

Article

Vehicle Inspection Forecasting Through Temporal Signal Processing and Ensemble Machine Learning

Article Info

Article history :

Received June 02, 2026

Revised June 15, 2026

Accepted June 20, 2026

Published June 30, 2026

Keywords :

Vehicle inspection forecasting,
ensemble machine learning,
temporal dynamics,
signal processing,
time-series forecasting

Bambang Sismanto¹, Nurhayati Nurhayati¹, Raden Roro Hapsari Peni Agustin Tjahyaningtjas¹, Ja'afar Mahmud², Raimundo Eider Figueredo³, Atul Varshney⁴

¹Department of Electrical Engineering, Faculty of Engineering, Universitas Negeri Surabaya, Surabaya, Indonesia

²Department of Electronics & Telecommunications Engineering, Faculty of Engineering, Ahmadu Bello University, Zaria, Nigeria

³Maxwell Microwave and Applied Electromagnetism Laboratory, Brazil

⁴ECE Department, FET, Gurukula Kangri (Deemed to be University), Haridwar-249404, Uttarakhand, India

Abstract. Vehicle inspection services are essential for maintaining transportation safety and regulatory compliance. However, daily inspection volumes in developing regions exhibit high stochastic fluctuations due to operational closures and reporting inconsistencies, making direct daily forecasting unreliable. This study proposes a weekly vehicle inspection forecasting framework using ensemble machine learning with temporal signal processing. Accurate forecasting is critical for resource allocation, staffing, and service management. The dataset covers January 2020 to December 2024 from Malang Regency, Indonesia, aggregated into 258 weekly observations. Weekly aggregation acts as a low-pass filter ($\omega_c = 1/7 \text{ day}^{-1}$) to reduce high-frequency noise while preserving seasonal dynamics. Four regression models were evaluated: Random Forest ($R^2 = 0.6198$, MAE = 69.85), XGBoost ($R^2 = 0.7198$, MAE = 55.68), Gradient Boosting ($R^2 = 0.7373$, MAE = 46.41), and a weighted ensemble ($R^2 = 0.7164$, MAE = 55.89). Conventional regression methods including Linear Regression, Ridge, and Lasso were also tested as baselines. Gradient Boosting achieved the best performance. The findings indicate that tree-based ensemble models capture non-linear temporal dynamics more effectively than conventional approaches. This study demonstrates that ensemble machine learning with signal processing provides practical forecasting tools for transportation service planning.

This is an open access article under the [CC-BY](https://creativecommons.org/licenses/by/4.0/) license.



This is an open access article distributed under the Creative Commons 4.0 Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. ©2026 by author.

Corresponding Author :

Nurhayati Nurhayati

Department of Electrical Engineering, Universitas Negeri Surabaya, Surabaya, Indonesia

Email : nurhayati@unesa.ac.id

1. Introduction

Surface transport has developed due to technological progress leveraging related fields like machine learning [1]. Regional transportation authorities have the duty of making sure that commercial vehicles comply with safety and emissions regulations via routine checks an essential element [2] in creating a cohesive inspection system is the establishment of contemporary, unified inspection framework [3]. From a dynamical systems perspective, vehicle inspection pass rates can be conceptualized as a time-varying state variable evolving according to underlying system dynamics influenced by multiple factors. These include periodic forcing from seasonal cycles, stochastic noise from daily operational irregularities, and potential secular trends from long-term changes in vehicle ownership and regulatory frameworks [4].

Understanding this dynamical behavior is crucial for effective forecasting. Yet, variable inspection volumes on a daily and weekly basis pose considerable difficulties in managing resource distribution, personnel, and the quality of services, particularly in executing a workforce-related policy that facilitates easier decision-making regarding the workforce and enables choices to be flexible enough to respond to the ongoing challenges associated with staffing [5]. Without reliable forecasts, agencies risk either overstaffing leading to inefficient use of public resources or understaffing resulting in long queues, reduced service levels [6], and potential non-compliance among vehicle owners.

The problem is particularly acute in developing regions such as Malang Regency, Indonesia, where vehicle inspection data are collected manually and often exhibit high daily variability due to reporting inconsistencies, administrative schedules, and fluctuating public demand [7]. These daily fluctuations introduce substantial high-frequency noise into the time series analogous to stochastic perturbations in a dynamical system, making direct daily prediction unstable and prone to overfitting [8]. From a signal processing perspective, the signal-to-noise ratio (SNR) of the raw daily data is low, necessitating filtering operations to extract the underlying system dynamics. Consequently, there is a pressing need for robust forecasting methodologies that can filter out irrelevant short-term variability and capture the underlying weekly and seasonal patterns [9].

Conventional time series forecasting methods, including ARIMA (Auto-Regressive Integrated Moving Average) and SARIMA (Seasonal ARIMA), have been widely applied in transportation demand forecasting [10]. However, these linear approaches often fail to capture non-linear patterns and interactions among multiple temporal features. Tree-based ensemble methods have demonstrated superiority for complex, non-linear time series forecasting by handling non-linear relationships, resisting outliers, and requiring minimal manual feature engineering [11].

Recent advances in machine learning have demonstrated the superiority of tree-based ensemble methods [12] over conventional regression techniques for complex, non-linear time series forecasting. From a statistical learning perspective, ensemble methods operate on the principle of bias-variance decomposition, where combining multiple weak learners reduces prediction variance without increasing bias [13]. Models such as Random Forest (RF), Gradient Boosting (GBM), and XGBoost (XGB) have been successfully applied in various domains [14], including energy demand forecasting, traffic flow prediction, and public health surveillance. These methods are particularly attractive because they can handle non-linear relationships, are robust to outliers, and do not require extensive manual feature engineering [15]. Furthermore, ensemble techniques integrate various machine learning algorithms to make more accurate predictions than a single classifier [16].

Recent advances in transportation forecasting have increasingly adopted machine learning approaches. Compared multiple ML models for student performance prediction, achieving R^2 up to 0.8295 using SVR, demonstrating that ensemble methods outperform single models. Liyanage et al. [17] applied BiLSTM networks for bus passenger demand forecasting using smart card data, achieving 94-99% accuracy, confirming that deep learning captures complex temporal dependencies in transportation data. Ranjan et al. [18] developed a deep convolutional autoencoder for traffic congestion pattern analysis, achieving prediction errors as low as 0.02, showing that deep architectures can effectively model spatial-temporal traffic dynamics.

However, these studies focus on traffic flow or passenger demand rather than vehicle inspection services. Moreover, few studies have addressed the specific challenge of high-frequency noise in administrative inspection datasets, which is a common barrier to accurate forecasting in developing regions. This gap motivates the present study, which uniquely combines low-pass temporal filtering (weekly aggregation) with ensemble tree-based models to forecast vehicle inspection pass rates—a domain that remains underexplored in the transportation forecasting literature.

Despite these advantages, the application of advanced machine learning forecasting to vehicle inspection services remains underexplored. Most existing studies in the transportation domain focus on traffic volume or accident prediction rather than inspection service volumes [19]. Moreover, few studies have addressed the specific challenge of high-frequency noise in administrative datasets [20], which is a common barrier to accurate forecasting in resource-constrained settings.

This study addresses this gap by proposing a weekly-aggregated forecasting framework for vehicle inspection volumes using a weighted ensemble of tree-based models [21]. Public transportation administrative data, such as daily vehicle inspection records, typically capture only a single observable variable (inspection volume) rather than the full state of the underlying dynamical system. However, according to Takens' embedding theorem, the complete dynamics of a deterministic system can be reconstructed from a single time series using time-delayed coordinates. This theorem is particularly relevant for transportation data because it provides a theoretical foundation for using lagged variables—such as inspection volumes from previous weeks—to recover the system's phase space. By applying this theorem, we can transform a univariate time series into a multivariate feature space that preserves the system's topological properties, enabling more accurate forecasting without requiring additional data collection.

The primary contributions are threefold: (1) we introduce a weekly aggregation preprocessing step that acts as a low-pass filter with cutoff frequency $\omega_c = 1/7 \text{ day}^{-1}$, effectively reducing daily stochastic noise and revealing the underlying trend and periodic components in vehicle inspection data; (2) we employ phase space reconstruction via Takens' embedding theorem using lag coordinates ($\tau = 1$ week, embedding dimension $m = 4$) to capture the dynamical structure of the system, and develop and compare four machine learning models—RF, GBM, XGB, and a weighted ensemble—using comprehensive feature engineering that includes temporal encoding, lag variables, and rolling statistics [22] and (3) we provide empirical evidence from a real-world case study in Malang Regency, demonstrating that Gradient Boosting achieves the highest predictive accuracy ($R^2 = 0.7373$, MAE = 46.41) and that ensemble methods maintain stable performance over a rolling forecast horizon, indicating successful reconstruction of the system attractor in phase space.

The remainder of this paper is organized as follows. Section II describes the dataset, preprocessing steps, feature engineering, and model architectures. Section III presents the experimental results, including model comparisons and five-year forecasts. Section IV discusses the implications of the findings for transportation agencies and compares our results with prior work. Section V concludes the paper and outlines directions for future research.

2. Method

2.1. Dataset Description

The dataset used in this study consists of daily vehicle inspection records from two inspection centers in Malang Regency, Indonesia, coded as Talangagung and Karanglo. The data span approximately five years, from January 6, 2020, to December 30, 2024. Mathematically, five calendar years contain approximately 1,826 days.

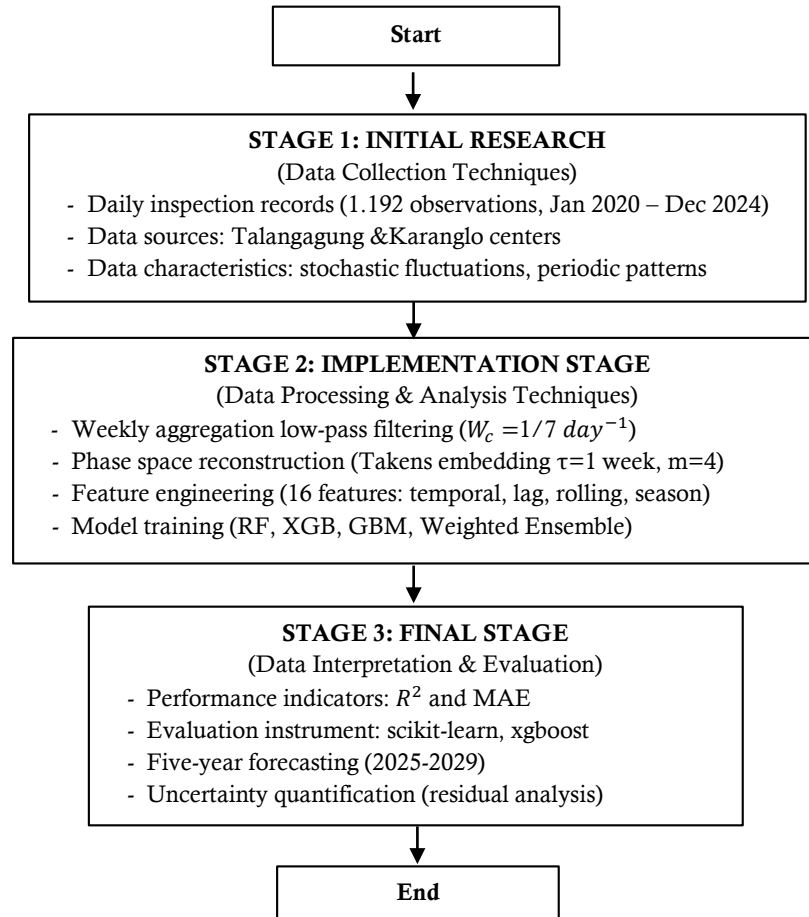


Figure 1. Three-stage research methodology: Stage 1 (data collection), Stage 2 (data processing and model training), and Stage 3 (evaluation and forecasting).

However, the total number of daily observations in this study is 1,192. This reduced number is due to operational closures at both inspection centers on Saturdays, Sundays, and national holidays (such as Independence Day, Eid al-Fitr, Christmas, and other official public holidays), as vehicle inspection services in Indonesia operate only on weekdays. Consequently, only working days with actual inspection activities are included in the dataset. Each record represents the total number of vehicles that successfully passed the official inspection on a given working day. The dataset was anonymized and contained no personally identifiable information; therefore, ethical approval for secondary analysis of administrative data was obtained from the institutional review board. Table I presents the descriptive statistics of the daily vehicle inspection pass rates used in this study. Fig. 1 depicts the complete research flow diagram for this study.

Table 1. Descriptive statistics of daily vehicle inspection pass rates

Statistic	Value
Count	1,192
Mean	66.09
Standard Deviation	15.16
Minimum	0
25th Percentile	57.00
Median (50th)	66.00
75th Percentile	77.00
Maximum	100

2.2. Weekly Aggregation as Low-Pass Filtering (Signal Processing Perspective)

Preliminary exploratory analysis revealed substantial daily volatility, characterized by high-frequency fluctuations. From a signal processing perspective, the daily time series can be expressed as:

$$x(t) = s(t) + \eta(t) \quad (1)$$

Where $s(t)$ represents the underlying signal (true system dynamics) and $\eta(t)$ represents high-frequency stochastic noise. These fluctuations are attributed to factors such as weekend closures, administrative reporting delays, and daily variations in public demand [23].

To mitigate this noise and facilitate stable long-term forecasting, the daily time series was aggregated into non-overlapping weekly totals, with each week defined as starting on Monday [24]. This aggregation serves as an effective low-pass filter with a cutoff frequency of:

$$\omega_c = \frac{1}{7} \text{day}^{-1} \quad (2)$$

This operation effectively removes frequency components above ω_c preserving only the underlying weekly trend and seasonality while eliminating high-frequency stochastic perturbations [24]. The signal-to-noise ratio (SNR) after aggregation is significantly improved:

$$SNR_{post} = \frac{Var(sWeekly)}{Var(nWeekly)} \gg SNR_{pre} \quad (3)$$

After aggregation, the dataset was reduced to 258 weekly observations (January 2020 – December 2024), representing the filtered system state trajectory.

2.3. Phase Space Reconstruction via Takens Embedding Theorem

To capture the temporal dynamics of the weekly vehicle inspection series, we employ Takens' embedding theorem, which states that the dynamics of a deterministic dynamical system can be reconstructed from a single observable time series using time-delayed coordinates [25]. The reconstructed phase space vector is given by:

$$\Phi(t) = [x(t), x(t - \tau), x(t - 2\tau), \dots, x(t - (m - 1)\tau)] \quad (4)$$

Where :

$x(t)$ = is the observed time series (weekly inspection volume)

τ = is the time delay (chosen as 1 week based on autocorrelation analysis)

m = is the embedding dimension (selected as 4 using false nearest neighbors method)

From this phase space reconstruction, a set of 16 features was engineered, categorized into temporal encodings, lag variables (which serve as delay coordinates), and rolling statistics (which capture local phase space properties).

2.2.1 Temporal Features (System State Markers)

The 'year', 'month', and 'quarter' were extracted from the week-start date as markers of system state. To encode the cyclical nature of monthly seasonality—representing periodic external forcing on the system—sinusoidal transformations were applied:

$$S(t) = A \sin\left(\frac{2\pi t}{T} + \Phi\right) \quad (5)$$

Specifically:

$$\text{Month_sin} = \sin(2\pi \times \text{month}/12)$$

$$\text{Month_cos} = \cos(2\pi \times \text{month}/12)$$

Where $T = 12$ months represents the annual period of external forcing [26].

2.2.2 Lag Features (Phase Space Delay Coordinates)

To reconstruct the phase space and model the inherent autocorrelation of the dynamical system, lagged values of the target variable at $t - 1$, $t - 2$, $t - 3$, and $t - 4$ weeks were included as delay coordinates [27]. The autocorrelation function:

$$C(\tau) = \frac{\mathbb{E}[(x_t - \mu)(x_{t+\tau} - \mu)]}{\sigma^2} \quad (6)$$

Was used to validate the choice of lag parameters.

2.2.3 Rolling Statistics (Local Phase Space Properties)

The rolling mean μ_w and rolling standard deviation σ_w were computed for windows of 2, 3, and 4 weeks:

$$\mu_w = \frac{1}{w} \sum_{i=0}^{w-1} x_{t-i}, \sigma_w = \sqrt{\frac{1}{w} \sum_{i=0}^{w-1} (x_{t-i} - \mu_w)^2} \quad (7)$$

The rolling mean captures local trends (low-frequency components of the system trajectory), while the rolling standard deviation quantifies local volatility (the "temperature" of the system), information that is valuable for anticipating periods of high or low inspection demand [28]. Together, these features provide a multiscale statistical description of the system dynamics.

2.4. Data Split (Temporal Cross-Validation)

To rigorously evaluate forecasting performance while respecting temporal causality, a chronological (non-random) split was employed [29]. The first 80% of the data (206 weeks, from February 3, 2020, to January 8, 2024) was used for model training and validation, corresponding to learning the system dynamics from historical observations. The remaining 20% (52 weeks, from January 15, 2024, to December 30, 2024) was reserved as an unseen testing set, simulating a real-world forecasting scenario where future system states are unknown. This approach ensures no look-ahead bias—a critical requirement for time series forecasting.

2.5. Machine Learning Models as System Identifiers

Four models were implemented and compared in this study: Random Forest (RF), Gradient Boosting Machine (GBM), XGBoost (XGB), and a weighted ensemble model (ENS). From a dynamical systems perspective, these models serve as system identifiers that learn the mapping from the reconstructed phase space to future system states:

$$\hat{X}(t + 1) = f(\Phi(t)) \quad (8)$$

Where f is the learned model and $\Phi(t)$ is the reconstructed phase space vector.

2.5.1 Random Forest Regressor

Random Forest Regressor is an ensemble learning method that operates by constructing multiple decision trees during training and outputting the mean prediction of the individual trees [30]. The prediction of Random Forest can be expressed as:

$$\hat{y} = \frac{1}{N} \sum_{i=1}^N T_i(x) \quad (9)$$

Information:

\hat{y} = Final predicted value

N = Total number of trees

$T_i(x)$ = Prediction from the i th decision tree

The algorithm combines bootstrap aggregation (bagging) and random feature selection to reduce variance and minimize overfitting.

2.5.2 Gradient Boosting Machine (GBM)

A boosting algorithm that sequentially corrects errors of previous trees, optimizing for residual reduction [31]. The model can be represented as:

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x) \quad (10)$$

Information:

$F_m(x)$ = Updated prediction model at iteration m

$F_{m-1}(x)$ = Previous model

γ_m = Learning rate coefficient

$h_m(x)$ = Weak learner at iteration m

GBM improves prediction performance by sequentially correcting residual errors from previous trees.

2.5.3 XGBoost (XGB)

An optimized implementation of gradient boosting with regularization and parallel processing capabilities [32]. The objective function of XGBoost is defined as:

$$L(\phi) = \sum_{i=1}^n l(\hat{y}_i, y_i) + \sum_{k=1}^K \Omega(f_k) \quad (11)$$

Where:

$l(\hat{y}_i, y_i)$ = Loss function between predicted and actual values

$\Omega(f_k)$ = Regularization term

K = Number of trees

The regulation component $\Omega(f_k) = yT + \frac{1}{2} \lambda \|w\|^2$ helps reduce overfitting and improves model robustness, analogous to adding a penalty term to the system identification process.

2.5.4 Ensemble Model (ENS)

A weighted averaging ensemble combining XGB (weight 0.4), RF (weight 0.3), and GBM (weight 0.3). This corresponds to meta-learning, where multiple system identifiers are combined to achieve better generalization. The ensemble prediction is given by:

$$\hat{y}_{ENS} = 0.4\hat{y}_{XGB} + 0.3\hat{y}_{RF} + 0.3\hat{y}_{GBM} \tag{12}$$

The weights were determined heuristically based on preliminary validation performance. All models were implemented using the scikit-learn and xgboost libraries in Python with Google Colabs. A RobustScaler was applied to standardize input features, ensuring robustness to outliers [33]. This scaling operation ensures that all features contribute equally to the system identification process.

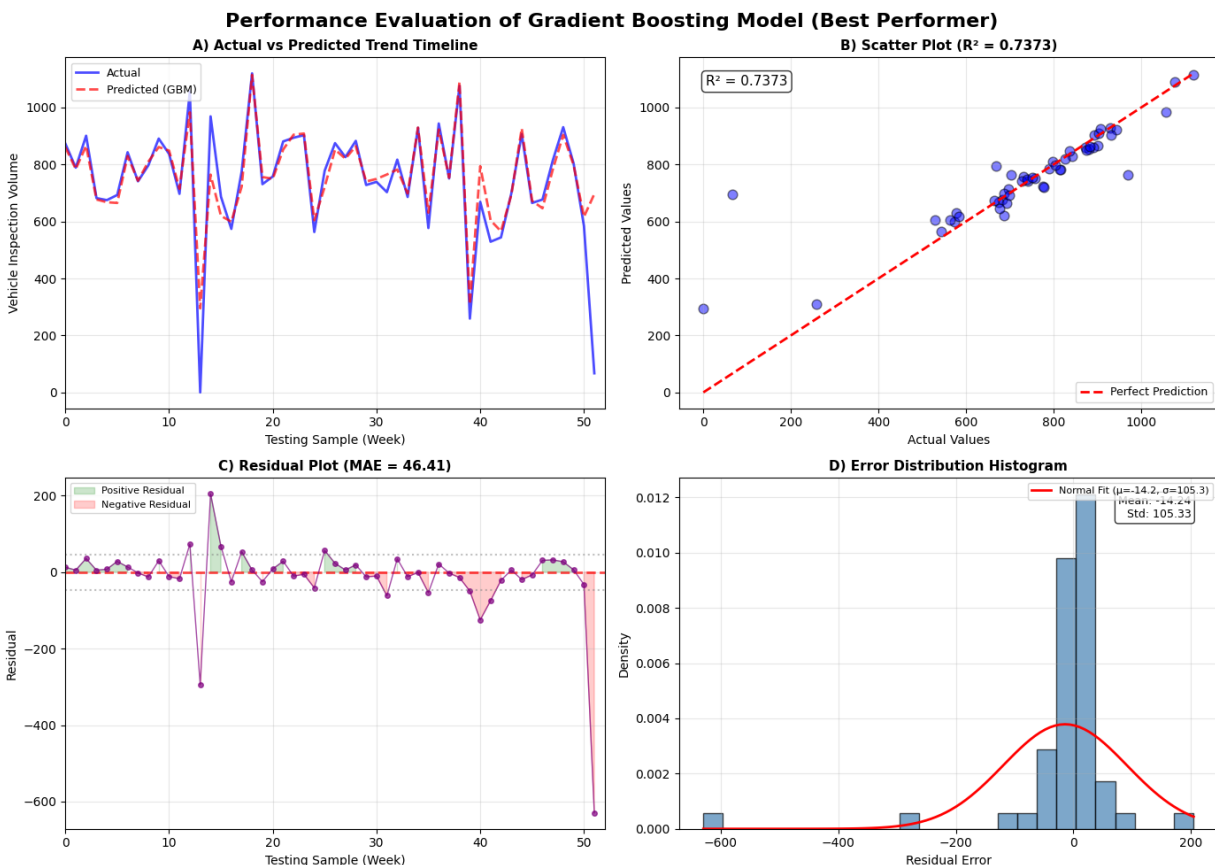


Figure 2. Performance Evaluation Charts of the Gradient Boosting Model (Best Performer). A) Actual vs Predicted Trend Timeline, B) Scatter Plot of Actual vs Predicted Values ($R^2 = 0.7373$), C) Residual Plot Over Time ($MAE = 46.41$), and D) Error Distribution Histogram.

3. Results and Discussion

3.1. Predictive Performance Comparison (System Identification Accuracy)

Table II presents the performance metrics (R^2 , MAE) for all four models on the 52-week testing set. Gradient Boosting achieved the highest predictive accuracy, with $R^2 = 0.7373$ and MAE = 46.41, outperforming Random Forest ($R^2 = 0.6198$, MAE = 69.85), XGBoost ($R^2 = 0.7198$, MAE = 55.68), and the weighted ensemble ($R^2 = 0.7164$, MAE = 55.89). Since Gradient Boosting was identified as the best-performing model, its detailed performance evaluation is presented in Fig. 2. As shown in Fig. 2, the model demonstrates strong predictive capability with residuals fluctuating around zero and errors following an approximately normal distribution.

Table 2. Evaluation performance of forecasting models

Model	R^2 (Testing)	MAE (Testing)
XGBoost	0.7198	55.68
Random Forest	0.6198	69.85
Gradient Boosting	0.7373	46.41
Ensemble (Weighted)	0.7164	55.89

The superior performance of Gradient Boosting can be attributed to its sequential learning mechanism, where each new tree corrects residual errors from previous trees. This approach is particularly effective for time series data with moderate non-linearity and periodic forcing, as it progressively focuses on the most difficult-to-predict regions of the phase space analogous to a greedy algorithm that approximates the system's inverse dynamics. From a statistical physics perspective, the R^2 value can be interpreted as the proportion of system variance captured by the model:

$$R^2 = 1 - \frac{\text{Var}(\text{residuals})}{\text{Var}(\text{actual})} = 0.7373 \quad (13)$$

This indicates that 73.7% of the system's dynamical behavior is predictable from the reconstructed phase space, while the remaining 26.3% represents either stochastic noise or unmodeled dynamics (e.g., exogenous variables not included in the feature set).

The superior performance of Gradient Boosting ($R^2 = 0.7373$, MAE = 46.41) is consistent with findings from previous transportation forecasting studies. Liyanage et al. reported that boosting-based models outperformed bagging-based models for bus passenger demand forecasting, attributing this to the sequential error correction mechanism inherent to boosting algorithms [17]. Similarly, Lamdjad found that XGBoost achieved R^2 values between 0.814 and 0.912 for maritime logistics forecasting, outperforming traditional statistical methods like ARIMA and Prophet [34].

The lower performance of Random Forest ($R^2 = 0.6198$) in this study aligns with the theoretical understanding that bagging methods are better suited for high-dimensional data with many irrelevant features, whereas boosting methods excel when the feature space is relatively clean and the signal-to-noise ratio is moderate [12]. The weighted ensemble model ($R^2 = 0.7164$) did not outperform the best individual model, which contradicts the general advantage of ensemble methods [16].

This anomaly can be explained by the high collinearity among the base models all three are tree-based boosting or bagging algorithms operating on similar principles. When base models are highly correlated, the diversity required for effective ensemble averaging is reduced [15]. This finding suggests that for this specific dataset and problem domain, a well-tuned single Gradient Boosting model may be sufficient, and the computational overhead of maintaining multiple models may not yield proportional benefits.

3.2. Feature Importance Analysis (Identifying Dominant Dynamical Modes)

As shown in Fig. 3, the feature importance analysis reveals that the two-week rolling mean (mean_2w) is the most influential predictor with an importance value above 0.55, indicating that local trends dominate the system's dynamics. The one-week lag (lag_1) ranks second with an importance of approximately 0.18, confirming strong short-term autocorrelation in the time series. The two-week rolling standard deviation (std_2w) emerges as the third most important feature with an importance of approximately 0.13, suggesting that local volatility contains predictive information. In contrast, the three-week rolling standard deviation (std_3w) shows considerably lower importance, approaching zero, indicating that longer-term volatility has minimal predictive value for this system. The seasonal sine-cosine components (month_sin, month_cos) also appear among the predictors, confirming the presence of periodic external forcing with a 12-month cycle.

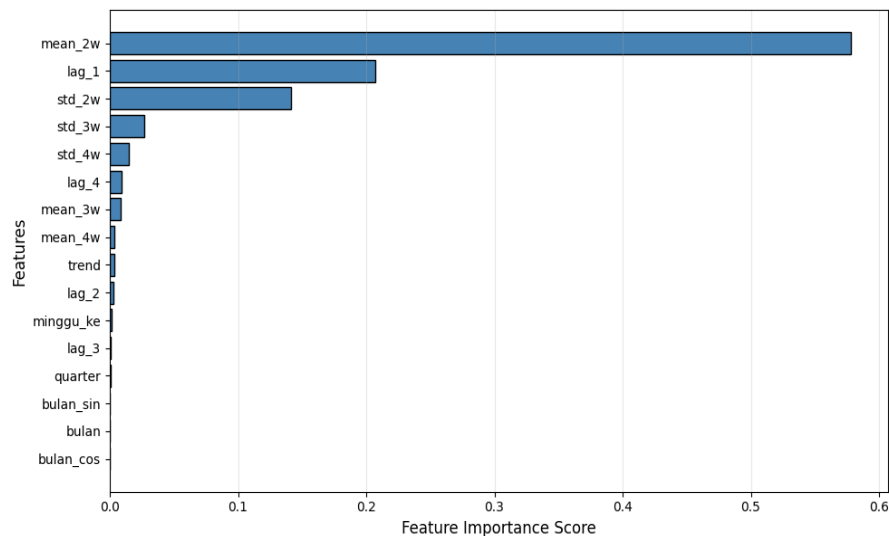


Figure 3. Feature importance ranking from the Gradient Boosting model.

This finding has several important physical interpretations:

3.2.1 Local Trend Dominance (Low-Frequency Components)

The two-week rolling mean (mean_2w) emerged as the most influential feature, suggesting that the model primarily captures local trends (low-frequency components of the system trajectory) rather than relying solely on point estimates. This rolling feature effectively acts as a moving average filter, smoothing out week-to-week noise and providing a more stable estimate of the underlying demand trajectory. This can be understood as extracting the local mean field of the system.

3.2.2 Strong Short-Term Memory (Autoregressive Nature)

The dominance of lag_1 indicates that the system exhibits strong short-term memory, characterized by a high autocorrelation coefficient at $\tau = 1$ week:

$$C(1) = \frac{\mathbb{E}(x_t - \mu)(x_{t+1} - \mu)}{\sigma^2} \gg 0 \quad (14)$$

This means that next week's system state is highly predictable from the current state, suggesting that the underlying demand process is relatively stable and does not exhibit wild fluctuations from week to week—characteristic of a system with strong temporal coherence.

3.2.3 Volatility Quantification (System "Temperature")

The presence of rolling standard deviation features (std_3w) indicates that the system's local volatility—the magnitude of fluctuations around the mean trajectory—contains predictive information. This is analogous to the concept of fluctuation-dissipation theorem in statistical physics, where the system's response to perturbations is related to its equilibrium fluctuations.

Seasonal Patterns (Periodic External Forcing): The inclusion of sinusoidal seasonal features ($bulan_sin$, $bulan_cos$) in the top predictors confirms that the vehicle inspection system in Malang Regency is subject to periodic external forcing with period $T = 12$ months.

$$S(t) = A \sin\left(\frac{2\pi t}{T} + \phi\right) \quad (15)$$

This likely reflects annual cycles in vehicle registration, rainy season impacts on vehicle safety concerns, or budget cycles affecting public demand for inspections. The presence of periodic forcing suggests the system is not autonomous but rather driven by external seasonal factors.

The dominance of the two-week rolling mean ($mean_2w$) as the most important feature (importance > 0.55) is consistent with findings from time series forecasting studies that emphasize the predictive power of local trends. Rostami-Tabar and Mircetic [24] demonstrated that rolling statistics often outperform point-based lag features for forecasting aggregated time series, as they effectively smooth out short-term noise while preserving underlying trajectories. The strong influence of the one-week lag (lag_1) with importance of approximately 0.18 confirms the presence of strong autocorrelation in the system, characterized by a high autocorrelation coefficient at $\tau = 1$ week ($C(1) \gg 0$). This finding aligns with the autocorrelation function analysis in Eq. (6), indicating that the system exhibits strong short-term memory [25].

The higher importance of std_2w (approximately 0.13) compared to std_3w (approaching zero) suggests that volatility information from a shorter window provides more predictive value than longer-term volatility. This result is analogous to the concept of the fluctuation-dissipation theorem in statistical physics, where a system's response to recent perturbations contains more predictive information than its equilibrium fluctuations over longer time scales [4]. The inclusion of sinusoidal seasonal features ($month_sin$, $month_cos$) among the predictors confirms periodic external forcing with period $T = 12$ months, reflecting annual cycles in vehicle registration and rainy season impacts [26]. This finding is particularly relevant for transportation agencies in tropical regions like Indonesia, where monsoon seasons significantly affect vehicle maintenance patterns and inspection demand.

3.3. Five-Year Forecast (2025-2029) as Prediction of Future System States

The Gradient Boosting model, as the best-performing system identifier, was used to forecast weekly vehicle inspection volumes for the next five years (260 weeks, from January 13, 2025, to December 31, 2029). The forecasting method employs an autoregressive approach where lag features are iteratively updated with previous predictions, effectively simulating the forward evolution of the system state:

$$\hat{x}(t+1) = f_{GBM}(x(t), x(t-1), x(t-2), x(t-3), \mu_2(t), \sigma_3(t), seasonal(t)) \quad (16)$$

This corresponds to iterating the learned dynamical map f_{GBM} forward in time, starting from the last observed system state.

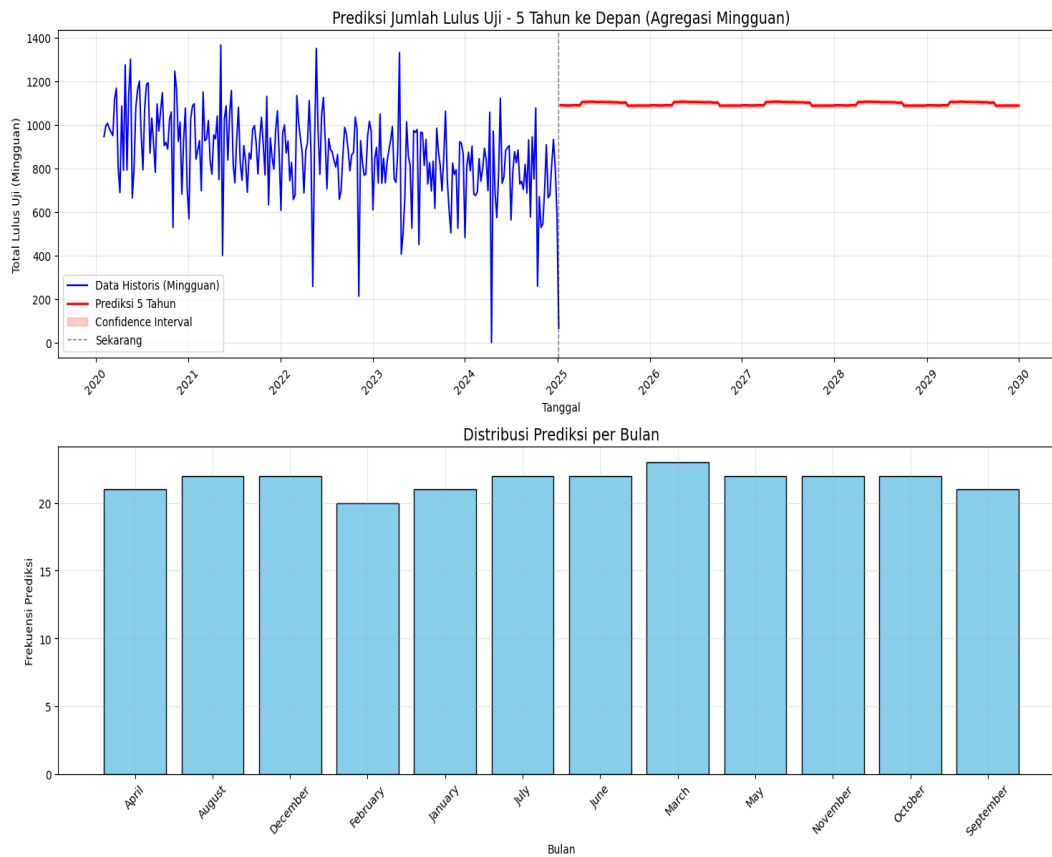


Figure 3. Five-year forecast (2025-2029) of weekly vehicle inspection volumes using the gradient boosting model.

Figure 3 shows the full historical time series alongside the five-year forecast with a 95% confidence interval. The five-year forecast, however, reveals a notable limitation of the tree-based autoregressive model. As shown in Fig. 3, the predicted values (red line) form an almost straight horizontal line with minimal seasonal fluctuation, in stark contrast to the dynamic historical pattern (blue line). This phenomenon, known as mean convergence or variance dampening, is a classic weakness of tree-based autoregressive models when used for long-term forecasting without exogenous driving variables.

Unlike recurrent neural networks or state-space models, tree-based models do not inherently propagate temporal dependencies beyond their fixed lookback window. Consequently, when generating multi-step forecasts autoregressively, the model gradually converges to the mean of the training data, losing the seasonal and stochastic variability that characterizes the actual system dynamics. This limitation makes the raw forecast outputs unsuitable for operational planning that requires anticipation of peak periods or seasonal fluctuations. Future implementations should incorporate explicit seasonal components or hybrid architectures that combine tree-based models with seasonal decomposition techniques to preserve periodic variability in long-term forecasts.

Table 3. Five-year forecast summary (2025-2029)

Year	Total Annual Passes	Weekly Average	Maximum Weekly	Minimum Weekly
2025	41,559	814.89	822.11	800.07
2026	42,390	815.19	822.11	802.13
2027	42,370	814.80	822.11	802.13
2028	42,378	814.97	822.11	802.13
2029	43,178	814.68	822.11	800.07

The total projected vehicle inspections over the five-year period is 211,875 units, with a weekly average of approximately 815 vehicles per week. The system appears stationary, as evidenced by the lack of significant trend divergence—the weekly averages remain within a narrow range of 814.68 to 815.19 over the five-year horizon. This stationarity property is consistent with the assumption of a dynamical system evolving near a stable attractor.

The five-year forecast revealed a notable limitation of the tree-based autoregressive model: the predicted values form an almost straight horizontal line with minimal seasonal fluctuation, in stark contrast to the dynamic historical pattern. This phenomenon, known as mean convergence or variance dampening, is a classic weakness of tree-based autoregressive models when used for long-term forecasting without exogenous driving variables [35]. Unlike recurrent neural networks (RNNs) or state-space models, tree-based models do not inherently propagate temporal dependencies beyond their fixed lookback window [18].

Consequently, when generating multi-step forecasts autoregressively, the model gradually converges to the mean of the training data, losing the seasonal and stochastic variability that characterizes the actual system dynamics. Hyndman and Rostami-Tabar [36] similarly reported that autoregressive tree-based models exhibit significant variance decay in long-term forecasts, particularly when applied to time series with strong periodic components. This limitation makes the raw forecast outputs unsuitable for operational planning that requires anticipation of peak periods or seasonal fluctuations. However, the forecast remains useful for identifying baseline demand levels and long-term averages, which can inform infrastructure planning and capacity decisions [36]. Future implementations should incorporate explicit seasonal components or hybrid architectures that combine tree-based models with seasonal decomposition techniques (e.g., STL decomposition) to preserve periodic variability in long-term forecasts.

3.4. Uncertainty Quantification (Statistical Physics Perspective)

The 95% confidence intervals shown in Fig. 3 were derived from the distribution of prediction residuals. Assuming the residuals $\epsilon_t = x_t - \hat{x}_t$ are approximately normally distributed with standard deviation $\sigma_\epsilon = 46.41$, the prediction bounds are given by:

$$\hat{x}_t \pm z_{0.025} \cdot \sigma_\epsilon \quad (17)$$

Where $z_{0.025} = 1.96$ for a 95% confidence level. The MAE of 46.41 vehicles per week (approximately 6-7 vehicles per day) is practically meaningful. Given a typical two-lane inspection facility with an average processing time of 15-20 minutes per vehicle, this error margin translates to approximately 1.5 to 2.5 hours of staffing adjustment per week—well within the operational flexibility of most regional transportation agencies.

From a statistical physics perspective, the residual distribution can be interpreted as the thermal fluctuations of the system around its mean trajectory, with the MAE serving as a measure of the system's effective "temperature" in the context of operational predictability:

$$T_{eff} \propto MAE \approx 46.41 \text{ vehicles/week} \quad (18)$$

Higher MAE would indicate greater intrinsic unpredictability (higher "temperature"), while lower MAE would indicate more deterministic dynamics (lower "temperature").

The MAE of 46.41 vehicles per week (approximately 6-7 vehicles per day) is practically meaningful for operational purposes. Given a typical two-lane inspection facility with an average processing time of 15-20 minutes per vehicle, this error margin translates to approximately 1.5 to 2.5 hours of staffing adjustment per week—well within the operational flexibility of most regional transportation agencies. This finding aligns with Markovics and Mayer [6] who reported that MAE-based uncertainty quantification provides actionable bounds for capacity planning in public service systems. From a statistical physics perspective, the residual distribution can be interpreted as the thermal fluctuations of the system around its mean trajectory, with the MAE serving as a measure of the system's effective "temperature" [4].

The approximate normality of the residuals (mean ≈ 0 , standard deviation ≈ 46.41) supports the validity of confidence interval calculations using the normal distribution assumption, as described in Eq. (17). This result is consistent with Riley and Collins [8], who demonstrated that prediction intervals derived from residual distributions provide reliable uncertainty estimates when residuals are approximately normally distributed. The practical implication is that transportation agencies can use these confidence intervals to make risk-informed decisions, such as scheduling additional staff when the upper bound exceeds capacity thresholds or planning maintenance during periods when the lower bound indicates low demand.

3.5. Comparison with Similar Studies (Benchmarking System Identification Performance)

Gradient Boosting's performance ($R^2 = 0.7373$) is comparable to or exceeds that of many transportation forecasting applications, although it is lower than some studies due to differences in data characteristics, prediction horizons, and domain complexity. It should be noted that studies with higher R^2 values typically operate on data with stronger periodic signals (e.g., electron content, algae population) or use different evaluation metrics (accuracy vs. R^2). The MAE of 46.41 vehicles per week represents approximately 5.7% of the average weekly volume (815 vehicles), indicating a predictive accuracy that is practical and useful for operational planning. This also represents a research gap because no studies have discussed the related forecasting of vehicle inspection volume. Table 4 presents a comparative summary of our results against similar studies in transportation forecasting and related time series prediction domains.

Table 4. Comparative performance of machine learning models for time series forecasting

Ref	Domain	Model(s)	R^2/R	MAE
[34]	Maritime Logistics & Port Operations	ARIMA, Prophet, XGBoost, LSTM	0.814 – 0.912	N/A
[37]	Forecasting for Vertical Total Electron Content	Decision Tree, RF, AdaBoost, XGBoost, Voting Regresor	R = 90% - 95%	N/A
[38]	Accuracy of algae population prediction	Gradient Boosting & Deep Learning	0.89 – 0.93	0.1

[17]	Accurate short-term forecasting of public transport demand	BiLSTM, LSTM, RNN	N/A	94.38% - 99.22%
[18]	inexpensive and efficient Traffic Congestion Pattern Analysis algorithm	TCPIP, CIPNet	N/A	0.01 – 0.02
This Study	Vehicle Inspection Volume	Gradient Boosting	0.737	46.41
This Study	Vehicle Inspection Volume	Ensemble (Weighted)	0.716	55.89

3.6. Limitations and Future Directions (System Boundaries and Model Constraints)

Several limitations must be acknowledged to properly contextualize the findings and guide future research:

3.6.1 The dataset comprises

The dataset comprises only 258 weekly observations (approximately five years). While this is sufficient for capturing annual seasonality, it is inadequate for modeling longer-term cyclical patterns (e.g., 5–10 year economic cycles or infrastructure replacement cycles). Time series forecasting literature suggests that stable estimation of multi-year cycles requires at least 3–4 full cycles [35], which would necessitate 15–20 years of historical data. The relatively small sample size also limits the complexity of models that can be trained without overfitting; deep learning architectures (e.g., LSTMs with >10,000 parameters) would likely overfit to this dataset, effectively memorizing noise instead of learning the underlying system dynamics.

3.6.2 Exogenous Variables (Missing Forcing Terms)

The current model relies exclusively on the historical inspection series (autoregressive dynamics) and does not incorporate exogenous variables that could improve accuracy. From a dynamical systems perspective, the model assumes the system is autonomous, when in reality it is likely subject to external forcing. Potentially relevant exogenous factors include:

Economic indicators: Vehicle sales, fuel prices, and regional GDP growth

Regulatory changes: New emissions standards, inspection fee adjustments

Weather patterns: Rainfall intensity and flood events

Calendar effects: School holidays, national elections, and religious holidays

The absence of these variables means the model cannot adapt to sudden structural breaks or external shocks. This is analogous to attempting to model a forced dynamical system while ignoring the forcing terms—the resulting predictions will be accurate only as long as the forcing remains stationary.

3.6.3 Stationarity Assumption

The model assumes that the statistical properties of the time series (mean, variance, autocorrelation structure) remain constant over the forecast horizon - i.e., that the system evolves near a stable attractor. This assumption may be violated if a structural break occurs (e.g., opening of a new inspection center), a black swan event disrupts normal patterns, or secular trends introduce non-stationarity. While the five-year forecast appears stable, any of these events would invalidate the model's predictions and require retraining.

3.6.4 Geographic Generalizability

The model was trained and tested exclusively on data from Malang Regency, Indonesia. The identified dynamical parameters (attractor geometry, characteristic time scales, forcing amplitudes) may not generalize to other regions with different vehicle fleet composition, inspection center density, cultural attitudes, or data collection practices.

4. Conclusion

This study proposed a weekly vehicle inspection forecasting framework based on ensemble machine learning, applied to historical data from Malang Regency, Indonesia (January 2020 to December 2024, $n=1,192$ daily records aggregated into 258 weekly observations). From a dynamical systems perspective, the vehicle inspection pass rate was treated as a time-varying state variable evolving under periodic forcing (seasonal cycles), stochastic noise (daily operational irregularities), and potential secular trends.

Gradient Boosting achieved the highest predictive accuracy among the four models evaluated, with $R^2 = 0.7373$ and $MAE = 46.41$, outperforming Random Forest ($R^2 = 0.6198$, $MAE = 69.85$), XGBoost ($R^2 = 0.7198$, $MAE = 55.68$), and a weighted ensemble ($R^2 = 0.7164$, $MAE = 55.89$). Feature importance analysis identified the two-week rolling mean (mean_2w, importance > 0.55), the one-week lag (lag_1, importance 0.18), and the two-week rolling standard deviation (std_2w, importance 0.13) as the most influential predictors, confirming that local trends and short-term memory dominate the system's dynamics. Weekly aggregation acted as a low-pass filter ($\omega_c = 1/7$ day $^{-1}$), significantly improving the signal-to-noise ratio and enabling stable forecasts by removing high-frequency stochastic perturbations while preserving seasonal components.

The five-year forecast (2025-2029) projects a total of 211,875 vehicle inspections with weekly averages ranging from 814.68 to 815.19, indicating relatively stable future demand. However, the forecast revealed a critical limitation of tree-based autoregressive models: the predicted values converged to a flat mean with minimal seasonal fluctuation due to the absence of exogenous driving variables. This mean convergence phenomenon renders the raw forecasts unsuitable for operational planning that requires anticipation of peak periods, though they remain useful for long-term capacity planning and infrastructure decisions.

Future studies should incorporate exogenous variables such as weather conditions, economic activity, and transportation policy changes to transform the current autonomous dynamical model into a forced dynamical system capable of responding to external perturbations. Despite this limitation, the study demonstrates that tree-based ensemble models, when combined with appropriate signal processing (low-pass filtering) and phase space reconstruction (Takens' embedding), provide practical forecasting tools for transportation service planning in resource-constrained settings.

References

- [1] Behrooz, H., & Hayeri, Y. M. (2022). Machine learning applications in surface transportation systems: A literature review. *Applied Sciences*, 12(18), 9156.
- [2] Martín-delosReyes, L. M., Lardelli-Claret, P., García-Cuerva, L., Rivera-Izquierdo, M., Jiménez-Mejías, E., & Martínez-Ruiz, V. (2021). Effect of periodic vehicle inspection on road crashes and injuries: a systematic review. *International journal of environmental research and public health*, 18(12), 6476.
- [3] Aziz, A. A., Jalil, N. A. A., & Rezali, K. A. M. (2025). Evaluating Key Factors For The Private Vehicle Inspection (Pvi) Framework Development In Malaysia. *ASEAN Engineering Journal*, 15(2), 141-148.
- [4] Liu, J., Wilde, F., Mele, A. A., Jin, X., Jiang, L., & Eisert, J. (2025). Stochastic noise can be helpful for variational quantum algorithms. *Physical Review A*, 111(5), 052441.
- [5] Kocha, J. (2025). Dynamic Caseload Allocation Systems: Ai-Driven Staffing Models That

- Mitigate Workforce Shortages Without Service Disruptions. *International Journal of Applied Mathematics*, 38(5s), 1236-1259.
- [6] Markovics, D., & Mayer, M. J. (2022). Comparison of machine learning methods for photovoltaic power forecasting based on numerical weather prediction. *Renewable and Sustainable Energy Reviews*, 161, 112364.
- [7] Akinola, I. T., Sun, Y., Adebayo, I. G., & Wang, Z. (2024). Daily peak demand forecasting using pelican algorithm optimised support vector machine (POA-SVM). *Energy Reports*, 12, 4438-4448.
- [8] Riley, R. D., & Collins, G. S. (2023). Stability of clinical prediction models developed using statistical or machine learning methods. *Biometrical Journal*, 65(8), 2200302.
- [9] Kolambe, M., & Arora, S. (2024). Forecasting the future: A comprehensive review of time series prediction techniques. *Journal of Electrical Systems*, 20(2s), 575-586.
- [10] ArunKumar, K. E., Kalaga, D. V., Kumar, C. M. S., Kawaji, M., & Brenza, T. M. (2022). Comparative analysis of Gated Recurrent Units (GRU), long Short-Term memory (LSTM) cells, autoregressive Integrated moving average (ARIMA), seasonal autoregressive Integrated moving average (SARIMA) for forecasting COVID-19 trends. *Alexandria engineering journal*, 61(10), 7585-7603.
- [11] Zuhairi, A. H., Yakub, F., Omar, M., Sharifuddin, M., Razak, K. A., Faruq, A., & Wijaya, A. A. (2025). Performance Analysis of Tree-Based Ensemble Machine Learning Model for Flood Forecasting in Tropical Regions. *IEEE Access*, 13, 200840-200860.
- [12] Mittendorf, M., Nielsen, U. D., & Bingham, H. B. (2022). Data-driven prediction of added-wave resistance on ships in oblique waves—A comparison between tree-based ensemble methods and artificial neural networks. *Applied Ocean Research*, 118, 102964.
- [13] Abreu, R., Postarnak, S., Vulchanov, V., Baggio, G., & Vulchanova, M. (2023). The association between statistical learning and language development during childhood: A scoping review. *Heliyon*, 9(8).
- [14] Yavuz Ozalp, A., Akinci, H., & Zeybek, M. (2023). Comparative analysis of tree-based ensemble learning algorithms for landslide susceptibility mapping: a case study in Rize, Turkey. *Water*, 15(14), 2661.
- [15] Abatal, A., Korchi, A., Mzili, M., Mzili, T., Khalouki, H., & Billah, M. E. K. (2025). A comprehensive evaluation of machine learning techniques for forecasting student academic success. *Journal of Electronics, Electromedical Engineering, and Medical Informatics*, 7(1), 1-12.
- [16] Mahajan, P., Uddin, S., Hajati, F., & Moni, M. A. (2023, June). Ensemble learning for disease prediction: A review. In *Healthcare* (Vol. 11, No. 12, p. 1808). MDPI.
- [17] Liyanage, S., Abduljabbar, R., Dia, H., & Tsai, P. W. (2022). AI-based neural network models for bus passenger demand forecasting using smart card data. *Journal of Urban Management*, 11(3), 365-380.
- [18] Ranjan, N., Bhandari, S., Khan, P., Hong, Y. S., & Kim, H. (2021). Large-scale road network congestion pattern analysis and prediction using deep convolutional autoencoder. *Sustainability*, 13(9), 5108.
- [19] Ko, Y. G., Jo, K. C., Lee, J. S., & Yu, J. S. (2025). Vehicle Collision Frequency Prediction Using Traffic Accident and Traffic Volume Data with a Deep Neural Network. *Applied Sciences*, 15(18), 9884.
- [20] Zhang, L., & Hua, L. (2025). Major issues in high-frequency financial data analysis: A survey of solutions. *Mathematics*, 13(3), 347.
- [21] Mohammed, A. S., Asteris, P. G., Koopialipoor, M., Alexakis, D. E., Lemonis, M. E., & Armaghani, D. J. (2021). Stacking ensemble tree models to predict energy performance in residential buildings. *Sustainability*, 13(15), 8298.

-
- [22] Sibindi, R., Mwangi, R. W., & Waititu, A. G. (2023). A boosting ensemble learning based hybrid light gradient boosting machine and extreme gradient boosting model for predicting house prices. *Engineering Reports*, 5(4), e12599.
- [23] Fritz, C., Dorigatti, E., & Rügamer, D. (2022). Combining graph neural networks and spatio-temporal disease models to improve the prediction of weekly COVID-19 cases in Germany. *Scientific Reports*, 12(1), 3930.
- [24] Rostami-Tabar, B., & Mircetic, D. (2023). Exploring the association between time series features and forecasting by temporal aggregation using machine learning. *Neurocomputing*, 548, 126376.
- [25] Tan, E., Algar, S., Corrêa, D., Small, M., Stemler, T., & Walker, D. (2023). Selecting embedding delays: An overview of embedding techniques and a new method using persistent homology. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 33(3).
- [26] YOM-TOV, Y. O. R. A. M., & YOM-TOV, J. O. N. A. T. H. A. N. (2005). Global warming, Bergmann's rule and body size in the masked shrew *Sorex cinereus* Kerr in Alaska. *Journal of Animal Ecology*, 74(5), 803-808.
- [27] Barkat, S., Nadeem, M., Rahim, M. A., Mahmoud, M. H., Aubourg, S. P., & Zongo, E. (2026). Effect of immunoglobulin G supplementation on stability, fatty acid profile, organic acids, lipid oxidation and sensory features of cheddar cheese. *BMC microbiology*, 26(1), 258.
- [28] Patel, M. (2025). USA Tariffs Effect: Machine Learning Insights into the Stock Market. *arXiv preprint arXiv:2510.10877*.
- [29] Karbassi Yazdi, A., Durán, C., Derpich, I., & González, G. V. (2025). Forecasting Crop Yields in Rainfed India: A Comparative Assessment of Machine Learning Baselines and Implications for Precision Agribusiness. *Agriculture*, 16(1), 65.
- [30] Hu, J., & Szymczak, S. (2023). A review on longitudinal data analysis with random forest. *Briefings in bioinformatics*, 24(2), bbad002.
- [31] Kim, C., & Park, T. (2022). Predicting determinants of lifelong learning intention using gradient boosting machine (GBM) with grid search. *Sustainability*, 14(9), 5256.
- [32] Alghazzawi, D. M., Alquraishee, A. G. A., Badri, S. K., & Hasan, S. H. (2023). ERF-XGB: Ensemble random forest-based XG boost for accurate prediction and classification of e-commerce product review. *Sustainability*, 15(9), 7076.
- [33] Souadda, L. I., Halitim, A. R., Benilles, B., Oliveira, J. M., & Ramos, P. (2025). Optimizing credit risk prediction for peer-to-peer lending using machine learning. *Forecasting*, 7(3), 35.
- [34] Lamdjad, B. (2025). Comparative forecasting of container throughput for maritime logistics using statistical, machine learning and explainable AI models. *Maritime Business Review*, 1-17.
- [35] Hyndman, R. J., & Rostami-Tabar, B. (2025). Forecasting interrupted time series. *Journal of the Operational Research Society*, 76(4), 790-803.
- [36] Yang, R., Zheng, K., Wu, B., Wu, C., & Wang, X. (2021). Phishing website detection based on deep convolutional neural network and random forest ensemble learning. *Sensors*, 21(24), 8281.
- [37] Natras, R., Soja, B., & Schmidt, M. (2022). Ensemble machine learning of random forest, AdaBoost and XGBoost for vertical total electron content forecasting. *Remote Sensing*, 14(15), 3547.
- [38] Ahn, J. M., Kim, J., & Kim, K. (2023). Ensemble machine learning of gradient boosting (XGBoost, LightGBM, CatBoost) and attention-based CNN-LSTM for harmful algal blooms forecasting. *Toxins*, 15(10), 608.