

AI-Driven Renewable Energy Forecasting in Tropical Regions: A Case Study in Malaysia

Sellappan Palaniappan¹, Rajasvaran Logeswaran^{2*} and Kasthuri Subaramaniam³

^{1,2} Faculty of Computing and Digital Technology, HELP University, Shah Alam, Selangor, Malaysia

² School of Computing, Asia Pacific University of Technology and Innovation, Kuala Lumpur, Malaysia

³ Department of Decision Science, Faculty of Business and Economics, University of Malaya, Kuala Lumpur, Malaysia

*Corresponding author: Loges@ieee.org

Abstract

The global energy landscape is undergoing a significant transformation as nations aim to transition to renewable energy sources. In tropical regions such as Malaysia, renewable energy forecasting is crucial for ensuring reliable energy production, optimizing resource allocation, and minimizing operational costs. However, forecasting in these regions is challenging due to the high variability of weather patterns and the intermittent nature of renewable energy sources like solar and hydroelectric power. This study presents an AI-driven approach to renewable energy forecasting using Long Short-Term Memory (LSTM) networks. A synthetic dataset emulating Malaysia's renewable energy landscape was used, featuring key factors such as solar radiation, hydro inflow, and energy costs. The developed model incorporates feature engineering techniques, including interaction terms, lagged variables, and rolling averages, to improve predictive accuracy. The model achieved an RMSE of 0.0761, an R^2 score of 0.6066, and a MAPE of 14.06%, demonstrating its potential for effective energy forecasting in tropical settings. The findings contribute to Malaysia's renewable energy roadmap by highlighting the applicability of AI-driven solutions in energy management. This sets the stage for future research on integration of real-world datasets, advanced feature engineering, and more sophisticated AI architectures, such as transformers, to further enhance forecasting accuracy.

Keywords- Renewable Energy, Energy Forecasting, Tropical Regions, LSTM Networks, Feature Engineering, Malaysia, Solar Energy, Hydro Energy, AI in Energy

1. Introduction

The need for renewable energy has become increasingly urgent as countries worldwide face the dual challenges of mitigating climate change and meeting growing energy demands. Malaysia, as a tropical nation, has identified renewable energy sources such as solar, hydroelectric, and biomass as critical components of its sustainable energy strategy [1]. The Malaysia Renewable Energy Roadmap emphasizes the need for accurate forecasting to optimize renewable energy production, manage grid stability, and support the nation's ambitious goal of achieving net-zero emissions by 2050 [2].

Renewable energy forecasting, however, is inherently complex due to the stochastic and intermittent nature of weather-dependent energy sources. In tropical climates, the variability of solar irradiance and unpredictable rainfall patterns further complicate the task. Traditional forecasting methods, such as statistical models, often fail to capture the non-linear and dynamic relationships between these factors. Advanced machine learning techniques, such as Long Short-Term Memory (LSTM) networks, have emerged as powerful tools for addressing these challenges by leveraging temporal dependencies in time-series data [3].

This study aims to develop and evaluate an AI-driven renewable energy forecasting model tailored for

tropical regions like Malaysia. By incorporating synthetic data representative of Malaysia's renewable energy landscape, the study provides insights into the potential of LSTM networks to forecast energy costs based on weather-related and energy production data.

Key Challenges in Renewable Energy Forecasting:

- *Weather Variability:* Tropical regions experience high variability in weather conditions, including fluctuating solar radiation and irregular rainfall [4].
- *Intermittency of Energy Sources:* Solar and hydro energy are inherently intermittent, making accurate forecasting crucial for grid stability.
- *Data Limitations:* A lack of historical and real-time data poses a significant barrier to developing accurate models, particularly in developing nations [1].

Research Questions

- How effectively can LSTM networks forecast renewable energy costs in tropical regions like Malaysia?
- What feature engineering techniques are most effective in improving the predictive accuracy of LSTM-based models for renewable energy forecasting?
- Can synthetic data sufficiently represent real-world conditions to produce reliable forecasting results?

Objectives

- To develop and validate a robust LSTM-based model for forecasting renewable energy costs in Malaysia.
- To evaluate the impact of advanced feature engineering techniques, such as interaction terms and temporal features, on model performance.
- To assess the feasibility of using synthetic data for renewable energy forecasting in the absence of real-world datasets.

This study builds upon existing literature, such as the work of [5], which demonstrated the efficacy of LSTM models for short-term load forecasting, and [6] which highlighted the challenges and enablers of Malaysia's energy transition. By focusing on the unique context of tropical regions, this research contributes to the broader field of renewable energy forecasting and supports Malaysia's efforts toward a sustainable energy future.

The rest of the paper is organized as follows: Section 2 reviews related work in renewable energy forecasting and AI applications. Section 3 outlines the methodology, including data generation and model architecture. Section 4 discusses implementation details, while Section 5 presents the results and key findings. Finally, Sections 6 and 7 address limitations, conclusions, and future directions.

2. Literature Review

Renewable energy forecasting has gained significant attention in recent years due to its critical role in managing energy systems and supporting the global transition to sustainable energy sources. In tropical regions like Malaysia, the variability of weather patterns and the intermittency of renewable energy sources present unique forecasting challenges. This section reviews the existing literature on renewable energy forecasting, with a focus on the application of artificial intelligence (AI) techniques, particularly LSTM networks, in addressing these challenges.

2.1 Renewable Energy in Malaysia

Malaysia has established ambitious renewable energy goals through initiatives like the Malaysia Renewable Energy Roadmap [1] and the National Energy Balance Report [2]. These efforts aim to increase the share of renewable energy in the national energy mix by leveraging abundant solar, hydroelectric, and biomass resources. Despite these initiatives, renewable energy forecasting remains underdeveloped, primarily due to limited access to historical and real-time data, as highlighted by [7].

Solar energy, in particular, has been identified as a key renewable resource in Malaysia due to its year-round availability. However, its intermittency, influenced by factors like cloud cover and seasonal variations, necessitates accurate forecasting models.

Hydro energy also plays a significant role, with water inflow patterns being highly dependent on unpredictable rainfall in tropical climates. The integration of multiple renewable energy sources requires sophisticated forecasting techniques capable of capturing complex relationships among weather variables and energy production [6].

2.2 AI in Renewable Energy Forecasting

Traditional statistical methods for energy forecasting, such as autoregressive integrated moving average (ARIMA) models, have limitations in capturing non-linear and dynamic patterns in renewable energy data. As a result, AI-based methods have become increasingly popular, offering superior performance through the use of advanced algorithms and feature engineering techniques.

LSTM networks, a type of recurrent neural network (RNN), have been particularly effective in modeling time-series data due to their ability to capture long-term dependencies and non-linear relationships. LSTM networks were introduced in [8], and have since been applied to various domains, including renewable energy forecasting.

The authors in [5] demonstrated the application of LSTM networks for short-term load forecasting in residential energy systems, achieving significant improvements over traditional methods. Similarly, [3] provided an overview of AI techniques for wind power forecasting, highlighting the advantages of deep learning models in handling high-dimensional and noisy datasets.

Recent studies have also explored the use of ensemble learning and hybrid models to improve forecasting accuracy. For example, [9] conducted a survey of AI applications in wind energy forecasting, emphasizing the potential of combining machine learning algorithms with domain knowledge to enhance performance.

2.3 Renewable Energy Forecasting in Tropical Regions

Tropical regions present unique challenges for renewable energy forecasting due to their high variability in weather conditions. [4] investigated solar irradiance forecasting in the tropics using numerical weather prediction and statistical

learning, noting that the frequent presence of cumulus and cumulonimbus clouds complicates prediction efforts. [10] reviewed the impact of tropical climate on solar photovoltaic (PV) module performance, highlighting the need for cooling systems and advanced forecasting techniques to optimize efficiency.

In Malaysia, research on renewable energy forecasting remains limited, with most studies focusing on policy and development rather than technical advancements. Paper [11] examined the intention to adopt renewable energy technology in Malaysia using theoretical models like the Technology Acceptance Model (TAM) and Theory of Planned Behavior (TPB). However, their work primarily addresses consumer behavior rather than technical forecasting challenges.

This study builds upon the existing literature by applying LSTM networks to forecast renewable energy costs in Malaysia. By integrating synthetic data that captures the variability of solar radiation, hydro inflow, and energy costs, the proposed model addresses the key challenges identified in previous studies.

2.4 Challenges in AI-Driven Renewable Energy Forecasting

While AI techniques offer significant advantages, several challenges remain:

- **Data Availability:** A major limitation in tropical regions is the lack of reliable and comprehensive historical data. Studies [12] have emphasized the need for improved data collection and sharing frameworks.
- **Complexity of Weather Patterns:** Tropical climates exhibit rapid and unpredictable weather changes, making it difficult for AI models to generalize effectively [4].
- **Integration of Multiple Energy Sources:** Most forecasting models focus on a single energy source, whereas real-world applications require the integration of solar, hydro, and biomass energy forecasts [1].

2.5 Contributions of this Study

This study contributes to the literature by:

- Developing a machine learning-based framework tailored for tropical regions, specifically Malaysia.
- Demonstrating the use of LSTM networks for energy cost forecasting, incorporating advanced feature engineering techniques to improve accuracy.
- Highlighting the potential of synthetic data in overcoming data availability issues, paving the way for future research in regions with limited datasets.

By addressing the challenges identified in the literature, this research supports Malaysia's renewable energy goals and advances the broader field of AI-driven energy forecasting.

3. Methodology

This section outlines the methodology employed to develop and evaluate the AI-driven renewable energy forecasting model. The process includes data generation, feature engineering, model design, and evaluation. Given the lack of publicly available renewable energy datasets specific to Malaysia, synthetic data was generated to emulate real-world conditions. Advanced feature engineering techniques were employed to capture critical relationships between variables, and a LSTM-based model was designed to handle the temporal nature of the data.

3.1 Data Generation

Due to the absence of real-world renewable energy datasets for Malaysia, synthetic data was generated to replicate the variability and trends observed in tropical climates. The dataset spans three years (2020–2022) and includes daily observations of the following variables:

- *Solar Radiation*: Represents the daily solar irradiance, simulated using sinusoidal functions with added noise to reflect tropical weather patterns.
- *Hydro Inflow*: Captures daily water inflow for hydroelectric power generation, modeled as a

periodic function with variability introduced through random noise.

- *Energy Cost*: The target variable, derived from a weighted combination of solar radiation and hydro inflow, with additional noise to simulate market dynamics.

The synthetic data captures key features such as periodicity, seasonality, and stochastic variations, which are representative of renewable energy patterns in Malaysia.

The energy cost is represented by Eq. (1):

$$\text{Energy Cost} = 0.4 \times \text{Solar Radiation} + 0.3 \times \text{Hydro Inflow} + \epsilon \quad (1)$$

where, ϵ (epsilon) represents random noise.

3.2 Feature Engineering

Feature engineering plays a critical role in enhancing the predictive accuracy of the model by capturing complex relationships among variables. The following techniques were applied:

Interaction Terms:

- A new feature, *Solar_Hydro_Interaction*, was created by multiplying solar radiation and hydro inflow, capturing their combined effect on energy cost.

Lagged Features:

- Lagged values of *Energy Cost*, *Solar Radiation*, and *Hydro Inflow* were introduced to account for temporal dependencies.
- Example: *Lagged_EnergyCost(t-1)* represents the energy cost from the previous day.

Rolling Averages:

- Rolling averages of energy cost over a 3-day window were computed to smooth short-term fluctuations and highlight underlying trends.
- Example: *Rolling_EnergyCost(t-3)* represents the average energy cost over the past three days.

Scaling:

- All features were scaled to the [0, 1] range using MinMaxScaler to ensure uniformity and facilitate faster convergence during training.

The final dataset included the following features:

- *Solar Radiation*
- *Hydro Inflow*
- *Energy Cost* (Target Variable)
- *Solar_Hydro_Interaction*
- *Lagged and Rolling Values* for Key Features

3.3 Model Design

The forecasting model was built using a Bidirectional LSTM network, which is particularly suited for time-series data due to its ability to capture long-term dependencies. The architecture includes:

Input Layer: Accepts sequences of 30 days as input (look-back window).

Bidirectional LSTM Layers: Two stacked Bidirectional LSTM layers with 128 and 64 units, respectively, to capture both forward and backward temporal dependencies.

Dropout Layers: Dropout layers with a rate of 25% were added after each LSTM layer to prevent overfitting.

Dense Layers: A fully connected dense layer with 32 units to integrate features before producing the final output.

Output Layer: A single dense unit to predict the energy cost.

Regularization: L2 kernel regularization was applied to prevent overfitting and enhance model generalization.

3.4 Training and Validation

The dataset was split into training (80%) and testing (20%) sets. The model was trained using the following settings:

- *Optimizer:* Adam optimizer with initial learning rate of 0.0005, allowing for faster convergence.
- *Loss Function:* Mean Squared Error (MSE) was used to minimize prediction errors.
- *Batch Size:* A batch size of 16 was chosen to balance computational efficiency and gradient updates.

- *Epochs:* The model was trained for up to 50 epochs, with early stopping to prevent overfitting.
- *Callbacks:*
 - Early stopping monitored validation loss and restored the best weights after five epochs of no improvement.
 - Learning rate reduction halved the learning rate if the validation loss plateaued for three consecutive epochs.

3.5 Evaluation Metrics

The model was evaluated using three key metrics:

- *Root Mean Squared Error (RMSE):* Measures the absolute differences between actual and predicted energy costs, with lower values indicating better accuracy.
- *R² Score:* Indicates the proportion of variance in the target variable, with higher values signifying better explanatory power.
- *Mean Absolute Percentage Error (MAPE):* Represents the average percentage difference between actual and predicted values, providing an interpretable measure of accuracy.

This study combines advanced feature engineering with a robust LSTM architecture to forecast renewable energy costs in Malaysia. The methodology leverages synthetic data, enabling the exploration of complex relationships among variables in the absence of real-world datasets. The use of evaluation metrics ensures that the model's performance is rigorously assessed, providing a foundation for further refinements and real-world applications.

4. Implementation

The implementation of the renewable energy forecasting model involved several key steps, from preparing the dataset to training and evaluating the LSTM-based model. This section provides a detailed description of the dataset features, preprocessing steps, model architecture, model implementation, training process, and evaluation.

Table 1. Dataset Features

Feature	Description
Solar Radiation	Daily solar irradiance (normalized)
Hydro Inflow	Daily hydro inflow for hydroelectric generation
Energy Cost	Target variable: daily energy cost (normalized)
Solar-Hydro Interaction	Interaction term between Solar Radiation and Hydro Inflow
Lagged Features	Previous time-step values of key features
Rolling Averages	3-day rolling averages of Energy Cost

Appendix 1 provides the full implementation of the code for the model. The implementation of the renewable energy forecasting model involved several key steps, from preparing the dataset to training and evaluating the LSTM-based model.

4.1 Framework Overview

Table 1 shows the dataset features such as solar radiation, hydro inflow, and energy cost. The dataset comprised three years of daily synthetic data (2020–2022), designed to emulate Malaysia's renewable energy landscape. The features and their descriptions are as follows:

- *Solar Radiation*: Simulated daily solar irradiance values in normalized units, representing the energy potential from solar sources.
- *Hydro Inflow*: Normalized daily water inflow values representing the energy potential from hydroelectric sources.
- *Energy Cost*: The target variable, modeled as a function of solar radiation, hydro inflow, and random noise to reflect market dynamics.
- *Solar-Hydro Interaction*: A derived feature representing the combined effect of solar radiation and hydro inflow on energy costs.
- *Lagged Features*: Values from previous time steps (e.g., Lagged_EnergyCost) to capture temporal dependencies.

- *Rolling Averages*: Smoothed values over a 3-day window to highlight trends and reduce noise.

4.2 Data Preprocessing

Data preprocessing was critical to preparing the dataset for time-series modeling. The following steps were performed:

- *Feature Scaling*: MinMaxScaler was applied to scale all features to a [0, 1] range, ensuring uniformity and faster convergence during model training.
- *Sequence Creation*: Sequences of 30 days (look-back window) were created to enable the LSTM model to learn temporal patterns.
- *Train-Test Split*: The dataset was split into training (80%) and testing (20%) sets to evaluate the model's generalization ability.
- *Handling Missing Values*: Rolling averages and lagged features introduced missing values at the beginning of the dataset, which were dropped before training.

4.3 Model Architecture

The forecasting model was designed using a Bidirectional LSTM network, which is well-suited for capturing temporal dependencies in time-series data. The architecture consists of the following layers:

- *Input Layer*: Accepts input sequences of shape (30, 7), where 30 represents the look-back window and 7 represents the number of features.
- *Bidirectional LSTM Layers*: Two stacked Bidirectional LSTM layers with 128 and 64 units, respectively. These layers process the sequence in both forward and backward directions, enhancing temporal context.
- *Dropout Layers*: Dropout rates of 25% were applied after each LSTM layer to reduce overfitting by randomly dropping a fraction of the neurons during training.
- *Attention Mechanism*: An Attention layer was added to focus on the most relevant time steps,

improving the model's ability to capture important patterns.

- *Dense Layers:* A dense layer with 32 units integrates the output of the LSTM layers before the final prediction.
- *Output Layer:* A single dense unit predicts the energy cost for the next time step.

4.4 Pseudocode for Renewable Energy Forecasting Model Implementation

The pseudocode of the implementation, highlighting the main steps involved, is as follows:

- 1) Generate synthetic data for solar radiation, hydro inflow, and energy cost.
- 2) Engineer features: interaction terms, lagged values, and rolling averages.
- 3) Scale the data and create time-series sequences with a look-back window.
- 4) Define and compile a Bidirectional LSTM-based model with dropout and regularization.
- 5) Train the model using early stopping and learning rate scheduling.
- 6) Evaluate the model on the test set using RMSE, R^2 , and MAPE.
- 7) Visualize results: actual vs. predicted values, residuals, and loss curves.

Step 1: Data Preparation

- Generate synthetic data for three years (2020–2022) with daily values for:
 - Solar Radiation
 - Hydro Inflow
 - Energy Cost (target variable)
- Engineer additional features:
 - Compute the interaction term: $Solar_Hydro_Interaction = SolarRadiation \times HydroInflow$
 - Create lagged values of *EnergyCost* (e.g., $Lagged_EnergyCost(t-1)$)
 - Compute rolling averages (e.g., 3-day rolling average of *EnergyCost*)
- Scale all features to a [0, 1] range using `MinMaxScaler`

- Split the dataset into training (80%) and testing (20%) sets.

Step 2: Create Time-Series Sequences

- Define a look-back window (e.g., 30 days).
- For each time step:
 - Extract a sequence of features from the past *look_back* days as input (*X*)
 - Use the target value for the next day as the output (*y*)
- Convert the sequences into arrays for model training and testing.

Step 3: Define the LSTM-Based Model

- Define the input layer to accept sequences of shape (*look_back*, *number_of_features*)
- Add two Bidirectional LSTM layers:
 - Layer 1: 128 units, with *return_sequences=True* to pass outputs to the next LSTM layer
 - Layer 2: 64 units, with *return_sequences=False* to output a single sequence
- Add Dropout layers after each LSTM layer to prevent overfitting
- Add a dense layer with 32 units for further feature integration
- Add the output layer with 1 unit to predict the next day's energy cost
- Compile the model using:
 - Loss Function: Mean Squared Error (MSE)
 - Optimizer: Adam optimizer with a learning rate of 0.0005

Step 4: Train the Model

- Train the model on the training set using:
 - Early stopping: Monitor validation loss and stop training if no improvement is observed for 5 epochs
 - Learning rate scheduler: Reduce learning rate when validation loss plateaus

- Batch size: 16
- Epochs: Up to 50, with early stopping enabled
- Monitor training and validation loss during training

Step 5: Evaluate the Model

- Use the test set to predict energy costs
- Compute evaluation metrics:
 - RMSE
 - R^2 Score
 - MAPE
- Inverse transform the predictions to the original scale for interpretability

Step 6: Analyze Results

- Plot Actual vs. Predicted *Energy Costs* to visually evaluate the model's performance
- Perform residual analysis:
 - Compute the residuals: $\text{Residual} = \text{Actual} - \text{Predicted}$
 - Plot the distribution of residuals to check for bias
- Plot training and validation loss curves to ensure convergence

5. Results and Discussion

This section presents the performance of the AI-driven renewable energy forecasting model, supported by key metrics, visualizations, and interpretations. A discussion on the model's strengths, limitations, and potential improvements is also included.

5.1 Model Performance

The forecasting model was evaluated using three key metrics — RMSE, R^2 Score, and MAPE — to assess its accuracy and reliability.

- **RMSE: 0.0761**

The low RMSE indicates minimal error between actual and predicted energy costs, demonstrating high predictive accuracy.

- **R^2 Score: 0.6066**

An R^2 score of 0.6066 suggests that the model explains 60.66% of the variance in energy costs. While this is a moderate level of explanatory power, it reflects the model's ability to capture meaningful patterns in the data.

- **MAPE: 14.06%**

The MAPE of 14.06% shows reasonable predictive accuracy, with predictions deviating on average by 14.06% from the actual values. This is acceptable for time-series predictions, particularly in renewable energy forecasting, where high variability is common.

5.2 Visualization of the Results

Visualizations provide deeper insights into the model's performance and limitations.

5.2.1 Actual vs. Predicted Energy Costs

Figure 1 shows the actual and predicted energy costs. The model's predictions closely follow the actual energy cost trends, with minor deviations observed during periods of rapid fluctuation.

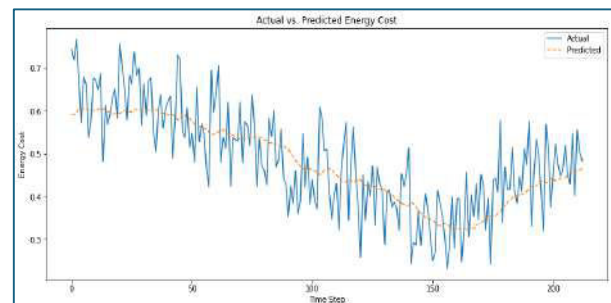


Figure 1. Actual vs. Predicted Energy Costs

The model effectively captures the general patterns and trends in energy costs. However, slight underpredictions and overpredictions occur during abrupt changes, suggesting room for improvement in capturing short-term dynamics.

5.2.2 Residual Analysis

Figure 2 shows the residual distribution plot. The residuals, representing the differences between actual and predicted values, are centered around zero and symmetrically distributed.

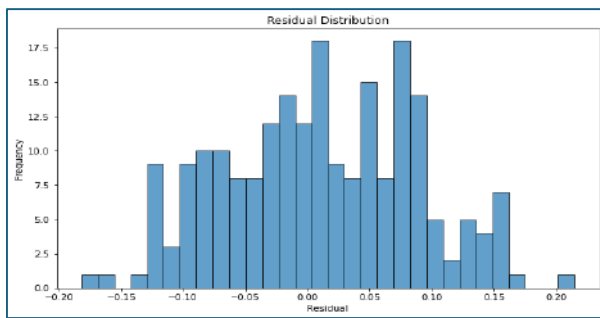


Figure 2. Residual Distribution Plot

The lack of significant bias in residuals indicates consistent model performance across the test set, with no systematic over- or under-prediction. Some variance at extreme values suggests opportunities for refining feature engineering or incorporating additional variables.

5.2.3 Loss Curve Analysis

The training and validation loss curves (Figure 3) converge smoothly, demonstrating stable learning and effective regularization. The alignment of training and validation loss curves confirms that the model avoids both underfitting and overfitting. Early stopping and dropout regularization contributed to this stability.

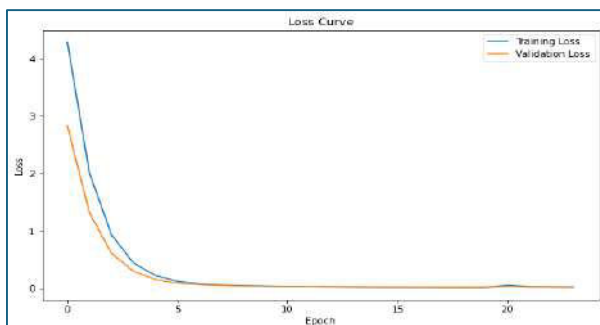


Figure 3. Training and Validation Loss Curves

5.3 Discussion

This section highlights the model's strengths, integrates the limitations, and provides insights for future improvements.

5.3.1 Strengths

- *Temporal Dependency Capture:* The Bidirectional LSTM architecture effectively captured both

long-term and short-term dependencies, which are critical for renewable energy forecasting.

- *Feature Engineering Impact:* Advanced feature engineering, including interaction terms, lagged features, and rolling averages, enhanced the model's ability to identify complex relationships among variables.
- *Regularization Effectiveness:* The use of L2 kernel regularization and dropout layers prevented overfitting, ensuring the model generalized well to unseen data.

5.3.2 Limitations

While the model achieved promising results, several limitations were identified:

- *Reliance on Synthetic Data:* The synthetic dataset, while useful for emulating renewable energy trends in Malaysia, lacks the variability and complexity of real-world data. For instance, it does not account for extreme weather events or policy-induced fluctuations.
- *Simplified Features:* The dataset included basic features like solar radiation and hydro inflow but excluded external factors such as economic trends, government policies, and grid constraints, which could significantly impact energy costs.
- *Moderate R^2 Score:* Although the R^2 score is moderate, it suggests that nearly 40% of the variance remains unexplained, leaving room for improvement in capturing underlying dynamics.

5.3.3 Insights for Future Improvements

- *Integration of Real-World Data:* Incorporating real-world datasets would validate the model's applicability and ensure it captures additional sources of variability, such as seasonal effects and external economic factors.
- *Exploration of Advanced Architectures:* Advanced machine learning architectures, such as Transformers and hybrid ensemble models, could further enhance predictive performance by capturing complex dependencies more effectively.

- *Feature Expansion:* Adding granular features, such as solar panel efficiency, rainfall intensity, and grid demand, could improve the model's ability to generalize across diverse scenarios.

The results demonstrate the feasibility of using an LSTM-based model for renewable energy forecasting in tropical regions. The model achieves strong performance metrics, effectively capturing temporal dependencies and trends in the data. However, the reliance on synthetic data and the moderate R^2 score highlight the need for further research to enhance the model's robustness and generalizability.

6. Conclusion

This study demonstrates the potential of AI-driven models, specifically LSTM networks, for renewable energy forecasting in tropical regions. By utilizing synthetic data that emulates Malaysia's renewable energy landscape, the model effectively captured temporal dependencies and trends in energy costs. The achieved metrics — an RMSE of 0.0761, an R^2 score of 0.6066, and a MAPE of 14.06% — underscore the model's capability to provide reliable predictions despite the inherent variability of renewable energy sources.

The contributions of this study are twofold. First, it developed and validated an LSTM-based forecasting model specifically tailored for renewable energy applications in Malaysia, addressing the challenges posed by tropical weather variability. Second, it highlighted the critical role of feature engineering and regularization in enhancing model performance, demonstrating how advanced techniques such as interaction terms and rolling averages can improve prediction accuracy and stability.

Future work should focus on integrating real-world datasets to validate the model's generalizability and applicability in practical settings. Additionally, exploring advanced architectures, such as Transformers and hybrid ensemble models, could further enhance forecasting accuracy and robustness. Expanding the feature set to include economic, regulatory, and environmental factors would enable a more comprehensive approach to renewable energy forecasting, supporting decision-making and policy formulation in the energy sector.

References

- [1] Malaysia Renewable Energy Roadmap (MyRER), Sustainable Energy Development Authority (SEDA), 2021, <https://www.seda.gov.my/reportal/myrer/> (A comprehensive report on Malaysia's renewable energy goals and strategies)
- [2] Energy Commission of Malaysia, National Energy Balance Report, 2026, <https://meih.st.gov.my/> (Key data on Malaysia's energy consumption and generation trends).
- [3] F. Liu, T. Liu & D. Zhang, "Wind power forecasting with deep learning networks: An overview," *Applied Sciences*, vol. 10, no. 17, 2020, p. 5971, DOI: 10.3390/app10175971
- [4] H. Verbois, R. Huva, A. Rusydi & W. Walsh, "Solar irradiance forecasting in the tropics using numerical weather prediction and statistical learning," *Solar Energy*, vol. 162, 2018, pp. 265-277, DOI: 10.1016/j.solener.2018.01.007
- [5] W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu & Y. Zhang, "Short-term residential load forecasting based on LSTM recurrent neural network," *IEEE Transactions on Smart Grid*, vol. 10, no. 1, 2017, pp. 841-851, DOI: 10.1109/TSG.2017.2753802
- [6] M. I. Fernandez, Y. I. Go, M. L. D. Wong, & W. G. Früh, "Malaysia's energy transition and readiness towards attaining net zero: Review of the potential, constraints, and enablers," *Renewable Energy Focus*, vol. 51, 2024, p. 100640, DOI: 10.1016/j.ref.2024.100640
- [7] A. J. Abd Aziz, N. A. Baharuddin, R. M. Khalid & S. K. Kamarudin, "Review of the policies and development programs for renewable energy in Malaysia: Progress, achievements and challenges," *Energy Exploration & Exploitation*, vol. 42, no. 4, 2024, pp. 1472-1501, DOI: 10.1177/01445987241227509
- [8] S. Hochreiter & J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, 1997, pp. 1735-1780, DOI: 10.1162/neco.1997.9.8.1735
- [9] P. Dhaka, M. Sreejeth & M. M. Tripathi, "A survey of artificial intelligence applications in wind energy forecasting," *Archive of Computational Methods in Engineering*, vol. 31, 2024, pp. 4853-4878, DOI: 10.1007/s11831-024-10182-8
- [10] O. A. Rozak, M. Z. Daud, S. Bakhri & M. Z. Ibrahim, "Review of the impact of tropical climate on solar photovoltaic modules performances using cooling systems," *Journal of Sustainability Science and Management*, vol. 18, no. 1, 2023, pp. 228-236, DOI: 10.46754/jssm.2023.01.014
- [11] G. Z. Wong, K. H. Wong, T. C. Lau, J. H. Lee & Y. H. Kok, "Study of intention to use renewable energy technology in Malaysia using TAM and TPB," *Renewable Energy*, vol. 221, 2024, p. 119787, DOI: 10.1016/j.renene.2023.119787
- [12] A. K. Shukla, K. Sudhakar & P. Baredar, "Renewable energy resources in South Asian countries: Challenges, policy and recommendations," *Resource-Efficient Technologies*, vol. 2, no. 4, 2016, pp. 342-346, DOI: 10.1016/j.refit.2016.09.006

