantifying whether the observed improvement could plausibly arise from sampling variability. Second, when validation does not respect temporal order, optimistic bias can inflate apparent gains in time-dependent data [18,19]. The lack of paired significance tests or effect-size reporting leaves open whether reported gains exceed sampling noise. Together, these gaps motivate a deliberate, statistically grounded re-examination of grid search as it is actually deployed in epidemic forecasting practice.

Against this background, the present study asks a focused question: Does grid-search hyperparameter tuning of SVR produce statistically significant accuracy gains for weekly dengue forecasting on the DengAI benchmark, when evaluated under a leakage-free, time-respecting validation protocol? We answer this question through three layered analyses consisting of chronological hold-out, rolling-origin cross-validation, and nested time-series cross-validation. This is coupled with paired nonparametric inference, effect-size estimation, multiplicity correction, and a sensitivity analysis over the number of retained principal components.

The contributions of this study are as follows:

How to cite this article:

1. We provide a controlled comparison of default versus grid-tuned SVR within a strictly leakage-free pipeline (imputation, scaling, and PCA fitted only on training partitions of each fold), conducted independently for the San Juan and Iquitos series.
2. We adopt nested time-series cross-validation as the inferential frame, in which inner-loop hyperparameter selection and outer-loop generalization estimation are fully decoupled. Statistical inference combines paired Wilcoxon signed-rank tests, Cohen's d_z effect sizes, percentile-bootstrap confidence intervals, and Holm-Bonferroni multiplicity correction.
3. We connect data-level variability to hyperparameter stability and computational overhead, and from these observations we articulate operational guidance for tuning, including a Smallest Effect Size of Interest (SESOI) threshold and a compute stop-rule, that is portable to other low-fold epidemic forecasting tasks.

Materials And Methods

The overall pipeline (Figure 1) consists of (i) data acquisition, (ii) leakage-free preprocessing, (iii) SVR model construction, (iv) hyperparameter optimization via grid search, and (v) evaluation under three time-respecting validation protocols, supported by formal statistical inference.

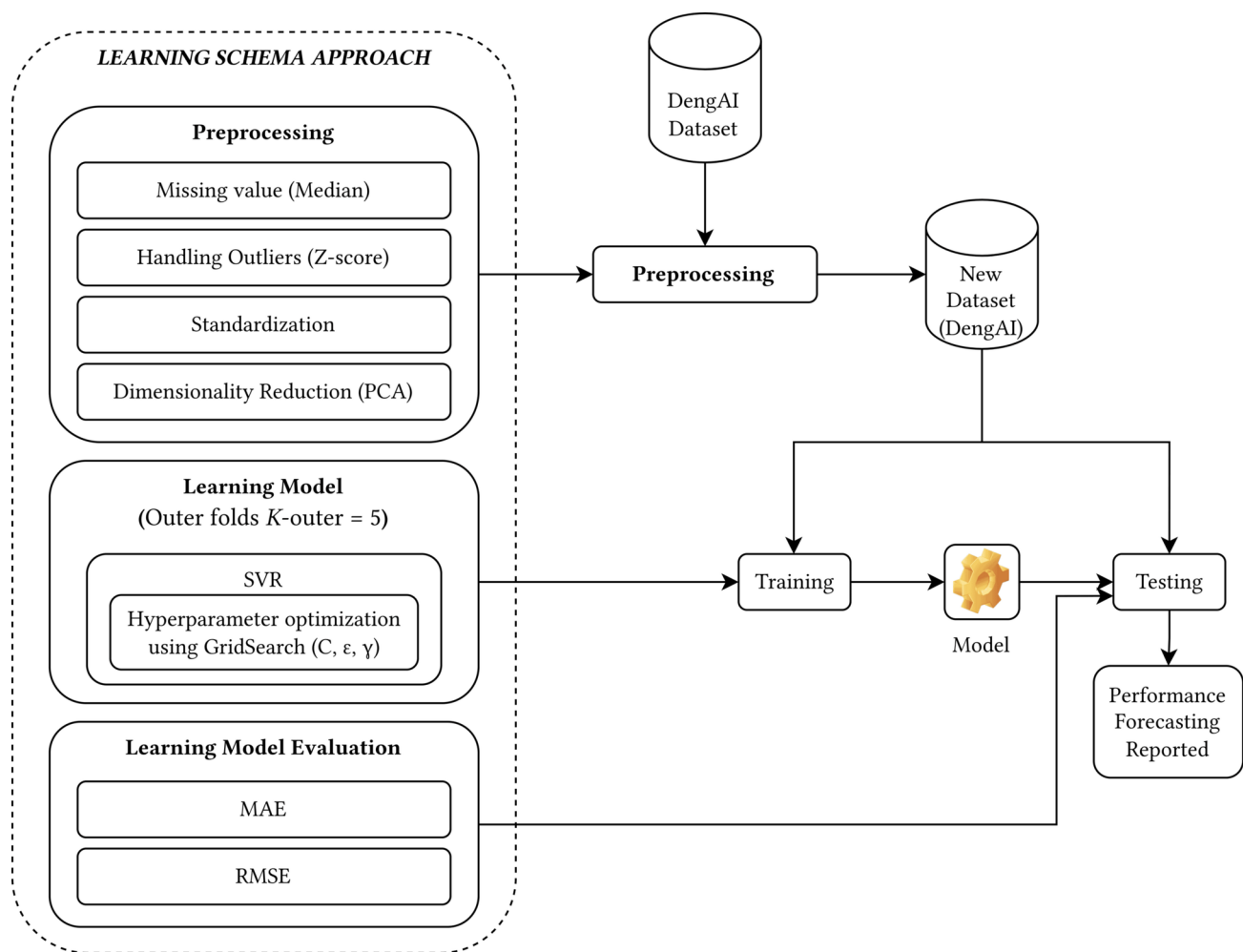


FIGURE 1: Block diagram of the proposed methodological pipeline

MAE, Mean Absolute Error; PCA, Principal Component Analysis; RMSE, Root Mean Squared Error; SVR, Support Vector Regression

Data source

We use the publicly available DengAI: Predicting Disease Spread benchmark distributed by DrivenData [20]. The dataset comprises weekly dengue case counts and environmental covariates for San Juan (Puerto Rico) and Iquitos (Peru) from 1990 to 2013, totaling 1,872 city-week observations (1,456 for training and 416 for testing in the original benchmark). Twenty-four numeric covariates span meteorological measurements (mean, dew-point, minimum, and maximum temperature; precipitation from station and satellite sources; relative and specific humidity; diurnal temperature range) and vegetation indices (Normalized Difference Vegetation Index for the four city quadrants), supplemented by temporal metadata. The target variable, `total_cases`, denotes the weekly count of reported dengue cases. A complete feature description is provided in Table 2.

How to cite this article:

Caesarizky O Y, Ilham A, Khikmah L, et al. (August 08, 2026) Evaluating the Statistical Significance of Grid Search Optimization in Support Vector Regression for Dengue Forecasting Using Nested Time-Series Cross-Validation. *Cureus J Comput Sci* 3 : es44389-026-00212-w. DOI <https://doi.org/10.7759/s44389-026-00212-w>

Evaluating the Statistical Significance of Grid Search Optimization in Support Vector Regression for Dengue Forecasting Using Nested Time-Series Cross-Validation

Feature	Description	Type	Range
city	City identifier (San Juan or Iquitos)	Categorical	sj/iq
year	Year of observation	Numeric	1990–2013
weekofyear	Week number of the year	Numeric	1–52
week_start_date	Start date of the week	Date	1990-04-30–2013-06-25
ndvi_ne/nw/se/sw	NDVI by quadrant (NE, NW, SE, SW)	Numeric	–1–1
reanalysis_avg_temp_k	Average temperature (K)	Numeric	≈270–310
reanalysis_dew_point_temp_k	Dew-point temperature (K)	Numeric	≈260–300
reanalysis_max_air_temp_k	Maximum air temperature (K)	Numeric	≈275–315
reanalysis_min_air_temp_k	Minimum air temperature (K)	Numeric	≈250–300
reanalysis_precip_amt_kg_per_m2	Precipitation amount (kg/m ²)	Numeric	0–350+
reanalysis_relative_humidity_percent	Relative humidity (%)	Numeric	0–100
reanalysis_sat_precip_amt_mm	Satellite-based precipitation (mm)	Numeric	0–350+
reanalysis_specific_humidity_g_per_kg	Specific humidity (g/kg)	Numeric	≈0–30
reanalysis_tdtr_k	Diurnal temperature range (K)	Numeric	≈0–15
station_avg_temp_c	Station average temperature (°C)	Numeric	≈15–35
station_diur_temp_rng_c	Station diurnal range (°C)	Numeric	≈0–15
station_max_temp_c	Station maximum temperature (°C)	Numeric	≈20–40
station_min_temp_c	Station minimum temperature (°C)	Numeric	≈10–30
station_precip_mm	Station precipitation (mm)	Numeric	0–350+

How to cite this article:

Caesarizky O Y, Ilham A, Khikmah L, et al. (August 08, 2026) Evaluating the Statistical Significance of Grid Search Optimization in Support Vector Regression for Dengue Forecasting Using Nested Time-Series Cross-Validation. *Cureus J Comput Sci* 3 : es44389-026-00212-w. DOI <https://doi.org/10.7759/s44389-026-00212-w>

total_cases (target)	Weekly reported dengue cases	Numeric	0–400+
----------------------	------------------------------	---------	--------

TABLE 2: Feature description of the DengAI dataset
NDVI, Normalized Difference Vegetation Index; NE, Northeast; NW, Northwest; SE, Southeast; SW, Southwest

Although the empirical evaluation is bounded by two distinct environments, the DengAI dataset serves as a widely recognized, high-variance baseline benchmark in epidemiological modeling. This localized focus allows for a controlled analysis of validation dynamics before extending the framework to broader, unstandardized regional data.

Exploratory analysis

A Pearson correlation analysis (Figure 2) reveals strong multicollinearity among temperature-related variables (for instance, *reanalysis_avg_temp_k*, *reanalysis_max_air_temp_k*, *station_avg_temp_c*, with pairwise $|r| > 0.95$) and similar redundancy across Normalized Difference Vegetation Index quadrants. By contrast, *station_precip_mm* and *satellite-derived* precipitation correlate only moderately, suggesting complementary information. These observations motivate a dimensionality reduction step prior to SVR fitting.

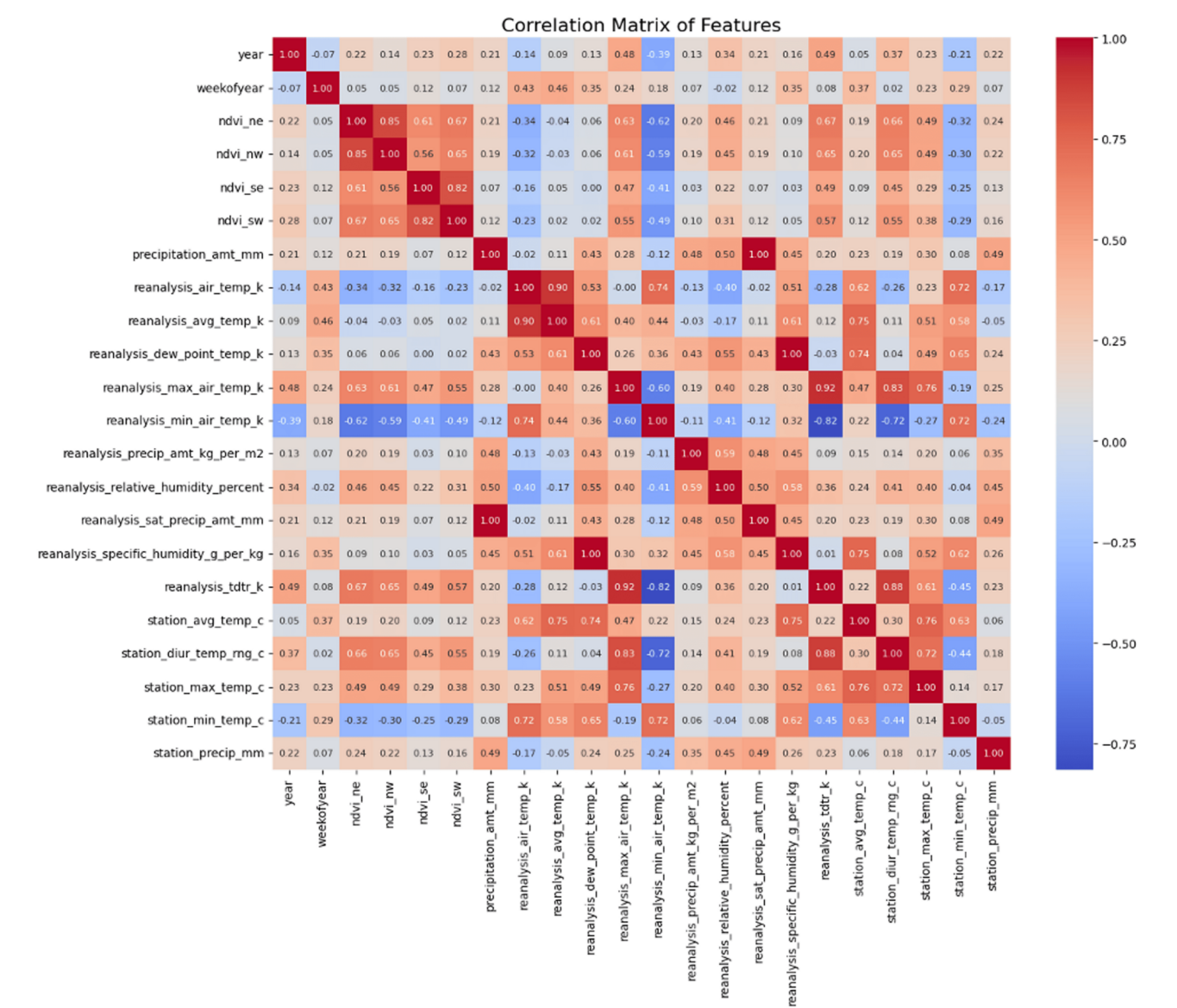


FIGURE 2: Pearson correlation heatmap of the environmental variables used in dengue case prediction

Leakage-free preprocessing pipeline

The critical role of systematic preprocessing in optimizing traditional machine learning classifiers is well-documented [21]. To ensure absolute methodological rigor, all preprocessing steps were embedded within a strictly leakage-free pipeline (Table 3). Specifically, every transformation was fit exclusively on the training partition of each validation fold or split before being applied to the corresponding validation or test partition. This stringent convention guarantees that no future information from held-out data inadvertently influences either the preprocessing parameters or the subsequent model selection. The pipeline itself was executed through four sequential stages, beginning with the imputation of missing numeric features. We employed a per-column median imputation strategy due to its inherent robustness against outliers and distributional skewness [22,23], which is essential for maintaining the theoretical consistency of supervised learning algorithms applied to incomplete datasets [24].

How to cite this article:

Following imputation, we implemented an outlier mitigation mechanism utilizing a Z-score filter ($|Z| > 3$). Crucially, this filter was estimated solely on the training partition within each temporal fold. Flagged training observations were systematically removed to stabilize the model fitting process, whereas the validation and test partitions were deliberately left intact to preserve strict chronological integrity (aggregate outlier counts on the full training set are detailed in Table 5, though fold-specific counts inherently vary with the advancing training window). Subsequently, all features were standardized to a zero mean and unit variance, relying strictly on scaling statistics derived from the training data. In the final stage, PCA was applied to orthogonalize the highly correlated environmental predictors, addressing severe multicollinearity issues that frequently degrade predictive accuracy [25]. This spatial transformation technique has been explicitly demonstrated to enhance machine learning stability and computational efficiency within health informatics contexts [15,26] and complex environmental exposure modeling [27]. We established a default configuration of $k = 12$ principal components, which empirically retained approximately 97.98% of the cumulative variance for San Juan and 96.09% for Iquitos (Table 6). To ensure that model performance was robust to this specific dimensionality choice and to systematically mitigate the risk of algorithmic overfitting [28], all experiments were systematically repeated across an evaluation space of $k \in \{8, 12, 16\}$.

Step	Description	Purpose
Missing-value imputation	Per-column median imputation on training partition only	Robust to outliers; insensitive to skewness
Outlier handling	Z-score filter ($ Z > 3$) on training partition only; flagged training rows removed	Stabilizes fitting without leakage to validation/test
Standardization	Z-score normalization with statistics fit on training only	Equalizes feature contribution to RBF kernel computation
Dimensionality reduction	PCA retaining $k = 12$ components ($\approx 97.98\%$ SJ, $\approx 96.09\%$ IQ)	Reduces dimensionality while preserving information

TABLE 3: Summary of preprocessing steps and their stated rationale

IQ, Iquitos; PCA, Principal Component Analysis; RBF, Radial Basis Function; SJ, San Juan

Imputation method	MAE
Median	21.419
Mean	21.426
Mode	24.681

TABLE 4: Comparison of missing-value imputation methods (lower MAE is better)

MAE, Mean Absolute Error

City	Outliers (count)
San Juan	87
Iquitos	51

TABLE 5: Number of outliers detected by the $|Z| > 3$ filter on the full training partition (per city)

Component	San Juan (%)	Iquitos (%)
PC1	42.72	30.48
PC2	14.54	21.95
PC3	9.77	13.60
PC4	7.60	7.42
PC5	6.46	4.96
PC6	4.14	4.37
PC7	3.20	3.52
PC8	2.83	3.11
PC9	2.32	2.13
PC10	1.85	1.74
PC11	1.32	1.48
PC12	1.22	1.33
Total	97.98	96.09

TABLE 6: Cumulative variance explained by PCA components (k = 12)

PCA, Principal Component Analysis

We deliberately did not perform ablations over preprocessing alternatives, lag engineering of the target or exogenous variables, or alternative outlier handling strategies. Such ablations would confound the attribution of error differences between hyperparameter search and pipeline design choices and, given the small number of outer folds in nested cross-validation, would further dilute statistical power. The interpretation of our results therefore strictly applies to this exact preprocessing design.

We explicitly standardized and fixed the preprocessing pipeline configuration to isolate the marginal effect of hyperparameter tuning. Introducing alternative preprocessing pathways or dynamic lag engineering within this specific focus would confound the error variances, making it unclear whether changes in forecasting accuracy stem from model optimization or data manipulation choices.

SVR: Formulation

To model the complex, nonlinear relationships inherent in epidemiological data, this study utilizes ϵ -SVR equipped with a Radial Basis Function (RBF) kernel [6]. Formally, given a training dataset $\{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^n$, where $\mathbf{x}_i \in \mathbb{R}^d$ denotes the d -dimensional input feature vector and $\mathbf{y}_i \in \mathbb{R}$ represents the corresponding continuous target variable, the formulation

How to cite this article:

Caesarizky O Y, Ilham A, Khikmah L, et al. (August 08, 2026) Evaluating the Statistical Significance of Grid Search Optimization in Support Vector Regression for Dengue Forecasting Using Nested Time-Series Cross-Validation. Cureus J Comput Sci 3 : es44389-026-00212-w. DOI <https://doi.org/10.7759/s44389-026-00212-w>

introduces w as the weight vector that defines the separating hyperplane and b as the bias term. The fundamental objective of ϵ -SVR is to determine a regression function that remains as flat as possible while strictly minimizing the empirical risk. The algorithm achieves this objective by solving the following primal optimization problem:

$$\min_{w, b, \xi, \xi^*} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \quad (1)$$

This minimization remains subject to the following set of constraints:

$$y_i - \langle w, \phi(x_i) \rangle - b \leq \epsilon + \xi_i \quad (2)$$

$$\langle w, \phi(x_i) \rangle + b - y_i \leq \epsilon + \xi_i^* \quad (3)$$

$$\xi_i, \xi_i^* \geq 0 \quad (4)$$

In this structural formulation, $\phi(\cdot)$ represents the implicit mapping of input data into a high-dimensional feature space via the RBF kernel, mathematically defined as $K(x, x') = \exp(-\gamma \|x - x'\|^2)$ [29]. Three critical hyperparameters jointly govern the behavior of the model. The regularization constant $C > 0$ dictates the trade-off between model capacity (flatness) and the tolerance for empirical errors, which the slack variables ξ_i and ξ_i^* explicitly capture. Simultaneously, $\epsilon \geq 0$ defines the width of the ϵ -insensitive loss tube, within which the model completely disregards forecasting errors. Furthermore, the kernel coefficient γ determines the spatial locality and influence radius of individual support vectors. Given that dengue epidemiological time series characteristically suffer from small sample sizes and stochastic environmental noise, the predictive stability of SVR remains acutely sensitive to these parameters. Consequently, even marginal misspecifications in the (C, ϵ, γ) configuration can drastically perturb the bias-variance trade-off, thereby necessitating a systematic and rigorous hyperparameter tuning protocol [7,9].

Grid-search hyperparameter optimization

To systematically determine the optimal model configuration, hyperparameters were optimized through an exhaustive grid search over the predefined ranges detailed in Table 7, yielding a comprehensive search space of $6 \cdot 7 \cdot 9 = 378$ distinct candidate configurations [9]. Within the architectural framework of the nested cross-validation, the inner loop executed this grid search utilizing a strict time-series splitter to rigorously preserve temporal dependencies [30]. Furthermore, the algorithm adopted negative Mean Absolute Error as the definitive scoring criterion, a choice explicitly motivated by its documented robustness in evaluating predictive forecast accuracy without excessively penalizing isolated extremes [31]. To rigorously prevent data leakage during this phase, all preprocessing transformers were computationally re-fitted exclusively on the inner training partitions. Upon identifying the optimal hyperparameter combination within the inner loop, the selected configuration was subsequently refitted on the complete outer training set prior to its final evaluation on the corresponding, unseen outer test partition. Furthermore, to maintain a precise methodological focus and explicitly isolate the empirical efficacy of exhaustive parameter tuning, alternative hyperparameter optimization algorithms, such as random search, Bayesian optimization, or ASHA-Hyperband, were deliberately excluded from the scope of this study.

How to cite this article:

Hyperparameter	Function	Tested values
C (regularization)	Trade-off between margin and error	{0.01, 0.1, 1, 10, 100, 1000}
ϵ (insensitive-loss width)	Width of the ϵ -tube in target space	{0.001, 0.01, 0.05, 0.1, 0.5, 1, 2}
γ (RBF kernel coefficient)	Kernel locality	{'scale', 'auto', 0.0001, 0.001, 0.01, 0.05, 0.1, 0.2, 1}

TABLE 7: Hyperparameter grid for SVR with RBF kernel

RBF, Radial Basis Function; SVR, Support Vector Regression

Alternative heuristic optimization techniques like Bayesian optimization or random search were intentionally excluded from this study's immediate scope. Our objective was specifically to critique exhaustive grid search, as it represents the de facto, computationally heavy baseline that practitioners routinely deploy under the assumption of automatic performance gains.

Time-respecting validation protocols

To rigorously quantify the empirical impact of grid search at progressively increasing levels of inferential rigor, this study adopts three distinct, time-respecting validation protocols. These protocols were executed independently for each city to preserve the distinct epidemiological and climatic regimes of San Juan and Iquitos. The first protocol involves a chronological 80:20 hold-out split, where the most recent 20% of weekly observations are strictly reserved as the test partition without any random shuffling. While this traditional split provides baseline comparability with prior dengue forecasting reports, it inherently produces only a single error estimate, thereby limiting statistical confidence. To address this limitation and obtain multiple paired error estimates suitable for nonparametric inference, a rolling-origin, or expanding window, cross-validation was subsequently implemented. This second protocol utilizes five expanding-window folds per city, wherein the training window iteratively grows over time while the validation window advances chronologically.

The choice of $K_{outer} = 5$ folds was strictly dictated by the temporal constraints of the DengAI dataset. In time-series cross-validation, increasing the number of outer folds excessively would either drastically shorten the initial training window or reduce the test fold sizes to a point where they no longer capture meaningful seasonal epidemic cycles.

Finally, to completely eliminate optimistic bias when assessing the true effect of hyperparameter optimization, we deployed a nested time-series cross-validation framework as the principal inferential design of this study. In this robust architecture, the inner loop performs the grid search using time-series folds with negative MAE as the definitive scoring criterion. Concurrently, the outer loop evaluates the optimized model to produce held-out test errors for direct between-model comparisons. Because inner-loop parameter selection and outer-loop generalization evaluation are entirely decoupled, this nested approach yields unbiased generalization error estimates. The complete procedural logic for this nested validation framework is formally delineated in Algorithm 1.

Algorithm 1: Nested Time-Series Cross-Validation for Hyperparameter Evaluation

Input: Time-ordered dataset $D = \{(x_t, y_t)\}$, SVR estimator, Hyperparameter space $Grid$, Number of outer folds K_{outer} , Number of inner folds K_{inner}

Output: Paired error sequences $MAE_{default}$ and MAE_{tuned} for statistical inference

How to cite this article:

Caesarizky O Y, Ilham A, Khikmah L, et al. (August 08, 2026) Evaluating the Statistical Significance of Grid Search Optimization in Support Vector Regression for Dengue Forecasting Using Nested Time-Series Cross-Validation. Cureus J Comput Sci 3 : es44389-026-00212-w. DOI <https://doi.org/10.7759/s44389-026-00212-w>

```

1: Generate  $K_{outer}$  time-series outer splits  $S_{outer} = \{(D_{train\_k}, D_{test\_k})\}_{k=1}^{K_{outer}}$ 
2: Initialize arrays  $MAE_{default} \leftarrow []$ ,  $MAE_{tuned} \leftarrow []$ 
3: For each outer fold  $k$  from 1 to  $K_{outer}$  do:
4: Fit preprocessing pipeline  $P_k$  exclusively on  $D_{train\_k}$ 
5:  $D_{train\_k}^{trans}$ ,  $D_{test\_k}^{trans} \leftarrow$  Transform  $D_{train\_k}$  and  $D_{test\_k}$  using  $P_k$ 
6:
7: // Baseline Model Evaluation
8: Fit default SVR on  $D_{train\_k}^{trans}$ 
9:  $err_{def} \leftarrow$  Compute MAE on  $D_{test\_k}^{trans}$ 
10: Append  $err_{def}$  to  $MAE_{default}$ 
11:
12: // Hyperparameter Optimization (Inner Loop)
13: Generate  $K_{inner}$  inner time-series splits within  $D_{train\_k}^{trans}$ 
14:  $BestScore \leftarrow -\infty$ 
15: For each parameter combination  $(C, \epsilon, \gamma) \in Grid$  do:
16:  $Score \leftarrow$  Compute mean inner-validation negative MAE
17: If  $Score > BestScore$  then:
18:  $BestScore \leftarrow Score$ 
19:  $(C^*, \epsilon^*, \gamma^*) \leftarrow (C, \epsilon, \gamma)$ 
20: End If
21: End For
22:
23: // Optimized Model Evaluation (Outer Loop)
24: Refit tuned SVR using optimal parameters  $(C^*, \epsilon^*, \gamma^*)$  on full  $D_{train\_k}^{trans}$ 
25:  $err_{tuned} \leftarrow$  Compute MAE on  $D_{test\_k}^{trans}$ 
26: Append  $err_{tuned}$  to  $MAE_{tuned}$ 
27: End For
28:
29: Return  $MAE_{default}$  and  $MAE_{tuned}$ 
30: Apply paired Wilcoxon signed-rank test, Cohen's  $d_z$  percentile bootstrap CI, and Holm-Bonferroni adjustment.

```

Metrics and statistical inference

To quantitatively evaluate the forecasting accuracy of the proposed models, MAE serves as the principal evaluation metric. MAE was specifically selected due to its direct interpretability in epidemiological case counts and its inherent robustness against heavy-tailed errors. Unless explicitly stated otherwise, the MAE values utilized for model comparisons correspond strictly to the outer-fold test errors derived from the nested time-series cross-validation. The mathematical formulation of MAE is defined in Equation 5.

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

As a secondary diagnostic, Root Mean Square Error (RMSE) is reported to assess the model's sensitivity to larger predictive deviations. By imposing a quadratic penalty on errors, RMSE is defined in Equation 6.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (6)$$

How to cite this article:

Furthermore, to facilitate a robust, scale-free comparison against a baseline seasonal naive forecasting approach (assuming a temporal period of $m = 52$ weeks), the Mean Absolute Scaled Error (MASE) is incorporated (Equation 7).

$$MASE = \frac{\frac{1}{n} \sum_{t=1}^n |e_t|}{\frac{1}{N-m} \sum_{t=m+1}^N |y_t - y_{t-m}|} \quad (7)$$

where $e_t = y_t - \hat{y}_t$ represents the forecast error at time t , N denotes the length of the in-sample data utilized for scaling, and n indicates the total number of forecasted observations. A resulting MASE value strictly less than 1.0 indicates that the predictive model achieves superior accuracy compared to the seasonal naive benchmark.

Beyond baseline metric comparisons, rigorous statistical inference between the default and grid-tuned SVR configurations was conducted utilizing the paired Wilcoxon signed-rank test, applied specifically to the per-fold outer MAE values. To provide a comprehensive evaluation of the results, we report Cohen's d_z as the standardized effect size for the paired differences, alongside percentile-bootstrap 95% confidence intervals generated from 10,000 resamples. To strictly control the family-wise error rate across the multiple tests conducted for the two distinct cities, Holm-Bonferroni adjusted p-values are applied. It is imperative to note that equivalence tests, such as Two One-Sided Tests, and predictive accuracy tests, such as the Diebold-Mariano test, were deliberately omitted from this methodology. Given that the nested validation architecture relies on only five outer folds, both procedures severely lack the requisite statistical power. Additionally, formal equivalence testing mandates a pre-registered SESOI, which was not specified prior to this epidemiological study. Consequently, the primary inferential objective of this research remains strictly directional rather than equivalence-based.

Computational cost and reproducibility

To rigorously quantify the computational demands of the proposed methodology, the final-fit training time during the chronological hold-out evaluation was captured by encapsulating the fitting function within a high-resolution monotonic timer. Concurrently, the total wall-clock overhead generated by the exhaustive grid search was logged independently. This strict methodological separation deliberately prevents the conflation of the inherent model fitting cost with the auxiliary hyperparameter search overhead. To guarantee computational determinism and exact reproducibility across all experimental iterations, global random seeds were permanently fixed to 42 throughout the environment. The entire analytical pipeline was engineered utilizing standard open-source Python libraries, with XGBoost explicitly integrated to serve as a standard baseline comparison. Furthermore, to ensure absolute methodological transparency and strictly prevent temporal data leakage, all preprocessing transformers were systematically re-estimated within each temporal split exclusively utilizing the respective training partitions. Ultimately, all generated empirical artifacts, including quantitative tables and graphical figures, have been meticulously preserved alongside the underlying source code to facilitate future replication and independent external validation.

Results And Discussion

Results

Results are presented at three levels of inferential rigor and complemented by a PCA-sensitivity analysis. We first establish the relative competitiveness of SVR among standard baselines, then test whether grid search improves SVR accuracy under chronological hold-out, under rolling-origin cross-validation, and under nested time-series cross-validation. We then examine fold-level hyperparameter stability and computational cost, followed by the PCA sensitivity analysis.

Experimental setup

All computational experiments were executed using Python 3.10.12, leveraging the scikit-learn library version 1.3.2 for the implementation of SVR, Principal Component Analysis, and standard preprocessing methodologies. We integrated the XGBoost library version 2.0.3 exclusively to construct the baseline comparison. To ensure a standardized evaluation of computational cost, we performed all analyses on a local computational workstation equipped with an Intel Core i7-

How to cite this article:

12700H processor and 16 GB of RAM, operating within a Windows 11 Pro environment. To guarantee exact computational reproducibility, we permanently fixed the global random seeds across all underlying numeric libraries. Furthermore, we strictly isolated the preprocessing pipeline within each cross-validation fold during execution to categorically prevent arbitrary data leakage.

Baseline model comparison

Table 8 summarizes the chronological 80:20 hold-out performance of SVR against four standard baselines - Linear Regression, Random Forest, XGBoost, and Decision Tree - under identical preprocessing. SVR achieved the lowest MAE in both cities: 7.129 in Iquitos (vs. 7.159 for Linear Regression; $\approx 0.42\%$ difference) and 19.696 in San Juan, where the gap to the nearest baseline (Linear Regression at 28.688) is substantially larger. The MASE values (0.777 in Iquitos; 0.495 in San Juan) confirm that SVR outperforms a seasonal-naïve forecast in both series. RMSE provides a more nuanced view: in Iquitos, Linear Regression attains a slightly lower RMSE (11.671) than SVR (12.504), indicating that SVR incurs occasional larger residuals although its average absolute error is smaller; in San Juan, SVR retains the best RMSE (31.188). Because operational early-warning systems prioritize robust average accuracy and direct interpretability in case counts, MAE is retained as the primary metric, with RMSE serving as a secondary diagnostic. These hold-out results justify a focused subsequent comparison between default and tuned SVR.

Model	Iquitos MAE	Iquitos MASE	Iquitos RMSE	San Juan MAE	San Juan MASE	San Juan RMSE
SVR	7.129	0.777	12.504	19.696	0.495	31.188
Linear Regression	7.159	0.780	11.671	28.688	0.721	35.022
Random Forest	7.586	0.827	12.553	31.935	0.803	43.832
XGBoost	8.383	0.914	13.145	33.588	0.844	46.431
Decision Tree	9.053	0.987	14.237	42.629	1.071	75.877

TABLE 8: Baseline model comparison on the chronological 80:20 hold-out (lower is better for all three metrics; bold indicates per-column best)

MAE, Mean Absolute Error; MASE, Mean Absolute Scaled Error; RMSE, Root Mean Squared Error; SVR, Support Vector Regression

Default vs grid-tuned SVR: Chronological hold-out

On the chronological 80:20 hold-out (Table 9), grid search produced only marginal MAE reductions: from 7.129 to 7.108 in Iquitos ($\Delta MAE = -0.30$) and from 19.696 to 19.581 in San Juan ($\Delta MAE = -0.59$). Final-fit times for the selected models were comparable across configurations (within 0.006 s). However, hold-out estimates rest on a single split and cannot, by themselves, support inference about the population-level effect of tuning. We therefore proceed to multi-fold protocols.

How to cite this article:

City	Model	MAE (hold-out)	Δ MAE vs default	Final-fit time (s)	Wilcoxon p (rolling)
Iquitos	SVR (default)	7.129	—	0.0135	—
Iquitos	SVR (grid search)	7.108	−0.30%	0.0191	0.4375
San Juan	SVR (default)	19.696	—	0.0343	—
San Juan	SVR (grid search)	19.581	−0.59%	0.0339	0.3125

TABLE 9: SVR performance before and after grid search on the chronological 80:20 hold-out, with rolling-CV Wilcoxon p-values

CV, Cross-Validation; MAE, Mean Absolute Error; SVR, Support Vector Regression

Default vs grid-tuned SVR: Rolling-origin cross-validation

Under rolling-origin cross-validation with five expanding-window folds per city, mean MAE was 7.114 ± 1.626 (default) versus 6.651 ± 1.436 (tuned) in Iquitos, and 25.514 ± 10.475 (default) versus 25.417 ± 10.473 (tuned) in San Juan (Table 10). The two-sided exact Wilcoxon signed-rank test on paired fold-level errors did not reject the null hypothesis of zero median difference (Iquitos $p = 0.4375$; San Juan $p = 0.3125$). These p -values are reported as descriptive evidence; rolling-origin CV does not fully decouple model selection from generalization estimation, and the headline inference of this study rests on the nested protocol reported next. The substantially larger fold-to-fold standard deviation observed in San Juan (≈ 10.47) compared with Iquitos (≈ 1.63) is itself informative, signaling regime heterogeneity that may render any tuning gains less stable in the larger metropolitan series.

City	SVR default—MAE (mean $\pm$ SD)	SVR tuned—MAE (mean $\pm$ SD)	Wilcoxon p
Iquitos	7.114 ± 1.626	6.651 ± 1.436	0.4375
San Juan	25.514 ± 10.475	25.417 ± 10.473	0.3125

TABLE 10: Rolling-origin cross-validation: mean $\pm$ SD of MAE and Wilcoxon p-values

MAE, Mean Absolute Error; SD, Standard Deviation; SVR, Support Vector Regression

Default vs grid-tuned SVR: Nested time-series cross-validation

Nested time-series cross-validation, our principal inferential design, reaches the same conclusion under stricter conditions. Across five outer folds per city, the mean Δ MAE (tuned – default) was -0.024 in Iquitos and $+0.494$ in San Juan (Table 11). Neither difference is statistically significant under the paired Wilcoxon test (Iquitos $p = 1.0000$; San Juan $p = 0.1875$). Holm-Bonferroni adjustment across the two cities yields adjusted p -values of 1.000 and 0.375 , respectively.

How to cite this article:

Caesarizky O Y, Ilham A, Khikmah L, et al. (August 08, 2026) Evaluating the Statistical Significance of Grid Search Optimization in Support Vector Regression for Dengue Forecasting Using Nested Time-Series Cross-Validation. Cureus J Comput Sci 3 : es44389-026-00212-w. DOI <https://doi.org/10.7759/s44389-026-00212-w>

Cohen’s d_z values are -0.085 (Iquitos) and 0.783 (San Juan); notably, the larger absolute effect size in San Juan favors the default model (tuned – default > 0 means tuned is worse), reinforcing that grid search did not deliver reliable improvements. Bootstrap 95% confidence intervals are reported as descriptive: with only five outer folds, percentile-bootstrap CIs can be unstable and may occasionally disagree with the paired test, which we therefore retain as the primary inferential statistic.

City	Outer folds	Mean ΔMAE (tuned – default)	95% CI low	95% CI high	Wilcoxon p	Holm–Bonf. p_adj	Cohen’s d_z
Iquitos	5	-0.024	-0.264	$+0.167$	1.0000	1.000	-0.085
San Juan	5	$+0.494$	$+0.033$	$+1.034$	0.1875	0.375	0.783

TABLE 11: Nested time-series cross-validation: between-model comparison of grid-tuned vs default SVR

MAE, Mean Absolute Error; SVR, Support Vector Regression

The contrast between rolling-origin and nested-CV estimates is methodologically instructive. Rolling-origin CV can appear marginally more optimistic because model selection and generalization estimation are not fully decoupled, whereas nested CV separates inner-loop selection from outer-loop estimation and therefore yields less biased generalization errors. Both protocols, nonetheless, agree on the qualitative conclusion: under our pipeline, grid search did not deliver statistically reliable accuracy gains.

Fold-level hyperparameter stability and computational cost

Table 12 reports per-fold MAE and the hyperparameters selected within each outer fold. In Iquitos, two folds favor the tuned model and three favor the default, producing a small and non-significant mean change. In San Juan, four of five folds favor the default model, yielding a positive mean ΔMAE that nonetheless remains non-significant. Hyperparameter selection is comparatively stable in Iquitos - frequently centered on $C \approx 1$ or 10 , $\epsilon \approx 1$, and $\gamma = 1$ - and visibly less stable in San Juan, where selected C spans $\{0.01, 0.1, 10\}$, ϵ spans $\{0.001, 10\}$, and γ spans $\{0.001, 1\}$. This instability across folds suggests that the inner loop is selecting from a relatively flat region of the validation-loss surface, where many configurations yield comparable performance. Furthermore, to explicitly address the preprocessing decisions leading to these outcomes, the empirical evaluation of missing-value strategies presented in Table 4 strictly justifies our adoption of median imputation over mean and mode alternatives.

How to cite this article:

City	Outer folds	MAE (Default)	MAE (Tuned)	Δ MAE (T–D)	Best hyperparameters (C, ϵ , γ)
Iquitos (IQ)	1/5	7.460	7.672	+0.212	(0.1, 0.001, 0.01)
	2/5	8.733	8.264	–0.469	(1, 1, 1)
	3/5	5.075	5.223	+0.148	(1, 1, 1)
	4/5	8.929	9.046	+0.117	(10, 1, 1)
	5/5	5.374	5.248	–0.126	(10, 1, 1)
Summary				mean Δ MAE = –0.024; 95% CI [–0.264, +0.167]; Wilcoxon p = 1.0000; Cohen's d_z = –0.085	
San Juan (SJ)	1/5	43.284	44.639	+1.355	(10, 10, 1)
	2/5	31.099	30.929	–0.170	(0.1, 0.001, 1)
	3/5	17.413	18.358	+0.945	(0.01, 10, 0.001)
	4/5	14.709	14.889	+0.180	(0.01, 0.001, 0.001)
	5/5	21.024	21.186	+0.162	(0.01, 0.001, 0.001)
Summary				mean Δ MAE = +0.494; 95% CI [+0.033, +1.034]; Wilcoxon p = 0.1875; Cohen's d_z = 0.783	

TABLE 12: Per-outer-fold MAE and selected hyperparameters in nested time-series cross-validation

MAE, Mean Absolute Error

Computational cost was assessed by separating final-fit time from grid-search wall-clock overhead. Final-fit times were nearly identical across configurations: 0.0135 s (default) and 0.0191 s (tuned) in Iquitos; 0.0343 s and 0.0339 s in San Juan. Grid search itself, however, added approximately 11.48 s to the wall-clock time in Iquitos and 14.80 s in San Juan - roughly two to three orders of magnitude greater than the final-fit cost. Because no statistically significant accuracy gain accompanies this overhead, the cost-benefit analysis under our pipeline favors the default configuration. Figure 3

How to cite this article:

Caesarizky O Y, Ilham A, Khikmah L, et al. (August 08, 2026) Evaluating the Statistical Significance of Grid Search Optimization in Support Vector Regression for Dengue Forecasting Using Nested Time-Series Cross-Validation. Cureus J Comput Sci 3 : es44389-026-00212-w. DOI <https://doi.org/10.7759/s44389-026-00212-w>

visualizes the rolling-CV impact of tuning on MAE; Figure 4 contrasts final-fit times by city. Figures 5 and 6 plot observed versus predicted weekly cases on the chronological hold-out for San Juan and Iquitos, respectively, where the visual similarity between tuned and default trajectories matches the negligible Δ MAE values reported above. Specifically, an in-depth analysis of Figure 5 demonstrates that both the default and tuned SVR models successfully capture the general seasonal epidemic peaks in San Juan, yet their overlapping trajectories visually confirm that the tuned model offers no significant deviation, as both models slightly underpredict the most extreme anomaly spikes. A nearly identical behavior is visually confirmed in Figure 6 for Iquitos, where the temporal alignment between the observed cases and both SVR configurations tightly corresponds to the stable model performance.

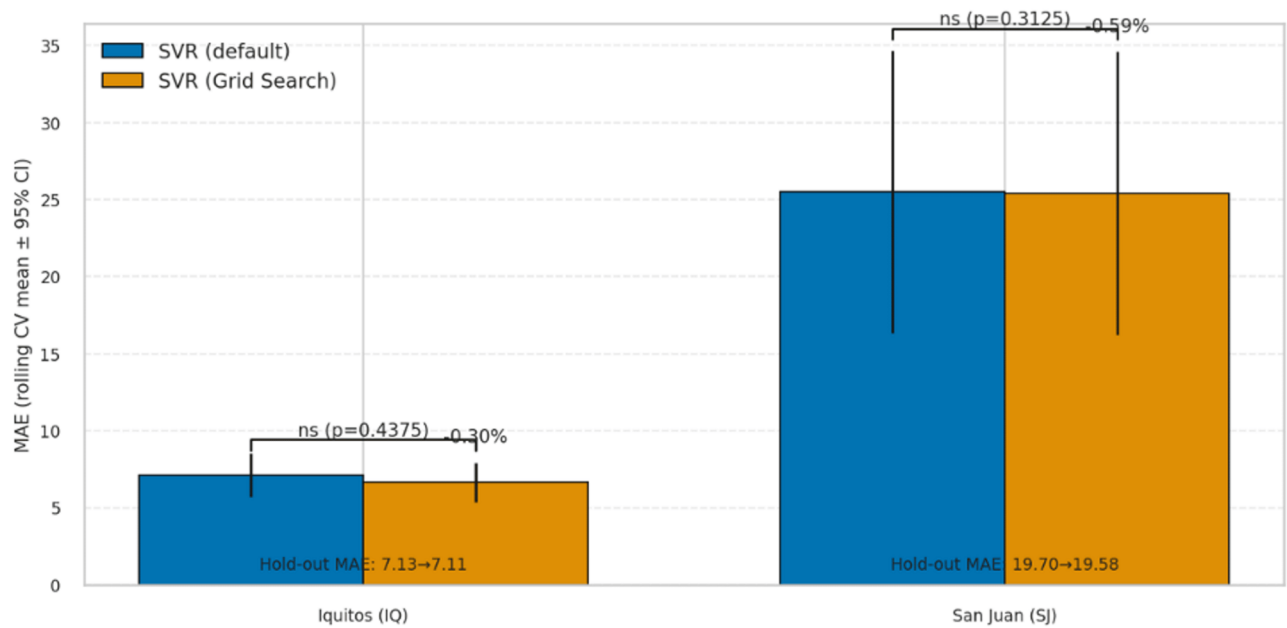


FIGURE 3: Impact of grid-search tuning on MAE by city (rolling-CV mean MAE; error bars = 95% percentile-bootstrap CIs)
CI, Confidence Interval; CV, Cross-Validation; MAE, Mean Absolute Error; SVR, Support Vector Regression

How to cite this article:

Caesarizky O Y, Ilham A, Khikmah L, et al. (August 08, 2026) Evaluating the Statistical Significance of Grid Search Optimization in Support Vector Regression for Dengue Forecasting Using Nested Time-Series Cross-Validation. Cureus J Comput Sci 3 : es44389-026-00212-w. DOI <https://doi.org/10.7759/s44389-026-00212-w>

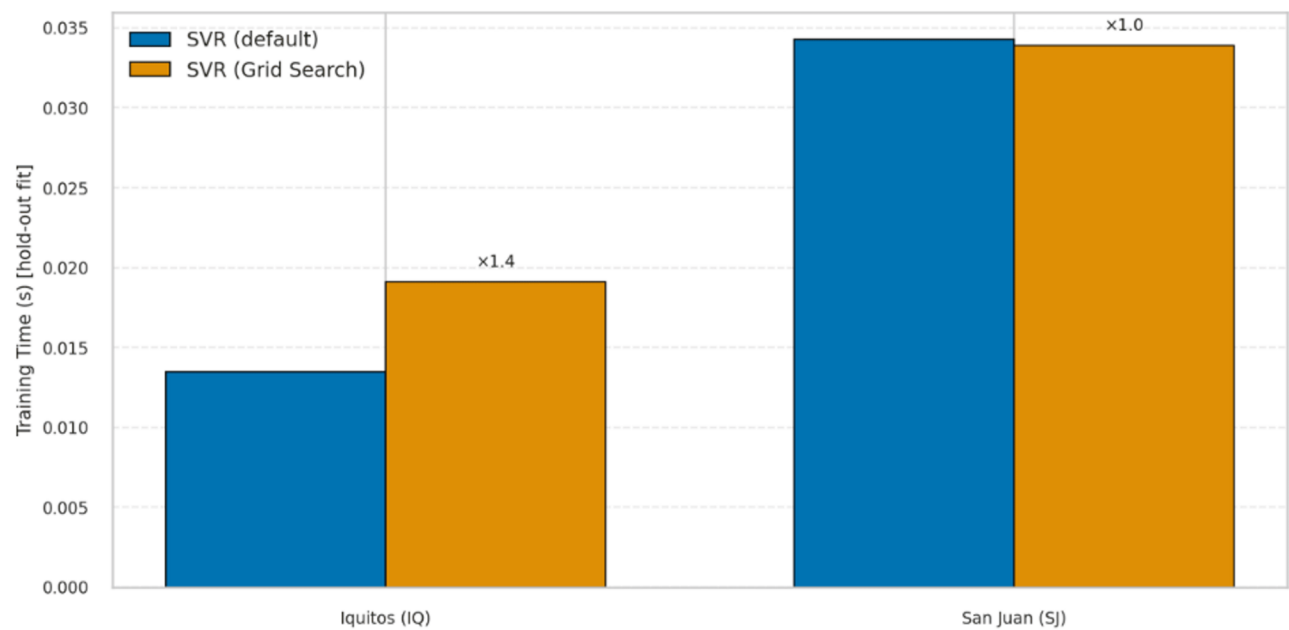


FIGURE 4: Final-fit training time by city for hold-out fits (grid-search overhead excluded; reported in main text)

SVR, Support Vector Regression

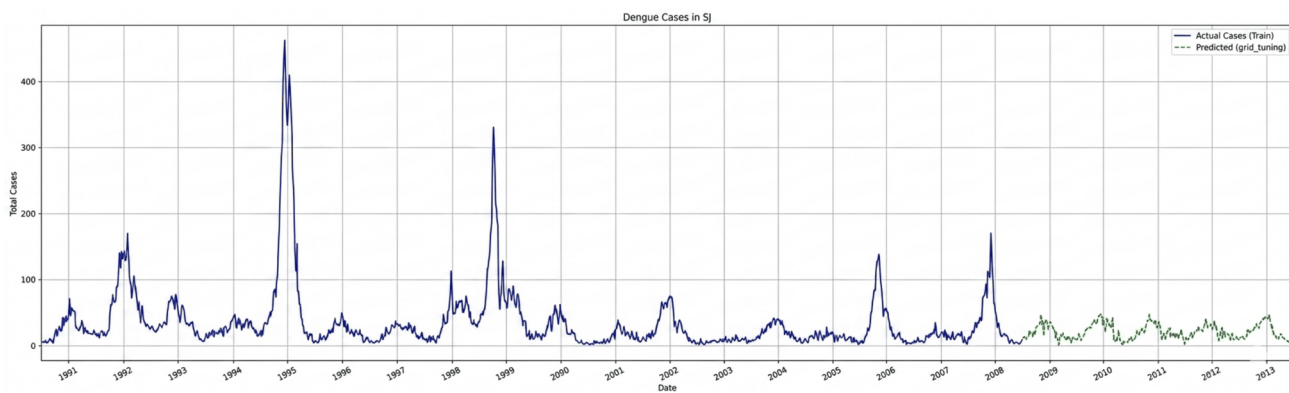


FIGURE 5: San Juan: observed vs SVR-predicted weekly cases on the chronological hold-out (most recent ~20% of weeks)

SVR, Support Vector Regression

How to cite this article:

Caesarizky O Y, Ilham A, Khikmah L, et al. (August 08, 2026) Evaluating the Statistical Significance of Grid Search Optimization in Support Vector Regression for Dengue Forecasting Using Nested Time-Series Cross-Validation. Cureus J Comput Sci 3 : es44389-026-00212-w. DOI <https://doi.org/10.7759/s44389-026-00212-w>

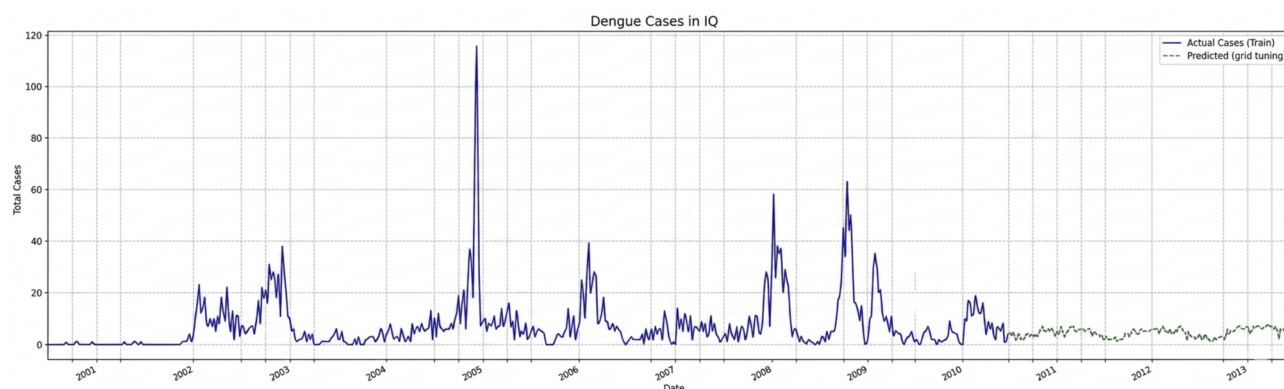


FIGURE 6: Iquitos: observed vs SVR-predicted weekly cases on the chronological hold-out (most recent ~20% of weeks)

SVR, Support Vector Regression

PCA sensitivity analysis

Robustness to dimensionality choices was examined by varying $k \in 8, 12, 16$ within the same leakage-free pipeline (Table 13). Qualitative conclusions are unchanged: tuned-versus-default differences remain small and statistically non-significant under both rolling-origin and nested time-series CV across all three values of k . In Iquitos, nested ΔMAE ranges from -0.124 to +0.187 (all non-significant); in San Juan, nested ΔMAE ranges from +0.449 to +0.537 (all non-significant). This sensitivity analysis confirms that the principal finding does not depend on the specific choice of $k = 12$.

How to cite this article:

Caesarizky O Y, Ilham A, Khikmah L, et al. (August 08, 2026) Evaluating the Statistical Significance of Grid Search Optimization in Support Vector Regression for Dengue Forecasting Using Nested Time-Series Cross-Validation. Cureus J Comput Sci 3 : es44389-026-00212-w. DOI <https://doi.org/10.7759/s44389-026-00212-w>

City	k	Rolling MAE default (mean±SD)	Rolling MAE tuned (mean±SD)	Rolling Wilcoxon p	Nested ΔMAE mean	95% CI low	95% CI high	Nested Wilcoxon p	Cohen's d _z	Grid time (s)
IQ	8	7.158 ± 1.820	6.655 ± 1.615	0.4375	+0.187	−0.248	+0.758	0.6250	+0.291	11.20
IQ	12	7.114 ± 1.626	6.651 ± 1.436	0.4375	−0.024	−0.264	+0.167	1.0000	−0.085	8.24
IQ	16	7.113 ± 1.794	6.590 ± 1.581	0.3125	−0.124	−0.560	+0.187	1.0000	−0.253	8.43
SJ	8	25.454 ± 11.697	25.446 ± 11.656	0.8125	+0.537	+0.080	+1.053	0.1250	+0.873	15.67
SJ	12	25.506 ± 11.721	25.396 ± 11.722	0.1250	+0.494	+0.033	+1.034	0.1875	+0.783	17.06
SJ	16	25.542 ± 11.684	25.428 ± 11.642	0.1250	+0.449	−0.047	+0.992	0.3125	+0.677	16.15

TABLE 13: PCA sensitivity: rolling-CV and nested-CV comparisons of grid-tuned vs default SVR for k ∈ {8, 12, 16}

CV, Cross-Validation; IQ, Iquitos; MAE, Mean Absolute Error; PCA, Principal Component Analysis; SD, Standard Deviation; SJ, San Juan; SVR, Support Vector Regression

Discussion

Why Default SVR Is Near-Optimal in This Setting

The consistent convergence of three distinct validation frameworks (chronological hold-out, rolling-origin cross-validation, and nested time-series cross-validation) upon the exact same qualitative conclusion necessitates a robust theoretical interpretation. Several underlying factors plausibly contribute to the near-optimality of the default SVR configuration within the proposed experimental pipeline. First, the rigorous leakage-free preprocessing chain, which incorporates median imputation, training-exclusive Z-score outlier mitigation, standardization, and Principal Component Analysis retaining approximately 96% to 98% of the dataset variance, functions as a remarkably strong inductive prior. This comprehensive preprocessing effectively neutralizes scale heterogeneity and structural multicollinearity, ensuring that the residual signal remains mathematically well-behaved within the default operational boundaries of the RBF kernel.

Second, the default kernel coefficient dynamically adapts to the variance of the standardized, PCA-transformed inputs. This autonomous algorithmic adjustment supplies a data-driven kernel bandwidth that is inherently suited for the processed feature geometry. Third, despite the presence of inherent epidemiological noise, the dengue surveillance series investigated in this study exhibits a stationary seasonal pattern once the exogenous climatic drivers are isolated via PCA. This spatial transformation leaves minimal residual nonlinearity for an exhaustive grid search to mathematically

How to cite this article:

Caesarizky O Y, Ilham A, Khikmah L, et al. (August 08, 2026) Evaluating the Statistical Significance of Grid Search Optimization in Support Vector Regression for Dengue Forecasting Using Nested Time-Series Cross-Validation. Cureus J Comput Sci 3 : es44389-026-00212-w. DOI <https://doi.org/10.7759/s44389-026-00212-w>

exploit. Finally, the observed instability of the selected hyperparameters across the temporal folds for San Juan (as detailed in Table 12) strongly implies the existence of a distinctly flat validation-loss surface. When multiple hyperparameter configurations yield nearly identical performance metrics during the inner-loop validation, the subsequent parameter selection acts primarily as stochastic noise. Therefore, any perceived predictive advantage observed on a singular data split is highly likely to be an artifact of random sampling variance rather than a genuine, systematic improvement in model generalization.

Practical Implications for Epidemic Forecasting Pipelines

These findings present two critical operational implications for the development of future epidemic forecasting pipelines. First, in the absence of a pre-specified Smallest Effect Size of Interest (SESOI), exhaustive grid search consumes valuable computational resources without delivering practically meaningful predictive gains. Within our experimental framework, allocating a computational search budget of 11 to 15 seconds per city yielded mean differences in Mean Absolute Error of merely a 0.024-case reduction for Iquitos and a 0.494-case increase for San Juan. These marginal statistical deviations are ultimately dwarfed by the typical reporting noise inherent in standard dengue surveillance systems. Consequently, we strongly recommend that public health practitioners and data scientists pre-register a SESOI grounded in tangible operational decision-making. Such a metric could represent, for instance, the specific case-count threshold required to modify a clinical early-warning vector control trigger. Researchers should subsequently proceed with hyperparameter tuning only when an inner-loop performance gain explicitly meets or exceeds this predefined operational threshold. Second, in scenarios where grid search is computationally performed, our empirical results strongly advocate for the implementation of nested time-series cross-validation rather than standard non-nested validation as the primary inferential design. The nested evaluation architecture transparently exposes the statistical cost of optimistic bias and fundamentally forces the analyst to critically confront the inherent temporal instability of the selected hyperparameters across evolving forecasting windows.

Comparison With Prior Literature

Our empirical results align consistently with the heterogeneous evidence previously summarized in Section 1. The existing reports of substantial performance gains derived from optimized SVR configurations in non-dengue domains [10,12,13,16,17] are not contradicted by our study. Instead, our findings clearly indicate that such algorithmic improvements are highly conditional on specific data characteristics, such as internal variance, background noise levels, and feature geometry. Furthermore, these reported gains heavily depend on the selected validation protocols, which may inadvertently introduce optimistic bias if temporal dependencies are ignored. In terms of absolute predictive accuracy, our point estimates for the default SVR configuration, which achieved an MAE of 7.129 in Iquitos and 19.696 in San Juan, demonstrate a numerically lower MAE than the Poisson regression baseline of 25.6 reported on the DengAI dataset by Baker et al. [16] and the SARIMAX error of 25.52 reported by Waseem et al. [5]. However, inherent differences in the respective validation protocols necessitate caution when making direct performance comparisons. Consequently, the principal contribution of the present work is fundamentally methodological. This study provides a highly transferable analytical template for rigorously evaluating the practical utility of hyperparameter optimization, comprehensively supported by paired Wilcoxon inference, effect-size reporting, multiplicity correction, and a robust dimensionality sensitivity analysis.

Threats to Validity

Several threats to validity merit explicit discussion to properly contextualize the findings of this study. First, regarding statistical power, the reliance on five outer folds per city while necessary to preserve meaningful seasonal epidemic cycles within the training windows inherently limits the sensitivity of the paired Wilcoxon test. Consequently, a null result does not establish statistical equivalence but merely indicates that the available evidence is insufficient to reject the null hypothesis. Formal equivalence testing was not performed because a Smallest Effect Size of Interest was not pre-specified prior to the experiments.

How to cite this article:

Second, although the DengAI dataset serves as a robust and widely recognized baseline in epidemiological modeling, the generalizability of this study is bounded by its empirical scope, which is strictly restricted to two cities at a weekly resolution within the single DengAI benchmark dataset. Therefore, these conclusions cannot be directly extrapolated to other infectious diseases, diverse geographical regions, or different temporal resolutions without explicit future replication.

Third, concerning pipeline specificity, the preprocessing architecture was deliberately fixed to isolate the precise marginal effect of the grid search on the SVR model without introducing confounding variables. Alternative preprocessing decisions, particularly the engineering of target or exogenous lags and the application of different outlier mitigation strategies, could potentially alter both the absolute error levels and the relative benefits derived from hyperparameter tuning.

Fourth, the scope of the optimization was inherently constrained by a bounded search grid. The exploration of finer or wider grid structures, alternative hyperparameter optimization algorithms such as random search or Bayesian methods, and different kernel functions was deliberately omitted, making the current conclusions strictly conditional on the predefined parameter configurations.

Finally, classical time-series baseline models, including SARIMA, Error-Trend-Seasonality frameworks, and Prophet, were not incorporated into the within-fold time-series cross-validation. The inclusion of these traditional forecasting architectures in future research would significantly strengthen the external comparability of the proposed machine learning methodology.

Ultimately, while these specific constraints were deliberately imposed to guarantee strict internal validity and perfectly isolate the marginal effects of hyperparameter tuning, they present clear and well-defined opportunities for future methodological extensions.

Conclusions

This study evaluated whether grid-search hyperparameter optimization yields statistically significant accuracy gains for Support Vector Regression in weekly dengue case forecasting utilizing the DengAI benchmark for San Juan and Iquitos. Across three time-respecting validation protocols, specifically chronological hold-out, rolling-origin cross-validation, and nested time-series cross-validation, as well as a sensitivity analysis regarding the number of retained principal components, the grid-tuned model was statistically indistinguishable from the default configuration. The mean differences in Mean Absolute Error under nested cross-validation demonstrated a marginal reduction of 0.024 cases in Iquitos ($p = 1.0000$; Holm-Bonferroni adjusted $p = 1.000$; $d_z = -0.085$) and an increase of 0.494 cases in San Juan ($p = 0.1875$; adjusted $p = 0.375$; $d_z = 0.783$). Furthermore, the grid search procedure added 11.48 seconds and 14.80 seconds of computational wall-clock overhead in Iquitos and San Juan, respectively, without delivering a corresponding accuracy benefit. The selected hyperparameters remained notably unstable across the outer folds, particularly in San Juan, which is highly consistent with the presence of a flat validation-loss surface.

Methodologically, this study contributes a transferable framework for assessing the true value of hyperparameter optimization in low-fold time-series settings. This template integrates a leakage-free pipeline embedded within a nested time-series cross-validation framework, supported by paired Wilcoxon inference, standardized effect sizes, percentile-bootstrap confidence intervals, and Holm-Bonferroni multiplicity correction. It also establishes an explicit pre-registration recommendation concerning a SESOI and a computational stopping rule. In practical terms, these results imply that in epidemic forecasting pipelines where the preprocessing phase already adequately controls for scale heterogeneity, multicollinearity, and outliers, the default Support Vector Regression configuration can remain highly competitive. Consequently, performing routine grid search acts as an inefficient allocation of computational resources unless a specific effect size threshold is pre-specified and empirically met.

How to cite this article:

The primary limitations of this research include the restricted focus on two cities at a weekly resolution, the utilization of a fixed preprocessing pipeline without extensive lag engineering, the exclusive reliance on grid search among available optimization methods, the absence of formal equivalence testing, and the modest statistical power afforded by the five outer folds. To address these boundaries, future work will incorporate classical baseline models, including SARIMA, ETS, and Prophet, evaluated under the identical nested protocol. Subsequent studies will also compare grid search against alternative optimization algorithms such as random search, Bayesian optimization, and successive-halving methods. Additionally, future research will introduce target and exogenous lag engineering to assess whether the observed null result regarding hyperparameter tuning persists. Finally, investigations will extend this experimental design to multi-city spatio-temporal settings utilizing blocked or rolling resampling with larger outer-fold counts, while fully pre-registering the effect size thresholds, search grids, and stopping rules to maximize reproducibility and operational relevance.

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Concept and design: Ahmad Ilham, Laelatul Khikmah, Ovien Y. Caesarizky

Acquisition, analysis, or interpretation of data: Ahmad Ilham, Laelatul Khikmah, Ovien Y. Caesarizky, Faiq Madani

Critical review of the manuscript for important intellectual content: Ahmad Ilham, Laelatul Khikmah, Faiq Madani

Supervision: Ahmad Ilham

Drafting of the manuscript: Ovien Y. Caesarizky, Faiq Madani

Disclosures

Human subjects: All authors have confirmed that this study did not involve human participants or tissue. **Animal subjects:** All authors have confirmed that this study did not involve animal subjects or tissue. **Conflicts of interest:** In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

Data Availability Statements

The data and code supporting this study are openly available in the GitHub repository at <https://github.com/pacekonathyo/Dengue-SVR-Optimization> (Release v1.0.0, Accessed: August 3, 2026)

Acknowledgements

The authors thank the Intelligent Data Science Research Group (IDSRG) at Universitas Muhammadiyah Semarang for constructive scientific discussions. For language refinement, Grammarly was utilized exclusively for grammatical correction and proofreading. No artificial intelligence (AI) tools were employed in the conceptualization, experimental design, data analysis, or interpretation of results.

References

1. WHO: Dengue. (2025). Accessed: May 9, 2026: <https://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue>.
2. Tansil MG, Rampengan NH, Wilar R: Risk factors for dengue hemorrhagic fever in children. Jurnal Biomedik JBM. 2021, 13:90-99. 10.35790/jbm.13.1.2021.31760

How to cite this article:

Caesarizky O Y, Ilham A, Khikmah L, et al. (August 08, 2026) Evaluating the Statistical Significance of Grid Search Optimization in Support Vector Regression for Dengue Forecasting Using Nested Time-Series Cross-Validation. Cureus J Comput Sci 3 : es44389-026-00212-w. DOI <https://doi.org/10.7759/s44389-026-00212-w>

3. Widyantoro W, Nurjazuli N, Darundianti YH: [The relationship between weather factors and the incidence of dengue fever in Bantul Regency \(Article in Indonesian\)](#). Jurnal Aisyah: Jurnal Ilmu Kesehatan. 2021, 6:823-830. [10.30604/jika.v6i4.863](#)
4. Juraphanthong W, Kesorn K: [Time series data enrichment using semantic information for dengue incidence forecasting](#). Science, Engineering and Health Studies. 2021, 15:21050013. [10.14456/sehs.2021.50](#)
5. Waseem MD, Nawaz A, Rasheed U, Raza A, Albarka MO: [A time series regression-based model for predicting the spread of dengue disease](#). 2023 International Conference on Robotics and Automation in Industry (ICRAI), Peshawar, Pakistan. 2023, 1-6. [10.1109/ICRAI57502.2023.10089545](#)
6. Roy A, Chakraborty S: [Support vector machine in structural reliability analysis: A review](#). Reliability Engineering & System Safety. 2023, 233:109126. [10.1016/j.res.2023.109126](#)
7. Hoque KE, Aljamaan H: [Impact of hyperparameter tuning on machine learning models in stock price forecasting](#). IEEE Access. 2021, 9:163815-163830. [10.1109/access.2021.3134138](#)
8. Ngoc TT, Van Dai L, Thuyen CM: [Support vector regression based on grid search method of hyperparameters for load forecasting](#). Acta Polytechnica Hungarica. 2021, 18:143-158. [10.12700/aph.18.2.2021.2.8](#)
9. Açikkar M: [Fast grid search: A grid search-inspired algorithm for optimizing hyperparameters of support vector regression](#). Turkish Journal of Electrical Engineering and Computer Sciences. 2024, 32:68-92. [10.55730/1300-0632.4056](#)
10. Safira AN, Warsito B, Rusgiyono A: [Analisis support vector regression \(SVR\) dengan algoritma grid search time series cross validation untuk prediksi jumlah kasus terkonfirmasi COVID-19 di Indonesia](#). Jurnal Gaussian. 2023, 11:512-521. [10.14710/j.gauss.11.4.512-521](#)
11. Seyam A, Mathew SS, Du B, El Barachi M, Shen J: [A stacking ensemble model for food demand forecasting: A preventative approach to food waste reduction](#). Cleaner Logistics and Supply Chain. 2025, 15:100225. [10.1016/j.clscn.2025.100225](#)
12. Shih CM, Yang KC, Chao WP: [Predicting stock price using random forest algorithm and support vector machines algorithm](#). 2023 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), Singapore, Singapore. 2023, 133-137. [10.1109/IEEM58616.2023.10406964](#)
13. Annas S, Rais Z, Aswi A, Indrayasari, Nurfajriani: [Implementation of Support Vector Regression \(SVR\) Analysis in Predicting Gold Prices in Indonesia](#). Proceedings of the 5th International Conference on Statistics, Mathematics, Teaching, and Research 2023 (ICSMTR 2023). Atlantis Press, The Netherlands; 2023. 97-107. [10.2991/978-94-6463-332-0_12](#)
14. Su X, Jiang H, Qin T, Lin G: [Particle swarm optimization-support vector regression \(PSO-SVR\)-based rapid prediction method for radiant heat transfer for a spacecraft vacuum thermal test](#). Applied Sciences. 2024, 14:9407. [10.3390/app14209407](#)
15. Agatra DFN, Marutho D, Xiong X: [Evaluating PCA and LDA to improve machine learning classification of consumer behavior in health informatics](#). Journal of Intelligent Computing & Health Informatics. 2025, 6:18-27. [10.26714/jichi.v6i1.16176](#)
16. Baker QB, Faraj D, Alguzo A: [Forecasting dengue fever using machine learning regression techniques](#). 2021 12th International Conference on Information and Communication Systems (ICICS), Valencia, Spain. 2021, 157-163. [10.1109/ICICS52457.2021.9464619](#)
17. Gupta G, Khan S, Guleria V, et al.: [DDPM: a dengue disease prediction and diagnosis model using sentiment analysis and machine learning algorithms](#). Diagnostics. 2023, 13:1093. [10.3390/diagnostics13061093](#)
18. Tashman LJ: [Out-of-sample tests of forecasting accuracy: an analysis and review](#). International Journal of Forecasting. 2000, 16:437-450. [10.1016/S0169-2070\(00\)00065-0](#)
19. Makridakis S, Kirkham R, Wakefield A, Papadaki M, Kirkham J, Long L: [Forecasting, uncertainty and risk; perspectives on clinical decision-making in preventive and curative medicine](#). International Journal of Forecasting. 2019, 35:659-666. [10.1016/j.ijforecast.2017.11.003](#)
20. DrivenData: [DengAI: Predicting Disease Spread](#). (2015). Accessed: May 9, 2026: <https://www.drivendata.org/competitions/44/dengai-predicting-disease-spread/>.
21. Siino M, Tinnirello I, La Cascia M: [Is text preprocessing still worth the time? A comparative survey on the influence of popular preprocessing methods on Transformers and traditional classifiers](#). Information Systems. 2024, 121:102342. [10.1016/j.is.2023.102342](#)
22. Hasan MK, Alam MA, Roy S, Dutta A, Jawad MT, Das S: [Missing value imputation affects the performance of machine learning: A review and analysis of the literature \(2010-2021\)](#). Informatics in Medicine Unlocked. 2021, 27:100799. [10.1016/j.imu.2021.100799](#)

How to cite this article:

Caesarizky O Y, Ilham A, Khikmah L, et al. (August 08, 2026) Evaluating the Statistical Significance of Grid Search Optimization in Support Vector Regression for Dengue Forecasting Using Nested Time-Series Cross-Validation. Cureus J Comput Sci : es44389-026-00212-w. DOI <https://doi.org/10.7759/s44389-026-00212-w>

23. Berkelmans GFN, Read SH, Gudbjörnsdóttir S, et al.: [Population median imputation was noninferior to complex approaches for imputing missing values in cardiovascular prediction models in clinical practice](#). Journal of Clinical Epidemiology. 2022, 145:70-80. [10.1016/j.jclinepi.2022.01.011](#)
24. Josse J, Chen JM, Prost N, Varoquaux G, Scornet E: [On the consistency of supervised learning with missing values](#). Statistical Papers. 2024, 65:5447-5479. [10.1007/s00362-024-01550-4](#)
25. Singh J, Goyal D, Vashist A: [Feature selection using correlation analysis for accurate breast cancer diagnosis](#). Applications of Synthetic High Dimensional Data. Sobczak-Michalowska M, Borah S, Polkowski Z, Mishra S (ed): IGI Global Scientific Publishing, Hershey, PA; 2024. 107-119. [10.4018/979-8-3693-1886-7.ch007](#)
26. BaniMustafa A, Almatarneh S, Bulkrock O, Alazaidah R, Almahasneh H, Samara G: [Investigating principal component analysis impact on the performance of machine learning classifiers: A health informatics application](#). 2024 25th International Arab Conference on Information Technology (ACIT), Zarqa, Jordan. 2024, 1-6. [10.1109/ACIT62805.2024.10876953](#)
27. Mamouei M, Zhu Y, Nazarzadeh M, Hassaine A, Salimi-Khorshidi G, Cai Y, Rahimi K: [Investigating the association of environmental exposures and all-cause mortality in the UK Biobank using sparse principal component analysis](#). Scientific Reports. 2022, 12:9239. [10.1038/s41598-022-13362-3](#)
28. Salam MA, Azar AT, Elgendy MS, Fouad KM: [The effect of different dimensionality reduction techniques on machine learning overfitting problem](#). International Journal of Advanced Computer Science and Applications. 2021, 12:641-655. [10.14569/ijacsa.2021.0120480](#)
29. Reichert TA, Balestra CET, Pansera WA, Carrazedo R: [Fick's law as a custom kernel in support vector regression: How to fully represent chloride concentration profiles in concrete structures](#). Journal of Building Engineering. 2026, 121:115691. [10.1016/j.jobbe.2026.115691](#)
30. Nguyen QH, Ly HB, Ho LS, et al.: [Influence of data splitting on performance of machine learning models in prediction of shear strength of soil](#). Mathematical Problems in Engineering. 2021, 2021:1-15. [10.1155/2021/4832864](#)
31. Hodson TO: [Root-mean-square error \(RMSE\) or mean absolute error \(MAE\): when to use them or not](#). Geoscientific Model Development. 2022, 15:5481-5487. [10.5194/gmd-15-5481-2022](#)

How to cite this article:

Caesarizky O Y, Ilham A, Khikmah L, et al. (August 08, 2026) Evaluating the Statistical Significance of Grid Search Optimization in Support Vector Regression for Dengue Forecasting Using Nested Time-Series Cross-Validation. Cureus J Comput Sci 3 : es44389-026-00212-w. DOI <https://doi.org/10.7759/s44389-026-00212-w>