Exploring Hybrid ARIMA-LSTM Models for Enhanced Short-Term Wind Power Production Forecasting

Mohammad Arif Rasyidi

Department of Informatics, Universitas Internasional Semen Indonesia, Gresik, Indonesia

mohammad.rasyidi@uisi.ac.id

**Abstract.**  This paper investigates the potential of hybrid ARIMA-LSTM models for enhancing short-term wind power production forecasting. Accurate forecasting is crucial for optimizing grid operations and integrating renewable energy sources. We begin by evaluating the performance of ARIMA and LSTM models individually, noting that ARIMA performs better for very short-term forecasts, while LSTM excels at longer prediction horizons. To leverage the strengths of both models, we develop several hybrid approaches, including the Simple ARIMA-LSTM Hybrid (SALH), Machine Learning Combined ARIMA-LSTM (MLCAL), Residual LSTM (RLSTM), ARIMA-LSTM Pipeline (ALP), and ARIMA-LSTM Augmented (ALA) models. Our results show that the hybrid models generally outperform the base models, with the ALA model achieving the lowest average RMSE across various prediction horizons, followed closely by the ALP model. Notably, MLCAL models using more sophisticated combiners, such as Multi-Layer Perceptron, also demonstrate significant improvements, although the choice of combiner critically affects performance. These findings underscore the effectiveness of hybrid models in capturing both linear and non-linear dependencies in wind power data, leading to more accurate predictions.

**Keywords:** Wind Forecasting, ARIMA, LSTM, Hybrid Models, Time Series Prediction.

# INTRODUCTION

As the world increasingly shifts towards renewable energy sources to combat climate change and reduce dependency on fossil fuels, wind energy has emerged as one of the most prominent and rapidly growing alternatives [1] [2]. According to recent global energy trends, wind power has significantly contributed to the global renewable energy portfolio, providing a clean and sustainable option to meet the rising energy demands [3]. However, the intermittent and variable nature of wind poses challenges for integrating wind energy into the power grid efficiently. Accurate forecasting of wind power production is, therefore, essential for maintaining grid stability, optimizing energy market operations, and reducing operational costs[4].

Accurate wind power forecasting is inherently challenging due to the stochastic and highly variable nature of wind patterns [5] [6]. These challenges are compounded by the influence of local weather conditions, geographic features, and temporal factors, making it difficult to predict wind power generation with high precision. Addressing these challenges requires robust forecasting models that can effectively capture the complex dynamics of wind behavior over different time horizons.

Among the various models developed for time series forecasting, the AutoRegressive Integrated Moving Average (ARIMA) model has been widely adopted in wind power forecasting due to its strong statistical foundation and ability to model linear dependencies in data [7] [8] [9]. ARIMA has demonstrated effectiveness in short-term forecasting, particularly when the underlying wind patterns exhibit linear characteristics. However, its reliance on linear assumptions limits its ability to capture non-linear dynamics and complex interactions, which are often present in wind data.

On the other hand, Long Short-Term Memory (LSTM) networks, a type of recurrent neural network, have gained popularity in recent years for time series forecasting tasks, including wind power forecasting [10]. LSTM networks are specifically designed to model sequential data and are capable of capturing long-range dependencies and non-linear patterns within the data. This makes LSTM particularly suitable for handling the complex and non-linear nature of wind power generation. Nevertheless, LSTM models typically require large datasets for training and can be computationally intensive, which may limit their effectiveness in certain scenarios, especially when dealing with shorter prediction horizons.

The strengths and weaknesses of ARIMA and LSTM highlight the potential for a complementary approach that leverages the advantages of both models. Recent research has indicated that while ARIMA often outperforms LSTM for shorter prediction horizons, LSTM tends to excel over longer periods where non-linear patterns dominate [11] [12]. This observation underscores the need for hybrid models that combine ARIMA's strengths in linear modeling with LSTM's ability to handle non-linear dynamics, thereby improving overall forecasting accuracy across different time scales.

Hybrid ARIMA-LSTM models have emerged as a promising solution to enhance wind power forecasting by integrating the best of both worlds [13] [14]. These hybrid models typically employ either series or parallel approaches to combine the outputs of ARIMA and LSTM [15]. In a series approach, ARIMA is used to model and remove linear components from the data, with the residuals then being fed into an LSTM model to capture the remaining non-linear patterns. In contrast, the parallel approach involves training both models independently on the same data and then combining their predictions to achieve a final forecast. While these approaches have shown potential, there is still a significant gap in the literature regarding the exploration and comparison of various hybridization strategies and their effectiveness across different forecasting horizons.

The primary objective of this paper is to explore and evaluate various strategies for combining ARIMA and LSTM models in the context of short-term wind power forecasting. Specifically, we aim to compare the performance of different hybrid models across various prediction horizons, shedding light on the conditions under which each hybridization strategy excels. By systematically analyzing these hybrid models, this research seeks to contribute to the growing body of knowledge on wind power forecasting and provide practical insights for enhancing the accuracy and reliability of wind power predictions.

This paper is organized as follows. In the subsequent section, we detail the methodology employed for this study. Next, we present and discuss the findings of our analysis, highlighting the most significant results and providing insights into their implications. Finally, we conclude by summarizing our key findings and outlining potential avenues for future research.

# METHODS

## Data Collection

The dataset used in this study consists of historical wind power generation data from the Belgian grid, sourced from Open Data Elia. This dataset was selected due to its comprehensive coverage of wind power production across various regions in Belgium, making it highly relevant for assessing the performance of forecasting models. The data comprises the load factor, defined as the percentage ratio between measured power generation and the total monitored capacity of the wind farms. Load factor is a critical measure as it standardizes power output, making it comparable across different wind farms regardless of their capacities.

The dataset includes time series data from five wind farms across different regions, recorded at 15-minute intervals, spanning from 2019 to 2023. The five wind farms are identified based on their proximity to the shore, geographic location, and grid connection type. Specifically, the time series are categorized as follows:

1. Onshore-Flanders-Dso
2. Onshore-Flanders-Elia
3. Offshore-Federal-Elia
4. Onshore-Wallonia-Dso
5. Onshore-Wallonia-Elia

Each time series comprises 175,296 observations, providing a rich dataset for training and testing forecasting models.

## Data Preprocessing

To ensure consistency across the different time series, each series was clipped to the (0, 1) interval, effectively handling any outliers and standardizing the data. Subsequently, Min-Max normalization was applied to scale the series uniformly between 0 and 1. This normalization step is crucial, particularly for neural network-based models like LSTM, which are sensitive to the scale of input data.

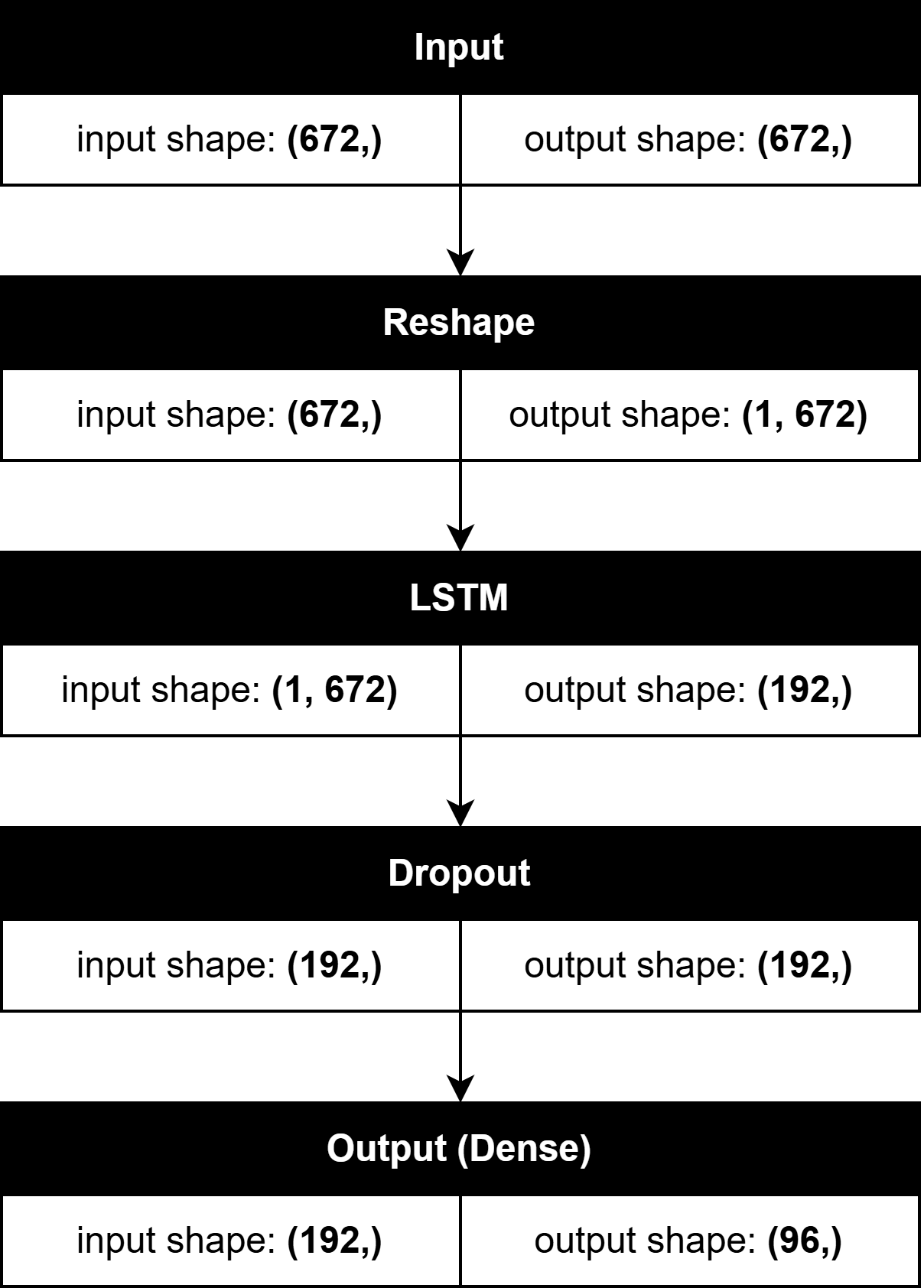
The dataset was then transformed using a sliding window approach to create the input features and corresponding target variables for the models. A window size of 672 (equivalent to one week) was chosen, with a stride of 1, enabling the model to capture weekly patterns in the wind power data. The prediction horizons ranged from 1 to 96 time steps (up to one day), allowing the models to forecast wind power production at various short-term intervals.

For model evaluation, the dataset was split into training and test sets. The last 35,040 data points (equivalent to approximately one year) were reserved as the test set, ensuring that the models are evaluated on recent data that was not used during training. The remaining data points were used for training the models, providing a substantial amount of data for model learning.

## Base Model Development

### ARIMA Model Development

The ARIMA model was developed using the auto ARIMA approach, as proposed by Hyndman and Khandakar, and implemented via the pmdarima library in Python [16]. The auto ARIMA method automates the selection of the optimal ARIMA parameters (p, d, q) by iteratively testing different combinations and selecting the model that minimizes information criteria such as the Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC). This approach is particularly useful for managing the complexity of tuning ARIMA models across multiple time series.



**FIGURE 1**. Architecture of the LSTM model

### LSTM Model Development

The LSTM model was designed with a simple yet effective architecture to forecast wind power production. The model, as shown in **FIGURE 1**, consists of the following layers:

1. Input Layer: Accepts a sequence of 672 values, representing the wind power data over the past week.
2. Reshape Layer: Reshapes the input data into a sequence format with dimensions (1, 672) suitable for LSTM processing.
3. LSTM Layer: Comprising 192 units, this layer is responsible for capturing temporal dependencies within the data.
4. Dropout Layer: A dropout rate of 0.2 is applied to prevent overfitting and improve model generalization.
5. Fully Connected (Dense) Layer: Outputs 96 values, corresponding to predictions for the next 96 time steps (up to one day).

The LSTM model was trained using the training dataset, with the last 20% of the training data reserved as a validation set. The model was optimized using the Adam optimizer, with Mean Squared Error (MSE) as the loss function. To further enhance training efficiency, a Cyclical Learning Rate (CLR) schedule, as proposed by Leslie N. Smith [17], was employed. CLR dynamically adjusts the learning rate within a specified range, promoting faster convergence and potentially avoiding local minima. The model was trained for up to 1,000 epochs, but Early Stopping was implemented to halt training once the validation loss ceased to improve, thereby preventing overfitting.

## Hybrid Model Development

To explore the potential of hybrid models in enhancing wind power forecasting accuracy, five variations of hybrid ARIMA-LSTM models were developed and evaluated:

### Simple ARIMA-LSTM Hybrid (SALH)

The SALH model adopts a straightforward approach by combining the predictions from ARIMA and LSTM models. Specifically, for the initial part of the prediction horizon, the ARIMA model’s forecasts are used, while the LSTM model’s predictions are applied for the subsequent part. This method leverages ARIMA’s strength in short-term forecasting and LSTM’s ability to model longer-term dependencies.

### ML-Combined ARIMA-LSTM (MLCAL)

In the MLCAL model, predictions from the ARIMA and LSTM models are combined using machine learning algorithms. Various machine learning models, including Linear Regression, Multi-Layer Perceptron (MLP), and K-Nearest Neighbors (KNN), were trained to learn the optimal weights for merging the ARIMA and LSTM forecasts. This data-driven approach allows for a more sophisticated combination of predictions, potentially improving overall accuracy.

### Residual LSTM (RLSTM)

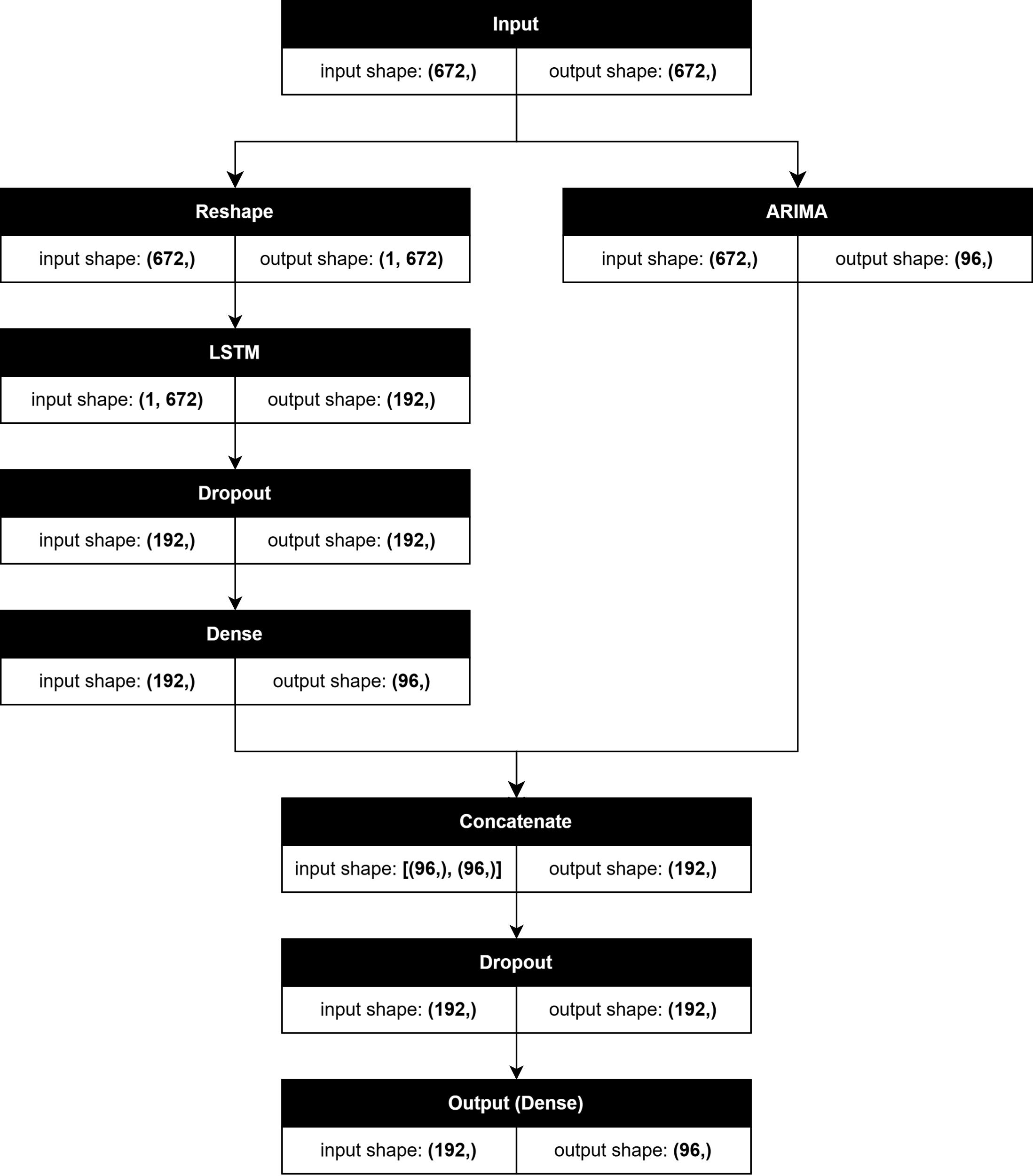
The RLSTM model integrates ARIMA and LSTM by first training an ARIMA model on the time series data. The residuals—differences between the actual values and ARIMA’s predictions—are then used as input for training the LSTM model. The final forecast is obtained by adding the ARIMA model’s predictions to the LSTM model’s predictions of the residuals. This approach aims to capture any non-linear dependencies that ARIMA may not fully model.

### ARIMA-LSTM Pipeline (ALP)

The ALP model utilizes a sequential process where ARIMA model predictions are fed as inputs into the LSTM model. Initially, the ARIMA model is trained and generates predictions based on the time series data. These predictions, along with the original data, are then used as features to train the LSTM model. This allows the LSTM to leverage the linear patterns captured by ARIMA, potentially improving its predictive performance.

### ARIMA-LSTM Augmented (ALA)

The ALA model is similar to the ALP approach but with a key difference: the ARIMA model’s predictions are used to augment the LSTM model’s input features rather than being the sole input. By training the LSTM on both the original time series data and ARIMA’s predictions, this model aims to capture both linear and non-linear patterns more effectively. As shown in **FIGURE 2**, this hybrid architecture combines the strengths of both ARIMA and LSTM models to improve forecasting accuracy.



**FIGURE 2**. Architecture of the ARIMA-LSTM Augmented (ALA) hybrid model

## Evaluation

The performance of each model was evaluated using the Root Mean Squared Error (RMSE), a widely used metric for measuring the accuracy of continuous predictions. RMSE, as shown in equation (1), was calculated across different prediction horizons to assess how well each model performed at forecasting wind power production over short-term intervals.

(1)

# RESULTS AND DISCUSSION

## Performance of Individual Models

The first phase of the study involved evaluating the performance of the ARIMA and LSTM models as standalone predictors for wind power generation. The results show that the average RMSE for ARIMA across all time series and prediction horizons is 0.1929. It was observed that the performance of the ARIMA model deteriorates as the prediction horizon extends, with RMSE increasing consistently. This is a clear indication of ARIMA's diminishing accuracy over longer-term forecasts, which aligns with its well-known limitations in capturing complex, non-linear patterns in time series data.

The LSTM model, on the other hand, produced an average RMSE of 0.1838 across all time series and prediction horizons, outperforming ARIMA overall. Like ARIMA, LSTM also exhibited increased RMSE with longer prediction horizons, though it generally provided more accurate forecasts than ARIMA across all horizons. This suggests that LSTM's ability to model non-linear dependencies and temporal patterns offers a significant advantage in wind power forecasting.

Interestingly, as illustrated in **FIGURE 3**, a more nuanced comparison reveals that ARIMA outperforms LSTM for shorter prediction horizons. Specifically, ARIMA maintains superiority for up to horizon 43 in Onshore-Flanders-Dso, horizon 32 in Onshore-Flanders-Elia, horizon 22 in Offshore-Federal-Elia, horizon 27 in Onshore-Wallonia-Dso, and horizon 17 in Onshore-Wallonia-Elia. On average, ARIMA is the better model for predictions up to a horizon of 28 (7 hours), after which LSTM consistently outperforms ARIMA. This result underlines the utility of ARIMA in very short-term forecasts, but also highlights the need for models that can combine ARIMA's short-term accuracy with LSTM's strength in longer-term predictions.

|  |
| --- |
|  |

**FIGURE 3**. Comparison of ARIMA and LSTM models' RMSE across different prediction horizons

## Performance of Hybrid Models

Given the complementary strengths of ARIMA and LSTM, various hybrid models were explored to see if they could outperform the base models by leveraging the strengths of both. The Simple ARIMA-LSTM Hybrid (SALH), which combines predictions from ARIMA and LSTM using a 50-50 split at horizon 48, demonstrated a notable improvement, yielding an average RMSE of 0.1804 across all series and horizons. Further refinement of SALH using a split at horizon 28 (SALH@28), based on the optimal horizon determined from the base model comparisons, reduced the average RMSE to 0.1788. This performance improvement was consistent across all series except Onshore-Flanders-Dso, where SALH@48 still performed slightly better.

The Machine Learning Combined ARIMA-LSTM (MLCAL) models also showed promising results. The MLCAL model using Linear Regression as the combiner (MLCAL-LR) achieved an average RMSE of 0.1772, better than both base models and SALH. Further improvement was seen with the MLCAL model using a Multi-Layer Perceptron (MLCAL-MLP) as the combiner, which reduced the RMSE to 0.1766. However, not all MLCAL variants performed well; the MLCAL model using KNN (MLCAL-KNN) as the combiner performed poorly, with an average RMSE of 0.1991. This underperformance, worse than both ARIMA and LSTM, underscores the importance of selecting appropriate hyperparameters and combiners in hybrid models.

The Residual LSTM (RLSTM) model, which attempted to capture the non-linear residuals left by ARIMA, produced an average RMSE of 0.1928. This result was only marginally better than ARIMA but significantly worse than LSTM, suggesting that the residuals might not fully capture the non-linear patterns that LSTM alone can model.

Among the hybrid models, the ARIMA-LSTM Pipeline (ALP) and ARIMA-LSTM Augmented (ALA) models demonstrated the most impressive performance. The ALP model, which uses ARIMA predictions as additional inputs for LSTM, achieved an average RMSE of 0.1753. The ALA model, which augments LSTM inputs with ARIMA predictions, further reduced the average RMSE to 0.1737. These results indicate that leveraging ARIMA's linear forecasting capabilities to enhance LSTM's non-linear modeling significantly boosts prediction accuracy, especially in longer horizons.

Overall, as shown in **TABLE 1**, the ALA model emerged as the best performer among all tested approaches, closely followed by the ALP model. This suggests that integrating the strengths of both ARIMA and LSTM through careful model design leads to superior forecasting performance. Other than the MLCAL-KNN and RLSTM models, which did not significantly outperform the base models, all hybrid methods demonstrated improved accuracy, reinforcing the value of hybrid approaches in wind power forecasting.

**TABLE 1**. Average RMSE of each model. The lowest average RMSE value, indicating the best performance, is shown in bold.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Method** | **Series** | | | | | **Average** |
| **Offshore-Federal-Elia** | **Onshore-Flanders-Dso** | **Onshore-Flanders-Elia** | **Onshore-Wallonia-Dso** | **Onshore-Wallonia-Elia** |
| ALA | **0.2545** | **0.1618** | **0.1548** | **0.1559** | 0.1415 | **0.1737** |
| ALP | 0.2549 | 0.1657 | 0.1560 | 0.1570 | 0.1430 | 0.1753 |
| MLCAL-MLP | 0.2573 | 0.1680 | 0.1606 | 0.1561 | **0.1412** | 0.1766 |
| MLCAL-LR | 0.2568 | 0.1715 | 0.1596 | 0.1568 | 0.1415 | 0.1772 |
| SALH@28 | 0.2593 | 0.1716 | 0.1612 | 0.1589 | 0.1432 | 0.1788 |
| SALH@48 | 0.2632 | 0.1710 | 0.1619 | 0.1605 | 0.1455 | 0.1804 |
| LSTM | 0.2635 | 0.1810 | 0.1679 | 0.1624 | 0.1442 | 0.1838 |
| RLSTM | 0.2860 | 0.1778 | 0.1721 | 0.1717 | 0.1562 | 0.1928 |
| ARIMA | 0.2862 | 0.1779 | 0.1722 | 0.1718 | 0.1563 | 0.1929 |
| MLCAL-KNN | 0.2907 | 0.1830 | 0.1798 | 0.1782 | 0.1636 | 0.1991 |

## Impact of Prediction Horizon on Model Performance

The results also reveal interesting insights regarding the impact of prediction horizons on model performance. For very short-term predictions (up to horizon 9, or 2 hours 15 minutes), the MLCAL-LR model performed best on average, suggesting that linear combinations of ARIMA and LSTM predictions are highly effective in the immediate future. However, for longer prediction horizons, the ALA model consistently provided the best performance, with the ALP model following closely behind. This consistent performance at longer horizons highlights the effectiveness of these hybrid approaches in capturing both the immediate linear trends and the more complex, longer-term dependencies in wind power data.

# CONCLUSIONS

This research explored the efficacy of hybrid ARIMA-LSTM models for enhancing short-term wind power production forecasting. By comparing the performance of ARIMA, LSTM, and various hybrid models, we observed that while ARIMA excels in short-term predictions, its accuracy diminishes with longer prediction horizons. LSTM models, although generally more accurate over a broad range of horizons, still struggle with the very short-term forecasts where ARIMA outperforms.

The hybrid models, particularly the ARIMA-LSTM Augmented (ALA) and ARIMA-LSTM Pipeline (ALP) models, demonstrated superior performance by effectively combining the strengths of both ARIMA and LSTM. The ALA model, in particular, achieved the lowest average RMSE, indicating its potential as a powerful tool for wind power forecasting. These findings underscore the value of hybrid modeling approaches in addressing the complexities of wind power generation, where both linear and non-linear patterns must be accurately captured.

While this study has made significant strides in improving wind power forecasting, several avenues for future research remain. Key areas include advanced hyperparameter optimization for both ARIMA and LSTM models, exploration of other hybrid architectures like GRU or Transformer, testing the models on different datasets, incorporating exogenous variables, investigating real-time implementation and scalability, and quantifying prediction uncertainty. By addressing these aspects, future research can further refine wind power forecasting models, enhancing the reliability and efficiency of renewable energy integration into power grids.

# Acknowledgments

This research was supported by Lembaga Penelitian dan Pengabdian Masyarakat (LPPM) Universitas Internasional Semen Indonesia (UISI).

# ReferenceS

1. Z. Gao *et al.*, "An overview on development of wind power generation," in *Chinese Control and Decision Conference (CCDC)*, 2016, pp. 435-439: IEEE.
2. S. Gandhar, J. Ohri, and M. Singh, "A critical review of wind energy based power generation systems," *Asian Journal of Water, Environment and Pollution,* vol. 17, no. 2, pp. 29-36, 2020.
3. M. Arshad and B. O’Kelly, "Global status of wind power generation: theory, practice, and challenges," *International journal of green energy,* vol. 16, no. 14, pp. 1073-1090, 2019.
4. M. Bhaskar, A. Jain, and N. V. Srinath, "Wind speed forecasting: Present status," in *International Conference on Power System Technology*, 2010, pp. 1-6: IEEE.
5. S. Hanifi, X. Liu, Z. Lin, and S. Lotfian, "A critical review of wind power forecasting methods—past, present and future," *Energies,* vol. 13, no. 15, p. 3764, 2020.
6. Y. Wang, R. Zou, F. Liu, L. Zhang, and Q. Liu, "A review of wind speed and wind power forecasting with deep neural networks," *Applied Energy,* vol. 304, p. 117766, 2021.
7. E. Yatiyana, S. Rajakaruna, and A. Ghosh, "Wind speed and direction forecasting for wind power generation using ARIMA model," in *Australasian universities power engineering conference (AUPEC)*, 2017, pp. 1-6: IEEE.
8. E. Grigonytė and E. Butkevičiūtė, "Short-term wind speed forecasting using ARIMA model," *Energetika,* vol. 62, no. 1-2, pp. 45-55, 2016.
9. S. Pasari and A. Shah, "Time series auto-regressive integrated moving average model for renewable energy forecasting," in *Indo-German Conference on Sustainability in Engineering*, 2020, pp. 71-77: Springer International Publishing Cham.
10. P. P. Prabha, V. Vanitha, and R. Resmi, "Wind speed forecasting using long short term memory networks," in *International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT)*, 2019, vol. 1, pp. 1310-1314: IEEE.
11. D. G. Taslim and I. M. Murwantara, "A Comparative Study of ARIMA and LSTM in Forecasting Time Series Data," in *International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE)*, 2022, pp. 231-235: IEEE.
12. D. Kobiela, D. Krefta, W. Król, and P. Weichbroth, "ARIMA vs LSTM on NASDAQ stock exchange data," *Procedia Computer Science,* vol. 207, pp. 3836-3845, 2022.
13. V. Bali, A. Kumar, and S. Gangwar, "A novel approach for wind speed forecasting using LSTM-ARIMA deep learning models," *International Journal of Agricultural and Environmental Information Systems (IJAEIS),* vol. 11, no. 3, pp. 13-30, 2020.
14. S. Sajid and S. R. Salkuti, "Short term wind speed forecasting using time series techniques," *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects,* vol. 44, no. 4, pp. 9861-9881, 2022.
15. Z. Hajirahimi and M. Khashei, "Hybrid structures in time series modeling and forecasting: A review," *Engineering Applications of Artificial Intelligence,* vol. 86, pp. 83-106, 2019.
16. R. J. Hyndman and Y. Khandakar, "Automatic time series forecasting: the forecast package for R," *Journal of statistical software,* vol. 27, pp. 1-22, 2008.
17. L. N. Smith, "Cyclical learning rates for training neural networks," in *Winter conference on applications of computer vision (WACV)*, 2017, pp. 464-472: IEEE.