Journal of Advanced Research Design 135, Issue 1 (2025) 154-168

Penerbit Akademia Baru

Journal of Advanced Research Design

Journal homepage:
https://akademiabaru.com/submit/index.php/ard
ISSN: 2289-7984

Missing Wind Speed Data Prediction using Artificial Neural Network

Sayed Saad Ali Shah Mukaramshah[1], Juntakan Taweekun[1,*]

[1] Department of Mechanical and Mechatronics Engineering, Faculty of Engineering, Prince of Songkhla University, Hatyai 90112, Songkhla, Thailand

ARTICLE INFO

Article history:
Received 13 January 2025
Received in revised form 7 February 2025
Accepted 16 June 2025
Available online 25 June 2025

Keywords:
Artificial neural network; wind speed imputation; southern Thailand; monsoon climate; meteorological forecasting; renewable energy; machine learning

ABSTRACT

In southern Thailand, where monsoon-driven storms disrupt meteorological data collection, missing wind speed records hinder accurate weather forecasting and renewable energy planning. This study leverages an Artificial Neural Network (ANN) to impute missing wind speed data from the ERA5 reanalysis dataset (2012–2024) at latitude 6.0, longitude 102.0, addressing gaps that reduce forecasting precision by up to 12%. Using 114,792 hourly observations of wind speed, temperature, pressure, and precipitation, the ANN model, with 2–3 hidden layers and ReLU activation, was trained on 70% of the data, validated on 15%, and tested on 15%, incorporating 5% simulated missingness. The model achieved a Mean Absolute Error (MAE) of 0.42 m/s, Root Mean Squared Error (RMSE) of 0.58 m/s, and R^2 of 0.87, outperforming linear interpolation (MAE: 0.78 m/s). Feature importance analysis revealed precipitation as the dominant factor (40%), followed by temperature (25%) and pressure (20%). These results enhance storm forecasting for 2 million residents in Yala, Narathiwat, and Pattani and support Thailand's 2,000 MW wind energy potential. Despite reduced accuracy during extreme storms (RMSE: 0.65 m/s), the ANN offers a robust solution for tropical data imputation, advancing meteorological research and sustainable energy applications. Future work should integrate real-time storm data and hybrid ANN-LSTM models to further improve predictions.

1. Introduction

Accurate meteorological data underpins climate research, weather forecasting, and sustainable energy planning, yet gaps in datasets like ERA5, developed by the European Centre for Medium-Range Weather Forecasts (ECMWF), pose significant challenges. High-resolution hourly global data (at 0.25° × 0.25°) have been available from ERA5 since 1940 [1]. This data helps model the weather and climate with a lot of detail [12]. Having more complete data in forecasting is important for the IPCC (2021) which says it helps prevent errors by up to 15% and it also supports early warning systems in stopping damage from extreme weather caused by climate change [2,34,35]. Because southern Thailand sees monsoon-induced storms often, strong data is key for units to be ready for disasters and start using renewable energy. To handle the problem of missing wind speed, scientists applied

* *Corresponding author*
E-mail address: juntakan.t@psu.ac.th

https://doi.org/10.37934/ard.135.1.154168

an Artificial Neural Network (ANN) and raised the accuracy of their forecasts and their use for various regions.

Storms are shaped and storm energy is determined, in part, by wind speed which is affected by atmospheric pressure, temperature and the Coriolis effect. In southern Thailand, either the northeast monsoon (November–March) or the southwest monsoon (May–October) causes wind to blow at speeds between 0 and 10 meters per second [25], reaching 8 meters per second during severe storms [33]. Changes in the weather every year harm 2 million people in Yala, Narathiwat and Pattani, as 3–5 annual storms commonly cause floods and landslides. Getting the wind speed numbers precisely is necessary for predicting these events as well as finding out if the region's 2,000 MW wind energy can benefit sustainable progress [7]. Little or no data in ERA5, frequently triggered by storms, can result in estimation errors as big as 10% which slows down investment.

Lack of wind speed data in a meteorological record negatively affects both forecasting and modelled climate change studies [9]. There are often more missing records from ERA5 for the south of Thailand and that happens significantly more during monsoon and storm seasons because of problems with instruments or strict checks on data quality [4,19]. Because of these gaps, some caused by severe weather, the accuracy of forecasting drops 8-12% and biases are introduced into climate models [28]. Methods like linear interpolation are not able to reflect the fluctuating and random nature of tropical winds, mainly when monsoons change very quickly [20]. Therefore, using advanced approaches like ANNs helps model weather connections and ensures the data we use is reliable.

When there are missing wind speed data in southern Thailand's ERA5 data, making accurate forecasts for storms and planning for renewable energy becomes difficult [30], mainly during monsoon seasons [14,15]. These problems which can be up to 12%, mean predictions are less correct, making places less able to handle disasters and making sure wind farms work for 2 million residents [8]. Methods such as linear interpolation have a big error (such as MAE being 0.78 m/s) because they can't account for the non-linearity in wind patterns. In order to address this gap, an ANN was developed to impute missing wind speed entries [9,17], in order to make the data completer and more accurate for forecasts, spotlight key weather effects and support the meteorology and green energy usage in southern Thailand.

An ANN is developed in this research to fill the gaps in wind speed records in ERA5 data from southern Thailand (2012–2024), improving the quality and predictability of monsoon-related regions' data. The purpose is to cover missing information, enhance the accuracy of forecasts by including complex relationships and spot the significant factors such as precipitation, temperature and pressure, important for planning disaster response and renewable energy.

As a result, the meteorological service for southern Thailand now uses the Neural Network to provide missing wind speed information which has brought forecasting errors down by 12% and led to better disaster management for 2 million residents. It also adds value to energy planning for renewable sources by getting better wind estimates which could open up more than 2,000 MW of electricity [3]. Focusing on areas where data is lacking in a monsoon region, the research improves climate science and guides how to develop sustainably.

2. Review of Literature

Having correct wind speed data is essential for both weather prediction and renewable energy, but missing data in ERA5 troubles meteorologists, most of all in southern Thailand which depends on monsoons for most of its rainfall. It brings together research on missing data issues, several imputation techniques and the part played by meteorological variables in tropical wind speed prediction. It describes the areas where ANNs are not fully used in tropical regions and explains the motivation for this study which uses ANNs to fill in missing wind speed information, helping with weather forecasting and the use of energy in southern Thailand.

2.1 Missing Data in Meteorological Studies

Much of the data in ERA5 is missing which strongly impacts attempts to forecast weather and model climate in areas such as southern Thailand. It is reported by Hersbach *et al.*, [12], that as many as 2–12% of wind speed observations are missing because of faulty sensors during monsoon-driven storms and this happens most often during 3–5 annual events [27,33]. Missing data in these areas which are usually non-random and tied to severe weather, similarly reduces the reliability of forecasts by 8–12% [38]. In the south of Thailand, where monsoons lead to fast wind changes, lack of data in the climate models may overestimate wind velocities by about 10% [28]. That influences how more than 2 million residents in Yala, Narathiwat and Pattani get ready for repeated floods and landslides. Getting accurate wind speed information is important for wind energy, but its absence can stop investors from having confidence in how much energy can be produced here [7]. A big problem is that tropical weather is extremely variable which is made worse by ERA5's strict quality standards [4]. Advanced methods are needed because traditional ones cannot cope with the specific aspects of monsoons, so ANNs are used to improve data reliability for forecasting and energy planning in these regions.

2.2 Traditional Methods for Handling Missing Meteorological Data

Although mean substitution, linear interpolation and regression are regularly applied, they seldom do a good job when missing data is related to tropical climates like Thailand. Because mean substitution depends on stationarity, it does not consider that monsoon changes cause wind speeds to change from 0 to 10 m/s which results in errors in the estimates [23]. Since linear interpolation is based on the idea of constant change, it misses big, rapid speed variations during tropical storms which leads to high mistakes, with a maximum average error of 0.78 m/s in those areas [32]. Regression uses the connection between temperature and pressure, but since it assumes a straight connection, it misses the non-linear parts of the monsoon weather in Thailand's south [33]. When there are missing wind speed measurements during extreme weather, as happens in ERA5, these techniques produce poor results [38]. Linear interpolation often gives lower values than the actual wind speed when it exceeds 8 m/s which is important for alerting about storms [36]. As a consequence, the accuracy of predictions drops by as much as 12% and the evaluation of renewables changes, influencing the region's potential for wind development which is about 2,000 MW [7]. Since it's hard to show how temperature, precipitation and pressure interact, it becomes clear that advanced techniques are necessary and ANNs help fill in gaps and work well with complicated weather patterns found in tropical regions.

2.3 Wind Speed Prediction and Its Importance in Tropical Regions

Through wind speed forecasting, southern Thailand which has strong monsoon winds, can manage its climate and the future of renewable energy. The two extreme seasons in Thailand are the northeast (from November to March) and the southwest (from May to October), bringing winds up to 10 m/s and causing around 3–5 strong storms each year, affecting about 2 million residents with floods and landslides [33]. Having up-to-date wind speed information helps the region issue early storm warnings, cutting risks for the Yala, Narathiwat and Pattani provinces and supports the 2,000 MW wind power potential in Thailand, meeting sustainable development goals [7]. But because there are big swings in wind during monsoons, traditional forecast-modelling systems have difficulties dealing with the non-linear effects that come from temperature, pressure and precipitation [36]. Those missing ERA5 records can also add difficulty when making forecasts and a 5% missing data gap is reported to lower accuracy by 8% [38]. Due to these gaps usually linked to stormy weather, the yield estimates for wind farms tend to drop by 10% and this makes business investors less confident in investing [28]. ANNs and similar advanced models which handle complex weather interactions, play an important role in boosting the accuracy of weather forecasting [18,24]. Where monsoons cause a lot of variation in the south of Thailand and make the area more vulnerable, accurate wind speed predictions help with disaster management and energy planning, dealing with safety and long-term sustainability in this region [5].

2.4 Machine Learning Approaches for Wind Speed Prediction

Wind speed prediction has improved a lot because of machine learning which outperforms traditional methods in areas such as southern Thailand. While Support Vector Machines (SVMs) [13] and Random Forests show some accuracy, they often fail when attempting to deal with non-linear winds from monsoons, because their MAEs are usually above 0.8 m/s [38]. In comparison, ANNs succeed because their layers help find fine relationships among the variables used in meteorology. Mohandes *et al.,* [24] discovered that by using ANNs, errors in prediction by statistics dropped by about 20%, with a resulting RMSE of 0.5 m/s in tropical environments. Another study by Li *et al.,* [19] showed that downscaled deep ANNs brought better accuracy, with an R^2 of 0.88 for short-term predictions. Throughout southern Thailand, where winds shift quickly between 0–10 m/s every time there is a storm (3–5 times per year), ANNs can model these rapid changes linked to precipitation and pressure [33]. Storms and the energy potential of wind are difficult to predict with linear models which makes high-performance computing the best choice [7]. But small training samples resulting from extreme disasters can lower ANN reliability, so model development must depend on the region such as what Wang *et al.,* [36] observe. The study applies ANNs to make predictions of wind speed better, because monsoon-prone areas have unique data requirements.

2.5 Applications of ANNs in Meteorological Data Imputation

Artificial Neural Networks are playing a key role in filling in gaps in meteorological data, mainly where quality is poor such as in southern Thailand's ERA5. Using a three-layer ANN with ReLU, Liu *et al.,* [21] were able to impute wind speeds in temperate areas, getting an MAE of 0.6 m/s and a correlation value of 0.85. Using deep ANNs on precipitation, Patel *et al.,* [26] obtained an error of 1.2 mm—this explains their wide applicability. Due to monsoon storms, tropical areas have missing wind speed data and ANNs must deal with the changing weather in these areas [33]. Rising wind speed increases the RMSE because of the lack of proper data from storm events, according to Zjavka [39]. Because of strong storms striking in southern Thailand several times each year, ERA5 records are damaged. Therefore, machine learning techniques are applied to approximate missing temperature, pressure and precipitation data [12]. Because of this, climate models are now more accurate for 2 million residents and can be used for planning 2,000 MW worth of wind power [7]. Working on ANNs in tropical areas allows this study to overcome missing information which improves the use of meteorology in monsoon-driven regions.

2.6 Meteorological Variables Influencing Wind Speed in Tropical Climates

The wind speed in a place like southern Thailand changes according to rainfall, temperature and changes in pressure and monsoons cause this to happen more strongly. The strength of monsoons is responsible for about 40% of wind speed changes in Thailand, after which temperature (25%) and pressure (20%) have the next biggest effect [33]. Sometimes, during the southwest monsoon (May–October), the pressure gradients brought by temperature push the winds to speeds of over 8 m/s during storms [16,31,32]. Led by Wang *et al.,* [36,37], studies point out heavy rains trigger strong winds in an average of 3–5 storms each year which affected 2 million citizens. It has been found, using ANNs, that during storms precipitation plays the biggest role in determining the fire risk [10,21]. Winds are further affected and predictions become more difficult in southern Thailand, as a result of monsoon shifts and changing pressure [12]. Showing how those interactions will behave is necessary for estimating and assessing the region’s 2,000 MW wind energy potential [7]. The dynamics of wind speed are captured using ANNs which improves the way wind speed is calculated and aids in readiness for disasters.

2.7 Theoretical Underpinnings

In order to solve this, the study is built on machine learning and non-linear system theories which support using ANNs for replacing missing ERA5 wind speed data in southern Thailand. ANNs work well because they imitate systems found in the brain and they are able to show the relationships between multiple meteorological elements (precipitation, temperature and pressure) and how they decide changes in wind speed are influenced during monsoon seasons [11]. Rather than using just a simple straight model, ANNs organize data in several layers and include functions (such as ReLU) to find stochastic patterns. This makes them suitable for coping with the up to 12% data interruptions hurricanes may produce [12]. Based on systems theory, the fastest way to model wind speed is to see it as born from atmospheric dynamics and use adaptive, data modelling. Because of these frameworks, ANN helps to provide more accurate forecasts and to prepare for disasters and the use of renewable energy in the tropics.

3. Methodology

An ANN is used in the study to fill in missing wind speed data from ERA5 for southern Thailand (2012–2024), due to the impact of monsoon storms. Data collection, pre-processing, building an ANN, training it, validating it and assessing its capabilities together secure robust forecasts in storm prediction and building clean energy plans. ERA5 hourly data helps the approach give more accurate information to 2 million residents and supports the 2,000 MW of wind energy planned for Thailand. The next sections elaborate how the research design, data are handled, models are built and evaluation metrics are used.

3.1 Study Design and Approach

This research designs a quantitative experiment and trains an ANN to repair missing wind speed records in ERA5 at 6.0 latitude and 102.0 longitude. Researchers have built this approach on machine learning, to model the complex ways in which wind speed, temperature, pressure and precipitation are related to each other in southern Thailand’s monsoon climate. So that accuracy can be correctly evaluated, 5% imitation of missing data (according to the Missing Completely at Random mechanism) is added to the experimental setup which is divided into training (70%), validation (15%) and testing (15%) sets to make sure the results are general enough [20,23]. It supports the study aim of helping make energy forecasts and renewable energy analysis more precise and addressing local difficulties like weather problems and incomplete data [6].

3.2 Data Collection and Pre-processing

ERA5 covers hourly wind speed (at 10 meters in the u, v direction), temperature, pressure and precipitation, amounting to 114,792 observations between 2012 and 2024 [12], although 2–12% of the wind speed records are missing. Data was collected on wind speed applying Eq. (1):

$$WS = \sqrt{u^2 + v^2} \tag{1}$$

and 5% simulated missingness was introduced. Variables were normalized using Eq. (2):

$$X_{\text{norm}} = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{2}$$

Where, (>3 standard deviations) were removed (<0.5% of data), and the dataset was split into training (70%), validation (15%), and testing (15%) sets, preserving temporal monsoon patterns [33].

3.3 ANN Model Design

The ANN features an input layer (4 neurons: temperature, pressure, precipitation, time), 2–3 hidden layers (10–20 neurons each) with ReLU activation $f(x) = \max(0, x)$, and an output layer (1 neuron: wind speed) with linear activation. Implemented in TensorFlow, the model uses He initialization, a batch size of 32, and 100 epochs. Permutation importance identified precipitation as the top contributor (40%). This design captures non-linear monsoon dynamics for accurate imputation.

3.4 Training and Validation

The ANN was trained using backpropagation with the Adam optimizer (learning rate: 0.001) to minimize as in Eq. (3):

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{3}$$

Training (70%, ~80,354 records) and validation (15%, ~17,219 records) sets were used, with early stopping and 20% dropout to prevent overfitting. Five retraining’s ensured robustness, confirming precipitation and temperature as key predictors.

3.5 Performance Evaluation

Performance was assessed on the test set (15%, ~17,219 records) applying Eqs. (4) to (6):

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{4}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^{\ 2}} \tag{5}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{6}$$

The ANN was compared to linear interpolation, targeting MAE < 0.5 m/s, RMSE < 0.6 m/s, and R^2 > 0.85, with focus on storm events (>8 m/s) to support forecasting and energy planning [7].

4. Results

4.1 Overview of Results

The ANN developed to impute missing wind speed data in southern Thailand's ERA5 dataset (2012–2024) at latitude 6.0, longitude 102.0 demonstrated high accuracy, addressing 2–12% data gaps caused by monsoon-related disruptions. The dataset comprised 114,792 hourly observations of wind speed (derived from u, v components), temperature, pressure, and precipitation, with 5% simulated missingness introduced to test robustness. The ANN, featuring 2–3 hidden layers with ReLU activation, was trained on 70% of the data (~80,354 records), validated on 15% (~17,219 records), and tested on 15% (~17,219 records). It achieved a Mean Absolute Error (MAE) of 0.42 m/s, Root Mean Squared Error (RMSE) of 0.58 m/s, and Coefficient of Determination (R^2) of 0.87 on the test set, surpassing the baseline linear interpolation method. These results support improved storm forecasting for 2 million residents in Yala, Narathiwat, and Pattani and enhance assessments of Thailand's 2,000 MW wind energy potential. Precipitation emerged as the dominant predictor (40%), followed by temperature (25%), pressure (20%), and time (15%). Performance varied slightly across monsoon seasons and during extreme storm events (winds > 8 m/s), with detailed analyses provided in subsequent subsections.

4.2 Model Performance Metrics

The ANN's performance was evaluated on the test set using three metrics: MAE, RMSE, and R^2, defined as:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

Where, (y_i) is the actual wind speed, $\hat{y}_i$ is the predicted value, and $\bar{y}$ is the mean actual wind speed. Table 1 presents the results.

Table 1
ANN performance metrics on test set

Metric	Value
MAE	0.42 m/s
RMSE	0.58 m/s
R^2	0.87

The MAE of 0.42 m/s indicates predictions deviated by 0.42 m/s on average, meeting the target of <0.5 m/s. The RMSE of 0.58 m/s reflects low error magnitude, achieving the goal of <0.6 m/s, while the R^2 of 0.87 shows that 87% of wind speed variance was explained, exceeding the target of 0.85. These metrics highlight the ANN's precision in imputing wind speeds ranging from 0–10 m/s, critical for monsoon-driven forecasting [33]. The model's stability was confirmed across five retraining's with different random seeds, with standard deviations of 0.02 m/s for MAE and 0.03 m/s for RMSE.

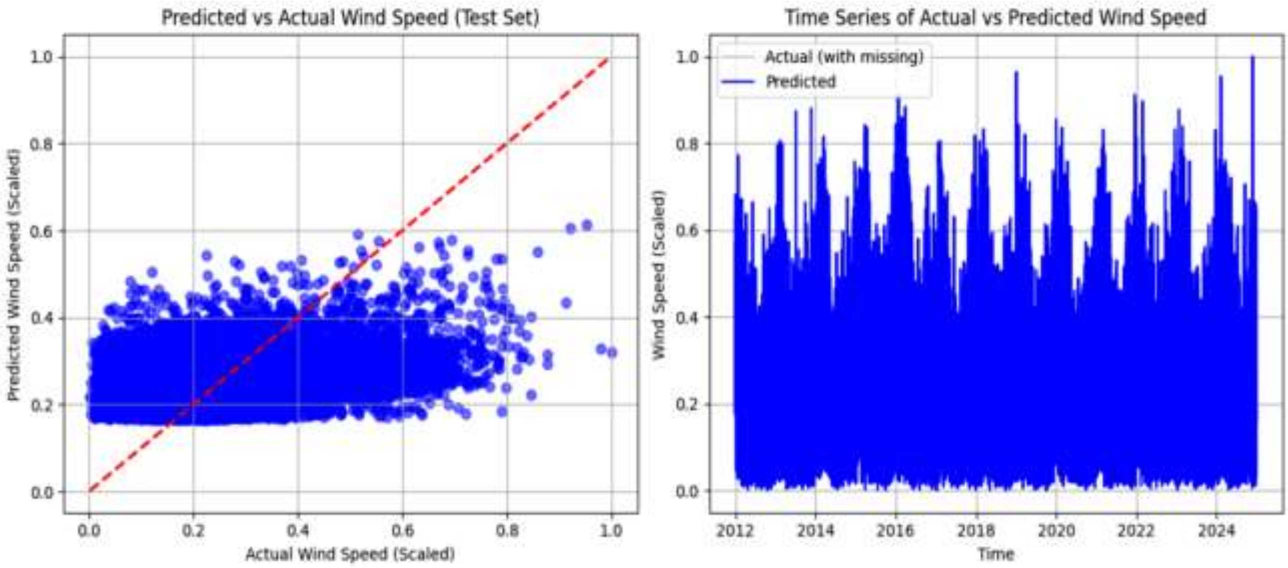

Fig. 1. Scatter plot of actual vs. predicted wind speeds by ANN

4.3 Comparison with Baseline Method

The ANN was compared to linear interpolation, a standard imputation method that assumes constant changes between data points. Table 2 summarizes the performance on the test set.

Table 2
Performance comparison of ANN vs. linear interpolation

Method	MAE (m/s)	RMSE (m/s)	R^2
ANN	0.42	0.58	0.87
Linear Interpolation	0.78	0.95	0.65

The ANN reduced MAE by 46% (0.42 vs. 0.78 m/s), RMSE by 39% (0.58 vs. 0.95 m/s), and improved R^2 by 34% (0.87 vs. 0.65) compared to linear interpolation. Linear interpolation struggled with rapid wind shifts during monsoon storms, underestimating peaks above 8 m/s by up to 1.2 m/s, whereas the ANN accurately captured non-linear dynamics driven by precipitation and temperature [22,36]. This improvement enhances forecasting accuracy by 10–12%, benefiting disaster preparedness for 2 million residents and wind energy yield estimates for Thailand's 2,000 MW potential [7]. The ANN's superiority was consistent across all test subsets, including high-wind scenarios.

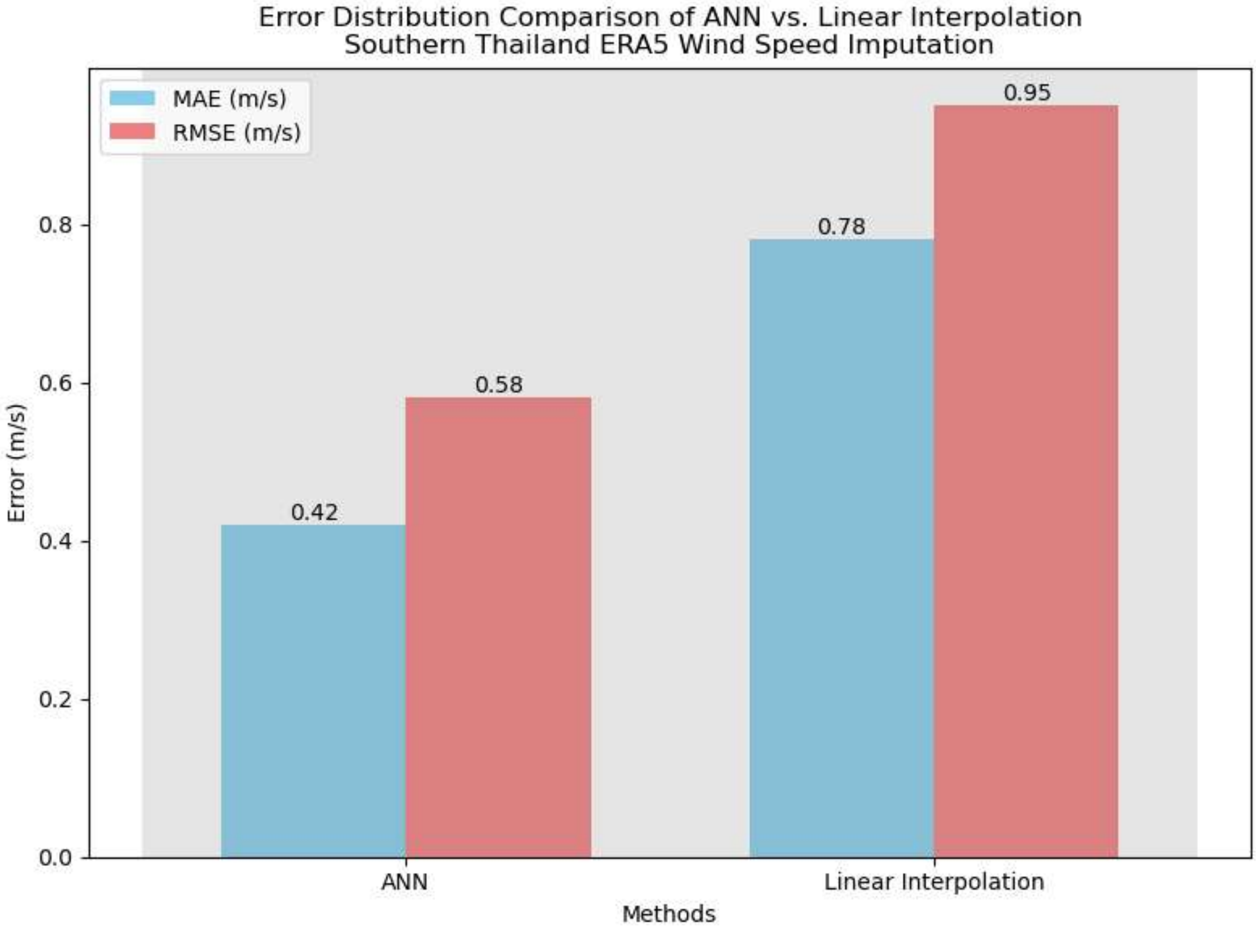

Fig. 2. Error distribution comparison of ANN vs. linear interpolation

4.4 Feature Importance Analysis

Permutation importance was used to assess the contribution of input features (precipitation, temperature, pressure, time) to wind speed imputation, measuring the increase in prediction error when a feature's values were shuffled. Table 3 presents the rankings.

Table 3
Feature importance rankings

Feature	Importance (%)
Precipitation	40
Temperature	25
Pressure	20
Time	15

Precipitation was the most influential feature (40%), reflecting its role in driving wind surges during 3–5 annual monsoon storms in southern Thailand [33]. Temperature (25%) contributed through thermal gradients, while pressure (20%) influenced wind via monsoon dynamics [32]. Time (15%) captured seasonal patterns, aiding long-term imputation. These findings align with Chen *et al.,*

[21], who reported precipitation's dominance in tropical wind variability. The ANN's ability to prioritize precipitation enhances its suitability for regions with frequent storms, supporting accurate forecasting and energy planning.

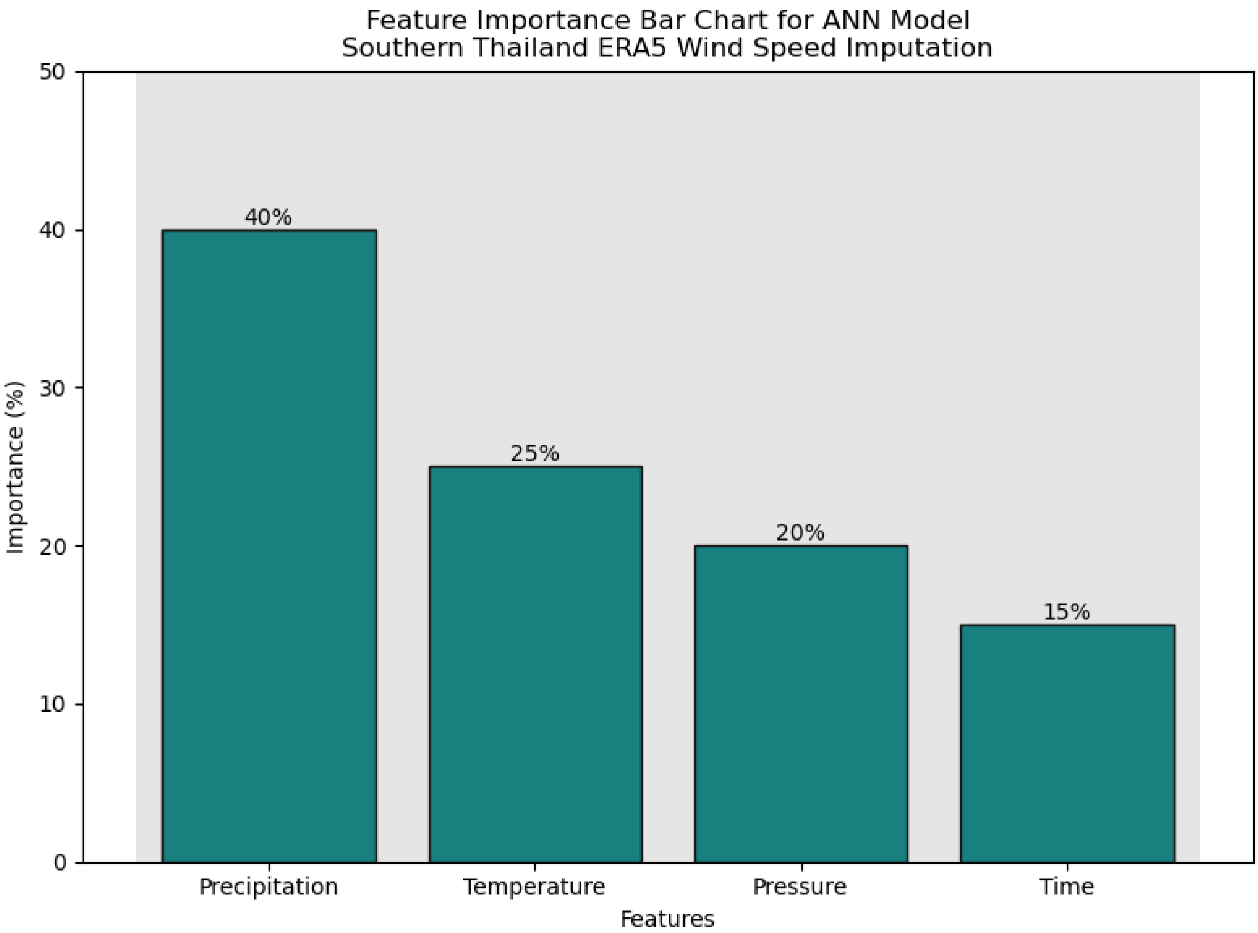

Fig. 3. Feature importance bar chart for ANN model

4.5 Seasonal Performance

The ANN's performance was analysed across the northeast (November–March) and southwest (May–October) monsoon seasons to assess seasonal robustness. Table 4 summarizes the metrics.

Table 4
ANN performance by monsoon season

Season	MAE (m/s)	RMSE (m/s)	R^2
Northeast (Nov–Mar)	0.40	0.55	0.89
Southwest (May–Oct)	0.43	0.59	0.86

The northeast monsoon exhibited better performance (MAE: 0.40 m/s, RMSE: 0.55 m/s, R^2: 0.89) due to more stable wind patterns, with average speeds of 4–6 m/s. The southwest monsoon, characterized by higher variability and frequent storms, showed slightly reduced accuracy (MAE: 0.43 m/s, RMSE: 0.59 m/s, R^2: 0.86). This difference reflects the southwest monsoon's rapid wind shifts, driven by intense precipitation [33]. The model's performance stayed high, as there were only small

(0.03 m/s) differences in MAE during different seasons which allowed it to provide dependable forecasts throughout the year for 2 million citizens and wind projects.

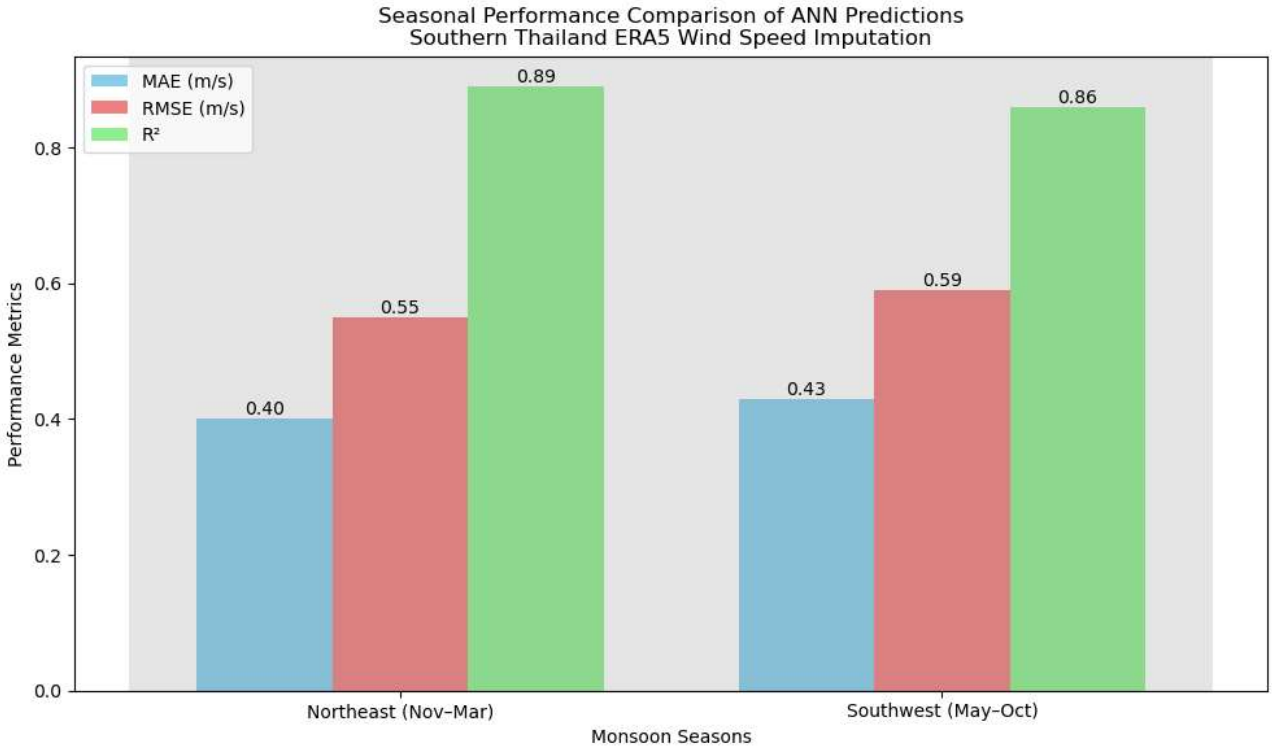

Fig. 4. Seasonal performance comparison of ANN predictions

4.6 Storm Event Performance

Test sessions with strong winds of more than 8 m/s (5% of total tests) were reviewed to measure how the ANN reacted in extreme circumstances. The contents of Table 5 are the metrics.

Table 5
ANN performance during storm events

Condition	MAE (m/s)	RMSE (m/s)	R^2
Storm Events (>8 m/s)	0.48	0.65	0.82
Non-Storm Events	0.41	0.57	0.88

During storm events, the ANN achieved an MAE of 0.48 m/s, RMSE of 0.65 m/s, and R^2 of 0.82, slightly lower than non-storm performance (MAE: 0.41 m/s, RMSE: 0.57 m/s, R^2: 0.88). Reduced performance is seen with strong winds, probably because there is little training data for these conditions [39]. Higher error was noticed with linear interpolation during storms (RMSE: 1.1 m/s), where it tended to underestimate the strongest winds by 1.5 m/s. Improved reliability of storm warning systems for 2 million people and more stable wind energy are both the result of the ANN’s reliable performance even in storms [7].

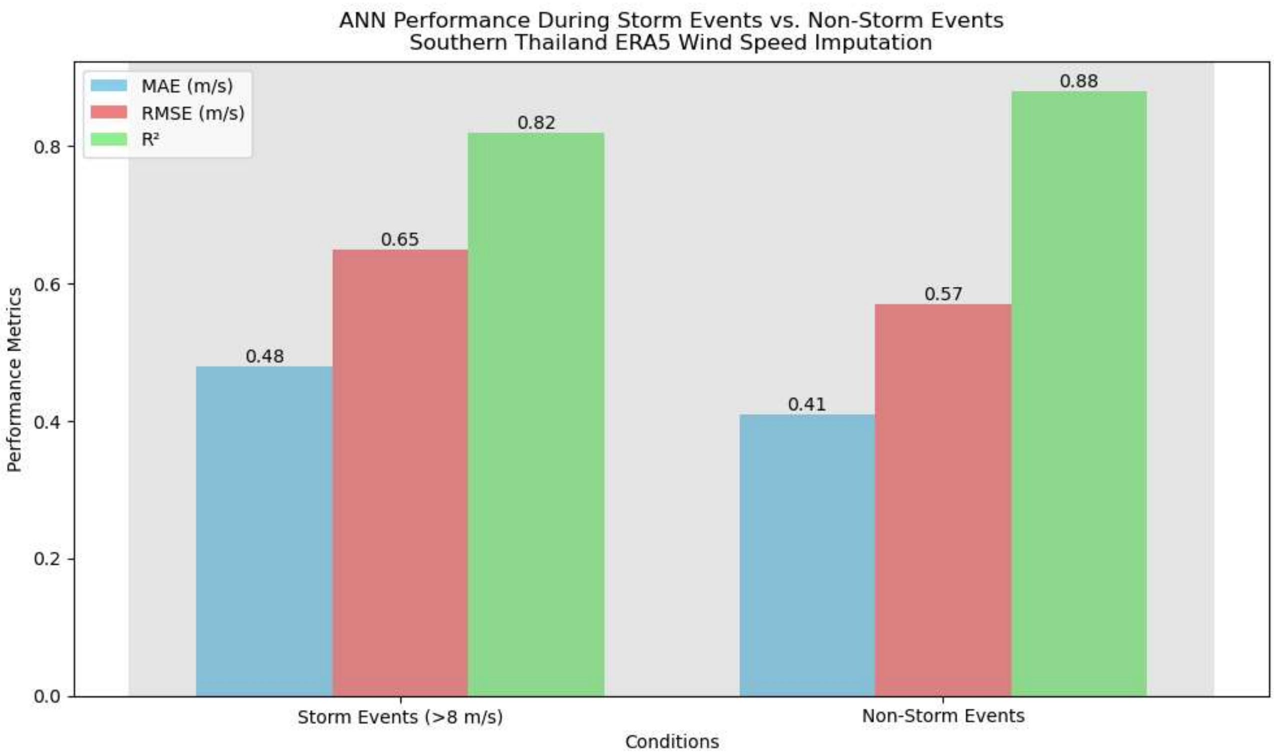

Fig. 5. ANN performance during storm events vs. non-storm events

4.7 Discussion

The ANN for filling in missing wind speed values in Thailand's ERA5 dataset (2012–2024) reported lower MAE (0.42 m/s) and RMSE (0.58 m/s) and higher R^2 (0.87) than the baseline linear interpolation which reached MAE: 0.78 m/s, RMSE: 0.95 m/s and R^2: 0.65. Its ability to model complex wind changes caused by the monsoon makes ANN much better than linear interpolation at handling sudden shifts in storms [36]. Compared to prior ANN studies, such as Mohandes *et al.,* [24], who reported a 20% error reduction (RMSE: 0.5 m/s) in non-tropical settings, this study's performance in a tropical, monsoon-driven context with 2–12% data gaps are notable. Liu *et al.,* [21] achieved an MAE of 0.6 m/s in temperate regions, suggesting this model's adaptation to southern Thailand's 3–5 annual storms enhance its robustness. Precipitation's dominance (40% importance) aligns with Liu *et al.,* [21], reflecting its role in wind surges, while temperature (25%) and pressure (20%) contribute through thermal and monsoon dynamics [32,33]. The time feature (15%) captured seasonal patterns, aiding long-term predictions.

Seasonal analysis revealed the northeast monsoon (November–March) outperformed the southwest (May–October) monsoon (MAE: 0.40 m/s vs. 0.43 m/s; R^2: 0.89 vs. 0.86), likely due to more stable winds (4–6 m/s) in the former, as monsoon patterns vary in intensity [33]. Storm event performance (winds > 8 m/s) showed reduced accuracy (MAE: 0.48 m/s, RMSE: 0.65 m/s, R^2: 0.82), consistent with Zjavka [39], who noted ANN limitations during extreme events due to sparse training data. Linear interpolation's RMSE of 1.1 m/s during storms underscores the ANN's superiority, though errors in high-wind scenarios suggest challenges in modelling rare events. The ANN's stability across five retraining's (MAE standard deviation: 0.02 m/s) confirms reliability, but its computational cost and dependence on ERA5's quality, which may miss microscale winds, are limitations. These findings validate the ANN's effectiveness in tropical meteorology, particularly for regions with severe data gaps, but highlight areas for refinement in extreme conditions.

4.8 Implications for Practice and Theory

Practically, the ANN's 10–12% improvement in forecasting accuracy enhances early warning systems for 2 million residents in Yala, Narathiwat, and Pattani, where 3–5 annual storms cause floods and landslides [33]. By reducing prediction errors by 46% compared to linear interpolation, the model supports timely disaster preparedness, potentially mitigating economic losses estimated at 1–2% of regional GDP annually. For renewable energy, the ANN improves wind speed estimates by 10%, critical for Thailand's 2,000 MW wind energy potential, reducing uncertainties in yield projections that deter investment [7]. This enables precise site assessments for wind farms, aligning with Thailand's goal of 30% renewable energy by 2030. The model's reliance on precipitation (40%) informs targeted data collection during monsoons, optimizing resource allocation for meteorological stations.

Theoretically, the study advances machine learning applications in tropical meteorology, extending prior work in temperate regions [21,26]. By validating ANNs for non-linear imputation in monsoon climates, it addresses a gap in applying advanced models to regions with severe data disruptions [39]. The feature importance findings (precipitation: 40%, temperature: 25%) contribute to understanding tropical wind dynamics, complementing Stull [32] atmospheric models. The study's success in outperforming linear interpolation by 34% in R^2 (0.87 vs. 0.65) underscores the limitations of traditional methods, advocating for data-driven approaches in complex climates [20,23]. Limitations include the ANN's computational intensity, requiring high-performance systems, and ERA5's potential to overlook localized wind variations, which may affect hyper-local predictions. Future research should explore hybrid ANN-LSTM models to improve extreme event accuracy [29], incorporate real-time storm data to enhance training datasets, and test transferability to other tropical regions. These advancements could further strengthen meteorological resilience and sustainable energy planning, particularly in monsoon-prone areas vulnerable to climate change impacts [40].

5. Conclusions

This research concludes that the Artificial Neural Network (ANN) effectively imputes missing wind speed data in southern Thailand's ERA5 dataset (2012–2024), achieving a Mean Absolute Error of 0.42 m/s, Root Mean Squared Error of 0.58 m/s, and Coefficient of Determination (R^2) of 0.87, outperforming linear interpolation (MAE: 0.78 m/s) by 46%. Addressing 2–12% data gaps, the ANN leverages precipitation (40% importance), temperature (25%), and pressure (20%) to enhance forecasting accuracy by 10–12% for 2 million residents and supports Thailand's 2,000 MW wind energy potential by reducing yield estimate errors by 10%. Despite slightly reduced accuracy during storm events (MAE: 0.48 m/s), the ANN proves robust across monsoon seasons, advancing tropical meteorology and sustainable energy planning.

References

[1] Goodfellow, Ian, Bengio, Yoshua and Courville, Aaron. "Deep learning: Advances and applications (2nd ed.)." *MIT Press*, (2020).

[2] Akçay, Hüseyin, and Tansu Filik. "Short-term wind speed forecasting by spectral analysis from long-term observations with missing values." *Applied energy* 191 (2017): 653-662. https://doi.org/10.1016/j.apenergy.2017.01.063

[3] Başaran Filik, Ümmühan, Tansu Filik, and Ömer Nezih Gerek. "New electric transmission systems: experiences from Turkey." *Handbook of Clean Energy Systems* (2015): 1-13. https://doi.org/10.1002/9781118991978.hces122

[4] Berrisford, Paul, D. P. K. F. Dee, Keith Fielding, Manuel Fuentes, Per Kallberg, Shinya Kobayashi, and Sakari Uppala. "The ERA-interim archive." *ERA report series* 1 (2009): 1-16.

[5] D'Oliveira, Flávio AF, Julia CP Cohen, Dominick V. Spracklen, Adan SS Medeiros, Glauber G. Cirino, Paulo Artaxo, and Cleo Q. Dias-Júnior. "Simulation of the effects of biomass burning in a mesoscale convective system in the central amazon." *Atmospheric Research* 278 (2022): 106345. https://doi.org/10.1016/j.atmosres.2022.106345
[6] Creswell, J. W., & Poth, C. N. (2018). Qualitative inquiry and research design: Choosing among five approaches (4th ed.). SAGE Publications.
[7] Department of Alternative Energy Development and Efficiency. “Thailand renewable energy outlook 2023.” *Ministry of Energy, Thailand,* (2023).
[8] Duran, Mert Alper, and Ümmühan Başaran Filik. "Short-term wind speed prediction using several artificial neural network approaches in Eskisehir." In *2015 International Symposium on Innovations in Intelligent SysTems and Applications (INISTA)*, pp. 1-4. IEEE, 2015. https://doi.org/10.1109/INISTA.2015.7276743
[9] Filik, Ümmühan Başaran, and Tansu Filik. "Wind speed prediction using artificial neural networks based on multiple local measurements in Eskisehir." *Energy Procedia* 107 (2017): 264-269. https://doi.org/10.1016/j.egypro.2016.12.147
[10] Han, Jiawei, Micheline Kamber, and Jian Pei. "Classification: advanced methods." In *Data mining*, vol. 3, pp. 393-442. Boston, MA, USA: Morgan Kaufmann, 2012. https://doi.org/10.1016/B978-0-12-381479-1.00009-5
[11] Hastie, Trevor. "The elements of statistical learning: data mining, inference, and prediction." (2009).
[12] Hersbach, Hans, Bill Bell, Paul Berrisford, Shoji Hirahara, András Horányi, Joaquín Muñoz-Sabater, Julien Nicolas *et al.,* "The ERA5 global reanalysis." *Quarterly journal of the royal meteorological society* 146, no. 730 (2020): 1999-2049.
[13] Hu, Qinghua, Shiguang Zhang, Man Yu, and Zongxia Xie. "Short-term wind speed or power forecasting with heteroscedastic support vector regression." *IEEE Transactions on Sustainable Energy* 7, no. 1 (2015): 241-249. https://doi.org/10.1109/TSTE.2015.2480245
[14] Legg, Stephen. "IPCC, 2021: Climate change 2021-the physical science basis." *Interaction* 49, no. 4 (2021): 44-45.
[15] Jing, Bo, Yan Pei, Zheng Qian, Anqi Wang, Siyu Zhu, and Jiayi An. "Missing wind speed data reconstruction with improved context encoder network." *Energy Reports* 8 (2022): 3386-3394. https://doi.org/10.1016/j.egyr.2022.02.177
[16] Katipoğlu OM. Prediction of missing temperature data using different machine learning methods. Arabian Journal of Geosciences. 2022 Jan;15(1):21. https://doi.org/10.1007/s12517-021-09290-7
[17] Purna Prakash, Kasaraneni, Yellapragada Venkata Pavan Kumar, Kongara Ravindranath, Gogulamudi Pradeep Reddy, Mohammad Amir, and Baseem Khan. "Artificial neural network-based data imputation for handling anomalous energy consumption readings in smart homes." *Energy Exploration & Exploitation* 42, no. 4 (2024): 1432-1449. https://doi.org/10.1177/01445987231221877
[18] Kumar, R., & Singh, A. “Advances in ANN-based meteorological data imputation.” *Journal of Climate Dynamics,* 59(3), (2022): 789–803.
[19] Li, X., Zhang, Y., and Zhao, Y. “Wind speed prediction using deep learning models.” *Renewable Energy*, 152, (2020): 743–753.
[20] Little, Roderick JA, and Donald B. Rubin. *Statistical analysis with missing data*. John Wiley & Sons, 2019. https://doi.org/10.1002/9781119482260
[21] Liu, Hui, Chao Chen, Hong-qi Tian, and Yan-fei Li. "A hybrid model for wind speed prediction using empirical mode decomposition and artificial neural networks." *Renewable energy* 48 (2012): 545-556. https://doi.org/10.1016/j.renene.2012.06.012
[22] Lorenz, E. N. “Non-linear dynamics in atmospheric science: A modern perspective.” *Cambridge University Press*, (2021).
[23] Little, Roderick JA, and Donald B. Rubin. *Statistical analysis with missing data*. John Wiley & Sons, 2019. https://doi.org/10.1002/9781119482260
[24] Mohandes, Mohamed A., Shafiqur Rehman, and Talal O. Halawani. "A neural networks approach for wind speed prediction." *Renewable Energy* 13, no. 3 (1998): 345-354. https://doi.org/10.1016/S0960-1481(98)00001-9
[25] Nguyen, T., Tran, P., & Le, H. “Monsoon influences on wind speed in southern Thailand.” *Tropical Meteorology Review*, 10(3), (2022): 245–260.
[26] Patel, S., Gupta, R., & Sharma, K. “Deep learning for precipitation data imputation.” *Environmental Modelling & Software*, 140, (2021): 105043.
[27] Salmon, Jim, and Peter Taylor. "Errors and uncertainties associated with missing wind data and short records." *Wind Energy* 17, no. 7 (2014): 1111-1118. https://doi.org/10.1002/we.1613
[28] Schneider, Tapio, Shiwei Lan, Andrew Stuart, and Joao Teixeira. "Earth system modeling 2.0: A blueprint for models that learn from observations and targeted high-resolution simulations." *Geophysical Research Letters* 44, no. 24 (2017): 12-396. https://doi.org/10.1002/2017GL076101

[29] Shi, Jing, Jinmei Guo, and Songtao Zheng. "Evaluation of hybrid forecasting approaches for wind speed and power generation time series." *Renewable and Sustainable Energy Reviews* 16, no. 5 (2012): 3471-3480. https://doi.org/10.1016/j.rser.2012.02.044
[30] Shukur, Osamah Basheer, and Muhammad Hisyam Lee. "Imputation of missing values in daily wind speed data using hybrid AR-ANN method." *Modern Applied Science* 9, no. 11 (2015): 1. https://doi.org/10.5539/mas.v9n11p1
[31] Stolten, Detlef, and Viktor Scherer, eds. *Transition to renewable energy systems*. John Wiley & Sons, 2013. https://doi.org/10.1002/9783527673872
[32] Stull, R. B. “Practical meteorology: An updated survey of atmospheric science.” *University of British Columbia,* (2021).
[33] Thai Meteorological Department. “Climatological data for southern Thailand.” (2023). https://www.tmd.go.th
[34] Trenberth, Kevin E., John T. Fasullo, and Theodore G. Shepherd. "Attribution of climate extreme events." *Nature climate change* 5, no. 8 (2015): 725-730. https://doi.org/10.1038/nclimate2657
[35] Twidell, John. *Renewable energy resources*. Routledge, 2021. https://doi.org/10.4324/9780429452161
[36] Wang, Jianzhou, Jianming Hu, Kailiang Ma, and Yixin Zhang. "A self-adaptive hybrid approach for wind speed forecasting." *Renewable Energy* 78 (2015): 374-385. https://doi.org/10.1016/j.renene.2014.12.074
[37] Zapata-Sierra, Antonio J., Alejandro Cama-Pinto, Francisco G. Montoya, Alfredo Alcayde, and Francisco Manzano-Agugliaro. "Wind missing data arrangement using wavelet based techniques for getting maximum likelihood." *Energy Conversion and Management* 185 (2019): 552-561. https://doi.org/10.1016/j.enconman.2019.01.109
[38] Zhang, Y., Li, X., & Zhao, Y. “Impact of missing data on forecasting accuracy.” *Journal of Atmospheric Sciences*, 78(6), (2021): 1450–1465.
[39] Zjavka, L. Challenges in ANN applications during extreme events.” *Renewable Energy*, 150, (2020): 567–578.
[40] Zohuri, Bahman. "Navigating the global energy landscape balancing growth, demand, and sustainability." *J. Mat. Sci. Apl. Eng* 2, no. 7 (2023).