Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 6, 431-443
2024
Publisher: Learning Gate
DOI: 10.55214/25768484.v8i6.2094

Hybrid CNN-GRU model for monthly electricity price forecasting: Performance evaluation on limited multivariable time-series Data

Fatma Yaprakdal[1]*
[1]Information System Engineering Department, Bucak Computer and Informatics Faculty, Burdur Mehmet Akif Ersoy University; fatmayaprakdal@mehmetakif.edu.tr (F.Y.)

Abstract: Understanding electricity pricing dynamics is crucial for market participants, as prices reflect supply and demand balances. Accurate medium-term price prediction aids in maintenance scheduling, expansion planning, and contracting, but poses challenges due to a long forecasting horizon and limited explanatory data. This paper proposes a hybrid forecasting system that merges Convolutional Neural Networks (CNN) and Gated Recurrent Units (GRU) for estimating monthly electricity prices. The CNN performs feature extraction, while the GRU handles temporal regression. We evaluate model performance using mean absolute percentage error (MAPE) and the coefficient of determination (R^2). Experimental results indicate that our model outperforms both popular deep learning (DL) methods (GRU, LSTM) and machine learning (ML) techniques (SVR, RF, XGBoost), confirming the feasibility and effectiveness of this approach for accurate electricity price prediction.
Keywords: *CNN, GRU, Electricity pricing, Hybrid model, Medium-term forecasting, SVR, RF, XG Boost.*

1. Introduction

The dynamics of electricity pricing are crucial in the electricity market, where trading activities are significantly influenced by current electricity prices. These prices reflect the delicate balance between supply and demand, serving as fundamental indicators of market value during transactions. Accurate forecasting of electricity prices is essential for market participants, as it enhances operational efficiency and informs strategic decision-making.

Electricity Price Forecasting (EPF) encompasses different timeframes tailored to specific needs. Short-term EPF typically covers periods ranging from one hour to a few weeks, focusing on immediate market fluctuations and operational decisions. Medium-term EPF spans several weeks to up to a year, facilitating strategic planning and resource management. In contrast, long-term EPF extends beyond a year, providing insights for broader investment and policy-making decisions. Each of these horizons presents unique challenges and necessitates distinct methodologies to accurately capture the relevant influencing factors [1, 2].

The literature on short-term load forecasting is extensive, spanning several decades. In comparison, medium-term forecasting has received less attention [3]. However, medium-term price prediction plays a pivotal role in the electricity sector, facilitating critical activities such as scheduling maintenance, planning for infrastructure expansion, and managing contractual agreements. Despite its importance, forecasting medium-term electricity prices presents significant challenges due to the complexity of the forecasting horizon and the multitude of variables influencing prices, including weather conditions, fuel costs, and market behavior. Various methodologies have been explored for EPF, ranging from traditional statistical approaches to advanced ML and DL techniques. Statistical methods, such as Autoregressive Integrated Moving Average (ARMA) and its variant with Exogenous Variables (ARMAX), have been widely applied in EPF and peak load estimation. For example, an improved

* Correspondence: fatmayaprakdal@mehmetakif.edu.tr

ARMAX model that integrates Hilbert operators was introduced to analyze the Moving Average (MA) components in time series data. However, traditional statistical approaches often struggle with the non-linear dynamics characteristic of electricity markets [4].

To overcome the limitations of traditional methods, ML models have become increasingly popular. Among the leading ML techniques for forecasting are Support Vector Machines (SVM), Random Forest (RF), and Extreme Gradient Boosting (XGBoost). For example, SVM has been effectively applied to predict electricity prices by analyzing historical data on electricity and gas prices in Germany [5]. Reference [6] explored various ML techniques, including Support Vector Regression (SVR) and RF, to improve the accuracy of electricity price predictions. Furthermore, XGBoost has demonstrated its potential for EPF in Ontario [7].

Given that ML methods are often deemed more suitable for smaller datasets, initial comparisons with these approaches were made in the context of this study. In recent years, DL methods, particularly Long Short-Term Memory (LSTM) networks, have gained popularity for their capacity to model complex temporal dependencies [8]. This study examines two recurrent neural network (RNN) models: LSTM and Gated Recurrent Unit (GRU), both of which are effective at capturing long-term dependencies and temporal dynamics in time series data. Their efficacy has been established in various energy-related studies [9, 10]. Notably, the combination of Convolutional Neural Networks (CNN) and GRU has shown promise in short-term EPF. [11] illustrates how CNN-GRU models can effectively capture short-term price fluctuations by leveraging both spatial and temporal data. Furthermore, [12] enhanced this approach by incorporating attention mechanisms, which improved prediction accuracy by allowing the model to dynamically focus on relevant features. However, the application of CNN-GRU models for medium-term predictions remains largely unexplored, creating a significant gap in the literature.

While CNN-GRU models have been successfully utilized in various financial contexts, such as stock market price predictions [13] and gold pricing [14], their potential for medium-term EPF has not yet been fully realized. This study seeks to bridge this gap by proposing a hybrid forecasting system that combines CNN and GRU to enhance prediction accuracy in the medium-term electricity market. By leveraging the strengths of both models—CNN for feature extraction and GRU for temporal regression—we aim to improve the accuracy of medium-term electricity price predictions. The entire study is conducted in the Python programming environment. This choice allows for the optimization of data analysis, model development, and result visualization processes by leveraging various libraries, such as Pandas, NumPy, Scikit-learn, and Matplotlib. Python's open-source nature and extensive community support make it a preferred choice, particularly in the fields of data science and ML [15]. Throughout this study, the efficiency provided by Python in data processing and analysis significantly contributes to the reliability and accuracy of the results. The model is validated using K-Fold Cross-Validation, and performance metrics such MAPE and the R^2 score are employed to evaluate accuracy. For this study, the hourly multivariate dataset sourced from the New England ISO from 2004 until 2008 is transformed into a monthly format. Our proposed method demonstrates superior performance compared to traditional ML techniques, which typically excel with larger datasets. This finding underscores the effectiveness of the CNN-GRU hybrid model in the context of medium-term forecasting. Ultimately, by demonstrating the feasibility and practicality of this combined model, we aim to advance the field of EPF and provide actionable insights for market participants, contributing to more effective decision-making in the electricity market.

The rest of this paper is structured as follows: Section 2 outlines the methodology adopted in this study, including details on data preprocessing, model architectures, and the evaluation metrics used for forecasting monthly electricity prices. Section 3 presents the experimental results, including a detailed comparison of different models' performances, followed by a discussion of the implications and insights drawn from the findings. Section 4 concludes the study by summarizing the main outcomes, addressing the significance of the results, and suggesting future research directions to further enhance forecasting accuracy in this domain.

2. Methodology

The methodology includes the following key steps: data analysis, the proposed CNN-GRU hybrid model architecture, a brief overview of benchmarking methods, training procedures, and evaluation metrics.

2.1. Data Analysis

2.1.1. Data Collection

The dataset utilized for this study is sourced from the New England ISO from 2004 until 2008, encompassing historical electricity prices and relevant features such as weather related conditions, natural gas prices, and historical electricity consumption. The weather data includes the dry bulb temperature, representing the ambient air temperature, and the dew point, which indicates the temperature at which air reaches saturation with water vapor. This valuable dataset is available on the MathWorks website and can be accessed through reference [16]. For this study, the original dataset has been resampled at monthly intervals to capture average values, instead of its original hourly resolution by using pyhton resampling code.

2.1.2. Correlation Analysis of Key Variables

The correlation matrix heatmap presented in Figure 1. summarizes the relationships between five critical variables: Dry Bulb Temperature (DryBulb), Dew Point Temperature (DewPnt), System Load (SysLoad), Natural Gas Price (NGPrice), and Electricity Price (ElecPrice). Each entry in the matrix indicates the strength and direction of the linear relationship between pairs of variables, with values between -1 and 1. A value that is close to 1 signifies a strong positive correlation, while a value that is close to -1 indicates a strong negative correlation. Key observations from the correlation map can be summarized as follows:

- ***Dry Bulb Temperature and Dew Point Temperature***: The correlation coefficient is 0.99, indicating an extremely strong positive relationship. This suggests that as the dry bulb temperature increases, the dew point temperature also tends to rise, reflecting their inherent connection as indicators of atmospheric conditions.
- ***Dry Bulb Temperature and Electricity Price***: A notable negative correlation of -0.26 is observed between Dry Bulb Temperature and Electricity Price. This indicates that higher dry bulb temperatures are associated with lower electricity prices, potentially reflecting reduced demand during warmer months when cooling systems are more prevalent.
- ***Dew Point Temperature and Electricity Price***: The correlation of -0.30 between Dew Point Temperature and Electricity Price further reinforces the inverse relationship observed with dry bulb temperature. The negative association may indicate that increased humidity (reflected in dew point increases) could lead to higher demand for cooling, thus impacting electricity prices.
- ***System Load and Natural Gas Price***: A weak positive correlation of 0.09 exists between System Load and Natural Gas Price. This suggests that changes in system load have a minimal direct impact on natural gas prices, which may be influenced by other factors such as supply chain dynamics or broader market trends.
- ***Natural Gas Price and Electricity Price***: A strong positive correlation of 0.79 between Natural Gas Price and Electricity Price is observed. This relationship indicates that fluctuations in natural gas prices significantly affect electricity prices, likely due to the reliance on natural gas as a primary fuel source for electricity generation.

- ***System Load and Electricity Price***: The correlation coefficient of 0.01 between System Load and Electricity Price suggests a negligible relationship, indicating that system load does not have a direct effect on electricity prices within the context of this dataset.

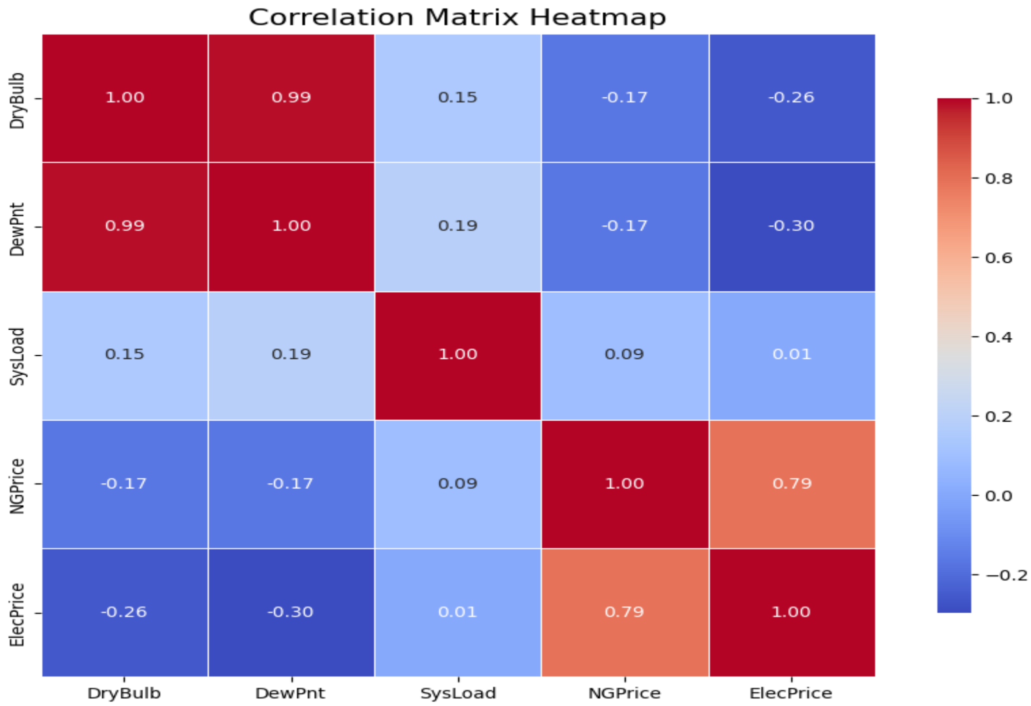


Figure 1.
The correlation matrix heatmap among all the features of the dataset.

This correlation analysis provides valuable insights into the interdependencies among key variables influencing electricity pricing and consumption. The strong correlations between temperature metrics and electricity prices highlight the impact of weather conditions on energy demand, while the robust relationship between natural gas prices and electricity prices underscores the importance of fuel costs in electricity market dynamics. Understanding these correlations can inform more effective forecasting models and strategic decision-making in energy management and policy formulation.

2.1.3. Data Processing

Before training, the dataset is normalized using Min-Max scaling formulea as follows:

$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

where x is the original feature value, and x_{min} and x_{max} are the minimum and maximum values of the feature, respectively. This transformation enhances model convergence during training [17]. Specifically, we set the min-max scaler range as [0, 1] for our analysis. Furthermore, sliding window approach is implemented to create sequences for training. The window size is set to 12 in order to forecast the 13^{th} data point, using the first 12 data points (x_1 to x_{12}) as input. The dataset consists of 60 rows and 5 columns, representing features such as historical monthly average loads, natural gas prices, electricity prices, dew point, and dry bulb temperature and it is splitted into 80% for training and 20% for testing.

2.2. The Proposed CNN-GRU Model Architecture

The proposed hybrid model integrates a 1D-CNN for feature extraction and a GRU for capturing temporal dependencies. This combination allows for the effective modeling of complex relationships within time series data.

2.2.1. CNN Component

The CNN component is designed to extract features from the input data through a series of convolutional operations. The fundamental operation of a convolutional layer is expressed by the equation:

$$\mathrm{Y} = f(\mathrm{W} * X + b)$$

where Y is the output, W is the weight matrix, b is the bias, X is the input, and f is the activation function, which is typically the Rectified Linear Unit (ReLU) for CNNs [13]. The architecture of the CNN component consists of the following layers:

I. *Input Layer*: Accepts the time series input data.
II. *Convolutional Layers*: Two 1D convolutional layers are employed, with the first layer containing 32 filters and the second containing 64 filters. Each layer utilizes a kernel size of 2 and applies the ReLU activation function to introduce non-linearity.
III. *Max Pooling Layer*: This layer reduces the dimensionality of the feature maps while retaining important features. The pooling size is set to 2, which aids in minimizing computational complexity.
IV. *Flatten Layer*: Converts the pooled features into a one-dimensional vector to facilitate processing by subsequent layers.

2.2.2. GRU Component

The GRU component processes the features extracted by the CNN layers and captures temporal dependencies within the data. The update equations for the GRU are expressed as follows:

$$z_t = \sigma(W_z . [h_{t-1}, x_t])$$
$$r_t = \sigma(W_r . [h_{t-1}, x_t])$$
$$\widetilde{h_t} = \tanh(W_h . [r_t . h_{t-1}, x_t])$$
$$h_t = (1 - z_t) . h_{t-1} + z_t . \widetilde{h_t}$$

In these equations, z_t represents the update gate, r_t denotes the reset gate, h_t is the hidden state, $\widetilde{\mathrm{h_t}}$ is the new hidden state computed using the current input and the reset hidden state, and x_t is the input at time *t* [18]. The GRU architecture comprises:

1. *Input Layer*: Similar to the CNN branch, it receives the time series data.
2. *GRU Layers*: The first GRU layer returns sequences to the second layer, which aggregates the information to generate the final output. The first layer consists of 50 units, while the second layer has 25 units.

2.2.3. Model Integration

The outputs from the CNN and GRU layers are concatenated before being passed through a fully connected layer, defined as:

$$Y_{output} = W_{fc} . h_t + b_{fc}$$

where W_{fc} and b_{fc} are the weights and bias of the fully connected layer, respectively. The integration of the CNN and GRU components occurs through the following steps:

1. *Concatenation*: The outputs from the CNN and GRU branches are concatenated, allowing the model to leverage both spatial and temporal features effectively.

2. *Output Layer*: A dense layer with a single unit is employed to produce the final prediction.
3. *Model Compilation*: The model is compiled using the Adam optimizer and the mean squared error for the loss function suitable for regression tasks.

2.3. Benchmarking Methods

To evaluate the efficacy of the proposed hybrid model, we compared it with established ML models, including SVR, RF, and LSTM.

2.3.1. SVR

SVR aims to find a function f(x) that deviates from the actual observed values *y* by a value no greater than ϵ. f(x) is formulated as in the following equation.

$$f(x) = \left(v.\varphi(x)\right) + b,$$

where x is the training data, v is the weight vector, b is the intercept, and ϕ is a mapping function. The optimization problem can be formulated as:

$$min\frac{1}{2}\|\mathrm{v}\|^2 + C\sum_{i=0}^{l}\Psi\left(f(x_i) - Y_i\right)$$

Subject to:

$$y_i - f(x)_i \le \epsilon + \xi_i$$
$$f(x)_i - y_i \le \epsilon + \xi_i^*$$

where C is the regularization parameter, and ξ_i are the slack variables [19].

2.3.2. RF

RF constructs numerous decision trees and combines them to achieve a more precise and reliable prediction. The output of the model (Y_m) is given by the average of predictions from all individual trees:

$Y_m = \frac{1}{T}\sum_{t=1}^{T} f_t\,(x)$

where T is the number of trees and $f_t(x)$ is the prediction from tree *t* [20].

2.3.2. LSTM

LSTM networks are a variant of recurrent neural networks (RNNs) designed to effectively capture and learn long-term dependencies. The cell state c_t and the hidden state h_t are updated as follows:

$$\widetilde{c_t} = \tanh\left(W_c[h_{t-1}, x_t] + b_c\right),$$
$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i),$$
$$f_t = \sigma\left(W_f[h_{t-1}, x_t] + b_f\right),$$
$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o),$$
$$c_t = (i_t \times \widetilde{c_t}) + (f_t \times c_{t-1}),$$
$$h_t = o_t \times \tanh\left(c_t\right)$$

where W_i, W_f and W_o, and b_i, b_f and b_o are the weights and biases that govern the behavior of the i_t, f_t, and o_t gates, respectively. W_c and b_c are the weights and bias of the memory cell candidate $\widetilde{c_t}$,respectively, and '*tanh*' is the hyperbolic tangent activation function [13].

ISSN: 2576-8484
Vol. 8, No. 6: 431-443, 2024
DOI: 10.55214/25768484.v8i6.2094

2.4. Training and Evaluation Procedure of The Models

To enhance the training of the models, two key callbacks are implemented [21]:

1. *Early Stopping*: This callback halts the training process if the validation loss does not show improvement for a specified number of epochs, known as "patience." It also restores the model weights to those from the epoch with the lowest validation loss, ensuring that overfitting is mitigated.
2. *Reduce Learning Rate*: This callback is activated when the validation loss plateaus, allowing for a gradual reduction of the learning rate. This adjustment can help fine-tune the model's performance, facilitating convergence and improving prediction accuracy.

The model is trained on the training dataset over a specified number of epochs (50), utilizing a batch size of 7. To ensure robust performance evaluation, 5-fold cross-validation is employed across all models. This method partitions the dataset into four subsets, allowing each subset to serve as a validation set while the remaining three subsets are used for training. Upon completion of the training process, predictions are generated on the validation set. The performance of the model is then assessed using two metrics: MAPE and R^2. The formulas for these metrics are provided below:

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right|,$$

$$R^2 = 1 - \frac{\sum_{t=1}^{n}(A_t - F_t)^2}{\sum_{t=1}^{n}(A_t - A_m)^2}$$

where A_t represents the actual values, F_t denotes the forecasted values, and A_m is the mean of the actual values [22]. Through these systematic training and evaluation processes, the study aims to ensure the reliability and accuracy of the model's predictions, ultimately contributing to improved decision-making in the relevant domain.

3. Experimental Results and Discussions

The performance of the proposed hybrid CNN-GRU forecasting model has been benchmarked against several other models, including SVR, XGBoost, RF, LSTM, and GRU, using two key evaluation metrics: MAPE and the R^2. In addition, to provide deeper insight into model training dynamics and generalization, the training and validation loss curves are plotted for LSTM, GRU, and CNN-GRU models over the cross-validation folds. Finally, prediction versus actual value plots for all models are given and discussed in detail.

3.1. Forecasting Results Based On Performance Metrics For Models' Evaluation

A summary of the results for all models is presented in Table 1, followed by a detailed analysis of the performance of each model. The performance of the forecasting models, evaluated using MAPE and R^2 metrics, reveals key insights about their effectiveness on a small dataset. SVR shows a relatively high MAPE of 28.436% and an R^2 of 0.803, indicating moderate predictive performance but substantial errors, particularly due to limited capacity in capturing complex temporal patterns. XGBoost achieves a MAPE of 21.300% and an R^2 of 0.832, showing better accuracy compared to SVR. It captures trends reasonably well but still has notable errors. RF performs well, with a lower MAPE of 16.157% and a high R^2 of 0.937, suggesting it models the variability effectively while keeping prediction errors relatively low. The LSTM model has a high MAPE of 29.793% and a lower R^2 of 0.705, indicating struggles in generalizing on small datasets, possibly due to the model's complexity and the limited amount of training data. The GRU model achieves a MAPE of 19.547% and an R^2 of 0.893, showing improved performance over LSTM, likely due to a more efficient architecture that works better with smaller datasets. The CNN-GRU hybrid model demonstrates the best performance with a MAPE of 15.738% and an R^2 of 0.949, highlighting its ability to effectively extract both spatial and temporal features, thereby reducing error and improving overall prediction quality.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 6: 431-443, 2024
DOI: 10.55214/25768484.v8i6.2094

In summary, the CNN-GRU hybrid and RF models perform the best on the small dataset, with the lowest MAPE and highest R^2 values, indicating their robustness and effectiveness in handling limited data. The GRU also performs well, while SVR, XGBoost, and LSTM are less effective, particularly struggling to minimize prediction errors and generalize well under data constraints.

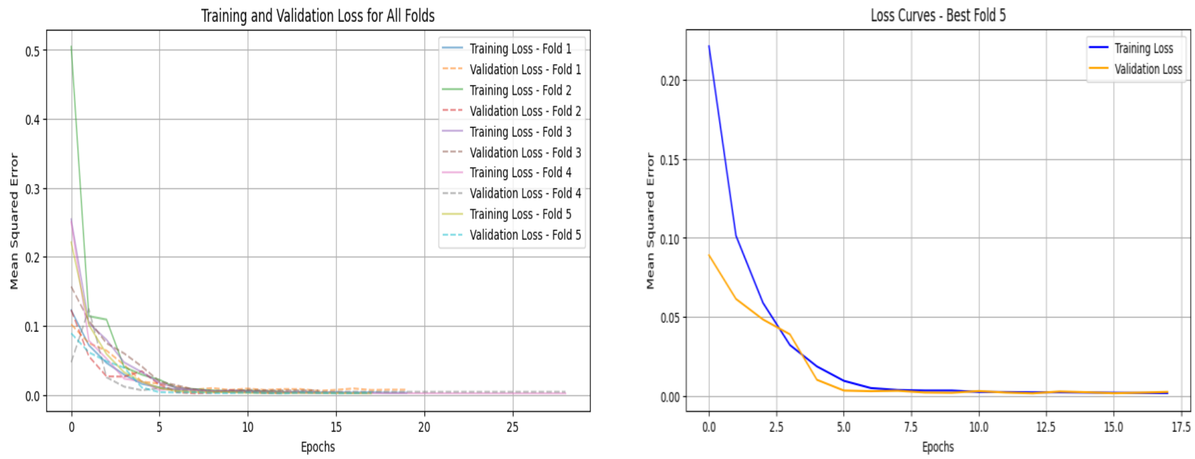
Table 1.
Monthly EP forecasting performances of all models.

Forecasting models	Mean MAPE (%)	Mean R^2
SVR	28.436	0.803
XGBoost	21.300	0.832
RF	16.157	0.937
LSTM	29.793	0.705
GRU	19.547	0.893
CNN-GRU	15.738	0.949

These findings suggest that ensemble methods (XGBoost and RF) and hybrid DL models (CNN-GRU) are particularly well-suited for complex forecasting tasks, while more traditional or standalone DL methods like LSTM may require more optimization or larger datasets to reach their full potential. The superior performance of XGBoost and CNN-GRU highlights their flexibility in capturing intricate patterns and relationships in the data, making them strong candidates for future forecasting applications.

3.2. Training and Validation Loss Curves for All Models

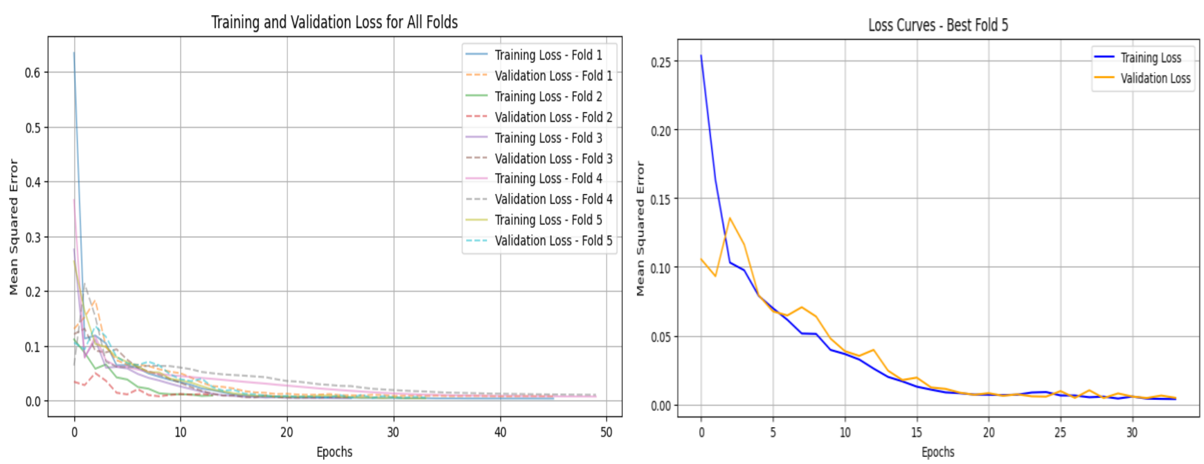
To provide deeper insight into model training dynamics and generalization, we plotted the **training** and validation loss curves for DL-based models over the cross-validation folds. These curves illustrate how each model converges during training and how well the model generalizes to unseen validation data. Figure 2 illustrate the training and validation loss of a CNN-GRU hybrid model across multiple cross-validation folds, as well as for the best-performing fold. Across all folds, the training and validation losses decrease rapidly during the initial epochs and stabilize at low values, indicating effective learning and generalization. The low variability across folds suggests consistent model performance, with minimal overfitting. For the best-performing fold, both training and validation losses converge to near-zero values, demonstrating effective model training and excellent generalization without significant overfitting. These results highlight the ability of the CNN-GRU hybrid model to capture both spatial and temporal features effectively, demonstrating robustness and stability across different data splits.



(a) **(b)**

Figure 2.
The training and validation loss curves for the CNN-GRU hybrid model across **(a)** five different folds and **(b)** the best fold.

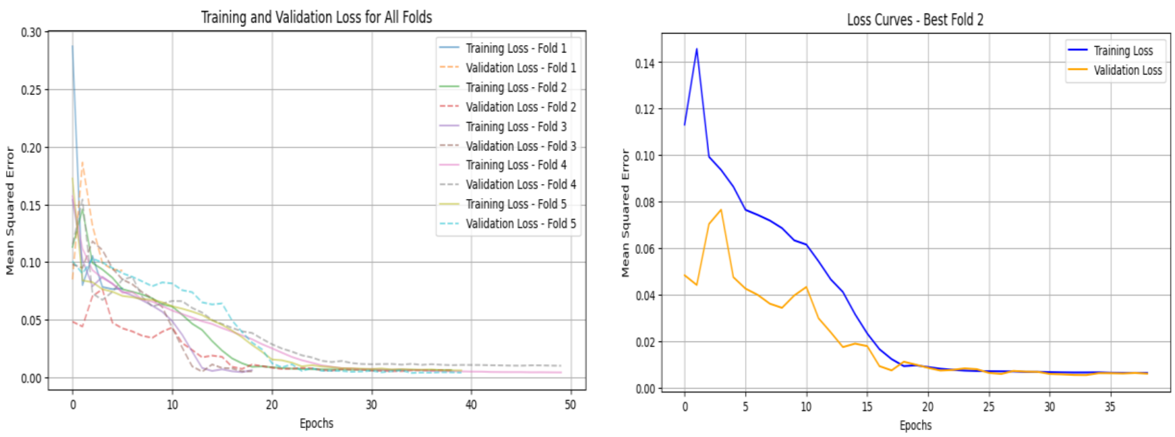
Figure 3 depicts the training and validation loss curves of a GRU model across multiple cross-validation folds, demonstrating effective learning of temporal features. The steep decline in losses across all folds, followed by convergence to low values, indicates robust learning and good generalization with minimal overfitting. The results for the best-performing fold (Fold 5) further confirm the model's strong ability to generalize, as evidenced by the close alignment of training and validation losses. Overall, the GRU model shows stability and effectiveness in consistently learning sequential patterns across different data splits.



(a) **(b)**

Figure 3.
The training and validation loss curves for the GRU model across **(a)** five different folds and **(b)** the best fold.

Figure 4 illustrates the training and validation loss (in Mean Squared Error) for an LSTM model across five cross-validation folds. Figure 4(a) presents the loss curves for all folds, showing a rapid decline in losses during the initial epochs followed by gradual stabilization, indicating effective learning of temporal dependencies and good generalization. Some variability among folds suggests different levels of complexity or noise in the data splits. Figure 4(b) highlights the best-performing fold (Fold 2), where both training and validation losses consistently reduce and converge to near-zero values, showing effective error minimization, stability, and minimal overfitting. Overall, the results demonstrate the robustness and efficacy of the LSTM model in capturing sequential patterns and generalizing well across multiple data splits.



(a) **(b)**

Figure 4.
The training and validation loss curves for the LSTM model across **(a)** five different folds and **(b)** the best fold.

Incorporating these training and validation loss plots allows for a detailed analysis of how each model performs across different training folds. These curves also highlight potential overfitting or underfitting problems, offering further insights into why certain models (e.g., LSTM) performed worse than others (e.g., RF, CNN-GRU).

3.3. Prediction vs Actual Value Plots for All Models

The provided figures in Figure 5 illustrate the comparison between predicted and true (actual) values for five different models—CNN-GRU hybrid, GRU, LSTM, SVR, XGBoost, and RF—on the best-performing folds in a cross-validation setting. Each graph evaluates the ability of these models to capture temporal dependencies from a very small dataset. The x-axis represents either time steps or sample indices, while the y-axis represents the normalized values of the target variable, such as electricity price.

When considering the metrics from Table 1, which presents the MAPE and R^2 for each model, we can reassess their performance. The CNN-GRU hybrid model demonstrates the best performance, achieving the lowest MAPE and highest R^2. In the graph, the predicted values closely align with the true values, suggesting that the model successfully captures both spatial and temporal features despite the limited data available. This strong alignment suggests that the CNN-GRU hybrid model is highly capable of generalizing well even with a small dataset. The GRU model also performs well, with a moderate MAPE and a high R^2. The graph shows that the GRU model effectively captures major trends and fluctuations, with minimal deviations between the predicted and true values. However, some slight discrepancies occur during sharp transitions, suggesting that the model faces challenges in fully generalizing in certain scenarios due to the limited data. The LSTM model, on the other hand, shows a higher MAPE and lower R^2, which indicate difficulties in generalizing from the small dataset. This is consistent with the graphical analysis, where noticeable deviations occur, particularly during abrupt changes. These findings suggest that the LSTM model struggles with variance, likely due to its complex architecture and the risk of overfitting with insufficient data. The SVR model has the second-highest MAPE and a moderate R^2, indicating significant limitations when applied to this dataset. The graph shows considerable deviations, especially in capturing peaks and valleys, which reflects SVR's inability to model non-linear and abrupt temporal changes effectively under limited data conditions. The XGBoost model performs moderately well, with a relatively high R^2 but still notable MAPE. The graph shows that XGBoost can capture the overall trends, although minor deviations remain at points of rapid change. This suggests that while XGBoost is adept at modeling structured data, it faces challenges in handling the complexities of small sequential datasets. The RF model shows strong performance, with a low MAPE and a high R^2. In the graph, the predictions align well with the true values, indicating that RF effectively captures general patterns. However, minor discrepancies at rapid transitions suggest that the model struggles slightly with more abrupt variations, especially given the limited data.

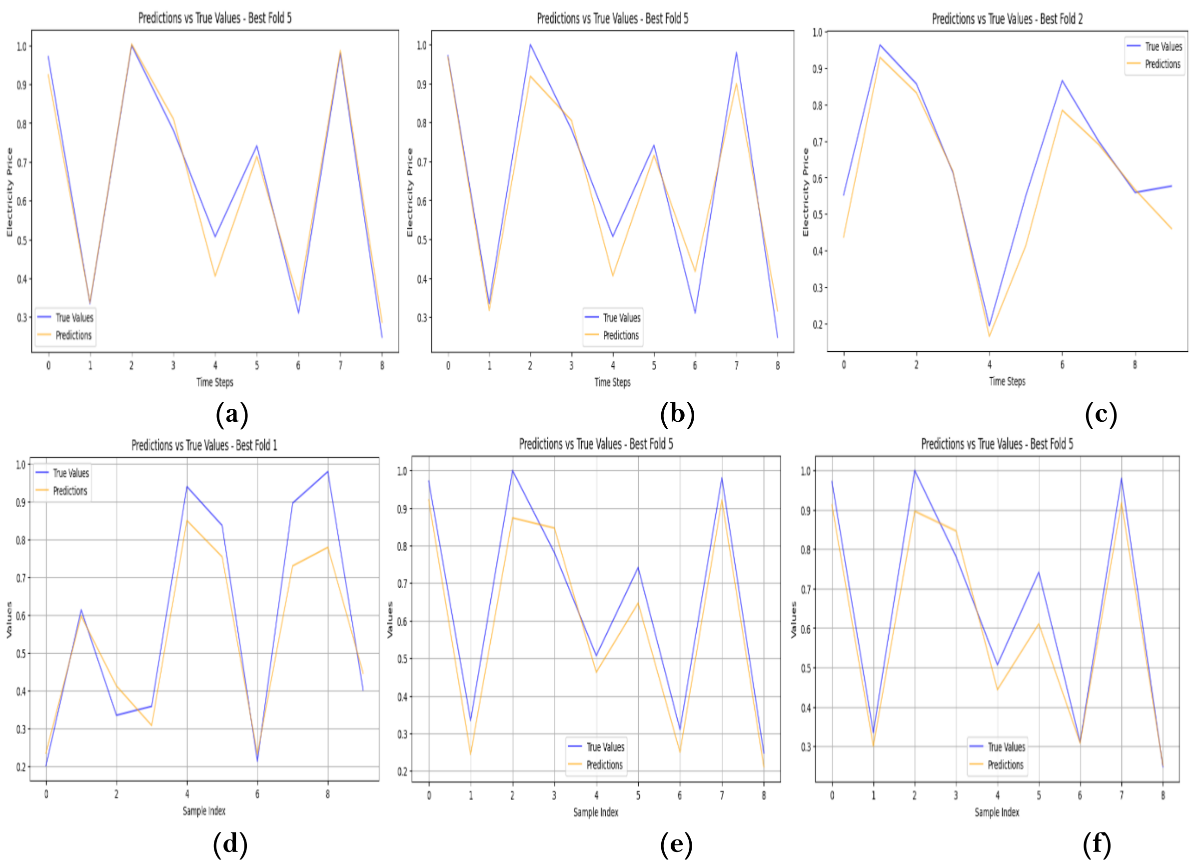


Figure 5.
The comparison between predicted and true values for five different models—**(a)** CNN-GRU hybrid, **(b)** GRU, **(c)** LSTM, **(d)** SVR, **(e)** XGBoost and **(f)** RF on the best-performing folds in a cross-validation setting.

In summary, the CNN-GRU hybrid and RF models emerge as the most effective for handling small datasets, demonstrating the lowest errors and strongest generalization capabilities. The GRU model also shows good performance, effectively capturing trends. In contrast, SVR, XGBoost, and LSTM demonstrate weaker performance, struggling with error minimization and generalization under data constraints. Overall, the CNN-GRU hybrid and RF models show the most promise in accurately forecasting with very limited data.

4. Conclusion

In this study, we evaluated the performance of various forecasting models—including CNN-GRU hybrid, GRU, LSTM, SVR, XGBoost, and RF—in predicting monthly electricity prices using a very small multivariable time-series dataset. The results demonstrate that the CNN-GRU hybrid model outperforms the other models, achieving the lowest MAPE and the highest R^2. This model's superior performance can be attributed to its unique combination of convolutional layers for spatial feature extraction and GRU layers for sequential learning, which enables it to effectively handle both spatial and temporal dependencies within the data. The RF model also showed strong predictive capabilities, demonstrating good generalization and low errors. Conversely, the LSTM, SVR, and XGBoost models struggled more, particularly in capturing sharp changes and maintaining generalization, likely regarding the temporal patterns' complexity and the limited dataset size.

The results address a gap in the literature regarding the application of CNN-GRU hybrid models to monthly EPF particularly under conditions of limited data. Existing literature often emphasizes the use of more conventional time-series models or DL models like LSTMs for forecasting, but few studies explore the advantages of hybrid models such as CNN-GRU for small datasets. The findings suggest that CNN-GRU is a promising approach for EPF when data is scarce, highlighting its robustness and ability to effectively extract meaningful patterns despite data limitations.

Finally, considering the emerging trend towards renewable energy and decentralized grids, expanding the forecasting horizon to accommodate renewable energy integration will be essential. Investigating hybrid approaches like CNN-GRU for different energy market structures, especially in combination with probabilistic forecasting methods to estimate uncertainty, could provide a comprehensive solution for stakeholders involved in electricity markets.

Overall, this study provides an important contribution to the field of EPF, demonstrating the value of hybrid DL models like CNN-GRU in conditions of limited data availability. The results encourage future exploration of hybrid and enhanced DL approaches to further improve the reliability and accuracy of electricity price predictions.

Copyright:

References

[1] Zhang, Y., et al. (2020). A comprehensive review on electricity price forecasting: An empirical analysis of the state of the art. Renewable and Sustainable Energy Reviews, 128, 109905.

[2] Wang, J., et al. (2021). Electricity price forecasting: A review of techniques and applications. IEEE Transactions on Power Systems, 36(2), 1234-1246.

[3] Ziel, F., & Steinert, R. (2018). Probabilistic mid- and long-term electricity price forecasting. Renewable and Sustainable Energy Reviews, 94, 251-266.

[4] Torbaghan, S. S., Motamedi, A., Zareipour, H., & Tuan, L. A. (2012). Medium-term electricity price forecasting. 2012 North American Power Symposium (NAPS), Champaign, IL, USA, 1-8.

[5] Shiri, A., Afshar, M., Rahimi-Kian, A., & Maham, B. (2015). Electricity price forecasting using Support Vector Machines by considering oil and natural gas price impacts. In 2015 IEEE International Conference on Smart Energy Grid Engineering (SEGE). IEEE, 1-5.

[6] Tschora, L., Pierre, E., Plantevit, M., & Robardet, C. (2022). Electricity price forecasting on the day-ahead market using machine learning. Applied Energy, 313, 118752.

[7] Albahli, S., Shiraz, M., & Ayub, N. (2020). Electricity price forecasting for cloud computing using an enhanced machine learning model. IEEE Access, 8, 200971-200981.

[8] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735-1780.

[9] Cho, K., Van Merriënboer, B., Bahdanau, D., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078.

[10] Wang, H.-z., Li, G.-q., Wang, G.-b., Peng, J.-c., Jiang, H., & Liu, Y.-t. (2019). Deep learning-based ensemble approach for probabilistic wind power forecasting.
[11] Sajjad, M., et al. (2020). A Novel CNN-GRU-Based Hybrid Approach for Short-Term Residential Load Forecasting. IEEE Access, 8, 143759-143768.
[12] Yang, G., Du, S., & Duan, Q. (2022). Short-term Price Forecasting Method in Electricity Spot Markets Based on Attention-LSTM-mTCN. Journal of Electrical Engineering & Technology, 17, 1009-1018.
[13] Song, H., & Choi, H. (2023). Forecasting Stock Market Indices Using the Recurrent Neural Network Based Hybrid Models: CNN-LSTM, GRU-CNN, and Ensemble Models. Applied Sciences, 13(7), 4644.
[14] Billah, M. M., & Das, S. (2021). Analysis and Prediction of Gold Price using CNN and Bi-GRU based Neural Network Model. 2021 24th International Conference on Computer and Information Technology (ICCIT), Dhaka, Bangladesh, 1-6
[15] Python, W. (2021). Python releases for windows. Python, 24.
[16] Deoras, A. (2024). Electricity Load and Price Forecasting with MATLAB. MathWorks Webinar Case Study. Available online: https://www.mathworks.com/matlabcentral/fileexchange/28684-electricity-load-and-price-forecasting-webinar-case-study.
[17] Rahul, K. A., Frankle, M., & Madan, M. T. (2019). Ensemble of relevance vector machines and boosted trees for electricity price forecasting. Applied Energy, 250, 540-548.
[18] Sajjad, M., et al. (2020). A Novel CNN-GRU-Based Hybrid Approach for Short-Term Residential Load Forecasting. IEEE Access, 8, 143759-143768.
[19] Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. Statistics and Computing, 14, 199-222.
[20] Breiman, L. (2001). Random Forests. Machine Learning, 45, 5-32.
[21] Vanitha K, R MT, Sree SS, Guluwadi S. Deep learning ensemble approach with explainable AI for lung and colon cancer classification using advanced hyperparameter tuning. BMC Med Inform Decis Mak. 2024 Aug 7;24(1):222.
[22] Yaprakdal, F. (2024). A comparative analysis of single learning and ensemble learning approaches for the forecasting of electricity prices. Edelweiss Applied Science and Technology, 8(5), 317-327.