A Multivariate Hybrid ARIMA-GA-SVR Approach for Short-Term Solar Radiation Forecasting

Andressa F. Arruda[1], **João F. L. de Oliveira**[1]

[1]Escola Politécnica de Pernambuco – Universidade de Pernambuco (UPE)
Recife – PE – Brazil

afa@ecomp.poli.br, fausto.lorenzato@upe.br

Abstract. *Accurate solar radiation forecasting is crucial for integrating photovoltaic energy into the electrical grid and addressing the challenges of intermittency and variability. In this work, a new multivariate hybrid approach was proposed that combines the linear modeling capabilities of ARIMA with the nonlinear learning ability of a Support Vector Machine (SVR). The key to innovation in this method is the design of a multivariate feature set that incorporates past residuals of the target variable, observed values of exogenous meteorological variables, and one-step-ahead linear forecasts for both the target and exogenous variables generated by ARIMA. By including these linear forecasts in the SVR's input, the model can better capture the dynamic relationship between solar radiation and future meteorological conditions. In addition, a GA-based feature selection step is employed to select the optimal subset of lagged features, aiming to reduce model complexity and computational cost.The proposed ARIMA–GA-SVR model was evaluated using hourly data from five Brazilian cities. Its performance was compared with hybrid models from the literature (OTL-ARIMAX-PSO-MC, ETS-ANN, ARIMA-ANN, KC-MLP) using RMSE, MAE, ARV, and IA as evaluation metrics. The results show that the ARIMA–GA-SVR model achieved the best overall performance across all the analysed cities, outperforming all compared models in most of them. These findings support the conclusion that incorporating linear forecasts of exogenous variables into the nonlinear residual modeling stage is an effective way to improve short-term solar radiation prediction.*

1. Introduction

In pursuit of a more sustainable energy matrix, renewable sources have emerged as alternatives to fossil fuels, which are associated with negative environmental impacts, including growing CO_2 emissions and global temperature rise [Krishnan et al. 2023]. Among renewable energy sources, solar energy is considered one of the most promising due to its abundance. Approximately 1367 W/m^2 of solar irradiance reaches the Earth's surface, equivalent to 1.74×10^{17} W per year [Kumari and Toshniwal 2021]. However, solar power generation is highly sensitive to weather conditions such as precipitation, temperature, and humidity, leading to intermittency and volatility. Since photovoltaic power generation is influenced by solar radiation, accurate solar radiation forecasting models are essential for balancing supply and demand and maintaining reliable power system operation [Aguiar et al. 2016].

Although the ARIMA model can describe time series through a linear combination of autoregressive and moving average components, by assuming a linear correlation between future forecasts and past observations, it may be inadequate for modeling real-world time series, such as solar radiation, which exhibits nonlinear patterns [Oliveira et al. 2022]. In contrast, machine learning models such as SVR (Support Vector Regressor), despite being able to capture nonlinear patterns, are sensitive to hyperparameter selection and can struggle with overfitting and underfitting when handling high-dimensional data [Wang et al. 2024].

Some studies have focused on machine learning and hybrid approaches that rely only on past observations of solar radiation [Solano et al. 2022]. This limits their ability to capture the multivariate and nonlinear relationships between solar radiation and exogenous variables such as temperature . For instance, residual hybrid approaches often adopt a univariate perspective, in which the residuals of the target series are predicted solely from their own past values, as in [Zhang 2003]. As a result, these methods may not capture relationships involving exogenous variables and endogenous series.

To address these limitations, this paper proposes a hybrid method that combines linear modeling using ARIMA with nonlinear modeling using a Support Vector Regressor (SVR). ARIMA is used to model linear components and generate one-step ahead forecasts, while SVR captures nonlinear relationships using lagged features from residuals and exogenous variables. Initially, a 24-hour lag structure is constructed and a Genetic Algorithm (GA) is used to select the most relevant lags with the objective of achieving a balance between computational efficiency and predictive performance. By incorporating linear forecasts of exogenous variables into the SVR input features, the model can anticipate meteorological changes that affect solar radiation. As a result, the SVR not only models the residuals, but also captures the dynamic relationships between solar radiation and the predicted future meteorological conditions. This approach is referred to as ARIMA-GA-SVR.

The main contributions and objectives of this work are:

- To propose a hybrid ARIMA–GA-SVR model that incorporates linear forecasts of exogenous variables into the nonlinear residual modeling stage,
- To implement a multivariate feature engineering strategy that combines lagged residuals of the target variable, lagged values of exogenous variables, and ARIMA predictions for both the target and exogenous variables, forming a feature set based on 24-hour lag window. A Genectic Algorithm is then applied to select the most relevant lags,
- To compare the performance of the proposed model with existing hybrid approaches in the literature using the RMSE, ARV, MAE, and IA metrics, and to apply the Friedman–Nemenyi test to verify whether there are statistically significant differences among the models and to rank them.

2. Related Works

Hybrid models improve forecasting performance by combining multiple algorithms. Among these hybrid approaches, three main categories are highlighted in this work: the General Ensemble Learning Approach (GELA), the Cluster-Based Ensemble Learning Approach (CELA), and the Residual-Based Ensemble Learning Approach (RELA) [Guermoui et al. 2020].

GELA-type hybrid models improve forecasting accuracy by combining diverse models, each offering a unique contribution. An effective strategy, as used in the HetDS architecture proposed by [Santos et al. 2022], is to build a pool of diverse models and dynamically select the best m models for each test instance based on their local performance (RMSE) within the k-nearest neighbors in the validation set. This approach outperformed individual models, except in two cases where SVR performed best.

CELA uses data mining methods, with a focus on unsupervised K-means algorithms, to group similar samples into clusters within the dataset. Each cluster is modeled using linear or nonlinear forecasting methods, and the final prediction is obtained by combining the outputs for all clusters. The method developed by [Ghofrani et al. 2016] used an improved K-means with a new initialization strategy and optimal cluster selection, and then trained an MLP model for each cluster. This method achieved lower MSE than other clustering methods in the literature and lower RMSE, rRMSE, MAE, and rMAE than classical models such as persistence, ARIMA, and MLP across various wind power datasets. Alternatively, the method SOM-SVR-PSO proposed by [Dong et al. 2015] uses a Self-Organizing Map (SOM) to divide the data into different regions, and an SVR (Support Vector Regression) tuned by Particle Swarm Optimization (PSO) was applied in each region to forecast the hourly global solar radiation data. Tested on datasets from Colorado and Singapore, the SOM–SVR–PSO hybrid outperformed classical models such as ARIMA, LES, SEM, and Random Walk.

Residual ensemble learning approaches (RELA) assume that solar radiation time series can be decomposed into linear and nonlinear components. A statistical model is used for the linear component, while a nonlinear model is fitted to the residuals. The final forecast combines both outputs. The model implemented by [Zhang 2003] is an example of RELA. The time series is divided into two components: an ARIMA model captures the linear part, while a neural network is trained on the residuals to learn any remaining nonlinear relationship. The final forecast is the sum of both estimates. In tests on classic datasets such as sunspots, Canadian lynx, and exchange rates, the hybrid consistently outperformed standalone models in all datasets in terms of MSE. In contrast, [Panigrahi and Behera 2017] followed the same framework but used an ETS model instead of ARIMA to capture the linear component. The ETS–ANN hybrid, tested in 16 real-world time series, achieved the lowest RMSE in seven datasets, outperforming ARIMA, ETS, MLP.

More recently, [Zhang et al. 2025] proposed a variation of the RELA in which the residual component is corrected using a Markov-chain instead of being modeled by a neural network. In this approach, the linear component is modeled by ARIMAX, which incorporates both endogenous and exogenous inputs, with the optimal lags of the exogenous variables determined using the OTL algorithm. In addition, particle swarm optimization (PSO) is employed to tune the hyperparameters of the Markov chain, which are the number of states and the correction coefficient. In a similar way, [Ahmadi et al. 2025] introduced a hybrid residual-based framework that decomposed the wind power time series into trend and residual components using dual-stage Kalman filtering module. In their methodology, the trend component was modeled by a multilayer perceptron (MLP) due to its more predictable nature, while the residual component, characterized by irregular fluctuations, is handled by a deep residual Bi-LSTM. Finally, a recurrent neural network

(RNN) is employed to combine the outputs of the components.

The proposed method ARIMA-GA-SVR belongs to the RELA category. Unlike ensemble-based approaches such as HetDS, it is not a combination of independent models, but rather adopts a residual-based hybrid approach in which a nonlinear model is applied to the residuals of a linear model.

3. Proposed Method

The proposed approach is a hybrid model that uses the strengths of ARIMA and Support Vector Regressor (SVR) to improve solar radiation forecasting. In this residual method, an ARIMA model is used to capture linear relationships and trends. In contrast, a SVR is used to model more complex, nonlinear patterns between exogenous variables and solar radiation. Figure 1 illustrates the main stages of the proposed method, which can be divided into three main steps: linear modeling using ARIMA, feature engineering and lag selection using Genetic Algorithm (GA), and nonlinear modeling using SVR. In this context, Y_t denotes the solar radiation at time t, and $X_t^{(j)}$ corresponds the j-th exogenous variable at time t, including meteorological variables such as temperature and wind speed.

First, independent ARIMA models are fitted to both the solar radiation series and each exogenous variable, generating one-step-ahead predictions. These forecasts are used as additional features for the SVR model, incorporating information from linear components into the input.

In the feature engineering and lag selection stage, to address the nonlinear patterns that ARIMA cannot capture, residuals are calculated, the differences between observed values and the ARIMA predictions, which are then used as the target for the SVR. The feature engineering process creates a set of features, including lagged residuals of solar radiation, lagged ARIMA forecasts for the target and exogenous variables, and lagged observed values of the exogenous variables. Initially a 24-hour sliding window is considered to generate lagged features from previous 24 time steps.

To reduce the computational cost associated with the feature engineering process and improve the model's performance, an additional Genetic Algorithm (GA)-based feature selection step is applied to identify the most relevant lags. The GA uses a binary chromosome, where each gene corresponds to a candidate temporal lag. If a gene's value is 1, that lag is selected; otherwise, it is not selected. The same selected subset of lags is applied to all input features. Next, the selected lag features are reshaped into a two-dimensional format and used as input for the SVR model. Candidate lag subsets are evaluated based on the validation RMSE of an SVR model as the fitness function of the Genetic Algorithm (GA). Figure 2 shows how these features are constructed and the optimal lag subset is selected using Genetic Algorithm.

To retain the two best-performing candidates of each generation, an elitist selection approach was applied. These individuals were used to produce new individuals through uniform crossover, and the resulting offspring were submitted to a bit-flip mutation according to the specified mutation probability.

In the nonlinear modeling stage, the selected features are used as input to train an SVR model. SVR predicts the nonlinear residual component, and the final solar radiation forecast is obtained by summing the SVR and ARIMA forecast. An overview of the entire

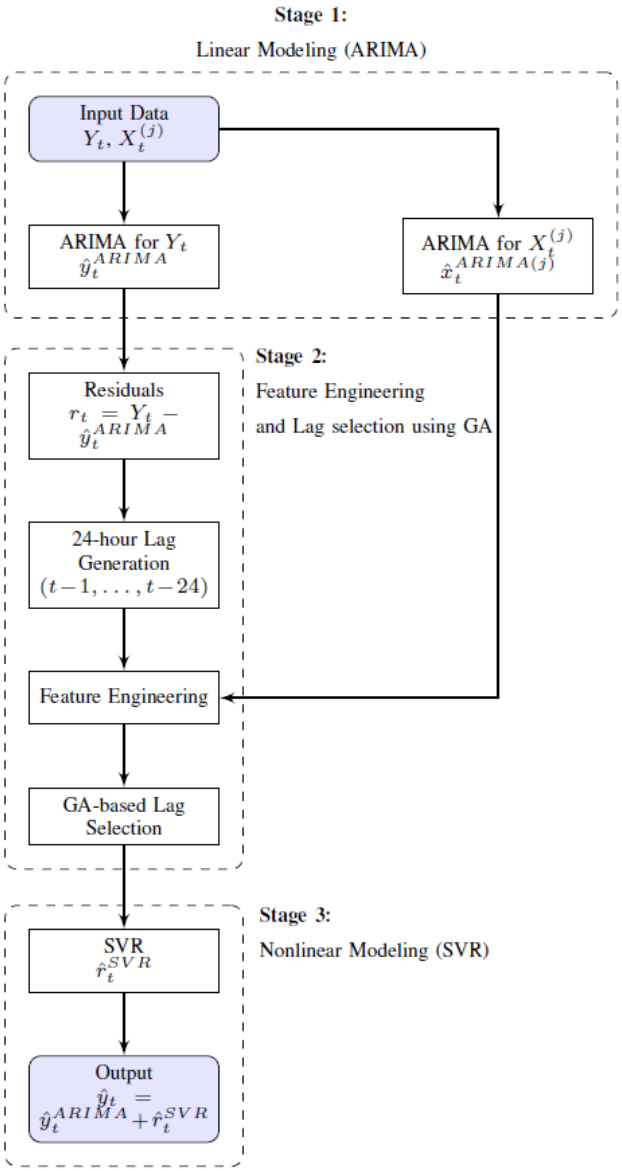


Figure 1. Hybrid ARIMA-GA-SVR forecasting methodology.

process is shown in Figure 1.

4. Experiments

A total of five datasets, collected from five different Brazilian cities, containing hourly solar irradiance and other meteorological variables, were used to evaluate the experiments. All 24-hour data information was kept. For nighttime hours (6:00 p.m. to 6:00 a.m. local time, or 21:00 to 9:00 UTC), solar irradiance was set to zero. In contrast, daytime measurements (6:00 a.m. to 6:00 p.m.) were reported to correspond with actual values observed. In addition, the data was split into three sets: training, validation, and test. Model parameters were determined using a grid search over the ranges given in Table 1. During this process, the validation set was used to select the optimal hyperparameters. For the SVR model, γ is the kernel parameter, C is the regularization constant, and ε is the

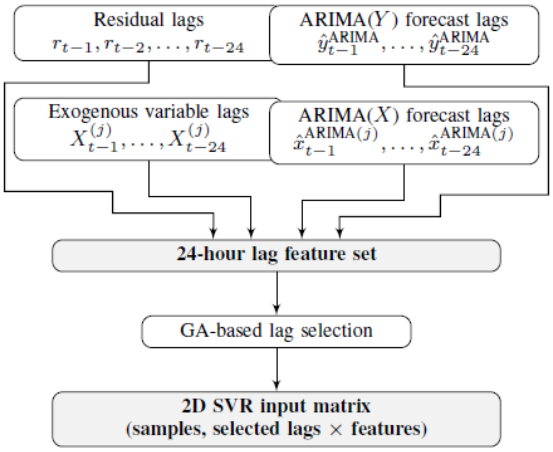


Figure 2. Detailed feature engineering and lag selection for the hybrid ARIMA-GA-SVR model.

value at which the maximum diverging value from the observed one occurs. In ARIMA, p, d, and q represent the autoregressive order, the degree of differencing, and the moving-average order, respectively. These parameters were selected automatically for each time series based on the Akaike Information Criterion (AIC). Additionaly, specifically for the KC-MLP baselinethe, silhouette score was used as a clustering validation criterion to guide the K-means process and select the optimal k.

For the prediction tasks, a total of 8,760 hourly solar irradiation samples were used, with 60% allocated for training, 20% for validation, and the remaining 20% for testing. Additionally, all datasets used in the experiments were scaled to the [0, 1] range.

In this work, the following metrics were used to compare the models' results: root mean squared error (RMSE), mean absolute error (MAE), average relative variance (ARV), and index of agreement (IA).

4.1. Data Description

Meteorological data were collected from five state capitals in the Northeast region of Brazil (Salvador, Teresina, São Luís, João Pessoa, and Fortaleza) based on information provided by the National Institute of Meteorology (INMET)[1]. The observation period covers January 1 to December 31, 2023, and includes 8,760 hourly solar radiation samples, with both daytime and nighttime hours included. The objective is to predict hourly solar radiation (kJ/m²) as a regression task.

Each dataset includes 16 features. To select the most relevant predictors, a correlation heatmap was used, and variables with a correlation greater than 0.7 with the target were selected. The chosen features were: hourly dry-bulb air temperature (°C), maximum air temperature at the previous hour (°C), minimum air temperature at the last hour (°C), hourly relative humidity (%); maximum relative humidity at the last hour (%), minimum relative humidity at the previous hour (%), and hourly wind speed (m/s). The selected features highlight strong relationships between solar radiation and the selected meteorological variables, thereby justifying their use as exogenous predictors in the model. A

[1] https://portal.inmet.gov.br/dadoshistoricos

Table 1. Hyperparameters evaluated for each model.

Model	Parameters	Values
ARIMA	Orders (p, d, q)	Hyndman Method[Hyndman and Khandakar 2008]
SVR	Kernel	RBF
	γ	{0.1, 0.01, 0.001}
	C	{10, 100, 1000}
	ε	{0.1, 0.01, 0.001}
MLP	Units in hidden layer	{20, 50, 100}
	Activation Function	Sigmoid
	Algorithm	Adam
RF	Number of estimators	{50, 100, 200}
	Max depth	{5, 10, 15}
	Max features	{0.6, 0.8, 1}
GB	Number of estimators	{50, 100}
	Learning rate	{0.1, 0.3}
	Max depth	{5, 10}
	Subsample	{0.6, 1.0}
	Max features	{0.6, 1.0}
ELM	Units in hidden layer	{20, 50, 100, 200, 500}
	Activation Function	Hyperbolic Tangent
	Algorithm	Moore–Penrose pseudo-inverse
GA	Number of generations	{5, 10}
	Population size	{10, 20}
	Mutation rate	{0.05, 0.1, 0.5}

24-hour temporal window was also applied to capture the daily seasonality of solar radiation.

4.2. Results

In Table 2, the results of the experiments conducted in the five cities analyzed are presented. The performance of the analyzed models is evaluated on the last 20% of each city's dataset. For all models, except the deterministic SVR and ARIMA, 30 independent runs were performed, and the results shown in the table are the average of these runs for each evaluated metric. The metrics used in the experiments were RMSE, MAE, ARV, and IA. The best results for each metric in each city are highlighted in bold.

In the experiments, hybrid models from the literature were used as baselines and are abbreviated for clarity. The residual hybrids by [Zhang 2003] and [Panigrahi and Behera 2017] are labeled as ARIMA-ANN and ETS-ANN, respectively. For simplicity, the cluster-based KC-MLP proposed by [Ghofrani et al. 2016] is referred to as KC-MLP in this work. The OTL-ARIMAX-PSO-MC model refers to the hybrid forecasting approach proposed by [Zhang et al. 2025]. Among all models evaluated, only KC-MLP, OTL-ARIMAX-PSO-MC and the proposed ARIMA–GA-SVR model include exogenous variables as input features.

According to the results presented in Table 2, the proposed ARIMA–GA-SVR model achieved the best overall performance across all evaluated cities in terms of the evaluated metrics. This method consistently obtained lowest RMSE and MAE, outperforming traditional and hybrid models. In João Pessoa, Fortaleza, and Salvador, the pro-

posed model achieved the best results across all evaluated metrics. Furthermore, in São Luis, it obtained the lowest values of RMSE and MAE, while, in Teresina, achieved the lowest RMSE, maintaining competitive performance across the remaining metrics in both cities.

To statistically assess the performance of the analyzed models, the Friedman–Nemenyi test was used at the 95% confidence level. Figure 3 presents the results of this test, highlighting the average ranking obtained by each model using RMSE as the comparison metric. The proposed method achieved the best average ranking among all models evaluated when considering the five datasets analyzed.

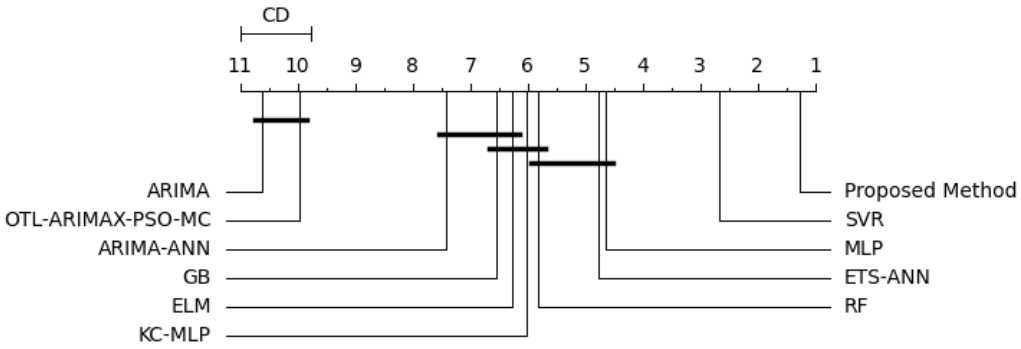


Figure 3. Results of the Friedman–Nemenyi test with 95% confidence.

Despite achieving statistically superior performance, as confirmed by the Friedman-Nemenyi test, the model's performance may vary according to the quality of the meteorological data, such as missing values and measurement inaccuracies. Moreover, since the region studied in this article for the experiments was the Northeast of Brazil, investigating how the model performs in other climatic conditions remains open for future research.

Figures 4 to 8 show the forecasts of the proposed method and the ARIMA model, compared to the actual solar radiation values, using the last three days of 2023 as a reference.

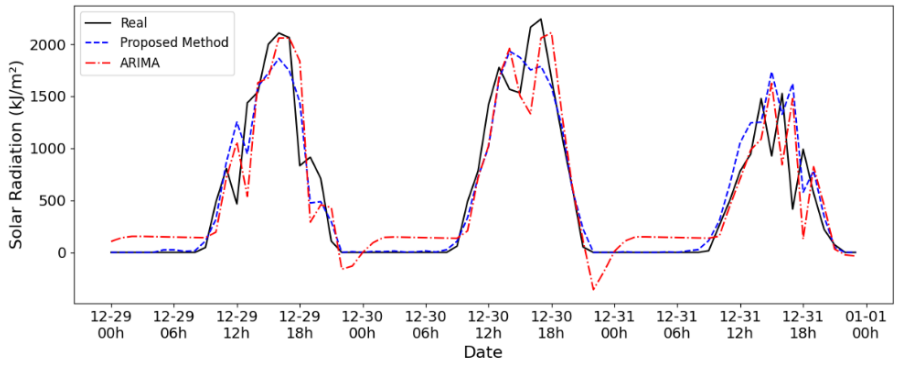


Figure 4. Forecast vs. actual solar radiation values — Teresina.

Table 2. Performance results of the models by city.

City	Model	RMSE	MAE	ARV	IA
Teresina	ARIMA	274.8478	208.0319	0.1030	0.9750
	SVR	197.3843	**103.1206**	**0.0489**	**0.9877**
	MLP	203.5149	114.1738	0.0564	0.9864
	ELM	215.5612	147.4810	0.0616	0.9849
	RF	213.6516	117.9659	0.0648	0.9847
	GB	207.3212	113.4393	0.0595	0.9857
	ARIMA-ANN	212.9292	130.2168	0.0599	0.9853
	ETS-ANN	216.2999	117.8712	0.0560	0.9856
	KC-MLP	265.7186	165.0338	0.1261	0.9735
	OTL-ARIMAX-PSO-MC	230.2841	151.6176	0.0733	0.9824
	Proposed Method	**196.2141**	108.1958	0.0505	0.9876
Fortaleza	ARIMA	307.9292	194.8115	0.0863	0.9766
	SVR	176.1232	95.2270	0.0335	0.9918
	MLP	181.5854	100.7907	0.0376	0.9910
	ELM	192.4636	117.6939	0.0420	0.9899
	RF	183.9838	96.1792	0.0403	0.9906
	GB	187.9344	100.5815	0.0419	0.9902
	ARIMA-ANN	187.6393	111.6506	0.0389	0.9905
	ETS-ANN	175.6055	96.7878	0.0321	0.9920
	KC-MLP	199.4051	116.6023	0.0376	0.9900
	OTL-ARIMAX-PSO-MC	244.5130	142.5150	0.0693	0.9834
	Proposed Method	**168.5435**	**86.6783**	**0.0301**	**0.9925**
Salvador	ARIMA	427.0798	289.5371	0.1399	0.9663
	SVR	317.0598	153.4281	0.0726	0.9821
	MLP	319.0990	164.3702	0.0775	0.9814
	ELM	322.1988	180.2548	0.0781	0.9812
	RF	331.4779	163.8206	0.0915	0.9791
	GB	338.7720	165.4474	0.0932	0.9784
	ARIMA-ANN	345.7941	203.7673	0.0925	0.9780
	ETS-ANN	322.1156	179.067	0.0779	0.9812
	KC-MLP	322.6953	168.2775	0.0805	0.9808
	OTL-ARIMAX-PSO-MC	389.1639	234.5723	0.1261	0.9710
	Proposed Method	**291.6498**	**146.7355**	**0.0599**	**0.9851**
São Luís	ARIMA	348.4662	269.1731	0.1151	0.9722
	SVR	282.1142	157.1611	0.0804	0.9807
	MLP	283.7657	165.6519	0.0836	0.9802
	ELM	288.5480	181.7681	0.0879	0.9793
	RF	290.2078	155.6084	0.0896	0.9790
	GB	290.5587	158.5654	0.0904	0.9789
	ARIMA-ANN	286.6585	180.2979	0.0810	0.9810
	ETS-ANN	261.0185	155.7109	**0.0579**	**0.9854**
	KC-MLP	275.9952	169.3605	0.0658	0.9834
	OTL-ARIMAX-PSO-MC	338.6826	246.5762	0.1142	0.9732
	Proposed Method	**260.6539**	**152.5425**	0.05993	0.9851
João Pessoa	ARIMA	268.0784	177.5489	0.0553	0.9864
	SVR	242.7867	134.4155	0.0453	0.9889
	MLP	251.6922	143.0017	0.0504	0.9878
	ELM	254.2948	156.5881	0.0510	0.9876
	RF	254.9432	138.0919	0.0532	0.9873
	GB	255.9489	139.3224	0.0539	0.9872
	ARIMA-ANN	267.5165	174.2956	0.0546	0.9865
	ETS-ANN	257.2287	155.8420	0.0502	0.9876
	KC-MLP	243.6437	133.9001	0.0469	0.9886
	OTL-ARIMAX-PSO-MC	273.3799	165.8006	0.0606	0.9854
	Proposed Method	**239.3044**	**129.7313**	**0.0436**	**0.9892**

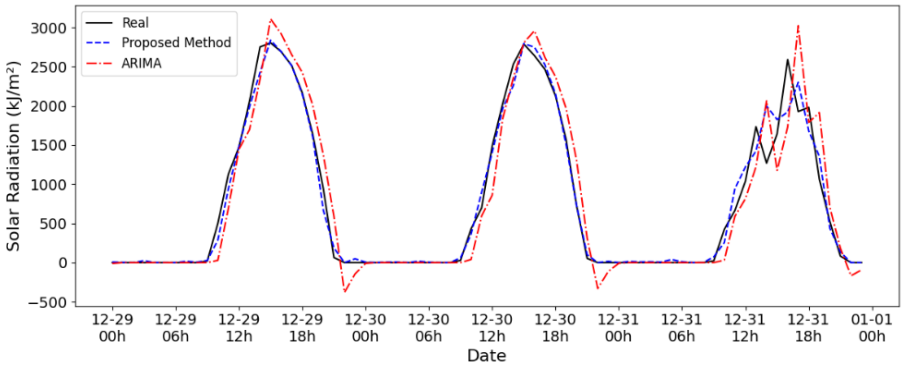


Figure 5. Forecast vs. actual solar radiation values — Fortaleza.

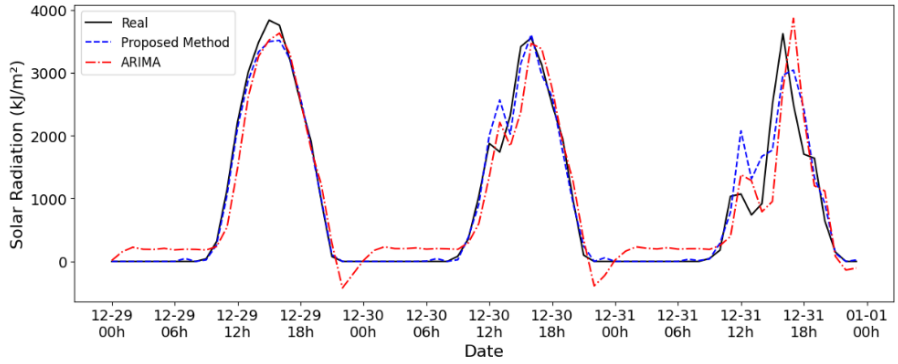


Figure 6. Forecast vs. actual solar radiation values — Salvador.

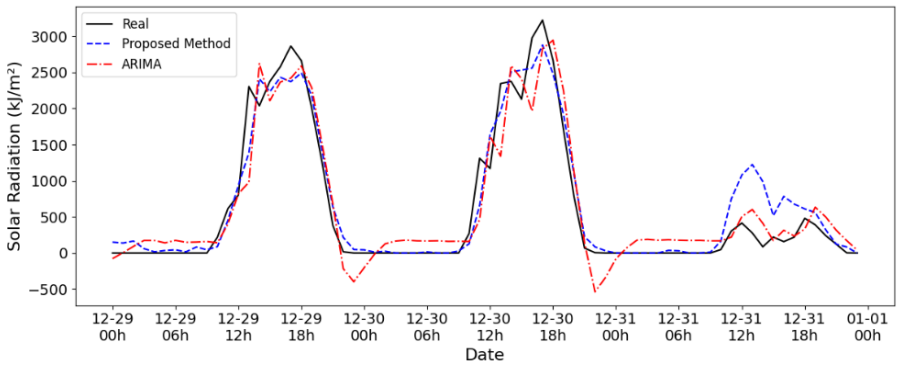


Figure 7. Forecast vs. actual solar radiation values — São Luís.

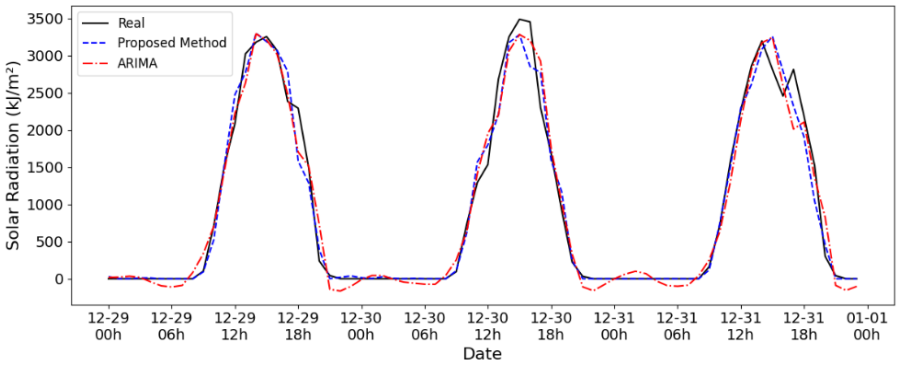


Figure 8. Forecast vs. actual solar radiation values — João Pessoa.

The selected lag subsets varied across runs and cities due to the stochastic nature of GA and SVR hyperparameter optimization. However, certain lags were repeatedly selected due to their temporal characteristics, such as short-term lags (1-4h) and those related to the daily cycle (23 and 24h). The repeated selection of these lags across runs and cities indicates that they are important inputs for accurate solar radiation forecasting.

5. Conclusion

In this work, the multivariate ARIMA-GA-SVR model was developed to forecast hourly solar radiation. This model, which combines ARIMA's ability to capture linear patterns with SVR's potential to learn complex nonlinear relationships from lagged features of residuals and exogenous variable. The initial feature structure, prior to the lag selection step, consisted of a 24-hour lag set that included past residuals, observations, and linear forecasts. A Genetic Algorithm (GA) was employed to identify the optimal subset of lags. In addition, this feature structure proved to be effective in representing the temporal dynamics and the relationship between meteorological data and solar radiation prediction. Furthermore, the lag selection process efficiently identified the most relevant past values contributing to the one-step-ahead forecast.

The experimental evaluation was conducted in five Brazilian cities. The proposed model achieved the best overall performance across most cities in terms of the evaluated metrics (RMSE, MAE, ARV, and IA), with the best results in Salvador, Fortaleza and João Pessoa, where it achieved the best results across all metrics. The results of the Friedman-Nemenyi statistical test showed that the proposed method achieved the best average ranking.

Regarding future work, the analysis of different forecasting horizons could be explored to examine how the model's performance changes with increasing time horizons. To evaluate the model generalization, the proposed method could be applied to other cities across different Brazilian regions with distinct climatic conditions.

Furthermore, other feature selection techniques, such as embedded methods or ensemble-based approaches, could be employed to further improve model performance. Alternative machine learning models, such as MLP (Multilayer Perceptron), could be investigated for the nonlinear modeling stage, along with ensemble strategies that combine multiple nonlinear learners to further enhance predictive performance.

References

Aguiar, L. M. et al. (2016). Combining solar irradiance measurements, satellite-derived data and a numerical weather prediction model to improve intra-day solar forecasting. *Renewable Energy*, 97:599–610.

Ahmadi, M., Aly, H., and Khashei, M. (2025). Enhancing power grid stability with a hybrid framework for wind power forecasting: Integrating kalman filtering, deep residual learning, and bidirectional lstm. *Energy*, 334:137752.

Dong, Z., Yang, D., Reindl, T., and Walsh, W. M. (2015). A novel hybrid approach based on self-organizing maps, support vector regression and particle swarm optimization to forecast solar irradiance. *Energy*, 82:570–577.

Ghofrani, M., de Rezende, M., Azimi, R., and Ghayekhloo, M. (2016). K-means clustering with a new initialization approach for wind power forecasting. In *IEEE Power and Energy Society General Meeting*.

Guermoui, M., Melgani, F., Gairaa, K., and Mekhalfi, M. L. (2020). A comprehensive review of hybrid models for solar radiation forecasting. *Journal of Cleaner Production*, 258:120357.

Hyndman, R. J. and Khandakar, Y. (2008). Automatic time series forecasting: The forecast package for r. *Journal of Statistical Software*, 27:1–22.

Krishnan, N., Kumar, K. R., and Inda, C. S. (2023). How solar radiation forecasting impacts the utilization of solar energy: A critical review. *Journal of Cleaner Production*, 388:135860.

Kumari, P. and Toshniwal, D. (2021). Deep learning models for solar irradiance forecasting: A comprehensive review. *Journal of Cleaner Production*, 318:128566.

Oliveira, J. F. L., Silva, E. G., and de Mattos Neto, P. S. G. (2022). A hybrid system based on dynamic selection for time series forecasting. *IEEE Transactions on Neural Networks and Learning Systems*, 33(8):3251–3263.

Panigrahi, R. and Behera, A. K. (2017). Exponential smoothing based ann model for time series forecasting. *Engineering Applications of Artificial Intelligence*, 66:49–63.

Santos, S. et al. (2022). Solar irradiance forecasting using dynamic ensemble selection. *Applied Sciences*, 12(7):3510.

Solano, E. S., Dehghanian, P., and Affonso, C. M. (2022). Solar radiation forecasting using machine learning and ensemble feature selection. *Energies*, 15(19):7049.

Wang, Z., Peng, T., Zhang, C., et al. (2024). Enhancing multi-step short-term solar radiation forecasting using hybrid deep learning models. *Applied Energy*, page 124708.

Zhang, G. P. (2003). Time series forecasting using a hybrid arima and neural network model. *Neurocomputing*, 50:159–175.

Zhang, H., Liu, X., Gao, X., and Liu, Z. (2025). Hybrid arimax model integrating optimal time-lag selection and markov chain error correction for residential heating load forecasting. *Energy Reports*, 14:5259–5273.