Jurnal E-Komtek
Vol. 9, No. 2. (2025) pp. 737-746
https://jurnal.politeknik-kebumen.ac.id/index.php/E-KOMTEK
p-ISSN : 2580-3719 e-ISSN : 2622-3066

A Support Vector Regression–Based Model for Multizone Electricity Consumption Forecasting

Nabila Farida[1], Indrianto Indrianto[2]
[1]Department of Master of Computer Science, Institut Teknologi PLN, Indonesia, 11750
[2]Department of the Engineering Profession, Institut Teknologi PLN, Indonesia, 11750

indrianto@itpln.ac.id.

https://doi.org/10.37339/e-komtek.v9i2.3028

Published by Politeknik Piksi Ganesha Indonesia

Artikel Info
Submitted:
03-02-2026
Revised:
11-02-2026
Accepted:
14-02-2026
Online first :
14-02-2026

Abstract

This study aims to develop and evaluate a multi-zone short-term electricity consumption prediction model based on weather factors using Support Vector Regression (SVR). Ten-minute resolution electricity consumption data from three zones is combined with variables such as air temperature, relative humidity, wind speed, and solar radiation. The research stages include data preprocessing, temporal feature engineering, time-based data partitioning, and SVR hyperparameter optimization with RBF kernel. Model performance is evaluated using RMSE, MAE, R^2, and MAPE, and compared with linear regression, Random Forest Regression, and Artificial Neural Network. The results show that SVR provides the best accuracy at high temporal resolution, with MAPE values of 4.78%, 4.11%, and 9.25% at 10-minute aggregation, respectively. Model performance decreases at higher time aggregation levels, indicating the influence of temporal scale and zone load characteristics on SVR effectiveness.

Keywords: Multizone Electric Power Consumption; Short-Term Load Forecasting; Weather; Support Vector Regression (SVR); Intelligent Energy Management Systems

Abstrak

Penelitian ini bertujuan mengembangkan dan mengevaluasi model prediksi konsumsi daya listrik jangka pendek multizona berbasis faktor cuaca menggunakan Support Vector Regression (SVR). Data konsumsi listrik beresolusi 10 menit dari tiga zona dikombinasikan dengan variabel suhu udara, kelembapan relatif, kecepatan angin, dan radiasi matahari. Tahapan penelitian meliputi praproses data, rekayasa fitur temporal, pembagian data berbasis waktu, serta optimasi hiperparameter SVR dengan kernel RBF. Kinerja model dievaluasi menggunakan RMSE, MAE, R^2, dan MAPE, serta dibandingkan dengan regresi linier, Random Forest Regression, dan Artificial Neural Network. Hasil menunjukkan bahwa SVR memberikan akurasi terbaik pada resolusi temporal tinggi, dengan nilai MAPE masing-masing sebesar 4,78%, 4,11%, dan 9,25% pada agregasi 10 menit. Kinerja model menurun pada tingkat agregasi waktu yang lebih besar, menunjukkan pengaruh skala temporal dan karakteristik beban zona terhadap efektivitas SVR.

***Kata-kata kunci**: Konsumsi Daya Listrik Multizona; Peramalan Beban Jangka Pendek; Meteorologis; Support Vector Regression (Svr); Sistem Manajemen Energi Cerdas*

1. Introduction

Increase in electricity demand is a phenomenon that is inseparable from economic growth, technological development, and increased social activity. Electricity has become a fundamental infrastructure that supports various strategic sectors, such as industry, transportation, public services, and households [1]. Therefore, managing a reliable and efficient electricity system is a major challenge for electricity providers, particularly in maintaining a balance between power generation and consumption [2].

As system scale increases, renewable energy penetration grows, and the need for operational reliability and efficiency rises, conventional approaches based on deterministic calculations become less adequate. Modern computer algorithms, especially those based on optimization and machine learning, are capable of processing large amounts of data, capturing nonlinear patterns, and adapting to the temporal and spatial dynamics of electrical systems [3].

As computational capabilities continue to grow and large volumes of data become more readily available, computer-based algorithms have become essential in facilitating the transition of electrical power systems toward smarter, more efficient, and sustainable operations. Recent studies indicate that data-driven models such as artificial neural networks, support vector regression, long short-term memory architectures, and convolutional neural networks are increasingly utilized for short- and medium-term electricity load forecasting, achieving higher accuracy compared to traditional forecasting approaches [4]. In particular, machine learning algorithms have proven effective in improving prediction performance by simultaneously processing historical data and external factors [5].

For this reason, this study applies the SVR method not only to one zone but on a multi-zone scale covering residential, commercial, and industrial zones, which increases the complexity of the predicted load patterns. Support Vector Regression (SVR) is used in this study because it has relatively better performance in modeling the non-linear relationship between meteorological variables and electricity consumption, which is complex and fluctuating. SVR is capable of producing optimal generalization through the use of the kernel trick and maximum margin so that the model remains robust even though the data size is not too large [6],[7]. SVR has also been proven to handle energy forecasting issues. This is because SVR is capable of handling abnormally distributed variables, small outliers, and non-linear seasonal dynamics. The multi-zone approach adds complexity to load patterns, requiring a model capable of capturing the heterogeneity of characteristics between zones [8],[9].

Research over the past five years has demonstrated the use of SVR in research on addressing challenges in electricity load forecasting [10],[11]. SVR in the field of renewable energy [12],[13],[14]. Research on control charts that monitor time series with conditional heteroscedasticity [15].

2. Method

2.1 Dataset Description

This research employs the Tetuan City Power Consumption Dataset sourced from Kaggle. The dataset contains 52,416 records collected at 10-minute intervals, yielding 144 observations per day. It spans almost one year of data and includes temporal information in the form of DateTime, meteorological features such as air temperature, relative humidity, wind speed, and solar radiation, as well as target variables representing electricity consumption in Zone 1, Zone 2, and Zone 3. To preserve the chronological structure of the time series, the dataset is split temporally into 80% training data and 20% testing data.

2.2 Feature Engineering

The features used in the model consist of temporal and meteorological features.

Tabel 1. Feature Engineering

Features	Description
Hour	Hours (0–23)
Day	Day
Month	Month
Temperature	Air temperature
Humidity	Humidity
Wind Speed	Wind speed
Solar Radiation	Solar radiation

Temporal features help models recognize daily and seasonal periodic patterns.

2.3 SVR Modeling

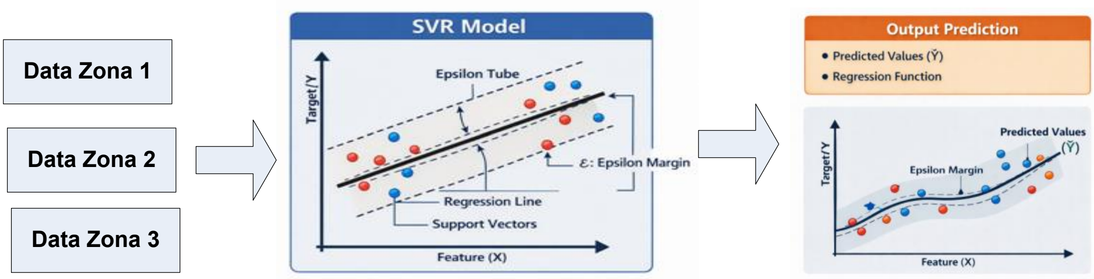

Figure 1. SVR Model

Figure of the Support Vector Regression (SVR) workflow in a multi-zone scheme, where data from the three zones are used as separate targets in the modeling scheme, with the same weather and temporal features. Each zone represents different load characteristics or data patterns (e.g., residential, commercial, and industrial), thereby enriching the variety of features

(X) and targets (Y) studied by SVR. In the modeling stage, SVR forms a regression function by maximizing the margin and minimizing errors using the concept of epsilon-insensitive loss, where only data outside the epsilon tube acts as support vectors. This mechanism makes the model more robust against noise and small fluctuations between zones. The final result is a prediction output (Y1) in the form of a regression curve that represents the estimated target value, while also capturing the combined nonlinear patterns of the three zones, making it suitable for predicting electricity load or energy requirements in complex multi-zone systems **[16].**

SVR regression function

$$f(x) = w^T \varphi(x) + b \tag{1}$$

with w as the weight vector, $\phi(x)$ as the nonlinear mapping function, and b as the bias

SVR optimization function

$$\min \left(\frac{1}{2}\right) ||w||^2 + C\, \Sigma\, (\xi_i + \xi_i *) \tag{2}$$

with restrictions:

$$\begin{gathered} y_i - (w^T \varphi(x_i) + b) \leq \varepsilon + \xi_i \\ (w^T \varphi(x_i) + b) - y_i \leq \varepsilon + \xi_i * \\ \xi_i, \xi_i * \geq 0 \end{gathered} \tag{3}$$

2.4 Kernel Function and Hyperparameters

The kernel used in this study is Radial Basis Function (RBF) because it is capable of mapping nonlinear relationships between weather features and electricity consumption. The RBF kernel function is formulated as:

$$K(x_i, x_j) = e^{\{-\gamma \backslash |x_i - x_j \backslash|^2\}} \tag{4}$$

where γ is the parameter that controls the range of influence of each support vector on the regression function.

Tabel 2. Kernel Function

Parameter	Nilai
Kernel	Radial Basis Function (RBF)
C	1.0
Epsilon	0.1

1. C = 1.0, as a regularization parameter to control the trade-off between model complexity and prediction error.

2. Epsilon = 0.1, as the error tolerance limit in the loss function.

To handle multizone predictions simultaneously, the SVR model is wrapped using MultiOutputRegressor, so that one SVR model is applied to each target variable. Before the training process, the data has undergone standardization because SVR is sensitive to differences in data scale. The model is then trained using training data and used to generate predictions on test data.

2.5 Model Evaluation

Model evaluation represents a vital step in the machine learning process, as it determines the model's ability to produce reliable and accurate predictions. The effectiveness of the Support Vector Regression (SVR) model is assessed using multiple performance metrics, namely MAE, MSE, RMSE, R^2, and MAPE.

Mean Absolute Error (MAE) represents the average absolute error between actual data and model predictions. (3):

$$MAE = \frac{1}{n} \sum_{\{i=1\}}^{\{n\}} |y_i - \hat{y_i}| \tag{5}$$

Mean Squared Error (MSE) calculates the average square of the difference between the actual value and the predicted value.(4):

$$MSE = \frac{1}{n} \sum_{\{i=1\}}^{\{n\}} |y_i - \hat{y_i}|^2 \tag{6}$$

The coefficient of determination (R^2) measures how much of the variation in the actual data can be explained by the model. (5):

$$MSE = \frac{1}{n} \sum_{\{i=1\}}^{\{n\}} |y_i - \hat{y_i}|^2 \tag{7}$$

Mean Absolute Percentage Error (MAPE) measures the average percentage error between actual values and predicted values. (6):

.

$$\text{MAPE} = \left(\frac{1}{n}\right) \times \Sigma \left|\frac{(yi - \hat{y}i)}{yi}\right| \times 100\% \tag{8}$$

3. Results and Discussion

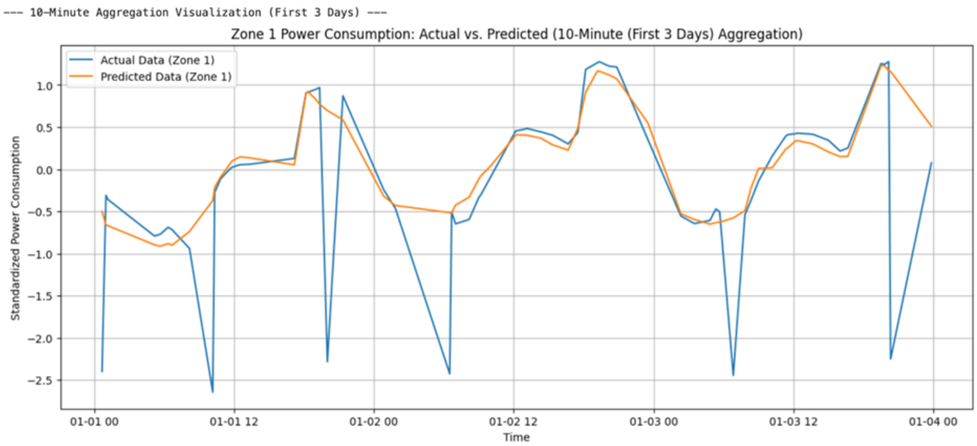

Figure 2. Zone 1 Chart Over 3 Days

The figure illustrates a comparison between measured power consumption and model predictions in Zone 1 using a 10-minute aggregation interval over the first three days of monitoring. Time is shown on the horizontal axis, while normalized power consumption is displayed on the vertical axis. The blue line represents the actual measurements, which exhibit significant variability at certain periods, including sharp declines that may reflect negative load spikes or operational irregularities. In contrast, the orange line corresponds to the predicted values, which appear smoother and more consistent, capturing the overall consumption trend but not fully reflecting short-term extreme variations. The close alignment of the two curves indicates that the model effectively captures the general daily consumption pattern. However, noticeable discrepancies at extreme points suggest limitations in modeling very short-term fluctuations. Overall, the results demonstrate that while the model successfully represents macro-level power consumption behavior, it remains constrained in accurately responding to rapid, short-duration changes, despite showing some sensitivity to outliers or abrupt shifts in demand.

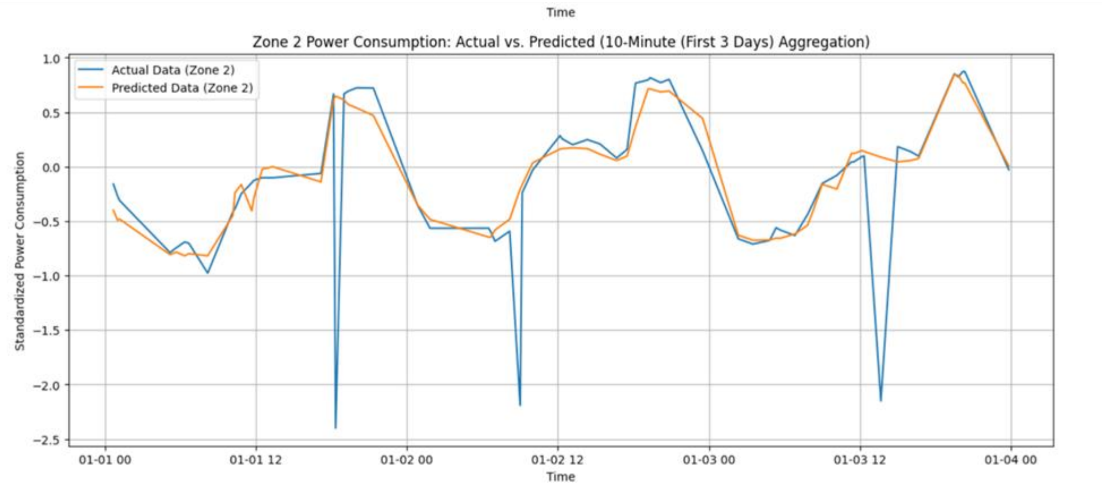

Figure 3. Zone 2 Chart Over 3 Days

The figure presents a comparison between observed electricity consumption and predicted values in Zone 2 using a 10-minute aggregation interval during the first three days of analysis. The horizontal axis illustrates the temporal progression of power consumption, while the vertical axis represents standardized consumption levels. The actual consumption pattern (blue line) exhibits relatively stable fluctuations that follow a daily cycle; however, several abrupt declines are evident at certain time intervals, suggesting anomalies, operational disruptions, or sudden load variations. In contrast, the predicted values (orange line) display a smoother and more stable behavior, effectively tracking the overall trend of actual consumption, particularly during periods of daily load increase and decrease. Discrepancies between the two curves are most pronounced at extreme points, indicating that the model tends to smooth sharp spikes or drops. Nevertheless, the results demonstrate that the model is capable of capturing the general characteristics and temporal dynamics of power consumption in Zone 2 with good performance.

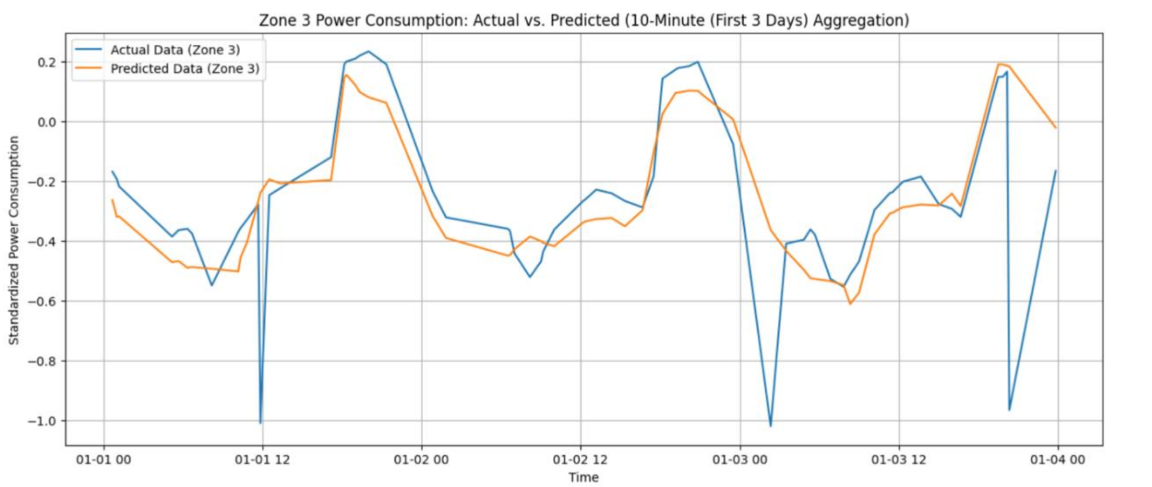

Figure 4. Zone 3 chart over 3 days

.

The figure presents a comparison between measured electricity consumption and predicted values in Zone 3 using a 10-minute aggregation interval over the first three days of observation. Normalized power consumption is shown on the vertical axis, while time is represented on the horizontal axis. The actual consumption pattern (blue curve) exhibits relatively smaller variations compared to those observed in Zone 1 and Zone 2, although several abrupt decreases are still present, indicating potential anomalies or sudden load changes. In contrast, the predicted values (orange curve) are smoother and more stable, closely following the overall trend of the actual data, particularly during periods of daily load rise and decline. The most pronounced discrepancies between the two curves occur at extreme points, where the model tends to smooth out sharp fluctuations. Overall, the results indicate that the prediction model effectively captures the general behavior of power consumption in Zone 3, despite its limited responsiveness to extreme increases or decreases in load.

Table 3. Metric Test

Aggregation Level	Zone	MAE	MSE	R-squared (R^2)	MAPE
10-Minute (First 3 Days)	Zone 1	0.3267	0.5644	0.3754	4.78%
10-Minute (First 3 Days)	Zone 2	0.2001	0.2862	0.3926	4.11%
10-Minute (First 3 Days)	Zone 3	0.1129	0.0422	0.4557	9.25%
Hourly (First 7 Days)	Zone 1	0.3838	0.5099	0.2701	12.32%
Hourly (First 7 Days)	Zone 2	0.2714	0.2418	0.4082	12.03%
Hourly (First 7 Days)	Zone 3	0.1723	0.0804	0.322	11.87%
Daily (First 30 Days)	Zone 1	0.2057	0.0602	0.0097	8.19%
Daily (First 30 Days)	Zone 2	0.1752	0.0494	0.3161	8.14%
Daily (First 30 Days)	Zone 3	0.1494	0.1985	-0.052	10.32%
Monthly	Zone 1	0.2263	0.0528	-3.0337	8.57%
Monthly	Zone 2	0.216	0.0482	-0.7897	9.08%
Monthly	Zone 3	0.0787	0.0114	0.8242	6.94%

The summary table of evaluation metrics shows the performance of the prediction model in three zones with various levels of time aggregation, namely 10-minute, hourly, daily, and monthly. In general, the model's performance is most optimal at 10-minute aggregation, as indicated by the lowest error value and relatively higher R^2 compared to other aggregations (first 3 days), especially in Zone 3, which produces the lowest MAE and MSE values and the highest R^2 (0.4557), indicating the model's relatively good ability to explain actual data variations at high time resolution. At the hourly aggregation, the model performance tended to decline across all zones, as reflected in the increase in MAE and decrease in R^2 values, especially in Zone 1. In daily aggregation, model performance becomes more unstable, with low to negative R^2 values indicating that the model is less able to explain data variation on a larger aggregation scale in Zone 3, suggesting that the model is less able to consistently represent daily patterns. The most significant decline is seen in the monthly aggregation, where the R^2 values in Zone 1 and Zone 2 become negative, indicating that the model's predictions are worse than the simple average approach, although Zone 3 still shows relatively better performance with an R^2 of 0.8242 and the lowest MAPE (6.94%). Overall, this table indicates that the effectiveness of SVR is greatly influenced by the time scale and characteristics of each zone, with optimal performance at higher time resolutions and in zones with more stable consumption patterns.

4. Conclusion

This study successfully developed a short-term electricity consumption prediction model based on Support Vector Regression (SVR) in a multizone scheme. The results show that SVR is capable of providing consistent and stable prediction performance across all zones, especially under conditions of high weather variability. At a 10-minute data resolution, the model showed the best performance, with low error and the highest R^2 value. The best performance was obtained at a 10-minute resolution, while the model's accuracy decreased as the time aggregation scale increased (hourly, daily, monthly), indicating that SVR is more effective at high time resolutions for capturing rapidly changing power consumption dynamics. Research findings show that SVR is effective in representing complex and varied electricity consumption patterns between zones in modeling nonlinear relationships and is known to be robust against noise in energy time series data. Differences in characteristics between zones cause different sensitivities to meteorological variables, and SVR is able to represent differences in consumption patterns between zones, even though sensitivity to meteorological variables differs depending on the characteristics of each zone. Overall, this study confirms that SVR shows good effectiveness for weather-based multizone electricity consumption forecasting at high time resolution, but its performance declines at larger aggregation scales, and it has the potential to be an important component in the development of smart and sustainable energy management systems.

References

[1] G. S. Mutumba, G. Mubiinzi, and D. Amwonya, "Electricity consumption and economic growth: Evidence from the East African community," *Energy Strateg. Rev.*, vol. 54, no. June, p. 101431, 2024, doi: 10.1016/j.esr.2024.101431.

[2] M. T. Noorzai, A. Bełdycka-Bórawska, A. Kutlar, T. Rokicki, and P. Bórawski, "Exploring the Link Between Energy Consumption, Economic Growth, and Ecological Footprint in the Major Importers of Poland Energy: A Panel Data Analysis," *Energies*, vol. 18, no. 13,

pp. 1–26, 2025, doi: 10.3390/en18133303.

[3] S. K. Panda, "Electrical load and solar power forecasting using machine learning techniques," *J. King Saud Univ. - Eng. Sci.*, vol. 37, no. 4, 2025, doi: 10.1007/s44444-025-00012-y.

[4] S. Akhtar, M. Adeel, M. Iqbal, A. Namoun, A. Tufail, and K. H. Kim, "Deep learning methods utilization in electric power systems," *Energy Reports*, vol. 10, pp. 2138–2151, 2023, doi: 10.1016/j.egyr.2023.09.028.

[5] M. Pournabi, M. Mohammadi, S. Afrasiabi, and P. Setoodeh, "Power system transient security assessment based on deep learning considering partial observability," *Electr. Power Syst. Res.*, vol. 205, no. December 2021, p. 107736, 2022, doi: 10.1016/j.epsr.2021.107736.

[6] Y. Jia, S. Zhou, Y. Wang, F. Lin, and Z. Gao, "A quadratic ν-support vector regression approach for load forecasting," *Complex Intell. Syst.*, vol. 11, no. 1, pp. 1–12, 2025, doi: 10.1007/s40747-024-01730-7.

[7] L. O. M. Yusuf, S. Handoko, and I. Setiawan, *Forecasting Short-Term Electrical Loads Using Support Vector Regression with Gaussian Kernel Functions*. Atlantis Press International BV, 2023.

[8] Y. Chen *et al.*, "Short-term electrical load forecasting using the Support Vector Regression (SVR) model to calculate the demand response baseline for office buildings," *Appl. Energy*, vol. 195, pp. 659–670, 2017, doi: 10.1016/j.apenergy.2017.03.034.

[9] J. Duan, X. Tian, W. Ma, X. Qiu, P. Wang, and L. An, "Electricity consumption forecasting using support vector regression with the mixture maximum correntropy criterion," *Entropy*, vol. 21, no. 7, 2019, doi: 10.3390/E21070707.

[10] Z. Zhang, Q. Zhang, H. Liang, and B. Gorbani, "Optimizing electric load forecasting with support vector regression/LSTM optimized by flexible Gorilla troops algorithm and neural networks a case study," *Sci. Rep.*, vol. 14, no. 1, pp. 1–20, 2024, doi: 10.1038/s41598-024-73893-9.

[11] R. Shang and Y. Ma, "Electric Vehicle Charging Load Forecasting Based on K-Means++-GRU-KSVR," *World Electr. Veh. J.*, vol. 15, no. 12, 2024, doi: 10.3390/wevj15120582.

[12] X. Xiang, X. Li, Y. Zhang, and J. Hu, "A short-term forecasting method for photovoltaic power generation based on the TCN-ECANet-GRU hybrid model," *Sci. Rep.*, vol. 14, no. 1, pp. 1–12, 2024, doi: 10.1038/s41598-024-56751-6.

[13] Y. Sun and S. Zhang, "A Multiscale Hybrid Wind Power Prediction Model Based on Least Squares Support Vector Regression–Regularized Extreme Learning Machine–Multi-Head Attention–Bidirectional Gated Recurrent Unit and Data Decomposition," *Energies* , vol. 17, no. 12, 2024, doi: 10.3390/en17122923.

[14] S. H. Moon, D. Y. Kim, and Y. H. Kim, "Post-Processing Maritime Wind Forecasts from the European Centre for Medium-Range Weather Forecasts around the Korean Peninsula Using Support Vector Regression and Principal Component Analysis," *J. Mar. Sci. Eng.*, vol. 12, no. 8, 2024, doi: 10.3390/jmse12081360.

[15] C. K. Kim, M. H. Yoon, and S. Lee, *Robust control chart for nonlinear conditionally heteroscedastic time series based on Huber support vector regression*, vol. 19, no. 2 February. 2024.

[16] S. Sitara, W. Fatima, and A. Rahimi, "A Review of Time-Series Forecasting Algorithms for Industrial," *mdpi*, 2024.