ITM Web of Conferences **70**, 03011 (2025) https://doi.org/10.1051/itmconf/20257003011
DAI 2024

Research for SARIMA and PatchTSMixer Models on the IEA Monthly Statistics Dataset

Yuwei Hu

Future Tech, South China University of Technology, Guangzhou, Guangdong, China

Abstract. The rapid evolution of intelligent algorithms has led to their extensive application in time-series forecasting, particularly in predicting electricity consumption. Accurate forecasting is crucial for energy management, policy-making, and ensuring a stable power supply. However, a significant gap exists in comparing the predictive performance of traditional machine learning methods with advanced deep learning models using real-world datasets. This study aims to address this gap by evaluating and comparing the prediction accuracy of machine learning and deep learning techniques using the Monthly Electricity Statistics dataset from the International Energy Agency (IEA). The research employs a rigorous experimental design, leveraging models such as ARIMA and PatchTSMixer, with an emphasis on model tuning and performance metrics like MAE, MAPE, and RMSE. The findings reveal that deep learning models, particularly PatchTSMixer, outperform traditional machine learning methods in terms of prediction accuracy, demonstrating their superior capability in capturing complex temporal dependencies in electricity consumption data. These results highlight the potential of deep learning models for enhancing predictive accuracy in time-series forecasting, providing valuable insights for future research and practical applications in energy management.

1 Introduction

Time-series forecasting has long been a subject of extensive research, particularly in the energy sector. Seasonal AutoRegressive Integrated Moving Average (SARIMA) is a well-established statistical approach commonly employed in forecasting tasks for its simplicity and effectiveness in handling seasonal variations and non-stationary trends in time series forecasting [1]. Studies like those by Utami et al. have demonstrated the efficacy of Seasonal ARIMA in various forecasting scenarios, such as international arrivals and consumer price indices [2]. However, the emergence of deep learning techniques has introduced more sophisticated models capable of capturing complex patterns and non-linear relationships in data [3]. PatchTSMixer, as explored by Ekambaram et al., represents a state-of-the-art deep learning model specifically designed for time-series data, offering promising results in predictive accuracy and computational efficiency [4].

Corresponding author: 202164690305@mail.scut.edu.cn

Despite the advancements in both methodologies, there remains a gap in the literature regarding their comparative performance on real-world electricity consumption datasets. Most existing studies either focus on a single method or lack comprehensive comparisons that consider the practical implications of their use in energy systems. For instance, Luo and Nie highlight the potential of various deep learning models but do not provide a detailed comparison with traditional models like SARIMA [5, 6]. This study seeks to address this gap by evaluating and contrasting the prediction performances of PatchTSMixer and SARIMA using the Monthly Electricity Statistics dataset from the International Energy Agency (IEA).

The significance of this research lies in its potential to inform stakeholders in the energy sector on the most effective forecasting tools. This study preprocesses the IEA dataset for data quality, implements PatchTSMixer and SARIMA models, and evaluates their forecasting accuracy using Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Root Mean Squared Error (RMSE) metrics. The analysis compares the models' predictive capabilities, computational efficiency, and scalability. The goal is to address the research gap by comparing PatchTSMixer and SARIMA in electricity consumption forecasting, aiming to identify the more accurate and reliable method for improving energy management and resource allocation, with potential economic and environmental benefits.

2 Methods

2.1 Data description

The monthly electricity statistics dataset used in this study was collected from the International Energy Agency (IEA) Database [7]. This dataset provides monthly data from January 2014 to February 2024 on electricity production and trade for 47 countries, including OECD member states and selected non-OECD economies like China. The data used includes net electricity production on total renewables (hydro, geo, solar, wind, and other) in China, as shown in Table 1.

Table 1. Net electricity production on total renewables in China

No.	Time	Value (GWh)
1	2015/1/1	86244.545
2	2015/2/1	71621.002
⋮	⋮	⋮
111	2024/3/1	253679.6074
112	2024/4/1	254704.6441

2.2 Analysis phase

The research process is depicted in the flowchart in Fig. 1, which outlines the critical steps undertaken to forecast electricity consumption using the PatchTSMixer and SARIMA models. The workflow begins with data preprocessing, followed by model training and evaluation, and ends with a comparison of the model's predictive performance.

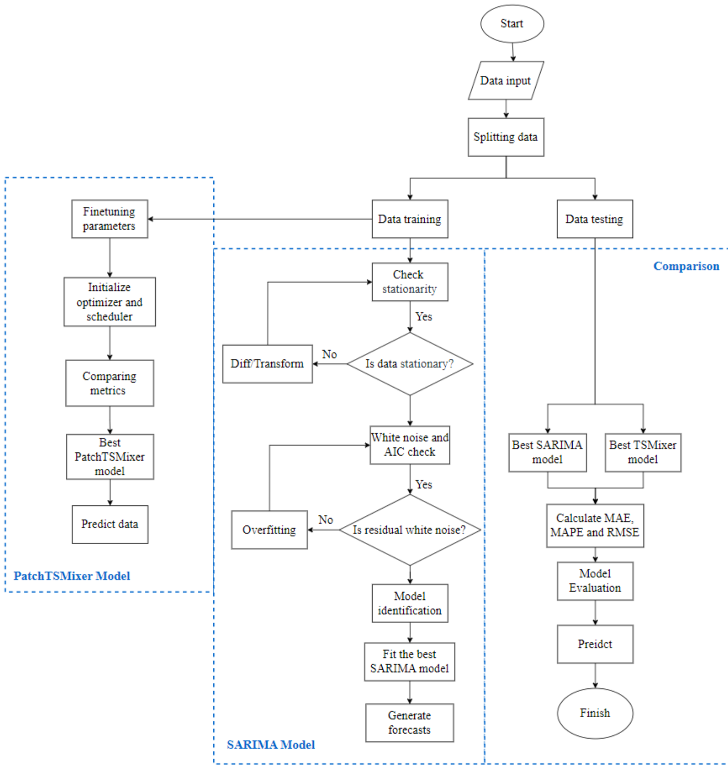


Fig. 1 Research workflow diagram (Photo/Picture credit: Original).

2.3 Seasonal ARIMA model

Seasonal ARIMA is a machine learning method that integrates both non-seasonal and seasonal components, specifically designed to capture the intricate seasonal patterns [8]. This model was employed as a benchmark to assess the effectiveness of other predictive models on monthly electricity statistics datasets. The SARIMA model is represented as $SARIMA\ (p, d, q)\ (P, D, Q, s)$, where (p, d, q) are the non-seasonal parameters:

p denotes the number of autoregressive (AR) terms, allowing the model to include the influence of previous values.

d represents the number of non-seasonal differences needed to achieve stationarity.

q indicates the number of lagged forecast errors in the moving average (MA) component, enabling the model to express the error as a linear combination of previous error values.

The seasonal parameters (P, D, Q) are analogous to the non-seasonal ones but are applied to the seasonal component of the time series. The parameter s refers to the periodicity of the time series. The mathematical representation of SARIMA model can be written as [9]:

$$(1-\phi_1 B)(1-\Phi_1 B^s)(1-B)(1-B^s)y_t = (1+\theta_1 B)(1+\Theta_1 B^s)\varepsilon_t \qquad (1)$$

Where ϕ_1 is the non-seasonal autoregressive coefficient, Φ_1 denotes the seasonal autoregressive coefficient, θ_1 represents the non-seasonal moving average coefficient, Θ_1 indicates the seasonal moving average coefficient, y_t is the observed time series at time t, B is the backward shift operator, also known as the lag operator, s refers to the seasonal period, while ε_t is the white noise error term at time t.

The model parameters were fine-tuned using the Akaike Information Criterion (AIC) to strike a balance between simplicity and the quality of fit, as determined by the following formula [10]:

$$AIC = \ln\left(\hat{\sigma}_a{}^2\right) + \frac{2(p+q+1)}{n} \tag{2}$$

2.4 PatchTSMixer model

PatchTSMixer model is a lightweight neural architecture specifically designed for multivariate time series forecasting [4]. It relies solely on Multi-Layer Perceptron (MLP) modules, making it a computationally efficient alternative to Transformer models [11]. The methodology of PatchTSMixer integrates several crucial components and design principles:

Patching-Based Architecture. Similar to patch-based methods in computer vision, PatchTSMixer divides the input time series $X \in \mathbb{R}^{T\times C}$ into patches to effectively capture temporal dynamics. Here, T represents the time steps, and C denotes the number of channels (or features). The model processes these patches using its backbone, which learns representations of the temporal sequences.

Backbone and Prediction Heads. The backbone is denoted by $f_{backbone}(X)$, which produces a latent representation $Z \in \mathbb{R}^{P\times D}$, where P is the number of patches, and D is the dimension of the feature space. The prediction head $\mathrm{h}_{\mathrm{head}}(Z)$ is then applied to produce the final output $\hat{Y} \in \mathbb{R}^{\mathrm{T}'\times C}$, where T' is the forecast horizon.

Online Reconciliation Heads. PatchTSMixer introduces online reconciliation heads to refine predictions:

Hierarchical Patch Reconciliation: This head aggregates information across patches, formally expressed as:

$$\widehat{\Upsilon}_{agg} = W_{agg} \cdot Z \tag{3}$$

Where W_{agg} is the aggregation weight matrix.

Cross-Channel Reconciliation: This captures interactions between different channels:

$$\widehat{\Upsilon}_{cross} = W_{cross} \cdot Z \tag{4}$$

Where W_{cross} facilitates the interaction between various time series.

Hybrid Channel Modeling. PatchTSMixer utilizes a hybrid approach to model channels. The backbone independently processes each channel, producing Z_c for each channel c. The reconciliation head models interactions between channels:

$$\widehat{\Upsilon}_c = h_{\mathrm{recon}}(Z_c) + \textstyle\sum_{j\neq c} h_{\mathrm{recon}}\left(Z_j\right) \tag{5}$$

Where h_{recon} is the reconciliation function.

Gated Attention Mechanism. To handle long sequences, PatchTSMixer incorporates a gated attention mechanism:

$$\mathrm{Attention}(Q, K, V) = \mathrm{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{6}$$

Where Q, K, and V are the query, key, and value matrices, respectively, and d_k is the dimension of the keys. This attention mechanism helps the model to focus on the most critical features.

Modular Design for Training. PatchTSMixer supports both supervised and self-supervised training methods. In supervised training, the model is optimized to minimize the MSE:

$$\mathrm{MSE} = \frac{1}{N}\textstyle\sum_{i=1}^{N}(\widehat{\Upsilon}_i - Y_i)^2 \tag{7}$$

Where N represents the number of samples, $\widehat{\Upsilon}_i$ denotes the predicted value, and Y_i refers to the actual value. In self-supervised training, the model is trained using a masked time series modeling (MTSM) task, where parts of the time series are masked and then predicted.

In this research, two models are trained on the IEA monthly electricity consumption datasets. The pre-trained PatchTSMixer model will be quickly fine-tuned on only 5% of the train split of the target data, and subsequently, evaluated on the test part of the target data.

For deep learning model PatchTSMixer, AdamW optimizer was used with the following hyperparameters: learning rate (lr) = 0.001, $\alpha = 0.001$, $\beta_1 = 0.9$, $\beta_2 = 0.999$ [12]. The learning rate was chosen based on preliminary experiments to ensure stable convergence and efficient training, while the AdamW optimizer was selected for its ability to incorporate weight decay directly into the update rule, which helps prevent overfitting by regularizing the model's parameters. The OneCycleLR scheduler is employed to manage the learning rate, typically aimed at facilitating rapid training with higher learning rates [13].

2.5 Performance evaluation

In this study, three evaluation metrics—RMSE, MAE, and MAPE—were employed to assess model performance. These metrics collectively provided a thorough evaluation of the models' predictive accuracy by quantifying the discrepancies between forecasted and actual values. Both RMSE and MAE measure the absolute magnitude of forecast errors, while RMSE places greater emphasis on larger errors. MAPE calculates the average percentage difference between predicted and observed values, offering an understandable measure of relative error [1].

3 Results

3.1 Data splitting

The model is constructed using 112 observations from training data spanning January 2015 to April 2024. The data partitioning was conducted to identify the optimal model by focusing on the most significant coefficients, minimizing the AIC value, and verifying that residual assumptions are satisfied. After selecting the model, its performance was assessed using a testing dataset comprising 24 observations from April 2022 to April 2024, as illustrated in Fig. 2. This approach ensures that both the SARIMA and PatchTSMixer models have a consistent number of observations for comparison.

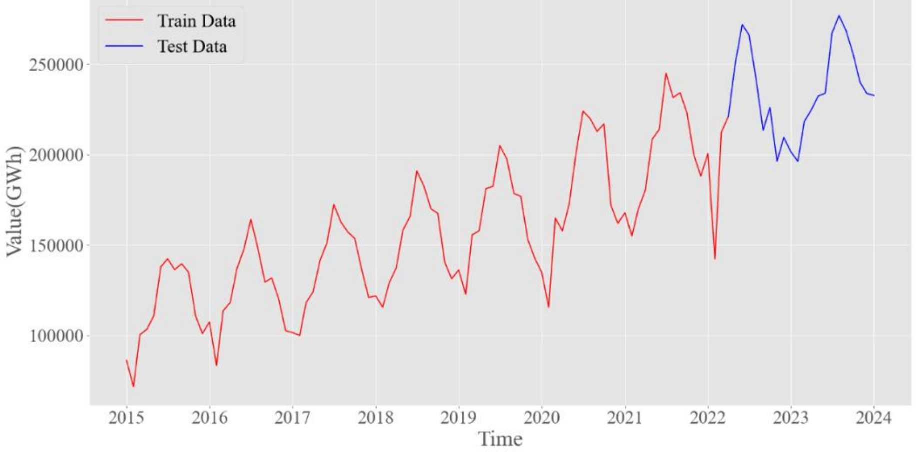
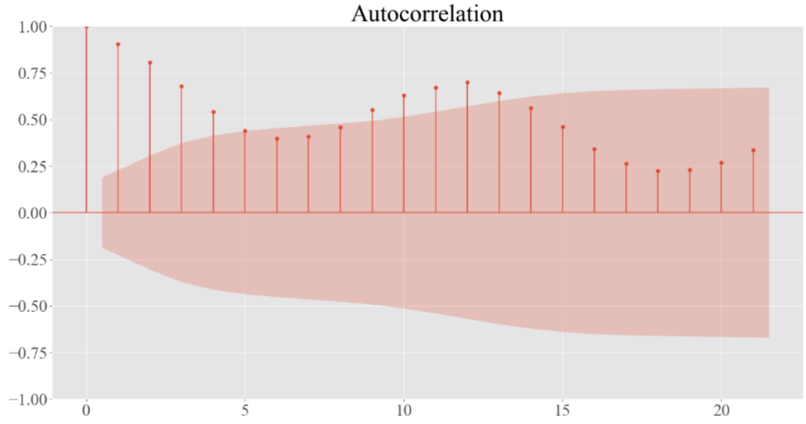


Fig 2. Division of testing and training data (Photo/Picture credit: Original).

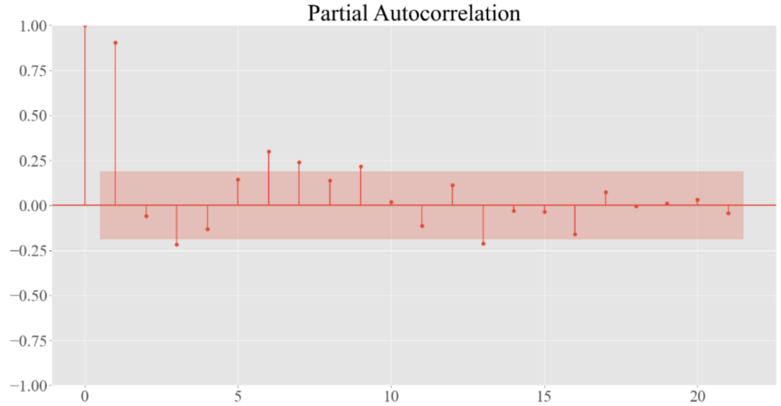
3.2 SARIMA

The SARIMA model assumes that the time series data should be stationary in both mean and variance to ensure accurate forecasting. Stationarity is typically evaluated through the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots, alongside statistical tests such as the Augmented Dickey-Fuller (ADF) test.

As shown in Fig. 3, the ACF plot (Fig. 3a) demonstrates a gradual decay, which is indicative of non-stationarity. Fig. 3b, the PACF plot, similarly reflects a non-stationary series. The ADF test on the original data yielded an ADF statistic of 1.0914 with a p-value > α (0.05), as shown in Table 2, confirming the non-stationarity of the series.



(a)



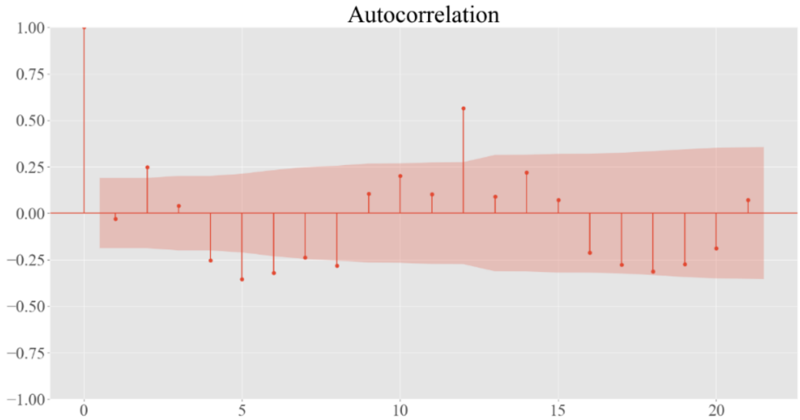
(b)

Fig. 3 Non-stationary data for (a) ACF and (b) PACF plots (Photo/Picture credit: Original).

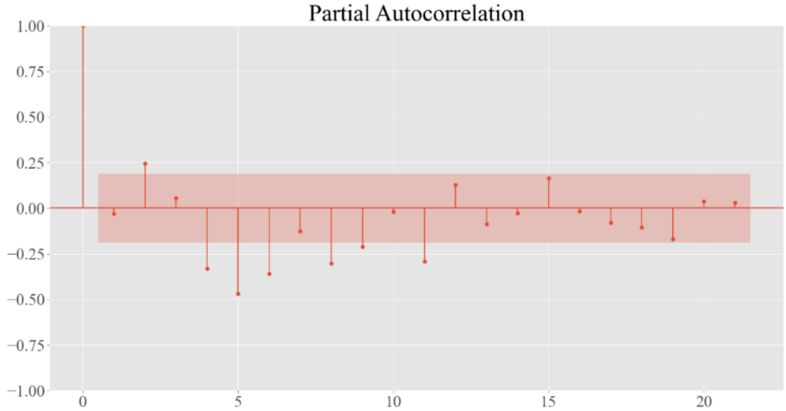
To address the non-stationarity, the data was differenced ($d = 1$). The results, presented in Fig. 4, reveal a notable spike at lag 1 in the ACF plot (Fig. 4a), which suggests strong autocorrelation. The PACF plot (Fig. 4b) also reflects this correlation, now suitable for model identification. The ADF test on the differenced data produced an ADF statistic of -8.1269 with a p-value near zero, as shown in Table 2, confirming stationarity post-differencing.

Table 2. Stationarity check

Data	ADF Statistic	p-value
Original Data	1.0914	0.9951
Differenced Data	-8.1269	1.1169×10^{-12}



(a)



(b)

Fig. 4 (a) ACF and (b) PACF plot of differencing data (Photo/Picture credit: Original).

AIC is determined by reducing the squared error of the prediction or estimate [14]. Consequently, the SARIMAX(3, 1, 0)x(3, 1, 1, 12) model, which had the lowest AIC value of 777.0669, was chosen as the optimal model for further residual diagnostic testing.

3.3 Comparison of SARIMA and PatchTSMixer model

Fig. 5 illustrates a comparison between the actual values and the predictions made by two models: SARIMA and PatchTSMixer, with a focus on forecasting China's monthly electricity consumption from renewable sources (in GWh) over time. The actual fit is depicted in red, the SARIMA and PatchTSMixer fit lines are marked by teal and gray respectively, and the out of sample prediction of PatchTSMixer is also represented by a gray line. Both models capture the underlying trend and seasonal patterns in the data quite well up to the year 2023. The actual data (red line) shows a clear seasonal pattern with peaks and troughs occurring annually, which both models seem to mimic effectively in the training period.

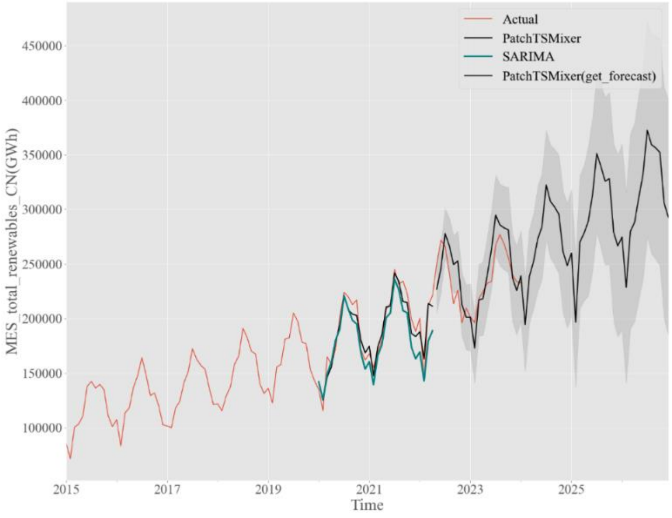


Fig. 5 Seasonal ARIMA and PatchTSMixer fit over the monthly electricity statistics dataset starting from 2020-01-01 (Photo/Picture credit: Original).

The SARIMA model depicted in the teal line closely follows the actual data throughout the observed period. The model's forecasts closely match the observed values, especially in terms of capturing seasonal fluctuations. However, as the forecasting horizon extends beyond 2023, the SARIMA model starts to deviate slightly from the actual values but still maintains a reasonable accuracy.

The PatchTSMixer model depicted in the gray line also tracks the actual values closely but introduces more variance and uncertainty in its predictions, especially in the forecasting horizon beyond 2023. The shaded area around the PatchTSMixer predictions represents the prediction interval, which widens significantly after 2023, indicating increasing uncertainty. This model seems to provide a more flexible forecast, which could be beneficial in capturing unexpected variations but also suggests a higher degree of uncertainty compared to SARIMA.

MAE, MAPE, and RMSE are used quantify the accuracy of the prediction for SARIMA and PatchTSMixer. The PatchTSMixer model outperforms the SARIMA model across all three key metrics. Specifically, the PatchTSMixer model achieves a significantly lower average MAE of 1.622 compared to 3.035 for the SARIMA model, as shown in Table 3. This suggests that, on average, PatchTSMixer's predictions align more closely with the actual values.

In terms of average MAPE, PatchTSMixer also shows superior performance with a value of 0.028, which is much lower than the 0.131 observed for SARIMA. This suggests that the PatchTSMixer model provides more accurate percentage-based predictions relative to the actual data. The average RMSE for PatchTSMixer is 2.754, less than half of SARIMA's 6.514, demonstrating that PatchTSMixer not only reduces error magnitude but also exhibits greater consistency in its forecasts. Overall, the PatchTSMixer model provides more accurate and reliable predictions than the SARIMA model for this dataset.

Table 3. Performance metrics of averaged SARIMA and PatchTSMixer models for monthly electricity consumption of renewables in China

Metrics	PatchTSMixer	SARIMA
Avg MAE	1.622	3.035
Avg MAPE	0.028	0.131
Avg RMSE	2.754	6.514

4 Discussion

4.1 Impact of SARIMA and PatchTSMixer on time series prediction performance

This study compares the performance of SARIMA and PatchTSMixer in the task of time series forecasting. The SARIMA model, as a traditional statistical method, is known for its stability and interpretability in handling seasonal time series data. However, with the increase of dataset complexity and dimensionality, its prediction accuracy and generalization ability are somewhat limited. In contrast, the PatchTSMixer model, leveraging deep learning techniques, can efficiently extract complex temporal features from large-scale data, demonstrating high prediction accuracy. Experimental results show that PatchTSMixer significantly outperforms SARIMA on multidimensional time series data, especially in capturing long- and short-term dependencies, and its prediction error is significantly reduced.

4.2 Study limitations and future research directions

Despite the results achieved in this study, there are still several limitations. Ajiboye suggests that utilizing a sufficiently large dataset for building predictive models can enhance accuracy and improve the model's generalization capability [15]. Therefore, the choice and size of the dataset may limit the generalization of the results, and future studies could incorporate more varied datasets to test the model's robustness. Secondly, the PatchTSMixer model has high training time and computational resource consumption, which may be limited in practical applications. Future research could explore more efficient model architectures or optimization strategies to reduce the computational cost. In addition, only two methods, SARIMA and PatchTSMixer, were explored in this study, which can be extended to other emerging deep learning models in the future to further compare and enhance the time series forecasting performance. Finally, how to better interpret the prediction results of deep learning models is also an issue that deserves in-depth research, especially in high-risk decision-making scenarios where model interpretability is crucial.

5 Conclusion

This research has demonstrated that the application of time-series prediction methods, particularly deep learning and machine learning approaches, which can significantly enhance forecasting accuracy in the context of the Monthly Electricity Statistics dataset from IEA. The analysis revealed that deep learning models, due to their ability to identify complex patterns and manage long-term dependencies, outperform traditional machine learning methods in predictive accuracy. Specifically, the PatchTSMixer model significantly outperformed the SARIMA model across key metrics, achieving an average MAE of 1.622, an average MAPE of 0.028, and an average RMSE of 2.754, compared to SARIMA's average MAE of 3.035, average MAPE of 0.131, and average RMSE of 6.514. These results emphasize the potential of deep learning models like PatchTSMixer in improving electricity demand forecasting, which is crucial for optimizing energy resource allocation and planning.

This finding highlights the importance of deep learning models in refining electricity demand forecasting, which is crucial for optimizing the allocation and planning of energy resources. Moreover, the successful application of the PatchTSMixer model suggests that migrating advanced techniques from image processing to time series analysis has broad applicability, offering new insights for integrating cross-domain methods. The research impacts not only the field of energy economics but also other industries where precise forecasting is essential, such as finance and supply chain management. Current models have already found extensive application in these areas, contributing to more effective decision-making and strategic planning.

Looking ahead, the continuous advancement of computational techniques and the growing availability of large datasets will further enhance the predictive capabilities of these models. Future research should focus on integrating hybrid models and exploring their applicability across diverse datasets to ensure more robust and generalized forecasting methods.

References

1. A. H. Ni, S. Meng, X. Geng, P. Li, Z. Li, X. Chen, ... & S. Zhang, Time series modeling for heart rate prediction: From ARIMA to transformers. arXiv preprint arXiv:2406.12199 (2024).
2. M.Y. Hu, S.L. Setyowati, K.A. Notodiputro, Y. Angraini, & L.N.A. Mualifah, Comparison of seasonal ARIMA and support vector machine forecasting method for international arrival in Lombok. Jambura J. Math. **6**, 212-219 (2024).
3. X. Liu, & W. Wang, Deep time series forecasting models: A comprehensive survey. Math. **12**, 1504 (2024).
4. V. Ekambaram, A. Jati, N. Nguyen, P. Sinthong, & J. Kalagnanam, Tsmixer: Lightweight MLP-mixer model for multivariate time series forecasting. Proc. 29th ACM SIGKDD Conf. Knowl. Discov. Data Min. 459-469 (2023).
5. Q. Luo, S. He, X. Han, Y. Wang, & H. Li, LSTTN: A long-short term transformer-based spatiotemporal neural network for traffic flow forecasting. Knowl. Based Syst. 293, 111637 (2024).
6. Y. Nie, N.H. Nguyen, P. Sinthong, & J. Kalagnanam, A time series is worth 64 words: Long-term forecasting with transformers. arXiv preprint arXiv:2211.14730 (2022).
7. IEA, Monthly electricity statistics. IEA, 2024/2/9. Paris, https://www.iea.org/data-and-statistics/data-product/monthly-electricity-statistics, Licence: Terms of Use for Non-CC Material.
8. M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, & J.W. Vaughan, Advances in neural information processing systems 34 (2021).
9. G.E. Box, G.M. Jenkins, G.C. Reinsel, & G.M. Ljung, Time series analysis: Forecasting and control. John Wiley & Sons (2015).
10. W.A. Woodward, B.P. Sadler, & S. Robertson, Time series for data science: Analysis and forecasting. Chapman and Hall/CRC (2022).
11. S.A. Chen, C.L. Li, N. Yoder, S.O. Arik, & T. Pfister, Tsmixer: An all-MLP architecture for time series forecasting. arXiv preprint arXiv:2303.06053 (2023).
12. R. Llugsi, S. El Yacoubi, A. Fontaine, & P. Lupera, Comparison between Adam, AdaMax and Adam W optimizers to implement a weather forecast based on neural networks for the Andean city of Quito. Proc. 2021 IEEE Fifth Ecuad. Tech. Chap. Meet. 1-6 (2021).
13. H. Qian, P. Ma, S. Gao, & Y. Song, Soft reordering one-dimensional convolutional neural network for credit scoring. Knowl. Based Syst. **266**, 110414 (2023).
14. K.R.A. Muslihin, & B.N. Ruchjana, Model autoregressive moving average (ARMA) untuk peramalan tingkat inflasi di Indonesia. Limits J. Math. Appl. **20**, 209-218 (2023).

15. A.R. Ajiboye, R. Abdullah-Arshah, H. Qin, & H. Isah-Kebbe, Evaluating the effect of dataset size on predictive model using supervised learning technique. Int. J. Comput. Syst. Softw. Eng. **1**, 75-84 (2015).