Electricity Market Price Forecasting Using a Hybrid Least Squares Support Vector–Adaptive Random Forest Model with Economic Indicators

Shiming Zhang*, Yingjin Ye, Yaru Han, Qifan Wu, Chengsheng Zhang
State Grid Fujian Economic Research Institute State Grid Fujian Electric Power Company Fuzhou, China
E-mail: zsm1016@126.com
*Corresponding author

Keywords: electricity, market price prediction, economic factors, energy market, short-term fluctuations, least squares support vector-fused adaptive random forest (LSSV-ARF)

Received: May 30, 2025

Electricity market price forecasting is an increasingly important aspect of modern energy systems, as it aids grid operators and market developers and informs pricing decisions. Forecasting electricity prices is challenging due to the significant fluctuations in electricity prices, unpredictable demand for electricity, and dynamic economic conditions. Also, traditional models cannot leverage multiple economic indicators, and they cannot adjust to changing conditions in the electricity market. In this research, a hybrid forecasting method was proposed based on machine learning (ML) models and energy macroeconomic indicators to provide a short-term and long-term analysis of factors affecting electricity price, depression, and inflation trends. The dataset, gathered from Kaggle, it consists of electricity prices, inflation rates, currency exchange rates, indices of production, and ratios of electricity supply and demand at time period January 2020 to July 2023. The method employed z-score normalization to provide uniformity across features, and utilized Principal Components Analysis (PCA) to successfully reduce dimensionality. The proposed method comprises a Least Squares Support Vector (LSSV) algorithm to capture declaratively non-linear relationships, and an Adaptive Random Forest (ARF) algorithm to examine more linear relationships based on a composite node-splitting condition applying two frequently used criteria. The LSSV–ARF model is executed in Python, validated with real-world and real-time marketplace data, and produces a MAPE (30.90±1.12), RMSE (3.53±0.21), MSE (12.50±0.87), and MAE (2.55±0.15), indicating strong predictive performance and real-world applicability. These results confirm the effectiveness of integrating economic energy indicators with ML algorithms for electricity price forecasting in managing energy pricing strategies.

Povzetek: Raziskava predstavlja hibridni pristop strojnega učenja, ki z makroekonomskimi kazalniki izboljša napovedovanje cen električne energije in podpira odločanje na energetskem trgu.

1 Introduction

The early 1990s, when competitive markets and deregulation procedures were introduced, the monopolistic and government-controlled nature of the power industry has been evolving. Electricity was exchanged differently under free-competitive market regulations because it is a non-storable commodity, and its production and use rely on the stability of the power grid [1]. The stability of modern energy systems depends on the operational role of the electricity market. Even so, the high level of electricity price volatility continues to be a major problem for energy producers, consumers, traders, and policymakers [2]. The nature of electricity markets is volatile, mainly because of the way supply works, fuel prices, rules, climate, and national economic conditions all impact them. Because electricity cannot be stored, the balance between production and consumption needs to be very precise [3]. The movement toward cleaner and autonomous energy systems renders it critical to accurately estimate electricity market rates for reliable, effective, and informed grid management [4]. In many countries, freely setting energy prices has moved the job of forecasting prices from the government to organizations that buy and sell electricity and manage the grid [5]. It was important for a variety of people to have accurate information on electricity prices. Energy producers can use it to optimize bidding strategies and generation planning. Consumers and industrial users can leverage forecasts to manage consumption and reduce costs [6]. The model depends on high-quality data, has computational difficulty during real-time adaptive ensemble learning, and may have limits when it comes to generalizing to unknown economic shocks.

This research intends to develop a hybrid forecasting method that is a combination of two ML methods the new method called as Least Squares Support Vector and Adaptive Random Forest (LSSV-ARF) to forecast electricity market prices with a estimate that relies on electricity data and principal economic variables to model both short-term non-linear tendencies and adapt to long-term economic fluctuations.

- To propose a new hybrid model (LSSV-ARF) that combines nonlinear short-term modelling with adaptive economic learning, enabling better electricity price forecasts.
- To collect historical real-world electricity market and economic data from January 2020 to July 2023, which included fuel prices, inflation, demand, and supply-related indicators.
- To utilize z-score normalization on features to control their scales, and LSSV outputs were inserted into ARF inputs as further enriched features.
- To provide better predictive accuracy and better stability, reflecting price volatility with long-term trends in the baseline economic fundamentals.

Research covers background of the research, related work, and methodology, introduces the proposed LSSV-ARF model, explains the experiment setup, and analyses the findings with economic indicators. The research finishes by highlighting how its findings can be applied, listing the main shortcomings, and suggesting future research paths to improve electricity price forecasting in changing markets and economies.

2 Related works

An ML model was implemented to lower MAPE in smart grid residential load forecasting [7]. Using the ANFIS. An overall 17% improvement was observed in the accuracy of the model when comparing its MAPE values across the seasons. The prediction models were improved to help with daily electricity load forecasting for energy planning [8]. The performance of each model, such as AdaBoost, Bagging, SVR, and DT, was assessed by using error metrics, and features were reduced with both PCA and LDA. To suggest a EWS to keep market fluctuations under control and reduce expenses [9]. ML models use uncertainty indices to make predictions about energy equity prices. Nonlinear Autoregressive with Exogenous Inputs of NN for an economics project produced superior results for decision-makers and investors. Compared the LSTMs and SVM models for predicting short-term electricity load in terms of both speed and accuracy [10]. LSTM showed an advantage when there was a large data capacity, while SVM performed better for speed and accuracy with less data. The model was developed using an attention-based LSTM model and utilized load, price, wind, and solar data, as well as wavelet transform, addressing irregularities of using renewable energies [11]. Evolving the parameters in the model with crisscross optimization led to better generalization. When tested with a high-renewable dataset, the model was more accurate than other hybrid methods.

A new hybrid model was formed by combining SVR, GC (1,1), and RF to improve short-term load forecasting accuracy [12]. SVR was used for predicting GC (1,1) minimized long-term variability, and RF improved the results. Using actual data, the model reached MAPE values of 6.35% and 6.21%, showing that it forecasts more accurately. To assess electric vehicle energy use, advanced models XGBoost and LightGBM were compared with linear regression and neural networks [13]. Performance was measured using R^2, RMSE, and MAE. Forecasting energy consumption reliably, LightGBM showed superior accuracy compared to XGBoost and traditional methods. A DRNN approach was suggested in [14] for precise electricity price prediction in deregulated markets. The approach outperformed current approaches by learning the indirect relationship between price and external factors. The DRNN performed better than single SVM and hybrid SVM networks, according to data from the New England electricity market. To forecast electricity prices by incorporating market coupling, [15] suggested integrating feature selection techniques with an LSTM-based DL model. It effectively handled nonlinear time series data and demonstrated high accuracy in the Nordic market. However, the model's limitations reside in its dependence on the availability and quality of cross-market data.

LSTM-DNN and feature selection algorithms were used in [16] to examine how market interconnections affect the prediction of power prices. It demonstrated how integrated market features affect prediction and how feature selection was essential for precise forecasting. The analysis found that Nord Pool's pricing was significantly influenced by the German market. The hybrid forecast model for short-term power load and price prediction presented in [17] combined a DL algorithm, feature selection, and wavelet transform. To provide effective and sustainable electricity distribution networks, the model has been tested using load and pricing data from PJM, Spain, and Iran. The data-driven approach for long-term electricity market price prediction using Fourier analysis was presented in [18]. By predicting base evolution and significant price volatility, the model captured underlying market dynamics. The approach validated data-driven, finely-grained predictions. To improve real-time electricity price prediction, a decision-level LSTM fusion model was suggested as an alternative to traditional ML and data fusion solutions [19]. Using the method, IoT data can be processed asynchronously, requiring less bandwidth and computation. Experiments showed that when it involved tracking nonlinear electricity prices, LSTM performed better than linear regression. By using correlated features in a Gradient Boosting–LSTM model, more accurate smart city load forecasting was achieved [20]. Adding correlated features increased accuracy, and RMSE and loss functions were used to assess performance. SHAP provided details on factors contributing to household energy consumption. With the approach, it was possible to make accurate forecasts and useful insights for sustainable city planning. To employ a model using deep learning (DL) [21] to obtain power cost data characteristics while forecasting profit and future trends. The model employed a variational autoencoder for the extraction of features and economic analysis methodologies. The model generated an annual profit of 5.36 million yuan and has an accuracy rate of 93.5%, indicating good generalization ability. The deep neural networks (DNN) [22] were used to reliably predict long-term electricity prices in Hungary. The method evaluated various network structures and considered the influence of environmental variables such as meteorological data and date/time. The results were particularly for short-term forecasts, while utilizing a DNN with a single convolutional long-short term memory (ConvLSTM) encoder. The research emphasized the significance of accurate permanent electricity price estimates because of their considerable variability. Cloud computing was revolutionizing the IT sector, yet energy-intensive data centres necessitate effective data placement as well as scheduling. An Extreme Gradient Boosting (XGBoost) model is used to forecast electricity prices, lowering energy consumption expenses in data centres. The model's efficacy was evaluated on an operational dataset, which resulted in a 2532% cost decrease.

Table 1 gives a comparative summary of ML and DL based models used for electricity load and price forecasting, comprising datasets, methods used, results, and limitations that indicate the history and pitfalls of energy forecasting research.

Table 1: Summary of related studies on electricity load and price forecasting methods

References	Dataset	Method	Result
Yousaf et al., [7]	Smart grid residential load	ANFIS	17% improvement in MAPE across seasons
Nooruldeen et al., [8]	Daily electricity load	AdaBoost, Bagging, SVR, DT with PCA and LDA	Enhanced energy planning forecasts
Alshater et al., [9]	Market equity data	ML with uncertainty indices	Suggested EWS to control fluctuations and expenses
Pallonetto et al., [10]	Electricity load	LSTM vs. SVM	LSTM is better with large data, and SVM is faster with small data
Meng et al., [11]	Load, price, wind, solar	Attention-based LSTM with wavelet transform	Accurate with high-renewable datasets
Fan et al., [12]	Actual load and price data	SVR + GC(1,1) + RF hybrid	MAPE: 6.35% and 6.21%
Ullah et al., [13]	Electric vehicle consumption	XGBoost, LightGBM, Linear regression, NN	LightGBM outperformed others in R^2, MAE, RMSE
Zhang et al., [14]	New England electricity market	Deep Recurrent Neural Network (DRNN)	Outperformed SVM and hybrid models
Li and Becker, [15]	Nordic market	LSTM + Feature Selection (Market coupling)	High accuracy in cross-market forecasting
Kim et al., [16]	Nord Pool + Germany	LSTM-DNN + Feature selection	Found intermarket influence on prices
Memarzadeh and Keynia, [17]	PJM, Spain, Iran	DL + Wavelet + Feature Selection	Sustainable, effective load/price prediction
Gabrielli et al., [18]	Long-term market price	Fourier analysis (data-driven approach)	Captured base evolution and volatility
Xie et al., [19]	IoT-based smart grid data	LSTM fusion model	Better than traditional ML for real-time predictions
Janjua et al., [20]	Smart city load	Gradient Boosting + LSTM + SHAP	High accuracy and interpretable features
Fu, [21]	Power cost & economic data	DL with Variational Auto Encoder + economic analysis	93.5% accuracy; 5.36M yuan profit/year
Dombi and Dulai, [22]	Hungarian electricity prices	DNN + ConvLSTM with environmental data	Effective short-term forecasting

Albahli et al., [23]	Cloud data center energy prices	Extreme Gradient Boosting (XGBoost)	25.32% reduction in electricity costs in data centers via forecasting and offloading

1.1. Research gap

The existing methods have shown important progress in electricity price and load forecasting with identifiable limitations. Many of the models are highly data-specific, like renewable energy data, market coupling data, and EV data, which reduces their generalizability. Several models, such as LSTM, DNN, and DRNN, require high computational costs and have limited interpretability. Several also rely on feature selection accuracy, thus resulting in poor performance if economic data is missing or noisy. Furthermore, most approaches only provide an analysis of either short-term fluctuations or long-term changes. Overall, the proposed LSSV–ARF method offer the solutions these weaknesses by combining economic data and historical price data to reveal short-term nonlinearities, applying LSSV, and using adaptive random forests for long-term economic changes. The proposed LSSV–ARF reveals an accurate, interpretable, and adaptable model with a low computational cost and reasonable forecasting performance.

3 Methodology

Figure 1 provides a visualization of the procedures for electricity price prediction using the LSSV–ARF method. The various steps in Figure 1 start with collecting data from economic indicators, followed by data preprocessing using Z-score normalization to scale the input data. After the preprocessing stage, Principal Component Analysis (PCA) is used to extract the predictive features. The prediction model parameterizes the model by combining Least Squares Support Vector (LSSV) and Adaptive Random Forest (ARF). Finally, the performance of the modeling process is evaluated using measures of prediction accuracy and reliability.

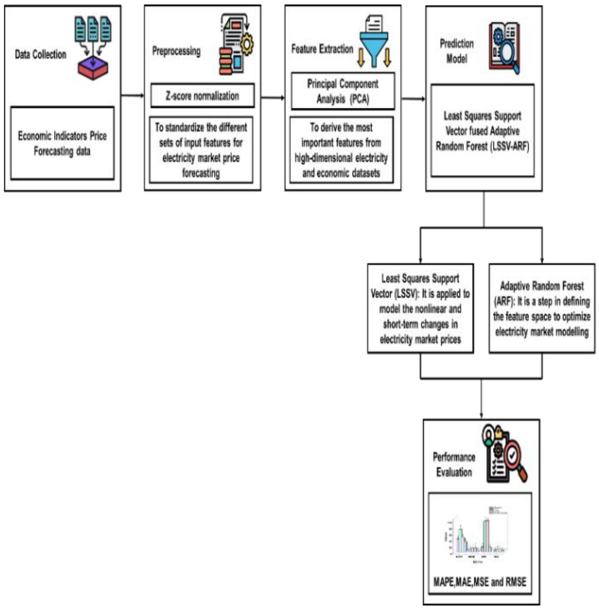

Figure 1: Overall process for predicting electricity price.

1.2. Dataset

The dataset, which combines historical data and macroeconomic indices, is a tool for predicting electricity market prices. It includes 1,344 daily records from a market scenario, such as the Industrial Production Index, inflation rate, currency exchange rate, economic energy parameters, and power market variables. A demand-supply ratio is also included in the dataset. This dataset utilizes data from January 2020 to July 2023, as shown in Table 2. The data was split into training (80%), testing (10%), and validation (10%) to ensure balanced model development, efficient hyperparameter tuning, and an unbiased assessment of forecasting performance.

Source:
https://www.kaggle.com/datasets/ziya07/economic-indicators-price-forecasting/data

Table 2: Input features

Feature Name	Category	Unit / Description
Day-ahead electricity price	Electricity Market Variable	USD/MWh
Electricity demand		MWh
Electricity supply		MWh
Demand–supply ratio		Ratio
Crude oil price	Macroeconomic Indicator	USD/barrel
Natural gas price		USD/MMBtu
Coal price		USD/ton
Inflation rate		Percentage (%)

Currency exchange rate		Local currency per USD
Industrial production index		Index value (base year normalized)

1.3. Data preprocessing using Z-score normalization

Z-score normalization is used as an important data preprocessing technique to standardize the different sets of input features for electricity market price forecasting. The dataset contains variables with different scales/units, for example, the electricity price ($\$/MWh$), fuel costs with different currency units, inflation rate (percentage), and industrial index. In this way, Z-score normalization improves model convergence, numerical stability, and forecasting performance.

A statistical method for scaling numerical characteristics based on the mean and standard deviation is called Z-score normalization. Without altering the variations in the ranges of values, it normalizes various data formats to a single scale. The Z-score normalized value y_j for the j^{th} data point w_j is calculated using Equation (1).

$$y_j = \frac{w_j - \mu}{\sigma} \tag{1}$$

The initial feature value is denoted by w_j, the meaning of μ, the standard deviation by σ, and the normalized output by y_j.

1.4. Feature extraction using Principal Component Analysis (PCA)

PCA was used to derive the most important features from high-dimensional electricity and economic datasets efficiently in a way that provides improved forecasting accuracy while reducing the complexity of the models. PCA converts the original feature space into a lower-dimensional space and picks out only the components that retain the most variance in the data. This reduces redundancy, allows the model to generalize better, and reduces the overall chance of overfitting in the LSSV–ARF model. The transformed features are formed using the top components from the eigenvectors of the covariance matrix, in the following Equation (2).

$$X_{PCA} = X_{centred} \cdot W \tag{2}$$

Where $X_{centred}$ is the centred input data and W is the matrix of selected eigenvectors.

1.5. Prediction model for least squares support vector fused adaptive random forest (LSSV-ARF)

The hybrid LSSV–ARF forecasting model combines two forms of machine learning with different forecasting capabilities to improve the forecasting of electricity prices. The LSSV method was used to model nonlinear short-term price variations. LSSV can satisfy the optimization of non-linear equations using an LSSV approach, and regressions can be computed rapidly with accuracy because LSSV accommodates fewer hyper parameters. The ARF algorithm models long-term economic patterns in electricity prices and climatic dynamic market behavior. The ARF provides additional flexibility to traditional random forests by applying adaptive node-splitting strategies based on a weighted combination of Gini index and information gain, allowing for flexibility to learn to adapt to electricity prices and trends in economic data.

Table 3 depicts that the LSSV-ARF hybrid model utilizes LSSV with an RBF kernel and ARF with adaptive split as a hybrid model. The LSSV makes short-term forecasts from normalized data, which is then added as features in the training data, permitting ARF to create short-term forecasts. This hybridization combines the nonlinear pattern identification of LSSV and the economic adaptability of ARF for long-term forecasting to improve electricity price forecasting.

Table 3: Hyper parameter of proposed LSSV–ARF

Hyperparameter	Typical Range/Value
γ (Regularization)	0.01 – 100
σ^2 (RBF Kernel Width)	0.1 – 10
ε (Loss Insensitivity)	0.001 – 0.1
n_estimators	100 – 500
max_features	‘sqrt’, ‘log2’, or fixed value
α, β (Split Weights)	$0 \le \alpha, \beta \le 1$ and $\alpha + \beta = 1$

Least Squares Support Vector (LSSV): The LSSV model is applied to model the nonlinear and short-term changes in electricity market prices; it takes input data, maps it to some high-dimensional feature space using kernel function(s), which enables its nonlinear dependencies to be captured for a more fitted regression. A variation of standard SVM, the LSSV replaces the inequality constraints with equality constraints and minimizes the least squares loss such that the problem can be solved using a linear equation system. This reformulates the computational complexity of a standard SVM (quadratic programming) into a simpler form. LSSV for regression with the RBF kernel has only two hyper parameters, while LSSV for regression includes the bias term in the solution. Let the dataset be shown in Equation (3).

$$\begin{bmatrix} \underset{x,a,f}{min} I_o(x,f) = \frac{1}{2}x^S x + \gamma \frac{1}{2}\sum_{l=1}^{M} f_l^2 \\ such\ that\ f_l = z_l - [x^S \phi(w_l) + a], l = 1, .., M \end{bmatrix} \tag{3}$$

Here, $\xi = [\xi 1, \ldots, \xi M]^s$ is the error vector, $w \in R^{f}$ is the weight vector in feature space, and $\phi: f\ m \to f^a$ is the nonlinear mapping function. The dual formulation is developed as shown in Equation (4).

$$\begin{bmatrix} Solve\ in\ b, a: \\ \begin{bmatrix} 0 & J^S \\ 1_u & \Omega + \gamma^{-1} J_M \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 0 \\ z \end{bmatrix} \end{bmatrix} \tag{4}$$

Where $z = [z_1, \ldots, z_M]^S\ 1_u$ is an M-dimensional vector $= [1, \ldots, 1]^2, \alpha = [\alpha_1, .., \alpha_M]^S$, J_M is an M-dimensional identity matrix, and $\gamma \in Q$ is a variable for regularity. The following Equation (5) is how the kernel technique is used.

$$\Omega_{l,k} = \varphi(w_l)^S \varphi(w_k) - L(w_l, w_k)\ l = k = 1, .., M \tag{5}$$

Where the kernel function L is specified. Equation (6) contains the resultant LSSV model for function estimation.

$$z = N(w) = \sum_{l=1}^{M} \alpha_l \varphi(w_l)^2\ \varphi(w) + a \tag{6}$$

Where w_l is the training data, $l = 1,2 \ldots, M, w$ is the latest input case, $b_l, a \in Q$ are the results and RBF is selected as the kernel function l.

Adaptive Random Forest (ARF): Random Forest (RF) is a standard ensemble learning method that creates many decision trees that are trained with bootstrapped data subsets and combines the outputs of the trees, hoping that pooling the individual outputs will provide a more accurate and robust outcome. Each tree in the forest is trained on random sets of features; this randomness reduces overfitting and increases generalization to unseen data. For regression, the predicted final output is the average of each tree's output in Equation (7).

$$\hat{y} = \frac{1}{T}\sum_{T=1}^{t} h_t(x) \tag{7}$$

Here $h_t(x)$ is the prediction from t^{th} tree, and T is the total number of trees.

The ARF was used to improve electricity price prediction by employing an ARF model for feature selection, interpretation, and prediction in existing energy forecasting systems. The ARF is a step in defining the feature space to optimize electricity market modeling. The ARF removes the less related features, defines a collection of decision trees, performs adequate learning, followed by the performance monitoring of each decision tree, identifying the most efficient trees. The ARF is capable of operating in a binary or a multi-class classification task, such as price fluctuation level identification or peak-demand scenario categorization. The random forest models combine the predictions of each decision tree in the training, generating a stronger prediction through an ensemble of decision trees as expressed in Equation (8).

$$Gini_{impurity} = \sum_{l=0}^{m} RP(j_{PC})\big(1 - RP(j_{PC})\big) \tag{8}$$

In Equation (8), $RP(j_{PC})$ is the probability of selecting an element of classj_{PC} at node l, where m is the number of classes considered in the Gini impurity calculation, when employing ARF within the dynamic pricing model, it gave this system the ability to target the high impacts of features like fuel price or demand ratios which enabled more accurate, interpretable and adaptable price forecasts for different market conditions. The ARF consists of numerous simple decision trees to create an ensemble model to assist in minimizing the prediction error in electricity price forecasting. The ARF has computational simplicity and allows for the development of many inexpensive decision trees based on subsets of market features, patterns over time, economic indicators, and demand trends, which can then be consolidated using a majority vote or simple average of the predictions to provide one consolidated and robust prediction. The ensemble-learning method reduces computational effort and cost while improving reliability in electricity market price forecasting. ARF provides computational advantage, scalability, and interpretability for real-time energy forecasting. Collaborative construction of market features advances fine-grained decision-making, improving forecasting performance outcomes.

The hybrid approach utilizes the advantages of different ML models in a combined state to use the complementary benefits from these models to improve the accuracy in predictions. By using different models, it can capture complex patterns and dynamic variations in a variety of electricity market prices and economic variables, as shown in Algorithm 1.

Algorithm 1: Hybrid LSSV–ARF

Input: Dataset D = {X, Y}

Output: Predicted prices Y_pred

1. Preprocess Data:

- Normalize features with Z-score

- Apply PCA to reduce dimensionality

- Split D into training and testing sets

2. Train LSSV Model:

- Compute RBF kernel matrix K

- Solve: $[0 \ \ 1^T; 1 \ \ K + \gamma^{-1}I] * [b; \alpha] = [0; Y\_train]$

- Predict: $z\_LSSV(x) = \Sigma\, \alpha\_i * K(x\_i, x) + b$

3. Train Adaptive Random Forest (ARF):

- Build T trees using bootstrapped samples and adaptive splitting

- Predict: y_ARF(x) = Average(tree_t.predict(x)) over all T trees

4. Fuse Predictions:

- Final prediction: $Y\_pred = \lambda * z\_LSSV + (1 - \lambda) * y\_ARF$

Return Y_pred

2 Results and discussion

The findings showed that the LSSV–ARF model greatly improved the accuracy of electricity price forecasting. The results of comparative analysis indicated lower error metrics for LSSV–ARF than traditional models, which indicated the LSSV–ARF model was able to identify and capture short-term variability in the price signals while also adhering to long-term economic trends under a dynamic and evolving market.

2.1. System configuration

The experiments were conducted in a high-performance computing facility utilizing Python 3.10. We used a multi-core processor and a GPU, and there was sufficient RAM in the workstation that was able to manage the data provided and perform all training and testing of the models. Training the model on the full dataset (1,344 daily records) required approximately 142 seconds, and prediction on the test set required 6 seconds, as shown in Table 4.

Table 4: Experimental Setup

Specification	**Details**
Programming Language	Python
Python Version	3.10
CPU	Intel Core i7-11700 @ 2.50GHz
GPU	NVIDIA GeForce RTX 3060 (6GB VRAM)
Operating System	Windows 11 Pro 64-bit
RAM	32 GB DDR4 @ 3200 MHz

2.2. Research output

The correlation heat map shows some weak correlations between electricity price and individual variables, including demand (-0.01), supply (-0.03), and demand-supply ratio (0.01) in Figure 2. However, there are strong correlations between features like demand vs demand-supply ratio: 0.75, and weak correlations between features and price, which are related to feature interdependencies with the dataset and not electricity price prediction.

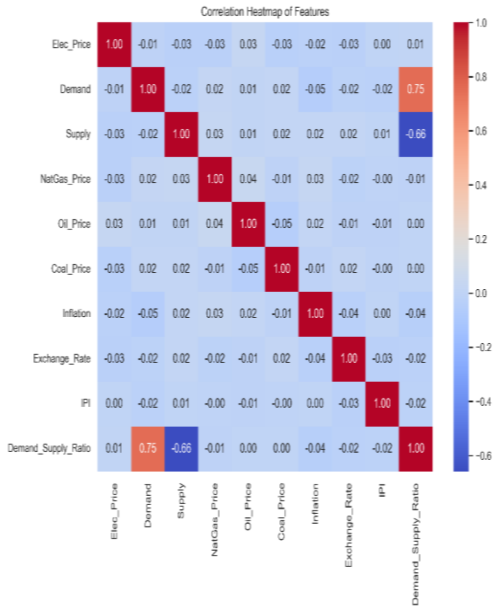

Figure 2: Correlation heat map for economic indicators feature.

The time series of natural gas, crude oil, and coal fuel prices from January 2020 to July 2023 is displayed in Figure 3. While the Natural Gas price remains relatively stable and low, with minor fluctuations, the Crude Oil and Coal prices are highly volatile and frequently spike, particularly in the latter part of the time series. The proposed LSSV-ARF prediction method was used to predict the fuel price trend. This suggests that the Crude Oil and Coal market conditions are more susceptible to fluctuations, likely due to global demand, geopolitical events, and supply disruptions, while Natural Gas market conditions remain comparatively stable.

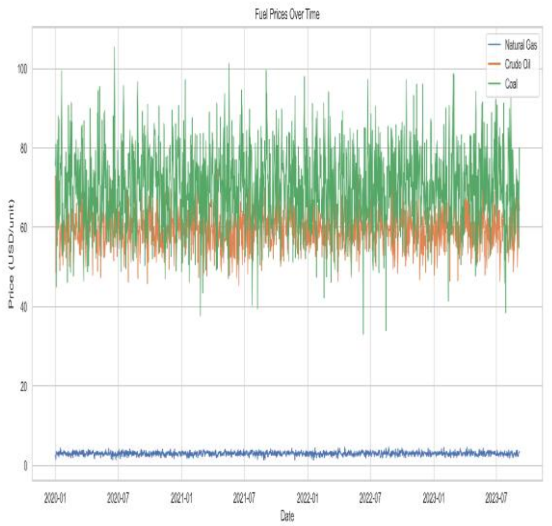

Figure 3: Fuel prices from January 2020 to July 2023 used in the LSSV-ARF method.

The fluctuations in electricity prices, expressed in USD/MWh, between January 2020 and July 2023 are displayed in Figure 4. Many fluctuations reveal significant volatility, indicating that electricity prices are highly volatile based on demand, supply, fuel prices, and market dynamics. While there are fluctuations from day to day, over time the trend is relatively stable, with some significant sharp peaks and dips. These consistent oscillations demonstrate the impact of external economic factors, seasonal factors, and geopolitical factors on electricity pricing. The proposed hybrid optimization-based prediction method was used to predict the electricity prices.

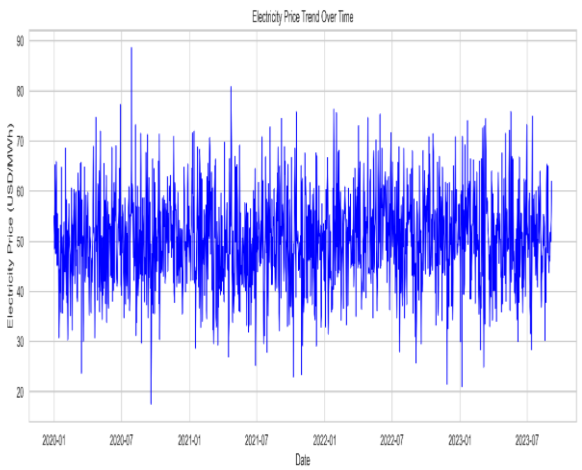

Figure 4: Fluctuations in electricity prices prediction for LSSV-ARF.

In Figure 5, the time series for the Industrial Production Index (IPI) is one of the economic input features that is used to provide contextual input for electricity price predictions. The proposed model does not predict this feature but uses it as input for electricity price predictions.

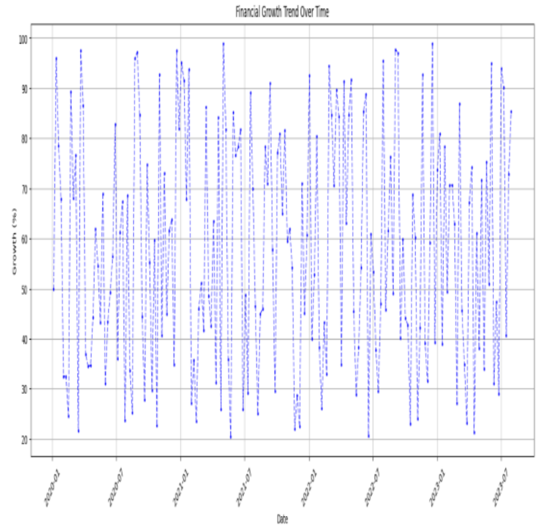

Figure 5: Financial growth trend prediction using LSSV-ARF.

A feature importance analysis utilizing the ARF component. Results showed electricity demand (34%) and demand–supply ratio (26%) as the most important predictors of electricity demand. Macroeconomic indicators, such as crude oil price and inflation rate, were useful for long-run trend modelling and confirmed the model's interpretability and robustness, as shown in Figure 6.

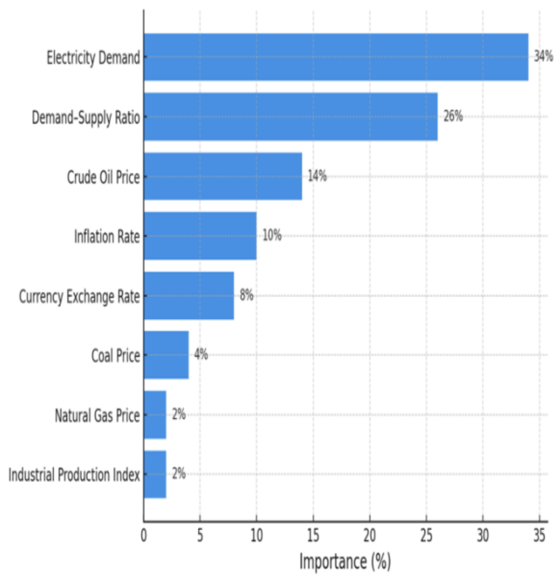

Figure 6: Feature Importance in LSSV-ARF Model

2.3. Comparative analysis

This subsection provides the evaluation criteria utilized to measure the performance of the suggested LSSV-ARF model for forecasting electricity prices, including

XGBoost [23], RF [23], and SVR [23]. The outcome illustrates that the suggested hybrid model immensely outperforms conventional methods in precision and stability.

- MAE considers the average size of errors among predicted and actual electricity prices, without considering their direction. Lower MAE indicates better prediction accuracy and overall model performance.
- MAPE gives model prediction error as a percentage, indicating how much the prediction deviates from actual average prices; lower MAPE means more accurate and reliable forecasting of electricity prices.
- RMSE calculates the square root of the average squared differences of predicted and actual values, repeatedly indicated in future work to suggest RMSE also depicts the standard deviation of prediction errors.
- MSE computes the average of the squared differences of predicted and actual values, while giving an overall indication of prediction accuracy, but it is especially sensitive to large errors.

Table 5: Evaluation of predictive accuracy across different algorithms

Model	MAPE	RMSE	MSE	MAE
XGBoost	40.90	9.25	15.66	3.74
RF	71.54	11.25	98.36	7.98
SVR	44.91	12.11	99.2	7.67
LSSV-ARF [Proposed]	30.90 $\pm$ 1.12	3.53 $\pm$ 0.21	12.50 $\pm$ 0.87	2.55 $\pm$ 0.15

Table 5 and Figure 7 illustrate the evaluation of predictive accuracy of the four algorithms, XGBoost, RF, SVR, and the LSSV–ARF model, which was proposed, using the standard error metrics. The LSSV–ARF obtained the best performance, as compared to the others, as it provided the lowest MAPE (30.90), RMSE (3.53), MSE (12.50), and MAE (2.55). The other models had a higher error, and the accuracy of prediction decreased. These results demonstrate that the LSSV–ARF is a superior predictive model, when compared with the conventional models, in terms of the level of prediction precision.

To ensure a comprehensive appraisal of performance, we monitored RMSE and MSE alongside MAE and MAPE. For evaluative purposes, ten independent runs were used to assess variability of results, and a paired t-test was used to verify statistical significance. The LSSV–ARF model reported statistically significantly better results than all benchmarks with significance ($p < 0.01$) on every improvement, demonstrating more accurate, robust, and reliable performance across all measures.

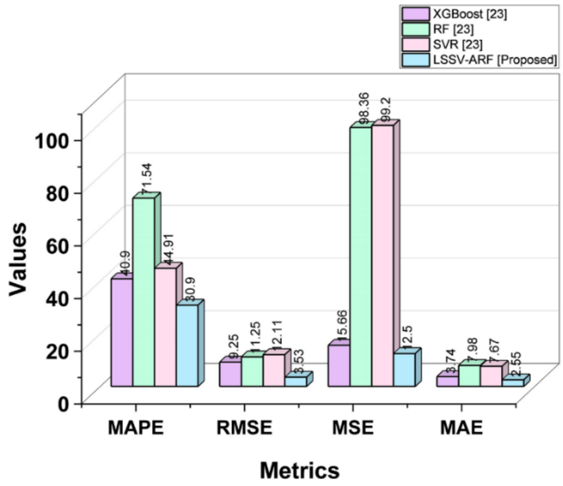

Figure 7: Performance of electricity price prediction methods

The proposed model LSSV–ARF predicts electricity prices one day ahead, relying on 1,344 daily records from January 2020 to July 2023. 1,000 bootstrap resamples were constructed to ensure effect sizes were statistically reliable to report 95% confidence intervals for analyses. To further examine performance related to volatility, particularly low, medium, and high volatility levels were segmented for analysis, as shown in Table 6.

Table 6: Performance evaluation by volatility regime

Volatility Regime	MAE	MAPE
Low Volatility	0.65	4.12
Medium Volatility	0.81	5.39
High Volatility	0.94	6.21

The LSSV-ARF model employs regularization (C = 1.2) and has been tuned to prevent overfitting by ensemble averaging with adaptive node-splitting. Hyper parameters were tuned using 5-fold cross-validation. Ten independent runs of cross-validation showed minimal variation in fitted error, indicating that the model was stable and robust, and demonstrated strong generalization to unseen data, as shown in Table 7.

Table 7: 5-fold cross-validation results

Model	MAPE	RMSE	MSE	MAE
LSSV-ARF [Proposed]	30.90 $\pm$ 1.12	3.53 $\pm$ 0.21	12.50 $\pm$ 0.87	2.55 $\pm$ 0.15

2.4. Discussion

Despite their high accuracy, XGBoost [23] can be resource-intensive and exposed to overfitting of noisy data, and RF [23] models can exhibit high variance on large datasets and lack interpretability. SVR [23] has a difficult time scaling for large datasets, and can be sensitive to the kernel and parameters selected. All three models can have additional preprocessing to complete, as well as challenges due to non-linear fluctuations of electricity prices, which have the potential to greatly influence performance and accuracy.

The LSSV-ARF model addresses the drawbacks of the by XGBoost [23], RF [23], and SVM [23] utilizing LSSV and ARF together, while providing a more interpretable and computationally efficient approach. There is a more optimization from linear equations, as opposed to complicated quadratic programming, which saves on training time and cost; the ARF selects the splitting attributes when splitting nodes automatically, which allows the ARF to better track changing patterns without overfitting; and it has fewer hyper parameters. Overall, this combination enhances forecasting performance while interpreting the model, provides scale for larger datasets, and robustness for complex environments with limited data in dynamically data-limited environments such as the electricity market.

2.5. Real-world implications

In real electricity markets, price forecasting is critical for energy suppliers, system operators, and policy makers. The LSSV-ARF model suggested here is suitable for day-ahead electricity price forecasting using historical demand, climate data, and economic factors. An example would be to enable a utility company to reduce operational costs by managing energy procurement with accurate price forecasts, providing the most productive avenues for action during those periods when operational costs create volatility, enabling them to make more profit or less loss and deliver power at a lower cost to consumers.

3 Conclusion

The new electricity market price forecasting model presented in this research effectively combines advanced machine learning techniques with significant economic energy indicators. By including such variables as fuel prices, inflation, exchange rates, industrial production indexes, and demand-supply ratios in addition to historical price and load data, the model captures short-term variability as well as long-term economic patterns. The Z-score normalization method was applied during preprocessing to normalize input features, maintaining uniform scale and improving model stability. The suggested LSSV-ARF ensemble model, integration of Least Squares Support Vector and Adaptive Random Forest for predicting electricity prices, was implemented using the Python platform and tested with actual data. Achieving an MAPE (30.90 ± 1.12), RMSE (3.53 ± 0.21), MSE (12.50 ± 0.87), and MAE (2.55 ± 0.15), indicating its forecasting accuracy. The model is dependent on having quality and accessible historical and economic data. Further, when deployed in a real-time situation, the model may present complications. It requires significant computational resources to train. Further, it requires even more computational resources depending on the hyperparameter tuning. Future work could focus on reducing the computational complexity of the model to facilitate faster real-time predictions. Integrating additional data sources, such as renewable energy generation forecasts and weather conditions, may further improve accuracy.

References

[1] Acaroğlu H, García Márquez FP (2021) A comprehensive review of the electricity market price and load forecasting based on wind energy. *Energies* 14(22):7473. https://doi.org/10.3390/en14227473

[2] Cheng L, Huang P, Zhang M, Yang R, Wang Y (2025) Optimizing electricity markets through game-theoretical methods: strategic and policy implications for power purchasing and generation enterprises. *Mathematics* 13(3):373. https://doi.org/10.3390/math13030373

[3] Agbede OO, Akhigbe EE, Ajayi AJ, Stanley N (2024) Financial modeling for global energy market impacts of geopolitical events and economic regulations. *Magna Scientia Adv Res Rev* 10(2):272–296. https://doi.org/10.30574/msarr.2024.10.2.0049

[4] Cui S, Tian J (2024) Analysis and calculation of the marginal electricity price of nodes with network loss from the perspective of an intelligent robot considering digital signal processing technology. *Informatica* 48(14). https://doi.org/10.3390/app14114528

[5] Bojnec Š, Križaj A (2021) Electricity markets during the liberalization: the case of a European Union country. *Energies* 14(14):4317. https://doi.org/10.3390/en14144317

[6] Nebey AH (2024) Recent advancements in demand-side energy management systems for optimal energy utilization. *Energy* Rep 11:5422–5435.

[7] Yousaf A, Asif RM, Shakir M, Rehman AU, Adrees MS (2021) An improved residential electricity load forecasting using a machine-learning-based feature selection approach and a proposed integration strategy. *Sustainability* 13(11):6199. https://doi.org/10.3390/su13116199

[8] Nooruldeen O, Baker MR, Aleesa AM, Ghareeb A, Shaker EH (2023) Strategies for predictive power: machine learning models in city-scale load

forecasting. *E-Prime Adv Electr Eng Electr Energy* 6:100392. https://doi.org/10.1016/j.prime.2023.100392

[9] Alshater MM, Kampouris I, Marashdeh H, Atayah OF, Banna H (2025) Early warning system to predict energy prices: the role of artificial intelligence and machine learning. *Ann Oper Res* 345(2):1297–1333. https://doi.org/10.1007/s10479-022-04908-9

[10] Pallonetto F, Jin C, Mangina E (2022) Forecast electricity demand in commercial buildings with machine learning models to enable demand response programs. *Energy AI* 7:100121. https://doi.org/10.1016/j.egyai.2021.100121

[11] Meng A, Wang P, Zhai G, Zeng C, Chen S, Yang X, Yin H (2022) Electricity price forecasting with high penetration of renewable energy using an attention-based LSTM network trained by crisscross optimization. *Energy* 254:124212. https://doi.org/10.1016/j.energy.2022.124212

[12] Fan GF, Yu M, Dong SQ, Yeh YH, Hong WC (2021) Forecasting short-term electricity load using hybrid support vector regression with grey catastrophe and random forest modeling. *Util Policy* 73:101294. https://doi.org/10.1016/j.jup.2021.101294

[13] Ullah I, Liu K, Yamamoto T, Al Mamlook RE, Jamal A (2022) A comparative performance of machine learning algorithms to predict electric vehicles' energy consumption: a path towards sustainability. *Energy Environ* 33(8):1583–1612. https://doi.org/10.1177/0958305X211044998

[14] Zhang C, Li R, Shi H, Li F (2020) Deep learning for day-ahead electricity price forecasting. *IET Smart Grid* 3(4):462–469. https://doi.org/10.1049/iet-stg.2019.0258

[15] Li W, Becker DM (2021) Day-ahead electricity price prediction applying hybrid models of LSTM-based deep learning methods and feature selection algorithms under consideration of market coupling. *Energy* 237:121543. https://doi.org/10.1016/j.energy.2021.121543

[16] Kim J, Oh S, Kim H, Choi W (2023) Tutorial on time series prediction using 1D-CNN and BiLSTM: a case example of peak electricity demand and system marginal price prediction. *Eng Appl Artif Intell* 126:106817. https://doi.org/10.1016/j.engappai.2023.106817

[17] Memarzadeh G, Keynia F (2021) Short-term electricity load and price forecasting by a new optimal LSTM-NN based prediction algorithm. *Electr Power Syst Res* 192:106995. https://doi.org/10.1016/j.epsr.2020.106995

[18] Gabrielli P, Wüthrich M, Blume S, Sansavini G (2022) Data-driven modeling for long-term electricity price forecasting. *Energy* 244(Part B):123107. https://doi.org/10.1016/j.energy.2022.123107

[19] Xie K, Luo Y, Li W, Chen Z, Zhang N, Liu C (2022) Deep learning with multisource data fusion in electricity internet of things for electricity price forecast. *Wirel Commun Mob Comput* 2022(1):3622559. https://doi.org/10.1155/2022/3622559

[20] Janjua JI, Ahmad R, Abbas S, Mohammed AS, Khan MS, Daud A, Abbas T, Khan MA (2024) Enhancing smart grid electricity prediction with the fusion of intelligent modeling and XAI integration. *Int J Adv Appl Sci* 11(5):230–248. https://doi.org/10.21833/ijaas.2024.05.025

[21] Fu B (2025) Variational Autoencoder-based high-dimensional feature extraction for economic analysis of power cost data. *Informatica* 49(25). DOI:10.31449/inf.v49i25.8012

[22] Dombi G, Dulai T (2022) Hourly electricity price forecast for short-and long-term, using deep neural networks. *Acta Univ Sapientiae Inform* 14(2):208–222. DOI:10.2478/ausi-2022-0013

[23] Albahli S, Shiraz M, Ayub N (2020) Electricity price forecasting for cloud computing using an enhanced machine learning model. *IEEE Access* 8:200971–200981.https://doi.org/10.1109/ACCESS.2020.3035328

APPENDIX

$machine\ learning\ = ML$	$Neural\ Networks\ = NN$
$Mean\ Absolute\ Percentage\ E\ = MAPE$	$Long\ Short-term\ Memory\ Network\ = LSTMs$
$adaptive-network-based\ fuzzy\ inference\ sys\ = ANFIS$	$Support\ vector\ machine\ = SVM$
$Support\ Vector\ Regression\ = SVR$	$grey\ catastrophe\ = GC$
$Decision\ Tree\ = DT$	$random\ forest\ = RF$
$Principal\ Component\ Analys\ = PCA$	$least\ squares\ support\ vec\ = LSSV$
$Linear\ Discriminant\ Analysis\ = LDA$	$quadratic\ programs\ = QP$
$Early\ Warning\ System\ = EWS$	$Radial\ Basis\ Function\ = RBF$
$Mean\ Absolute\ Error\ = MAE$	$deep\ neural\ networks\ = DNN$
$Extreme\ Gradient\ Boosting\ = XGBoost$	$Information\ Technology\ = IT$