國立臺灣科技大學

工業管理系

博士學位論文

學號： D10501813

Improving Energy Forecast Accuracy through Hybrid Deep Learning and Advanced Data Preprocessing Methods

研 究 生： Irene Karijadi

指導教授： Shuo-Yan Chou

共同指導： Po-Hsun Kuo

中華民國一一二年十一月

D10501813

博士學位論文指導教授推薦書
Doctoral Dissertation Recommendation Form

系所： Department/Graduate Institute	工業管理系 Department of Industrial Management
姓名： Name	IRENE IRENE KARIJADI
論文題目： (Dissertation Title)	透過混合深度學習和先進的數據預處理方法提高能源預測準確性 Improving Energy Forecast Accuracy through Hybrid Deep Learning and Advanced Data Preprocessing Methods

係由本人指導撰述，同意提付審查。

This is to certify that the dissertation submitted by the student named above, has been written under my supervision. I hereby approve this dissertation to be applied for examination.

指導教授簽章：
Advisor's Signature ____________

共同指導教授簽章（如有）：
Co-advisor's Signature (if any) ____________

日期：
Date(yyyy/mm/dd) ______/______/______

D10501813

博士學位考試委員審定書

Qualification Form by Doctoral Degree Examination Committee

系所： Department/Graduate Institute	工業管理系 Department of Industrial Management
姓名： Name	IRENE IRENE KARIJADI
論文題目： (Dissertation Title)	透過混合深度學習和先進的數據預處理方法提高能源預測準確性 Improving Energy Forecast Accuracy through Hybrid Deep Learning and Advanced Data Preprocessing Methods

經本委員會審定通過，特此證明。

This is to certify that the dissertation submitted by the student named above, is qualified and approved by the Examination Committee.

學位考試委員會

Degree Examination Committee

委員簽章：

Member's Signatures

召集人簽章：

Committee Chair's Signature

指導教授簽章：

Advisor's Signature

共同指導教授簽章（如有）：

Co-advisor's Signature (if any)

系所（學程）主任（所長）簽章：

Department/Study Program/Graduate Institute Chair's Signature

日期：

Date(yyyy/mm/dd) _____/_____/_____

ABSTRACT

Energy forecasting is paramount in today's world, given the ever-increasing demand for energy across various sectors and its crucial role in daily life. This study explores the integration of machine learning and data decomposition techniques to enhance the accuracy of energy forecasting. While machine learning, particularly deep learning architectures such as Long Short-Term Memory (LSTM), has shown promise in time-series forecasting, real-world energy data often exhibits nonstationary and nonlinear patterns, challenging traditional forecasting methods. To address this, scholars have proposed hybrid approaches incorporating data decomposition techniques to mitigate the effects of non-stationarity and non-linearity. However, the highest frequency components resulting from decomposition remain challenging, leading to forecasting inaccuracies.

Therefore, to address these limitations, this study introduces two novel hybrid approaches for energy forecasting, each tailored to specific energy applications. The first approach, CEEMDAN-RF-LSTM, focuses on enhancing the precision of predictions for building energy consumption. The second approach, CEEMDAN-EWT-LSTM, is designed for wind power forecasting. Both methods incorporate advanced data decomposition techniques to effectively handle the highest frequency components. These approaches offer a promising avenue for achieving more accurate and reliable energy forecasts, benefiting industries, investors, and governments in making informed decisions and planning for the future. The effectiveness of these proposed approaches is evaluated using actual time series datasets. Finally, the study concludes by discussing the advantages and potential future developments of deep learning and data preprocessing strategies in the context of time series forecasting for energy-related applications.

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to all those who have contributed to the completion of this dissertation.

First and foremost, I would like to thank Jesus Christ who has given me grace to pursue the doctoral program at Industrial Management department of National Taiwan University of Science and Technology.

I am deeply thankful to my advisor, Prof. Shuo-Yan Chou and my co advisor Prof. Po-Hsun Kuo, for their support, guidance, and insightful feedback throughout the entire research process. Their expertise and encouragement have been instrumental in shaping the direction of this work.

I extend my appreciation to Prof. Tiffany Hui-Kuang Yu, Prof. Yu-Ling Hsu, Prof. Shih-Wei Lin, and Prof Jen-Ming Chen as the members of my dissertation committee for their constructive feedback, expertise, and time invested in reviewing and evaluating this research.

Special thanks go to my family and friends for their unwavering encouragement, understanding, and patience during this demanding period. Their support has been a constant source of motivation.

I am truly grateful to everyone who has been a part of this academic journey.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

This chapter presents the study's background and motivation, objectives, scopes, and limitations.

1.1 Background

It is widely acknowledged that energy plays a crucial role in our society. It serves multiple purposes, such as automating processes and facilitating daily activities, making it an indispensable component of human existence [1]. The anticipated increase in worldwide energy requirements is projected to rise by 48% over the next two decades [2], driven by the transformative advancements in various industrial domains and the expansion of infrastructure development [3]. This escalating demand underscores the continued significance of energy in our lives, making the protection of energy availability a crucial target on a global scale [4]. The growth in renewable and non-renewable energy needs has reached various sectors of our society, transforming it into one of the most challenging issues of the 21st century [3]. Thus, searching for a technique that can accurately forecast energy supplies and demand becomes increasingly important [5].

Energy forecasting is crucial for modern power systems, also known as smart grids (SGs), by facilitating planning, investments, and decision-making and overcoming operational and management difficulties [6]. Implementing smart meters and advanced metering infrastructure (AMI) in smart grid (SG) systems has led to a notable rise in the exchange of energy and data between the grid and end-users in both directions. Consequently, this has given rise to various data analytics applications, including energy forecasting, within smart grids. These applications offer valuable advantages for scheduling power generation, executing demand response strategies, and optimizing financial outcomes through well-informed energy market bidding [6]. For example, from a demand-side perspective, accurate energy demand forecasts assist power producers in scheduling their generation, enabling them to efficiently distribute power and mitigate the chances of the utility company generating too little or too much power. The integration of accurate demand forecasting algorithms into BEMS has emerged as a critical advancement in building energy management, enabling proactive optimization of energy consumption based on anticipated needs. Its importance lies in its capacity to facilitate various building applications, including Demand Side Management (DSM) [7], intelligent control decisions, and safety monitoring [8]. From a supply viewpoint, an accurate power generation forecast empowers power systems to intelligent decisions regarding power generation, storage, and distribution [9]. Accurate renewable energy forecasts aid grid operators in managing the integration of renewable resources into the grid. Grid operators can plan for the varying output of renewables, ensuring grid stability and reducing the need for backup generation.

Many research efforts have been dedicated to forecasting various aspects of the energy industry, such as oil production [9] and consumption [10], wind energy [11,12], energy consumption [8,13] and solar energy [14]. In recent times, machine learning (ML) has exhibited significant potential in energy data forecasting, offering accurate and reliable forecasts of future energy demand and supply through historical time-series data [15,16]. A more sophisticated method, such as deep learning, has attracted considerable interest in the past few years due to its capability to handle complex data and its computational power advancement [17]. The utilization of deep learning techniques in the energy sector has greatly expanded, emerging as a powerful method for forecasting energy prices, supplies, and demands [5].

Deep Learning models, including Long Short-Term Memory (LSTM), have been effectively utilized for their excellent performance in time-series forecasting [18]. However, nonstationary and nonlinearity patterns often characterize real-world time series data [19]. While LSTM and GRU perform better in modeling temporal relationships in time series data, their performance may degrade when dealing with non-stationarity and nonlinearity in the data [19]. Some scholars have proposed hybrid approaches to enhance forecasting accuracy to reduce the negative effects of the non-stationarity characteristic of time series data. One prominent approach involves integrating the data decomposition technique into the forecasting method to decrease nonstationary and nonlinearity effects associated with time series data [20]. Researchers have integrated hybrid data decomposition techniques, resulting in a notable enhancement in accuracy and forecasting performance [21,22]. For instance, Zhang et al. [23] used Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) to preprocess the original wind data. Then, they used deep learning as the forecasting method. The proposed approach outperformed other techniques in terms of accuracy.

The framework of decomposition-based hybrid models typically breaks down the nonstationary initial time series data into several more stable components. Then, individual forecasting is developed for each component. The final forecast outcome is obtained by aggregating all these individual forecasts. The adoption of this decomposition-based hybrid model has the capacity to improve forecast accuracy significantly. The decomposition techniques can help convert non-stationary data into more stationary components. This preprocessing step can enhance accuracy because forecasting methods typically exhibit superior performance on stationary data.

Despite the significant improvement achieved through the integration of decomposition techniques with the machine learning method, there is still a need for further enhancements to enhance the accuracy of energy forecasting. This need arises because the highest frequency component, generated by the decomposition technique, remains highly volatile and contains noises [24]. Forecasting this particular series is especially challenging [25] and can negatively impact forecasting accuracy. Many energy forecasting methods traditionally depend on a single decomposition technique, often overlooking the complexities associated with the highest frequency component. For example, Zhang et al. [23] exclusively applied CEEMDAN for data decomposition to forecast wind power data. Consequently, forecasting performance is constrained due to the inadequate treatment of the intricacies associated with the highest frequency component not being dealt with appropriately.

Considering the limitations of the works mentioned above, there is still a need to build an appropriate forecasting approach to deal with the highest frequency component generated from the decomposition process. This study introduces novel methods to improve energy forecast accuracy in response to this challenge. These approaches take a dual perspective, targeting energy's demand and supply sides. On the demand side, we present a hybrid CEEMDAN-Random Forest (RF) and Long-Short Term Memory (LSTM) approach designed to elevate building energy consumption forecasting performance. The highest frequency component is modeled and forecasted using RF in this approach. The other components are forecasted using LSTM. This study also addresses the supply aspect by presenting a hybrid approach known as CEEMDAN-Empirical Wavelet Transform (EWT)-LSTM to enhance the precision of the wind power forecast. The highest frequency component generated from CEEMDAN is denoised in this approach using the EWT denoising approach. By incorporating deep learning with decomposition techniques and giving special attention to the highest frequency component, these methods offer the potential for significant advancements in energy forecasting.

1.2 Research objective and contribution

The research objective of this thesis is to develop accurate hybrid machine learning methods for building energy consumption forecasting and wind power forecasting. The main contributions of this thesis are as follows:

1. Two novel hybrid machine learning methods are proposed for building energy consumption and wind power forecasting.
2. Advanced data preprocessing methods were developed to enhance forecast accuracy in the two proposed approach. These methods introduce a data preprocessing approach that utilizes the Complete Ensemble Empirical Mode Decomposition with Adaptive Noise as the data decomposition method to improve forecasting accuracy. Additionally, a specialized approach is incorporated to address the challenges posed by the highest frequency components generated during the decomposition process. This involves the integration of Random Forest and Empirical Wavelet Transform Denoising, effectively managing the intricacies associated with high-frequency data components and contributing to the overall enhancement of forecasting accuracy.
3. The proposed approach are evaluated on real-world datasets, and the results show that they outperform state-of-the-art methods in terms of forecasting accuracy.

In addition to the aforementioned contributions, this thesis also comprehensively reviews the related literature on building energy consumption and wind power forecasting. This review can be a valuable resource for researchers and practitioners interested in these fields.

1.3 Scope and limitations

This thesis aims to develop and evaluate hybrid machine learning methods for building energy consumption forecasting and wind power forecasting. The main limitations of this thesis are as follows:

1. The proposed approaches have only been evaluated on real-world datasets from a few countries. The performance of the proposed approach may vary on datasets from other countries.
2. This thesis exclusively concentrates on forecasting techniques for univariate time series and does not include multivariate analysis. Complete Ensemble Empirical Mode Decomposition with Adaptive Noise and other decomposition approaches can effectively break down individual time series into multiple components within the frequency domain, enhancing their applicability for subsequent machine learning. However, these methods are specifically designed for decomposing single time series into multiple components and are not tailored for handling multivariate time series. When applied independently to process multivariate time series, these decomposition methods may generate components that lack correlation. Consequently, in this study, only univariate analysis is employed. As a result, the potential influence of external factors and interrelationships among multiple variables on the forecasted outcomes remains unexplored within the confines of this research.

1.4 Organization of thesis

Chapter 1 provides an overview of the thesis, including the research objective, contributions, scope, and limitations. Chapter 2 presents the theoretical background that supports this study. Chapter 3 introduces the proposed hybrid CEEMDAN-Random Forest-LSTM method for energy consumption in the building forecasting. Chapter 4 introduces the proposed hybrid CEEMDAN-EWT-LSTM deep learning method for wind power forecasting. Chapter 5 presents the comparative evaluation and analysis of both proposed approach. Finally, Chapter 6 ends the thesis with a conclusion and a future study.

CHAPTER 2

THEORETICAL BACKGROUND

This part comprehensively describes the conceptual foundations of the methods incorporated in our proposed approach.

2.1 CEEMDAN

Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) was created to break down data that is both non-linear and non-stationary into multiple Intrinsic Mode Functions (IMF) series, providing increased stability and stationarity to the data. CEEMDAN is an extension of the Empirical Mode Decomposition algorithm [26]. EMD, developed by Huang et al. [27], faced limitations like mode mixing, leading to the emergence of EEMD as a solution [28]. Despite its effectiveness, EEMD suffers from increased computational complexity [29]. CEEMDAN addresses this issue by strategically injecting adaptive noise into the residual signal after the EMD process instead of applying it directly to the original data [30]. This approach leverages the iterative nature of CEEMDAN to achieve noise averaging, eliminating unnecessary modifications to the original signal and reducing computational burden while enhancing decomposition accuracy [31]. The flowchart of CEEMDAN is illustrated in Figure 1.

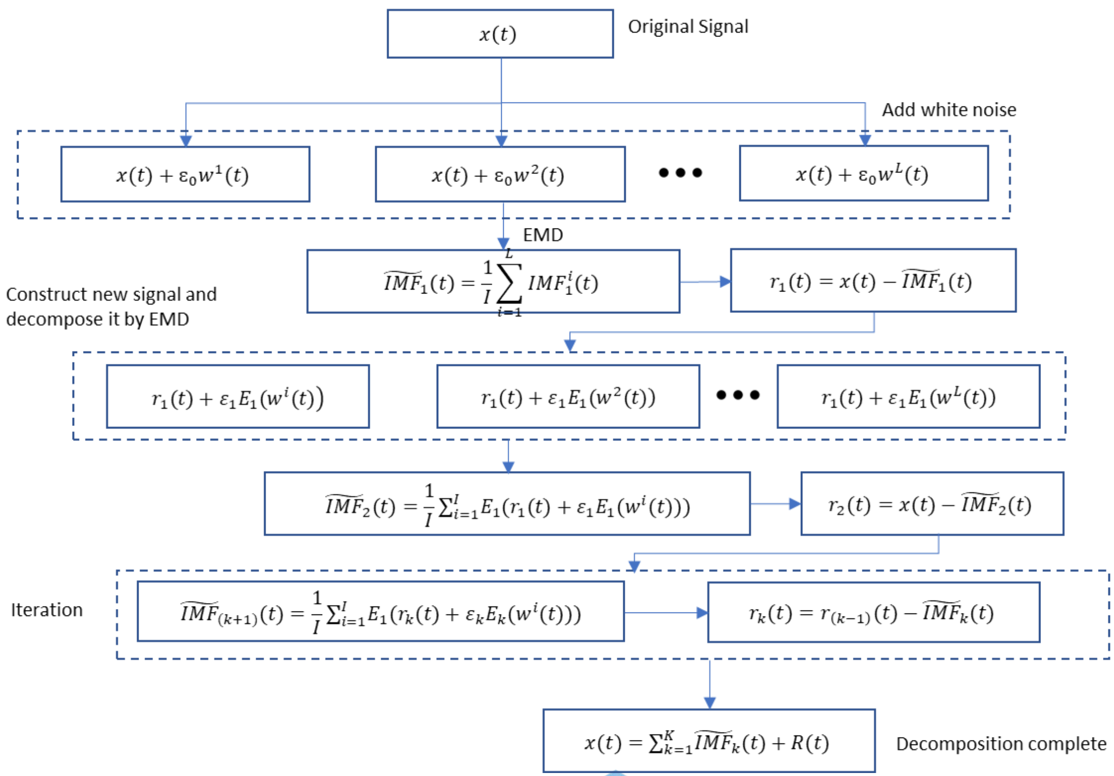

Figure 1 Flowchart of CEEMDAN

2.2 Empirical Wavelet Transform (EWT)

This study employs the Empirical Wavelet Transform (EWT), a data-driven preprocessing approach. EWT dynamically builds empirical scaling and empirical wavelet functions from the signal's inherent frequency spectrum. Its core principle involves calculating the Fourier spectrum [33] and subsequently generating wavelet filters to decompose the original signal into distinct modes [34]. The flowchart of the Empirical Wavelet Transform technique is depicted in Figure 2 [35].

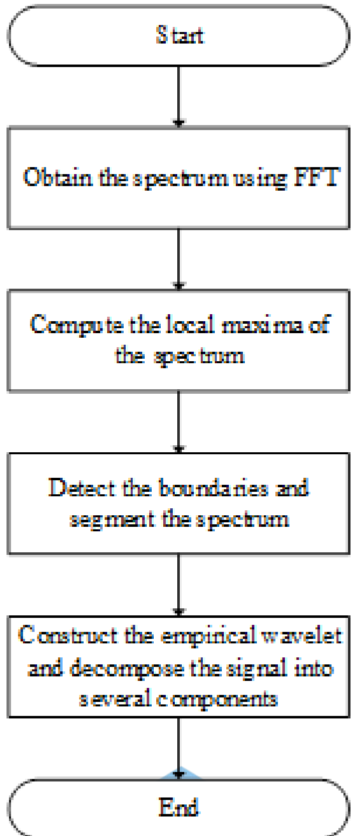

Figure 2 Flowchart of the EWT method [37]

2.3 Random Forest (RF)

RF is a composite of numerous decision trees designed to address classification and regression problems [38]. In contrast to machine learning approaches such as ANN or SVR, which construct a comprehensive model derived from the initial data, RF employs an ensemble learning method by creating multiple models and combining their results. This approach can enhance accuracy, particularly when dealing with complex systems [39]. Figure 3 illustrates the arrangement of RF for regression

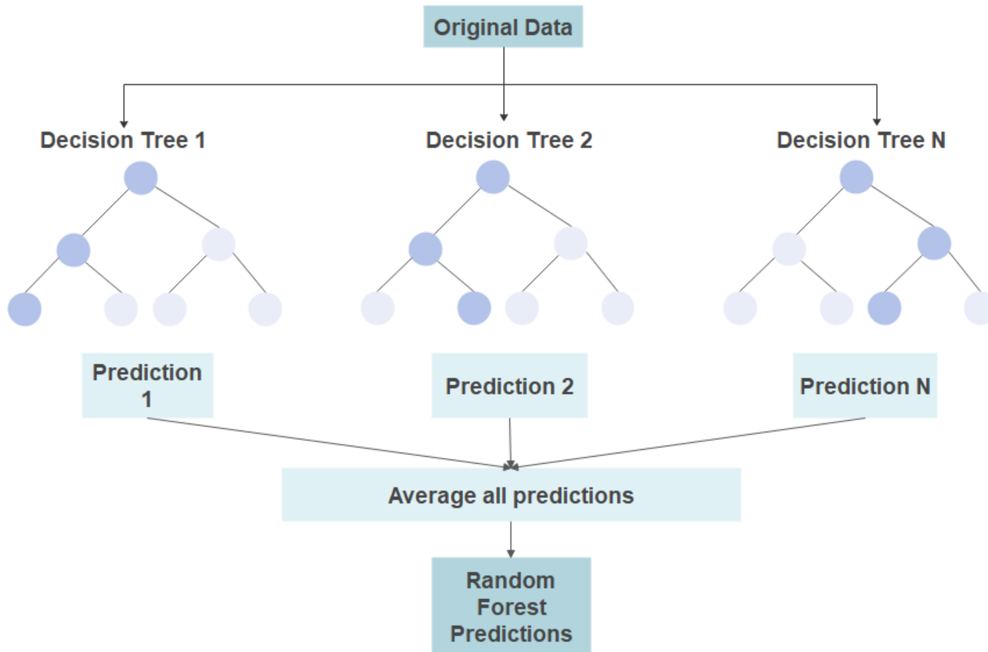

Figure 3 The configuration of Random Forest designed for regression tasks.

2.4 Long Short-Term Memory (LSTM)

The Long Short-Term Memory (LSTM) method represents a notable aspect of deep learning technique widely utilized in time series forecasting [44]. It is characterized by its exceptional memory capacity and proficiency in discerning regular patterns from historical data [45]. LSTM's incorporation of a "gates" mechanism sets it apart, which enhances the recurrent cell memory's core functions [48]. Through this gate mechanism, LSTM acquires the ability to control the information flow[46]. As a result, it can retain essential information over extended periods while disregarding less relevant historical data within the time series. Figure 4 illustrates the structural components of the LSTM cell, including the input gate, output gate, and forget gate [47].

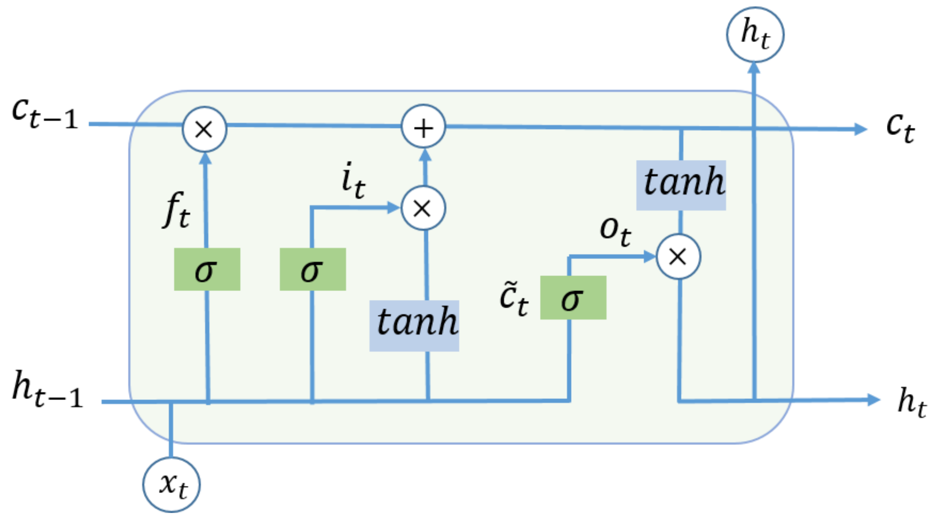

Figure 4. The configuration of LSTM cell

The formulations for the LSTM model are provided below [48]:

$$f_t = \sigma\left(W_f.\left[h_{t-1}, x_t\right] + b_f\right) \tag{1}$$

$$i_t = \sigma(W_i.\left[h_{t-1}, x_t\right] + b_i) \tag{2}$$

$$o_t = \sigma(W_o.\left[h_{t-1}, x_t\right] + b_o) \tag{3}$$

$$\tilde{c}_t = tanh(W_C.\left[h_{t-1}, x_t\right] + b_C) \tag{4}$$

$$c_t = f_t \otimes c_{t-1} + i_t \otimes \tilde{c}_t \tag{5}$$

$$h_t = o_t \otimes \tanh(c_t) \tag{6}$$

where $W_f, W_t,\ W_c\ ,W_\sigma$ are the set of weights, $b_f, b_i, b_c,\ b_o$ are the corresponding bias vectors and $\otimes$ are element-wise multiplication.

CHAPTER 3

BUILDING ENERGY CONSUMPTION FORECASTING

This chapter introduces a novel hybrid approach to improve energy forecast accuracy on the demand side, particularly for building energy consumption forecasting. In this section, we introduce a combined approach that integrates CEEMDAN-Random Forest (RF) and Long-Short Term Memory (LSTM) techniques to enhance the precision of forecasting building energy consumption.

3.1 Introduction to Building Energy Consumption Forecasting

The building sector plays a dominant role in global energy consumption[49], accounting for a substantial share (35%) and contributing significantly to CO2 emissions (38%) [50]. Since 2000, building energy use has experienced a steady climb, averaging an annual rise of 1.1%, a trajectory projected to continue in the foreseeable future [51,52]. This escalation is driven by population growth and desires for comfortable living spaces [53,54], prompting concerns about potential supply constraints and environmental ramifications [55]. In light of these conditions, fostering energy efficiency within the building sector becomes critical. Precisely predicting energy consumption is fundamental, providing the cornerstone for building energy management systems [56]. The field of energy forecasting encompasses a spectrum of temporal horizons, ranging from long-term (exceeding one year) to very short-term (spanning minutes to less than an hour)[57–59]. This work focuses specifically on enhancing building energy consumption forecasts for one hour ahead.

Throughout the years, a variety of methods have been employed for forecasting energy consumption, encompassing statistical methods [60]. However, these methods encounter limitations when applied to non-linear time series data. Therefore, machine learning approaches have emerged as powerful alternatives [61,62], have emerged as powerful alternatives. Predicting building-level energy consumption poses a notable difficulty due to the data's inherent nonstationary and highly volatile nature [77]. As previously discussed, this characteristic renders single forecasting techniques insufficient for capturing all relevant patterns, ultimately hindering their ability to achieve accurate predictions. To address this limitation, researchers have explored the potential of hybrid methods that integrate decomposition techniques with forecasting techniques to enhance forecasting accuracy. The effectiveness of this approach has been demonstrated in various studies [63–65]. Empirical mode decomposition (EMD) has established itself as an effective method for handling complex data[27]. Preprocessing the original data using EMD can significantly improve the performance of subsequent forecasting methods [63–65]. EMD's efficiency is hampered by mode mixing [66] and to overcome the challenge, Ensemble Empirical Mode Decomposition (EEMD) was developed [67]. EEMD resolves the mode mixing problem by introducing additional noise. However, the resulting decomposition suffers from residual noise contamination, especially when the quantity of ensemble trials is limited [68]. In response, Torres et al. [32] proposed the Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) algorithm, which effectively improves decomposition results and reduces computational time by incorporating dynamic noise at each step of process EMD [32].

While previous research has explored hybrid forecasting methods leveraging decomposition techniques, these approaches often neglect the unique characteristics of individual components. They typically employ a single forecasting method for all components, potentially hindering accuracy. This study addresses this research gap by taking into accounts the distinct properties of each series and strategically utilizing various forecasting methods for optimal results. We propose a hybrid approach based on Long Short-Term Memory networks and Random Forest with CEEMDAN to improve building energy consumption forecasting accuracy. This method decomposes the initial data into several distinct series. Subsequently, we analyze utilize the attributes of each component tailored forecasting methods for each, maximizing the accuracy of the overall prediction.

Our approach initiates with decomposing the initial data into multiple components using the Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) algorithm. Compared to other decomposition methods like EMD and EEMD, CEEMDAN has demonstrated superior performance in handling non-stationary data[69], leading to its selection for this study. Following the decomposition, distinct forecasting methods are employed based on the characteristics of the extracted components. The first component, characterized by its highest frequency, is modeled and forecasted using Random Forest (RF). Long Short-Term Memory (LSTM) networks are then utilized to forecast the remaining components. Finally, the features of each component are combined through summation to generate the ultimate prediction. As far as we know, this is the first investigation into applying a hybrid CEEMDAN-RF-LSTM methodology for building energy consumption forecasting.

3.2 Building Energy Consumption Data

This research leverages a publicly available dataset curated by the Building Data Genome Project [70]. The project measured total electricity usage data from the building's main meter., encompassing the heating system's energy consumption [86]. Energy consumption data with hourly intervals from five distinct buildings for March-May 2015 was analyzed. Figure 5 depicts the hourly energy consumption profiles of each building. As evident from Figure 5, each building exhibits unique energy consumption patterns characterized by randomness and nonlinearity.

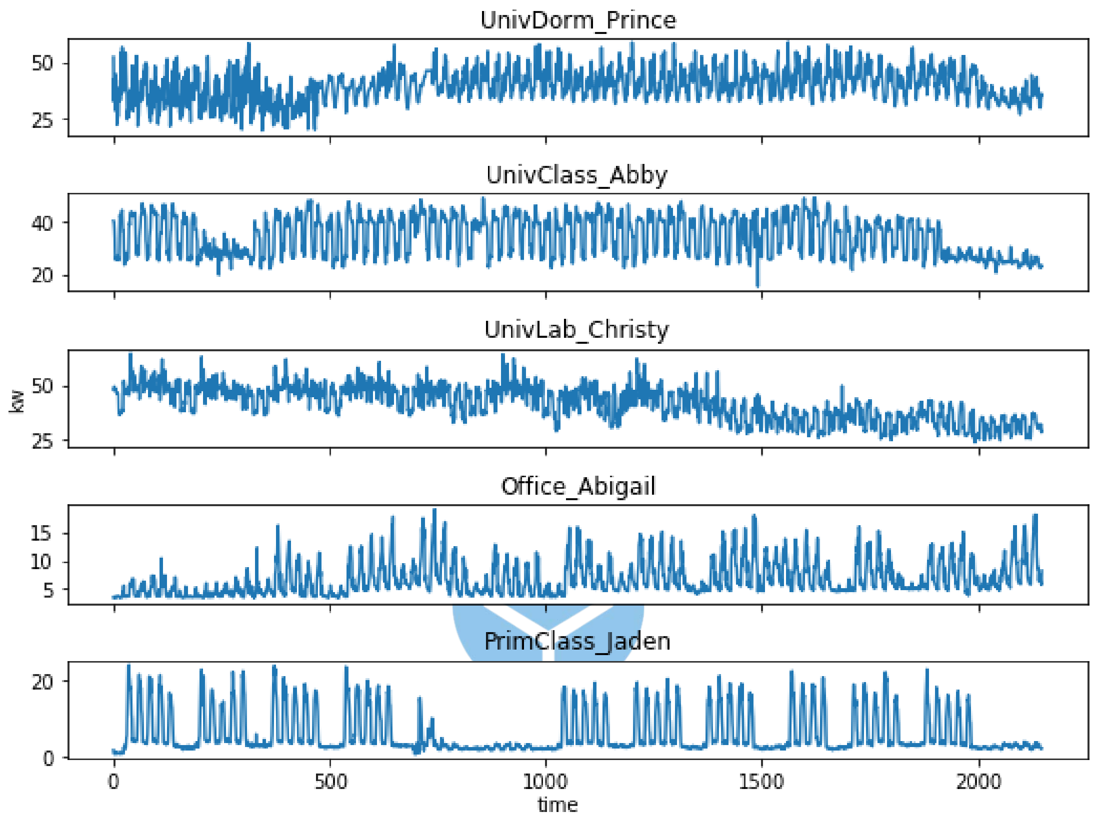

Figure 5 Energy consumption profiles for five distinct buildings, measured at hourly intervals.

3.3 Framework of the Proposed Hybrid CEEMDAN-RF-LSTM Approach

This study introduces a hybrid forecasting approach for building energy consumption, leveraging the combined power of three techniques: Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN), Random Forest (RF), and Long Short-Term Memory (LSTM). CEEMDAN first decomposes the original data into a set of Intrinsic Mode Function (IMF) components representing different frequency bands. RF, with its robust handling of complex and noisy data [39,71], is employed to forecast the most volatile component (IMF1). LSTM, known for its proficiency in handling periodic patterns [72,73], is then utilized to predict the remaining periodic components. Finally, LSTM, adept at learning trends [74], also forecasts the long-term trend represented by the residual component. The individual forecasts are then combined to acquire the ultimate forecast. Figure 6 depicts the diagram of the proposed hybrid CEEMDAN-RF-LSTM.

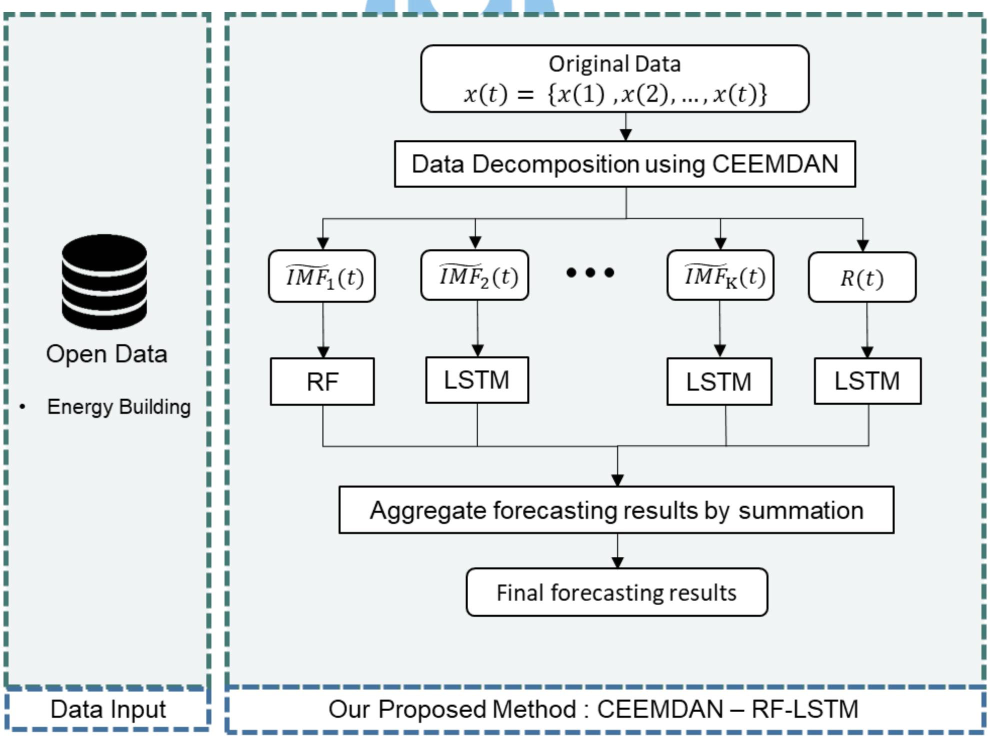

Figure 6 Diagram of Proposed Hybrid CEEMDAN-RF-LSTM method

3.4 Evaluation Metrics for Energy Building Forecasting

We evaluate the performance of our model using three common metrics: Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE):

$$MAPE\% = \frac{100}{n}\sum_{t=1}^{n}\left|\frac{y_t{}' - y_t}{y_t}\right| \tag{7}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t{}' - y_t)^2} \tag{8}$$

$$MAE = \frac{1}{n}\sum_{t=1}^{n}\left|y_t{}' - y_t\right| \tag{9}$$

where y_t is the actual value at time t and $y_t{}'$ is the predicted value at time t. The variable n indicates the total number of data points.

3.5 Experimental Setting for Energy Building Forecasting

This work develops a combination of forecasting method combining CEEMDAN, Random Forest (RF), and Long Short-Term Memory (LSTM) for building energy consumption. CEEMDAN decomposes the original data into intrinsic mode functions (IMFs) and residuals. We implemented CEEMDAN using the pyEMD package [75]. The RF and LSTM models were implemented using scikit-learn [76] and Keras [77], respectively. For RF, we used the default number of trees (100) [64],and the suggested number of features per node (8) [78]. For LSTM, we chose the Adam optimizer with the recommended learning rate (0.001) [79] due to its efficiency and performance [80,81]. We trained the model for 100 epochs [82] and used a batch size 64 [83]. After testing different configurations using grid search, it was observed that an LSTM with 64 hidden neurons yielded the optimal results. This study focuses on one-hour forecasting.(X_t), using the previous one-day energy consumption as input (X_{t-1} to X_{t-24}). The data was split into 80% training and 20% testing sets.

3.6 Experimental Results for Energy Building Forecasting

The proposed approach decomposes the initial hourly energy consumption data using CEEMDAN. Figure 7 visually depicts the results of decomposition for one of the buildings. Notably, the extracted intrinsic mode functions (IMFs) exhibit decreasing frequency, with IMF1 showcasing the highest frequency and IMF8 reflecting the overall trend. We used permutation entropy (PE) [100] to measure the complexities of each component generated from CEEMDAN. The PE values range from 0 to 1, with 0 indicating a highly regular time series with a repeating pattern.

Conversely, higher PE values imply increased randomness and complexity. As shown in Table 1, IMF 1 exhibits the highest PE at 1, suggesting its data is the most complex and unpredictable among the components. Consequently, we focus our special treatment solely on IMF 1. Therefore, we give special treatment solely to IMF 1 and not to other IMF components.

Table 1 PE values for IMFs generated from CEEMDAN

	IMF#1	IMF#2	IMF#3	IMF#4	IMF#5	IMF#6	IMF#7	IMF#8
Entropy	1.000	0.841	0.671	0.563	0.479	0.429	0.405	0.225

The number of Intrinsic Mode Functions (IMFs) produced by CEEMDAN can vary depending on the dataset characteristics. The local extrema drive the decomposition process and adaptively adjusts to the underlying oscillatory modes in the data. Different datasets with distinct characteristics will result in different sets of IMFs because the algorithm adapts to the unique features of each dataset.

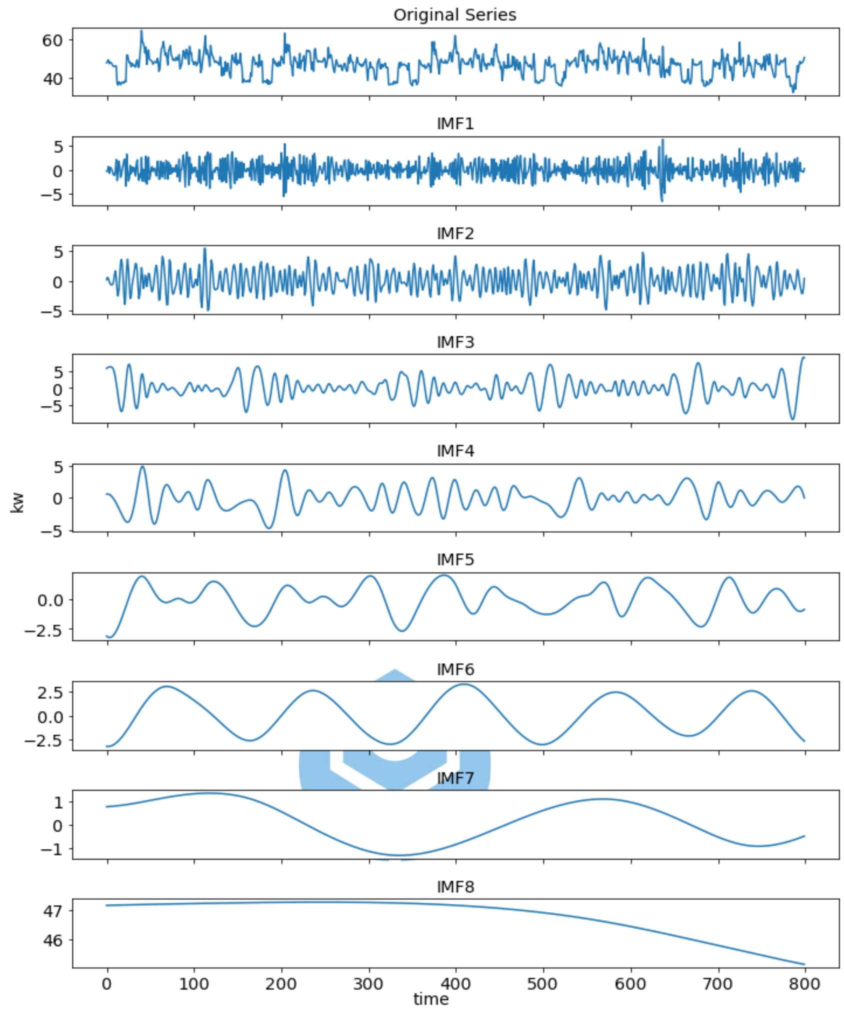

Figure 7 Result of decomposition for one of the buildings (University Laboratory)

Following the decomposition of the original data via CEEMDAN, the first Intrinsic Mode Function (IMF) is forecasted using Random Forest (RF). At the same time, the remaining IMFs are predicted using Long Short-Term Memory (LSTM). Ultimately, the predictions for each element are aggregated through summation to produce the ultimate forecast. As visualized in Figure 8, the proposed approach's forecast closely aligns with the actual values, demonstrating its high accuracy.

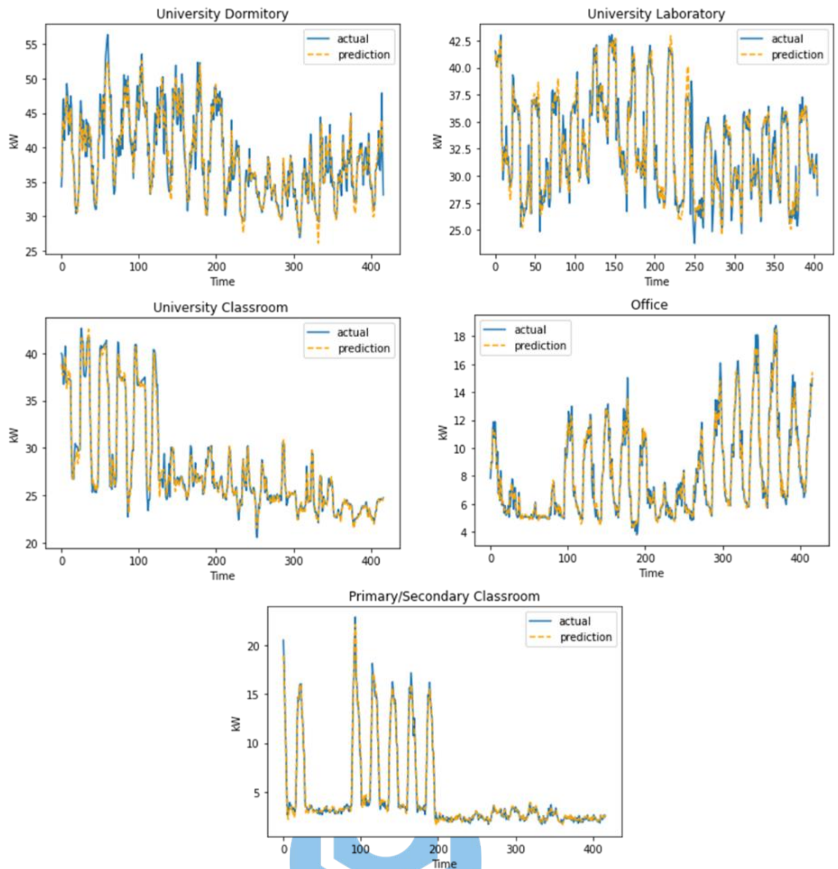

Figure 8 Forecasting results through the proposed hybrid approach combining RF and LSTM with the CEEMDAN method

The effectiveness of the proposed approach was evaluated against various established methods. As evidenced by Table 2, the proposed approach achieved the best performance and demonstrably outperformed all benchmark methods in forecasting accuracy.

Table 2 Energy Forecasting Results of different forecasting method

Building Type	**Evaluation Metrics**	**Methods**							
		Linier Regression	**Support Vector Regression**	**Artificial Neural Network**	**Random Forest**	**Long Short-Term Memory**	**CEEMDAN Random Forest**	**CEEMDAN LSTM**	**Proposed approach (CEEMDAN-RF-LSTM)**
Dormitory	MAPE (%)	6.10	6.49	6.72	6.15	6.83	3.96	4.08	**3.51**
	RMSE	3.09	3.39	3.39	3.17	3.47	2.01	2.07	**1.76**
	MAE	2.40	2.59	2.64	2.44	2.68	1.55	1.59	**1.37**
	Running Time (s)	*0.34*	*0.38*	*1.94*	*1.12*	*11.99*	*37.39*	*105.33*	*103.59*
Laboratory	MAPE (%)	5.86	6.42	6.14	6.27	6.39	3.57	3.51	**3.19**
	RMSE	2.58	2.71	2.66	2.72	2.63	1.47	1.43	**1.29**
	MAE	1.87	2.09	2.01	2.05	2.06	1.13	1.12	**1.01**
	Running Time (s)	*0.24*	*0.38*	*1.53*	*1.12*	*6.68*	*31.27*	*86.68*	*75.81*
Classroom	MAPE (%)	3.66	5.81	5.98	6.87	5.17	3.08	2.12	**1.97**
	RMSE	1.63	2.53	2.40	2.68	2.17	1.33	0.87	**0.82**
	MAE	1.06	1.78	1.75	2.05	1.53	0.92	0.61	**0.57**
	Running Time (s)	*0.54*	*0.35*	*1.94*	*1.16*	*9.04*	*32.14*	*85.87*	*81.06*
Office	MAPE (%)	11.17	9.31	11.29	10.25	9.79	6.17	6.35	**5.33**
	RMSE	1.22	1.11	1.15	1.08	1.12	0.65	0.67	**0.57**
	MAE	0.90	0.77	0.88	0.81	0.81	0.49	0.50	**0.43**
	Running Time (s)	*0.28*	*0.29*	*1.64*	*1.16*	*6.34*	*32.36*	*83.04*	*81.80*
Primary/ Secondary Classroom	MAPE (%)	15.51	14.67	14.46	18.40	16.65	10.94	8.55	**7.16**
	RMSE	1.04	1.15	0.90	1.34	0.96	0.75	0.52	**0.47**
	MAE	0.65	0.65	0.58	0.79	0.62	0.46	0.34	**0.30**
	Running Time (s)	*0.32*	*0.17*	*0.83*	*1.22*	*7.89*	*42.32*	*98.65*	*95.25*

**The optimal outcome for each metric for each building is represented by bold values.*

In order to assess the enhanced performance of the proposed approach compared to other benchmark methods, this study employs three improvement percentage metrics [84]. The calculation of the enhancement in MAPE, RMSE, and MAE between two methods is as follows:

$$P_{\mathrm{MAPE}} = \left|\frac{MAPE_1 - MAPE_2}{MAPE_1}\right| \tag{10}$$

$$P_{\mathrm{RMSE}} = \left|\frac{RMSE_1 - RMSE_2}{RMSE_1}\right| \tag{11}$$

$$P_{\mathrm{MAE}} = \left|\frac{MAE_1 - MAE_2}{MAE_1}\right| \tag{12}$$

An overview of the error improvement percentages in comparison to other benchmarking techniques can be found in Table 3.

Table 3 Percentage enhancement of Hybrid RF-LSTM using CEEMDAN in comparison to alternative benchmark methods

Building Type	**Improvement Percentage Metrics**	**Proposed approach vs. Linier Regression**	**Proposed approach vs. Support Vector Regression**	**Proposed approach vs. Artificial Neural Network**	**Proposed approach vs. Random Forest**	**Proposed approach vs. Long Short-Term Memory**	**Proposed approach vs. CEEMDAN RF**	**Proposed approach vs. CEEMDAN LSTM**
Dormitory	P_{MAPE} (%)	42.4	45.87	47.24	42.86	48.58	11.22	13.95
	P_{RMSE} (%)	43.03	47.99	48.04	44.48	49.22	12.28	14.91
	P_{MAE} (%)	42.97	47.13	48.14	43.79	48.86	11.7	13.
Laboratory	P_{MAPE} (%)	45.54	50.3	47.99	49.14	50.03	10.54	9.08
	P_{RMSE} (%)	49.86	52.22	51.35	52.37	50.77	11.78	9.49
	P_{MAE} (%)	45.6	51.54	49.4	50.41	50.84	10.37	9.09
Classroom	P_{MAPE} (%)	46.25	66.17	67.16	71.01	61.99	36.29	7.46
	P_{RMSE} (%)	49.93	67.76	66.05	69.22	62.41	38.79	5.84
	P_{MAE} (%)	46.32	67.98	67.40	71.75	62.76	38.03	7.06
Office	P_{MAPE} (%)	52.27	42.72	52.79	48.00	45.56	13.53	16.10
	P_{RMSE} (%)	53.17	48.44	50.34	47.38	49.24	12.21	14.70
	P_{MAE} (%)	52.27	44.08	50.89	46.64	46.58	12.40	13.79
Primary/ Secondary Classroom	P_{MAPE} (%)	53.81	51.17	50.47	61.06	56.98	34.54	16.21
	P_{RMSE} (%)	55.14	59.43	48.26	65.23	51.24	37.75	10.89
	P_{MAE} (%)	54.07	53.81	48.26	61.96	52.01	34.40	12.62

Derived from the performance outcomes presented in Table 2 and Table 3, several key observations can be made:

1. Hybrid methods leveraging CEEMDAN consistently outperform single forecasting methods. This demonstrates the effectiveness of CEEMDAN in enhancing forecasting

accuracy by decomposing the original data and addressing its non-stationary behavior. Forecasting accuracy can be significantly improved by transforming the non-stationary data into more predictable components.

2. Our proposed approach outperforms other hybrid CEEMDAN methods that utilize identical forecasting methods for all components. This highlights the effectiveness of combining diverse forecasting methods tailored to the specific characteristics of each IMF component, further enhancing overall forecasting accuracy.
3. Hybrid CEEMDAN methods require longer running times compared to single methods. This is primarily due to the additional processing required for data decomposition and the use of multiple forecasters for each component.
4. Our proposed approach achieves faster execution and superior forecasting accuracy compared to CEEMDAN-LSTM. This suggests its suitability as a tool for building energy consumption forecasting. While this study utilizes a CPU-based environment, the development of GPU computing offers the potential to significantly accelerate deep learning methods like LSTM (up to 45 times faster than CPU implementation) [85]. Therefore, utilizing GPUs can further improve the running time of our proposed approach.

3.7 Energy Building Forecasting Summary

Predicting building energy consumption is crucial for optimizing energy efficiency and informed decision-making. This study introduces a hybrid forecasting method RF and LSTM powered by CEEMDAN. This method decomposes the original hourly data into multiple components, allowing different forecasting approaches tailored to each component's characteristics. RF is utilized for the highly volatile first component, while LSTM handles the remaining components. Finally, the individual component forecasts are combined to obtain the final prediction. This strategy leverages the strengths of both RF and LSTM, leading to improved forecasting accuracy. The effectiveness of the proposed approach was validated using hourly data from five diverse buildings, surpassing the performance of benchmark methods. Currently, the forecasting relies solely on past energy values (univariate). Future work will explore incorporating exogenous variables like weather, building operation schedules, and temporal factors into the model. Additionally, further research will investigate the impact of data set size and granularity (daily, minutely) on forecasting performance.

CHAPTER 4

WIND POWER FORECASTING

In this chapter, we present an innovative hybrid method to enhance wind power forecast accuracy, especially within the context of energy forecasting on the supply side. The approach we introduce combines CEEMDAN-Empirical Wavelet Transform (EWT) with Long-Short Term Memory (LSTM) models, strategically crafted to improve the accuracy of wind power forecast.

4.1 Introduction to Wind Power Forecasting

Driven by a surging population, global energy consumption is anticipated to witness a 48% increase over the next two decades [2]. However, the reliance on fossil fuels, the current predominant energy source, is constrained by limitations and has led to elevated energy costs and scarcities[86]. Consequently, renewable energy sources are gaining prominence as alternatives to conventional fossil fuels[87], aiming to reduce dependency and mitigate global warming risks [88]. The global shift towards renewables, coupled with energy efficiency enhancements, is predicted to elevate the portion of renewable energy in the worldwide primary energy supply from 14% in 2015 to 63% by 2050. Among these, wind energy stands out as a swiftly expanding renewable source, driven by cost reductions [89] and ease of large-scale installation [90].

While wind energy offers notable advantages, it is not without its challenges. Wind power output is heavily influenced by natural factors, resulting in highly intermittent and fluctuating energy generation [91]. These characteristics pose technical challenges, such as issues related to grid connectivity, reliability of power, and control of generation [91], thereby increasing the susceptibility of power systems [92]. Accurate wind power forecasting is essential for handling the variability and unpredictability associated with wind energy. This functionality enables power systems to make informed decisions regarding the generation, storage, and dispatch of power. Therefore, precise wind power predictions yield substantial economic benefits and technical advantages [94].

Wind power data exhibits significant unpredictability and fluctuation because of the intermittent characteristics of wind energy [95]. Given these characteristics, obtaining accurate wind power predictions using a single method is challenging. Consequently, researchers have proposed hybrid approaches, integrating data preprocessing strategies employed in conjunction with artificial intelligence methods. The decomposition-based method is widely utilized, yielding favorable forecasting outcomes [96]. Within hybrid methods that rely on decomposition, the decomposition techniques are applied to break down initial data into multiple comparatively stationary components. Subsequently, a prediction model is constructed for each component, and their forecasting outcomes are aggregated to yield the final results. This decomposition strategy and separate prediction of each subseries significantly improve the precision of wind forecasting.

Various decomposition-based hybrid approaches have been applied in wind power forecasting, including techniques such as EMD) [97,98], EEMD [99], CEEMD [100,101], and CEEMDAN [21,23,102]. The experimental results indicate the effectiveness of these decomposition-based hybrid methods in enhancing forecasting accuracy, surpassing the performance of individual models. For example, Wang et al. [98] employed EMD as a preprocessing tool, combining it with ENN for wind speed forecasting. EMD breaks down the original data, and ENN constructs prediction models for every component. Their results indicated that EMD-ENN improves wind speed forecasting accuracy. Ren et al. [22] utilized EMD and its enhanced version to decompose wind data, employing ANN and SVR to construct forecasting methods. Their study demonstrated that CEEMDAN outperformed EMD, EEMD, and CEEMD[22].

Despite advancements in integrating decomposition techniques with AI algorithms for wind power forecasting, further performance improvements remain crucial. The high volatility and noise present in the highest-frequency component generated by CEEMDAN decomposition [24] pose a significant challenge, as its unpredictable nature hinders accurate forecasting [25]. Existing methods often overlook this issue, which relies on a single decomposition technique, such as EMD [98] or EEMD [101]. This results in limitations in forecasting accuracy due to inadequate handling of the complexities within the highest frequency component. To address this challenge, we propose applying the EWT denoising method specifically to the highest frequency series obtained from CEEMDAN. EWT's ability to remove noise and extraneous information from data (e.g., fault diagnosis [103], ECG denoising [36], seismic data analysis [104]) has been well established across various applications. Its effectiveness in denoising wind speed data is further evidenced by studies [105–107], suggesting its potential for enhancing wind power forecasting accuracy.

Hence, to address the previously mentioned challenges, this study proposes a novel method of CEEMDAN, EWT, and deep learning with LSTM. The original wind power data is decomposed into multiple subseries using CEEMDAN, followed by applying EWT to remove noise specifically from the highest frequency component. This targeted noise reduction further facilitates the forecasting process. Subsequently, each individual component undergoes separate forecasting using an LSTM network. Finally, the forecasting outcomes of each component are combined to produce the ultimate prediction. Notably, our implementation of this novel hybrid approach, combining CEEMDAN, EWT, and LSTM, represents a unique contribution to the field of very-short-term wind power prediction.

4.2 Wind Power Dataset

The efficacy of the proposed hybrid approach is assessed utilizing wind power from real-world data of two geographically diverse countries: France and Turkey. Data from a 2050 kW French wind farm and a 3600 kW Turkish wind farm is utilized and collected at 10-minute intervals over a year [108]. The first 80% of data for each dataset constitutes the training set, while the remaining 20% serves as the testing set. The model is trained on the former and then evaluated on the latter to assess its forecasting performance. Figure 9 presents sample wind power observations from both datasets for July. These plots highlight the non-stationary and non-linear characteristics inherent to wind power data, which pose significant challenges for accurate forecasting.

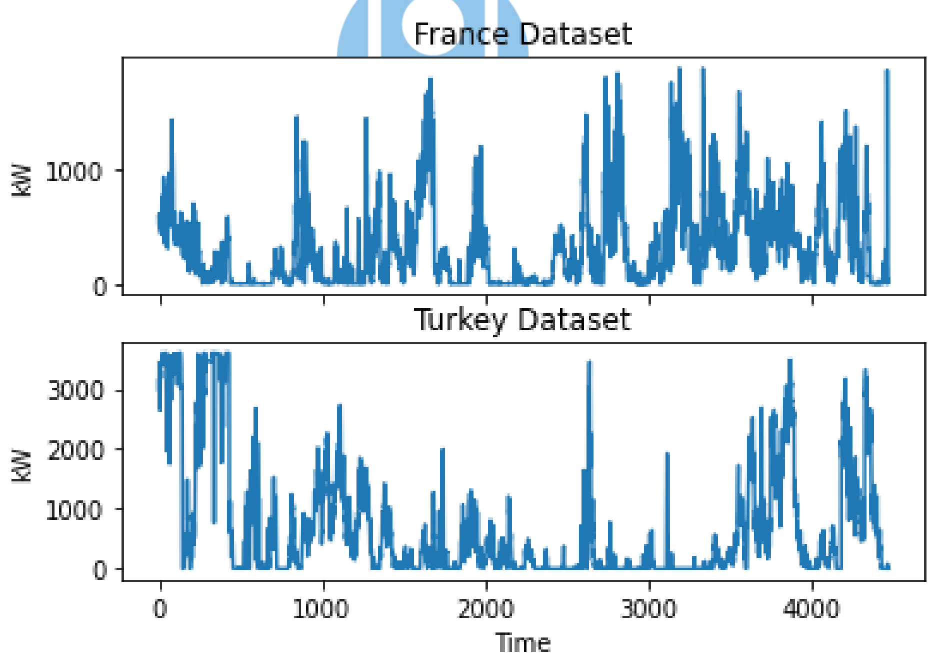

Figure 9 Wind energy data extracted from the July datasets of France and Turkey.

4.3 Structure of the Proposed CEEMDAN-EWT-LSTM Approach

This research introduces a hybrid approach for very-short-term wind power prediction, employing CEEMDAN EWT-LSTM. Figure 9 illustrates the overall structure of the proposed approach.

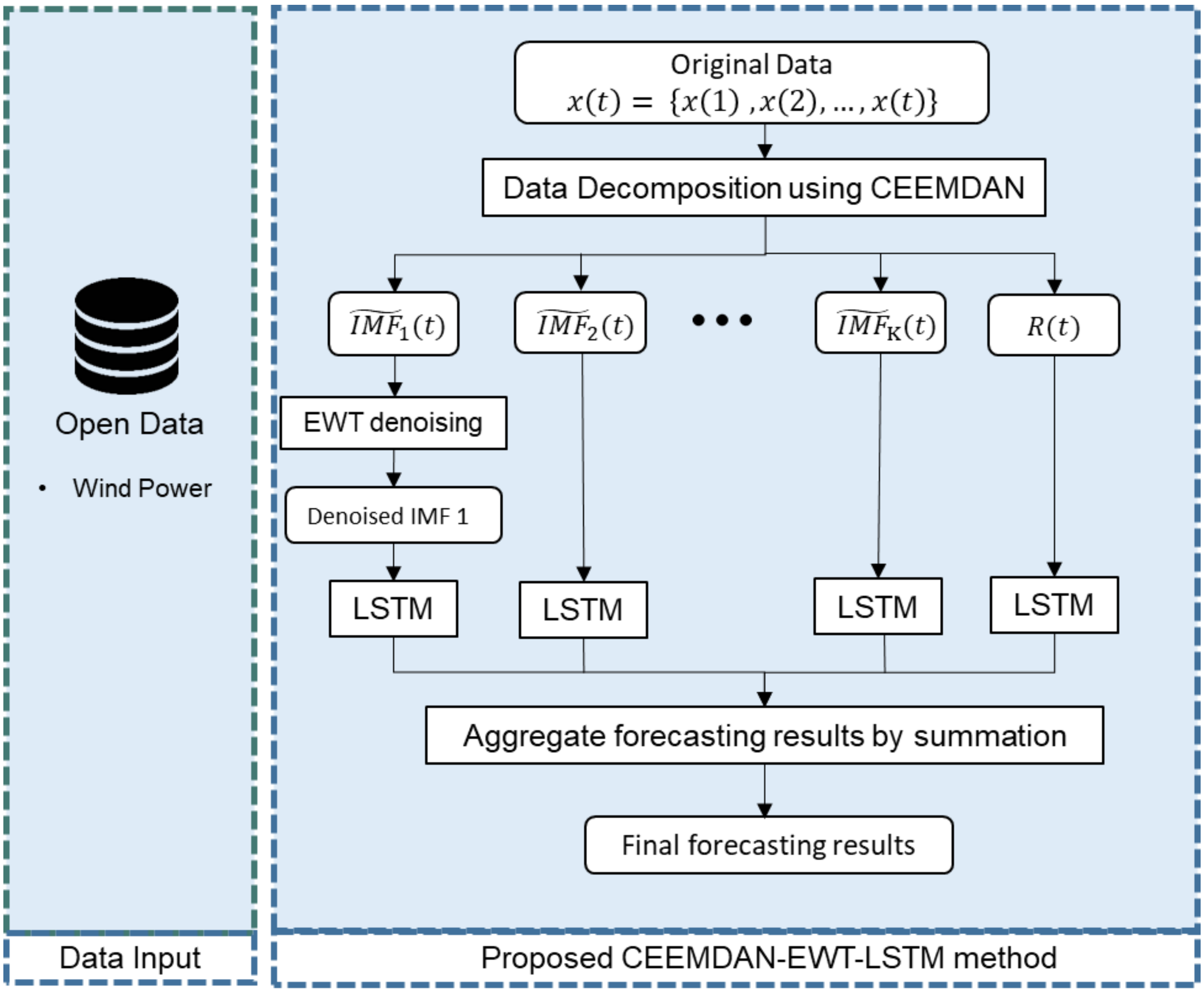

Figure 10 The general structure of the proposed CEEMDAN-EWT-LSTM method

The proposed approach employs a multi-stage framework for wind power forecasting:

- Stage 1 utilizes the CEEMDAN decomposition technique to breakdown the original data into multiple components, known as Intrinsic Mode Functions (IMFs). This approach is chosen due to its superior performance and computational efficiency compared to alternative techniques [109].

- In Stage 2, the first IMF, characterized by its high level of irregularity, undergoes specific denoising using Empirical Wavelet Transform (EWT). This process separates meaningful information from noise within the first IMF, ultimately yielding a smoother and more predictable series.
- Stage 3 leverages the Long Short-Term Memory (LSTM) forecasting model to predict all the components obtained through the CEEMDAN-EWT process.
- Finally, Stage 4 aggregates the individual forecasts to generate the final wind power prediction.

4.4 Experimental Setting Wind Power Forecasting

In this study, CEEMDAN [46] was implemented using the pyEMD package, while EWT was implemented using the ewtpy [47]. Keras [48] served as the tool for LSTM implementation. For LSTM configuration, the Adam optimizer was chosen due to its demonstrated effectiveness compared to the alternative stochastic optimization method [49,50]. Initial grid search experiments were executed to identify the optimal hidden neurons and size for the batch of the proposed approach [110,111]. Different neurons and batch size configurations were evaluated, including 32, 64, and 128 values. The findings indicated that the setup comprising 128 hidden neurons and 64 as the batch size exhibited superior performance compared to other configurations. Consequently, this configuration was chosen for training the proposed approach. The forecasting in this study is conducted for ten minutes ahead (X_t) utilizing the preceding one-hour data $(X_{t-1}$ to $X_{t-6})$ as the input for the prediction model

4.5 Evaluation Metrics Wind Power Forecasting

In this investigation, three prevalent assessment criteria were employed, namely Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and Mean Absolute Error (MAE) [112]. The definitions for these three metrics are provided as follows [112]:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2} \tag{13}$$

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_i^{pre} - y_i}{y_{max}}\right| \tag{14}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|y_i^{pre} - y_i\right| \tag{15}$$

Here y_i represents the actual wind power value at time i and $\hat{y}_i$ is the forecasted wind power value at a time and represents the predicted value at time i. N signifies the total number of data points, and y_{max} is the maximum value within the entire dataset. The calculation for percentage improvements is outlined as follows [72]:

$$P_{RMSE} = \frac{RMSE_1 - RMSE_2}{RMSE_1} * 100\% \tag{16}$$

$$P_{MAPE} = \frac{MAPE_1 - MAPE_2}{MAPE_1} * 100\% \tag{17}$$

$$P_{MAE} = \frac{MAE_1 - MAE_2}{MAE_1} * 100\% \tag{18}$$

4.6 Experimental Results Wind Power Forecasting

Given wind data's pronounced nonlinearity and nonstationary nature [131], CEEMDAN decomposes the initial data into several Intrinsic Mode Functions (IMFs) characterized by reduced nonlinearity and nonstationary traits. Figure 11 illustrates the IMFs derived from CEEMDAN in the France dataset. As depicted in Figure 11, the initial IMF exhibits the highest frequency, while the final IMF component represents the lowest frequency, resembling the overall trend of the data.

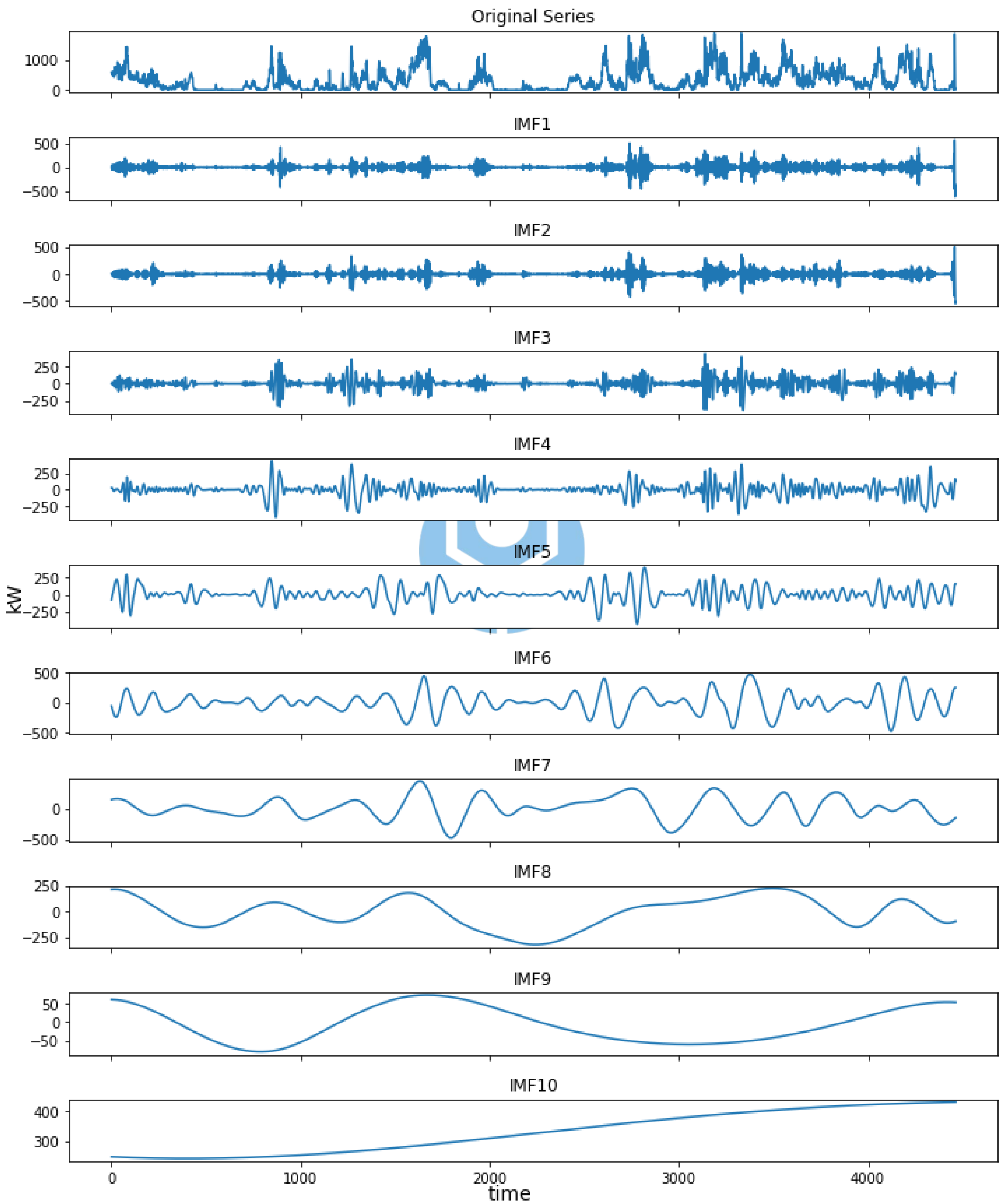

Figure 11 Results of the decomposition process for the July dataset in the France dataset

As IMF 1 represents the most chaotic and irregular component, it can adversely impact the forecasting model's precision and stability. Then, to alleviate forecasting challenges, the EWT denoising technique is implemented to refine and reduce the randomness and fluctuations in the first IMF. This denoising process aims to enhance the model's learning capability and facilitate more effective series modeling. Following the data preprocessing phase, LSTM is employed to predict all the components, and their cumulative sum yields the final forecasting outcomes. Figure 12 and Figure 13 illustrate the outcomes derived from the proposed approach. Notably, the forecasting lines in both figures closely align with the actual values, displaying minimal deviations, indicative of the method's accurate forecasting capabilities.

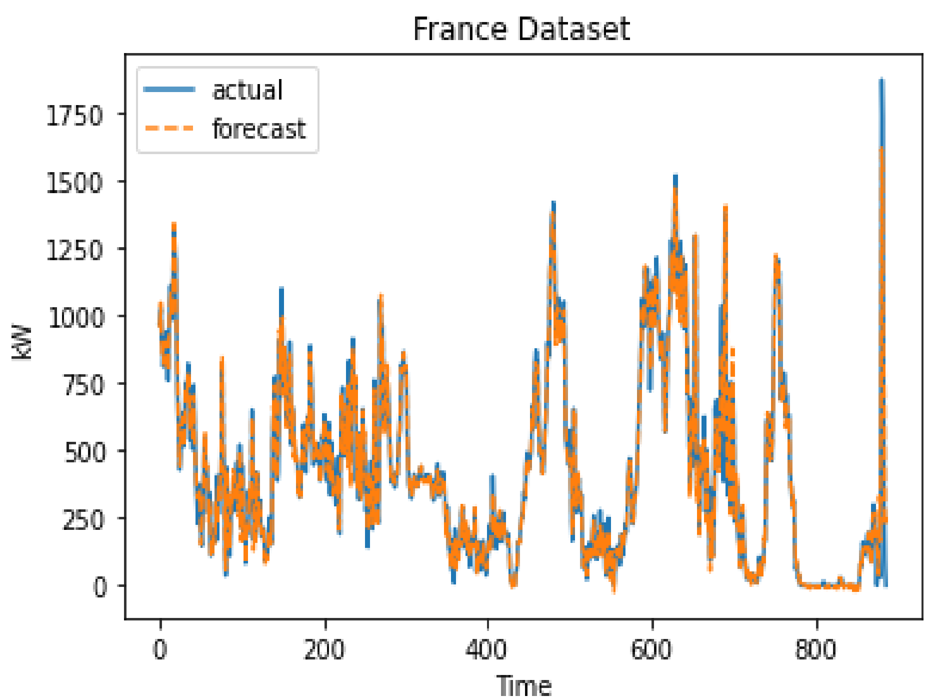

Figure 12 Actual and forecast values in the France dataset for the month of July

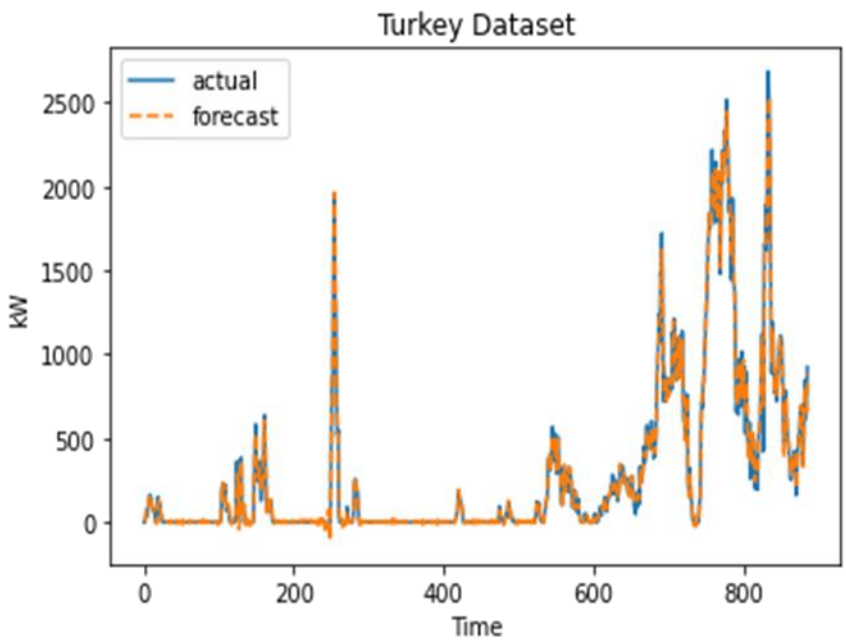

Figure 13 Actual and forecast values in the France dataset for the month of July

To gauge the effectiveness of our proposed approach, we compared its performance to seven other forecasting methods. The first four methods (SVR, ANN, RF, and LSTM) are standalone techniques, while the remaining three are hybrid decomposition approaches. These hybrid methods employ a single decomposition technique followed by independent LSTM predictions for each decomposed component. Table 4 and Table 5 present a comparative analysis of the results obtained for the France and Turkey datasets. In addition, Table 6 andTable 7 showcase the improvement percentages over other benchmarks, visually illustrating the superior forecasting capabilities of our proposed approach.

Table 4 Accuracy of various forecasting methods within the dataset in France

		Forecasting Methods							
Month	**Evaluation Metrics**	**Support Vector Regression**	**Artificial Neural Network**	**Random Forest**	**Long Short-Term Memory**	**EMD LSTM**	**EEMD LSTM**	**CEEMDAN LSTM**	**Proposed approach**
Jan	MAPE	5.46	3.72	4.41	3.44	1.66	1.68	1.58	**1.19**
	RMSE	165.74	112.39	122.09	106.85	53.00	53.22	58.54	**35.12**
	MAE	111.95	76.30	90.36	70.50	33.96	34.48	32.35	**24.36**
Running Time (s)		*1.20*	*3.45*	*4.98*	*37.48*	*810.61*	*971.60*	*832.52*	*1805.65*
Feb	MAPE	8.92	6.08	8.54	5.87	3.02	2.90	2.84	**2.54**
	RMSE	232.90	179.06	216.59	169.86	98.89	91.40	94.51	**76.19**
	MAE	183.00	124.58	175.16	120.35	61.91	59.47	58.15	**52.14**
Running Time (s)		*0.65*	*2.11*	*2.10*	*22.61*	*589.73*	*667.69*	*680.7*	*872.34*
Mar	MAPE	3.26	3.16	3.19	2.84	1.25	1.35	1.16	**1.03**
	RMSE	119.93	98.79	105.40	89.87	38.58	43.23	38.72	**34.90**
	MAE	66.93	64.73	65.35	58.21	25.58	27.66	23.78	**21.13**
Running Time (s)		*0.87*	*1.54*	*2.59*	*22.01*	*623.05*	*665.18*	*771.78*	*788.24*
Apr	MAPE	2.98	3.01	2.97	2.96	1.83	1.49	1.53	**1.13**
	RMSE	118.17	111.57	112.85	115.68	62.55	51.99	61.69	**41.24**
	MAE	61.06	61.69	60.93	60.66	37.45	30.47	31.28	**23.17**
Running Time (s)		*0.95*	*1.61*	*2.34*	*20.78*	*614.71*	*706.67*	*4750.5*	*836.33*
May	MAPE	3.69	3.38	3.81	3.20	1.88	1.73	1.82	**1.21**
	RMSE	106.82	104.98	111.25	103.98	67.35	61.98	65.54	**38.57**
	MAE	75.73	69.22	78.11	65.67	38.61	35.57	37.40	**24.76**
Running Time (s)		*2.13*	*2.29*	*2.89*	*22.85*	*640.60*	*721.05*	*849.10*	*721.346*
Jun	MAPE	6.49	5.13	6.18	4.90	2.92	2.73	2.76	**2.16**
	RMSE	180.15	165.79	172.94	160.27	109.33	98.41	103.01	**70.94**
	MAE	133.11	105.27	126.76	100.47	59.90	55.90	56.65	**44.37**
Running Time (s)		*0.89*	*2.19*	*2.62*	*25.01*	*684.86*	*771.77*	*786.21*	*913.81*
Jul	MAPE	5.72	4.59	5.41	4.40	2.17	2.31	2.26	**1.88**
	RMSE	164.08	142.84	151.88	137.11	68.91	89.03	91.22	**60.03**
	MAE	117.30	94.07	111.04	90.23	44.59	47.47	46.29	**38.62**
Running Time (s)		*0.89*	*2.19*	*2.62*	*25.01*	*684.86*	*771.76*	*786.22*	*913.81*
Aug	MAPE	2.48	2.53	2.47	2.31	1.17	1.23	1.09	**0.92**
	RMSE	90.13	90.75	90.62	87.00	45.25	46.19	47.18	**35.10**
	MAE	50.85	51.88	50.64	47.38	23.93	25.20	22.39	**18.86**
Running Time (s)		*1.26*	*3.62*	*4.91*	*35.95*	*726.84*	*863.85*	*873.41*	*4281.89*
Sep	MAPE	1.47	1.65	1.48	1.35	0.90	0.87	0.66	**0.57**
	RMSE	71.90	70.36	68.90	62.88	37.24	30.80	27.95	**23.80**
	MAE	30.18	33.85	30.41	27.73	18.48	17.87	13.52	**11.59**
Running Time (s)		*0.70*	*1.76*	*3.35*	*21.79*	*530.28*	*663.86*	*685.64*	*721.77*
Oct	MAPE	3.27	3.41	3.33	3.08	2.24	1.58	1.40	**1.20**
	RMSE	102.03	103.86	108.91	100.84	72.91	50.67	50.79	**40.58**
	MAE	67.03	69.90	68.29	63.26	45.90	32.43	28.65	**24.64**
Running Time (s)		*0.76*	*1.31*	*2.01*	*20.78*	*543.74*	*665.22*	*700.54*	*775.85*
Nov	MAPE	3.88	3.88	4.23	3.83	1.90	1.74	1.54	**1.47**
	RMSE	113.71	114.06	124.93	112.41	57.06	54.32	50.35	**44.07**
	MAE	79.67	79.51	86.68	78.64	38.87	35.62	31.51	**30.04**
Running Time (s)		*0.58*	*1.73*	*1.75*	*17.83*	*576.74*	*689.12*	*681.91*	*691.54*
Dec	MAPE	9.97	6.01	8.37	6.93	3.10	2.82	2.69	**2.20**
	RMSE	306.32	168.39	230.08	196.20	92.61	83.19	82.43	**64.74**
	MAE	204.36	123.33	171.57	142.13	63.60	57.90	55.21	**45.19**
Running Time (s)		*1.87*	*4.61*	*4.56*	*31.76*	*880.57*	*681.03*	*927.71*	*677.93*
Average	MAPE	4.80	3.88	4.53	3.76	2.00	1.87	1.78	**1.46**
	RMSE	147.66	121.90	134.70	120.25	66.97	62.87	64.33	**47.11**
	MAE	98.43	79.53	92.94	77.10	41.07	38.34	36.43	**29.91**

Table 5 Accuracy of various forecasting methods within the Turkey dataset

Month	Evaluation Metrics	Forecasting Methods							
		Support Vector Regression	**Artificial Neural Network**	**Random Forest**	**Long Short Term Memory**	**EMD LSTM**	**EEMD LSTM**	**CEEMDAN LSTM**	**Proposed approach**
Jan	MAPE	5.05	3.18	4.31	2.61	5.64	1.76	1.65	**1.41**
	RMSE	533.82	281.36	395.40	269.38	370.98	145.18	167.75	**121.68**
	MAE	182.66	115.16	155.96	94.57	203.93	63.73	59.85	**50.90**
Running Time (s)		*0.59*	*2.16*	*2.39*	*18.99*	*447.95*	*624.72*	*801.40*	*968.97*
Feb	MAPE	7.04	4.65	6.83	3.62	2.42	2.10	2.19	**1.64**
	RMSE	480.75	309.29	410.40	266.11	150.48	137.91	140.58	**102.72**
	MAE	254.80	168.44	247.07	131.10	87.48	76.10	79.12	**59.44**
Running Time (s)		*0.81*	*2.17*	*2.29*	*26.02*	*749.06*	*645.28*	*676.96*	*865.68*
Mar	MAPE	6.28	4.75	6.70	4.54	2.51	2.55	2.23	**1.83**
	RMSE	335.30	304.62	354.01	287.65	167.25	156.50	147.69	**115.76**
	MAE	227.22	171.77	242.38	164.13	90.82	92.20	80.65	**66.33**
Running Time (s)		*0.73*	*2.82*	*2.80*	*21.50*	*600.46*	*749.83*	*1074.62*	*840.16*
Apr	MAPE	4.29	3.12	4.87	2.71	1.56	1.41	1.19	**1.09**
	RMSE	206.86	189.64	228.97	171.71	94.01	81.74	72.49	**65.31**
	MAE	155.42	112.89	176.25	98.11	56.56	51.03	43.18	**39.42**
Running Time (s)		*0.61*	*1.95*	*2.15*	*20.93*	*571.06*	*694.41*	*708.09*	*716.77*
May	MAPE	9.71	6.32	8.66	6.24	3.11	2.84	2.72	**2.29**
	RMSE	473.05	322.12	406.70	318.82	164.97	161.55	156.07	**122.68**
	MAE	351.44	228.72	313.50	225.73	112.53	102.82	98.25	**82.80**
Running Time (s)		*0.53*	*1.37*	*1.48*	*16.87*	*559.41*	*651.76*	*796.86*	*704.89*
Jun	MAPE	10.17	7.12	9.18	6.87	3.17	3.10	3.27	**2.60**
	RMSE	520.86	381.57	434.38	354.32	159.36	158.45	187.77	**130.95**
	MAE	368.08	257.75	332.25	248.49	114.89	112.17	118.35	**93.96**
Running Time (s)		*1.46*	*9.21*	*9.34*	*35.43*	*560.06*	*483.48*	*538.01*	*589.81*
Jul	MAPE	1.81	1.81	1.78	1.85	1.17	0.99	0.83	**0.66**
	RMSE	140.61	131.74	134.84	127.04	71.18	57.95	58.57	**45.57**
	MAE	65.37	65.35	64.30	66.77	42.39	35.96	29.97	**24.00**
Running Time (s)		*0.62*	*1.28*	*1.77*	*15.94*	*508.37*	*558.76*	*635.84*	*630.21*
Aug	MAPE	6.71	4.31	6.41	3.87	2.21	1.80	1.56	**1.45**
	RMSE	420.21	225.34	341.90	209.83	111.19	105.23	92.73	**78.01**
	MAE	242.93	155.81	231.95	139.90	79.99	64.97	56.43	**52.60**
Running Time (s)		*0.72*	*1.41*	*2.12*	*15.97*	*495.88*	*567.39*	*587.74*	*605.62*
Sep	MAPE	10.23	4.72	8.58	4.45	2.43	1.98	1.66	**1.51**
	RMSE	637.89	249.90	469.70	241.14	146.52	110.71	99.24	**86.88**
	MAE	370.06	170.93	310.51	161.19	87.87	71.50	60.20	**54.58**
Running Time (s)		*0.54*	*1.19*	*1.54*	*15.05*	*468.14*	*511.31*	*493.16*	*505.71*
Oct	MAPE	3.81	4.22	3.98	3.49	2.44	1.76	1.51	**1.15**
	RMSE	204.47	235.81	215.02	200.51	123.42	93.83	88.45	**62.26**
	MAE	137.70	152.71	143.93	126.46	88.27	63.74	54.59	**41.46**
Running Time (s)		*0.52*	*1.48*	*1.61*	*14.55*	*416.28*	*446.46*	*528.75*	*586.41*
Nov	MAPE	10.03	2.50	6.03	2.18	1.48	1.47	1.25	**1.04**
	RMSE	944.06	226.74	475.99	212.76	105.67	108.54	99.66	**81.22**
	MAE	362.78	90.39	218.15	78.95	53.54	53.37	45.10	**37.64**
Running Time (s)		*0.56*	*1.35*	*1.58*	*14.16*	*442.97*	*524.07*	*538.07*	*612.31*
Dec	MAPE	2.16	1.22	2.74	1.03	1.00	0.88	0.57	**0.45**
	RMSE	342.58	125.07	286.66	120.84	81.45	56.91	46.03	**39.64**
	MAE	78.22	44.05	99.28	37.17	36.03	31.95	20.68	**16.14**
Running Time (s)		*0.41*	*0.74*	*1.37*	*17.51*	*469.97*	*621.38*	*619.85*	*665.46*
Average	MAPE	6.44	3.99	5.84	3.62	2.43	1.89	1.72	**1.43**
	RMSE	436.70	248.60	346.16	231.68	145.54	114.54	113.09	**87.72**
	MAE	233.06	144.50	211.29	131.05	87.86	68.30	62.20	**51.61**

Table 6 The percentage enhancement of the proposed approach compared to alternative benchmarking methods in the France dataset

Month	Improvement Percentage Metrics	Proposed approach vs Support Vector Regression	Proposed approach vs Artificial Neural Network	Proposed approach vs Random Forest	Proposed approach vs Long Short Term Memory	Proposed approach vs EMD LSTM	Proposed approach vs EEMD LSTM	Proposed approach vs CEEMDAN LSTM
Jan	P_{MAPE}	78.24%	68.08%	73.04%	65.45%	28.28%	29.35%	24.71%
	P_{RMSE}	78.81%	68.75%	71.23%	67.13%	33.74%	34.01%	40.00%
	P_{MAE}	78.24%	68.08%	73.04%	65.45%	28.28%	29.35%	24.71%
Feb	P_{MAPE}	71.51%	58.15%	70.23%	56.67%	15.78%	12.32%	10.32%
	P_{RMSE}	67.29%	57.45%	64.82%	55.15%	22.96%	16.65%	19.39%
	P_{MAE}	71.51%	58.15%	70.23%	56.67%	15.78%	12.32%	10.32%
Mar	P_{MAPE}	68.44%	67.36%	67.67%	63.71%	17.41%	23.62%	11.17%
	P_{RMSE}	70.90%	64.67%	66.89%	61.16%	9.53%	19.27%	9.86%
	P_{MAE}	68.44%	67.36%	67.67%	63.71%	17.41%	23.62%	11.17%
Apr	P_{MAPE}	62.06%	62.44%	61.98%	61.81%	38.14%	23.97%	25.95%
	P_{RMSE}	65.10%	63.04%	63.46%	64.35%	34.07%	20.67%	33.15%
	P_{MAE}	62.06%	62.44%	61.98%	61.81%	38.14%	23.97%	25.95%
May	P_{MAPE}	67.30%	64.22%	68.30%	62.29%	35.86%	30.39%	33.79%
	P_{RMSE}	63.89%	63.25%	65.33%	62.90%	42.72%	37.76%	41.14%
	P_{MAE}	67.30%	64.22%	68.30%	62.29%	35.86%	30.39%	33.79%
Jun	P_{MAPE}	66.67%	57.85%	65.00%	55.84%	25.93%	20.63%	21.69%
	P_{RMSE}	60.62%	57.21%	58.98%	55.74%	35.12%	27.91%	31.13%
	P_{MAE}	66.67%	57.85%	65.00%	55.84%	25.93%	20.63%	21.69%
Jul	P_{MAPE}	67.08%	58.95%	65.22%	57.20%	13.38%	18.64%	16.57%
	P_{RMSE}	63.41%	57.97%	60.47%	56.21%	12.88%	32.57%	34.19%
	P_{MAE}	67.08%	58.95%	65.22%	57.20%	13.38%	18.64%	16.57%
Aug	P_{MAPE}	62.91%	63.65%	62.75%	60.19%	21.20%	25.16%	15.76%
	P_{RMSE}	61.06%	61.33%	61.27%	59.66%	22.43%	24.01%	25.62%
	P_{MAE}	62.91%	63.65%	62.75%	60.19%	21.20%	25.16%	15.76%
Sep	P_{MAPE}	61.59%	65.75%	61.88%	58.19%	37.27%	35.13%	14.28%
	P_{RMSE}	66.89%	66.17%	65.45%	62.14%	36.08%	22.72%	14.83%
	P_{MAE}	61.59%	65.75%	61.88%	58.19%	37.27%	35.13%	14.28%
Oct	P_{MAPE}	63.24%	64.75%	63.92%	61.05%	46.32%	24.02%	13.99%
	P_{RMSE}	60.23%	60.93%	62.74%	59.76%	44.34%	19.91%	20.11%
	P_{MAE}	63.24%	64.75%	63.92%	61.05%	46.32%	24.02%	13.99%
Nov	P_{MAPE}	62.29%	62.21%	65.34%	61.80%	22.71%	15.67%	4.67%
	P_{RMSE}	61.25%	61.37%	64.73%	60.80%	22.77%	18.87%	12.49%
	P_{MAE}	62.29%	62.21%	65.34%	61.80%	22.71%	15.67%	4.67%
Dec	P_{MAPE}	77.89%	63.35%	73.66%	68.20%	28.95%	21.94%	18.14%
	P_{RMSE}	78.87%	61.56%	71.86%	67.01%	30.10%	22.18%	21.46%
	P_{MAE}	77.89%	63.35%	73.66%	68.20%	28.95%	21.94%	18.14%
Average	P_{MAPE}	67.41%	62.75%	66.48%	60.74%	26.51%	23.04%	17.51%
	P_{RMSE}	66.29%	61.67%	64.44%	60.63%	27.66%	25.32%	25.97%
	P_{MAE}	67.41%	62.75%	66.48%	60.74%	26.51%	23.04%	17.51%

Table 7 The percentage enhancement of the proposed approach compared to alternative benchmarking methods in the Turkey dataset

Month	Improvement Percentage Metrics	Proposed approach vs Support Vector Regression	Proposed approach vs Artificial Neural Network	Proposed approach vs Random Forest	Proposed approach vs Long Short Term Memory	Proposed approach vs EMD LSTM	Proposed approach vs EEMD LSTM	Proposed approach vs CEEMDAN LSTM
Jan	P_{MAPE}	72.14%	55.80%	67.36%	46.18%	75.04%	20.13%	14.96%
	P_{RMSE}	77.21%	56.75%	69.23%	54.83%	67.20%	16.19%	27.46%
	P_{MAE}	72.14%	55.80%	67.36%	46.18%	75.04%	20.13%	14.96%
Feb	P_{MAPE}	76.67%	64.71%	75.94%	54.66%	32.05%	21.90%	24.87%
	P_{RMSE}	78.63%	66.79%	74.97%	61.40%	31.74%	25.52%	26.93%
	P_{MAE}	76.67%	64.71%	75.94%	54.66%	32.05%	21.90%	24.87%
Mar	P_{MAPE}	70.81%	61.38%	72.63%	59.59%	26.96%	28.06%	17.76%
	P_{RMSE}	65.48%	62.00%	67.30%	59.76%	30.79%	26.03%	21.62%
	P_{MAE}	70.81%	61.38%	72.63%	59.59%	26.96%	28.06%	17.76%
Apr	P_{MAPE}	74.64%	65.08%	77.63%	59.82%	30.30%	22.75%	8.70%
	P_{RMSE}	68.43%	65.56%	71.48%	61.97%	30.53%	20.11%	9.91%
	P_{MAE}	74.64%	65.08%	77.63%	59.82%	30.30%	22.75%	8.70%
May	P_{MAPE}	76.44%	63.80%	73.59%	63.32%	26.42%	19.47%	15.73%
	P_{RMSE}	74.07%	61.92%	69.84%	61.52%	25.64%	24.07%	21.40%
	P_{MAE}	76.44%	63.80%	73.59%	63.32%	26.42%	19.47%	15.73%
Jun	P_{MAPE}	74.47%	63.55%	71.72%	62.19%	18.22%	16.23%	20.61%
	P_{RMSE}	74.86%	65.68%	69.85%	63.04%	17.83%	17.36%	30.26%
	P_{MAE}	74.47%	63.55%	71.72%	62.19%	18.22%	16.23%	20.61%
Jul	P_{MAPE}	63.28%	63.27%	62.67%	64.05%	43.37%	33.26%	19.91%
	P_{RMSE}	67.59%	65.41%	66.21%	64.13%	35.98%	21.37%	22.20%
	P_{MAE}	63.28%	63.27%	62.67%	64.05%	43.37%	33.26%	19.91%
Aug	P_{MAPE}	78.35%	66.24%	77.32%	62.40%	34.24%	19.04%	6.78%
	P_{RMSE}	81.43%	65.38%	77.18%	62.82%	29.84%	25.86%	15.87%
	P_{MAE}	78.35%	66.24%	77.32%	62.40%	34.24%	19.04%	6.78%
Sep	P_{MAPE}	85.25%	68.07%	82.42%	66.14%	37.89%	23.66%	9.34%
	P_{RMSE}	86.38%	65.23%	81.50%	63.97%	40.71%	21.53%	12.46%
	P_{MAE}	85.25%	68.07%	82.42%	66.14%	37.89%	23.66%	9.34%
Oct	P_{MAPE}	69.89%	72.85%	71.19%	67.21%	53.03%	34.95%	24.04%
	P_{RMSE}	69.55%	73.60%	71.05%	68.95%	49.56%	33.65%	29.62%
	P_{MAE}	69.89%	72.85%	71.19%	67.21%	53.03%	34.95%	24.04%
Nov	P_{MAPE}	89.63%	58.36%	82.75%	52.33%	29.71%	29.48%	16.55%
	P_{RMSE}	91.40%	64.18%	82.94%	61.83%	23.13%	25.17%	18.50%
	P_{MAE}	89.63%	58.36%	82.75%	52.33%	29.71%	29.48%	16.55%
Dec	P_{MAPE}	79.37%	63.37%	83.75%	56.59%	55.21%	49.49%	21.97%
	P_{RMSE}	88.43%	68.31%	86.17%	67.20%	51.33%	30.35%	13.88%
	P_{MAE}	79.37%	63.37%	83.75%	56.59%	55.21%	49.49%	21.97%
Average	P_{MAPE}	74.94%	63.83%	73.97%	59.89%	38.91%	27.05%	17.01%
	P_{RMSE}	76.23%	65.09%	73.38%	62.73%	36.17%	23.73%	20.95%
	P_{MAE}	74.94%	63.83%	73.97%	59.89%	38.91%	27.05%	17.01%

Table 4 and Table 5 highlight the superior performance of the LSTM method in contrast to other individual forecasting methods across both datasets. Lower MAPE, RMSE, and MAE values for LSTM demonstrate its effectiveness as the fundamental forecasting technique for our proposed approach. In comparison, hybrid approaches (Empirical Mode Decomposition-LSTM, Ensemble Empirical Mode Decomposition-LSTM, Complete Ensemble Empirical Mode Decomposition with Adaptive Noise-LSTM, and our proposed CEEMDAN-EWT-LSTM) achieve significantly lower average MAPE values than the single forecasting methods due to their ability to handle the inherent volatility and non-stationarity of wind data. This is achieved by breaking down the initial data into components that exhibit greater stationarity, improving the input quality for forecasting models and leading to more accurate predictions. Additionally, CEEMDAN surpasses EMD and EEMD in its decomposition capabilities, as evidenced by its lower average MAPE values compared to EMD-LSTM and EEMD-LSTM. Applying CEEMDAN-EWT-LSTM further enhances the accuracy of CEEMDAN-LSTM by over 17%, demonstrating the positive impact of denoising on forecasting performance. By mitigating the adverse effects of uncertainty and irregularities in the IMF 1 through denoising, the proposed CEEMDAN-EWT-LSTM method consistently produces the most favorable forecasting outcomes among all benchmark methods, achieving average MAPE values below 1.5%. This exceptional performance highlights the effectiveness of the proposed approach in ultra-short-term wind power prediction.

4.7 Wind Power Forecasting Summary

Accurate wind power forecasting is crucial for reliable and efficient power grid operation, but the inherent non-linearity and non-stationarity of wind data make it challenging. This study presents a novel hybrid deep learning-based LSTM method for wind power forecasting, leveraging CEEMDAN decomposition and EWT denoising. The method decomposes the original data using CEEMDAN, followed by targeted EWT denoising applied specifically to the first IMF. Subsequently, LSTM forecasts all subseries derived from the CEEMDAN-EWT process. Finally, the individual forecasts are aggregated to generate the final prediction. Real-world data from two geographically diverse wind farms validate the proposed approach's superior performance in contrast to existing approaches. This promising approach offers a significant advancement in ultra-short-term wind power forecasting. Future research will explore expanding the scope of multivariate forecasting by incorporating relevant influencing factors as input data.

CHAPTER 5

COMPARATIVE EVALUATION

In this section, we compare the performance of the two proposed approach: the hybrid CEEMDAN-RF-LSTM method and the hybrid CEEMDAN-EWT-LSTM method. We assess the performance of the proposed approach using both the building energy consumption dataset and the wind power dataset. Furthermore, we explore the effectiveness of these methods in multistep building energy consumption forecasting and wind power forecasting within this section. Multistep forecasting entails a model making predictions for a sequence of values rather than a single value. In this section, the building energy consumption and wind power are forecasted for various horizons or periods ranging from 1 step to 6 steps ahead.

5.1 Building Energy Consumption Dataset

A comparative evaluation of the proposed hybrid CEEMDAN RF-LSTM method and the proposed CEEMDAN-EWT-LSTM method for building energy consumption forecasting has been conducted on a public dataset from the Building Data Genome project [70]. Table 8 shows the comparative evaluation of the proposed approach for one-step building energy consumption forecasting in different buildings.

Table 8 A comparative evaluation of the proposed approach using building energy consumption dataset for one-step forecasting (1h ahead)

Building	Evaluation Metrics	Proposed CEEMDAN-RF-LSTM method	Proposed CEEMDAN-EWT-LSTM method
University Dormitory	MAPE (%)	3.511	**2.978**
	RMSE	1.761	**1.481**
	MAE	1.369	**1.159**
Running time(s)		*103.58*	*95.91*
University Laboratory	MAPE (%)	3.191	**2.561**
	RMSE	1.293	**1.073**
	MAE	1.014	**0.816**
Running time(s)		*75.807*	*89.43*
University Classroom	MAPE (%)	1.965	**1.693**
	RMSE	0.815	**0.697**
	MAE	0.570	**0.490**
Running time(s)		*81.056*	*101.84*
Office	MAPE (%)	5.331	**4.851**
	RMSE	0.570	**0.527**
	MAE	0.430	**0.390**
Running time(s)		*81.803*	*101.79*
Primary/ Secondary Classroom	MAPE (%)	7.164	**6.729**
	RMSE	0.467	**0.423**
	MAE	0.299	**0.275**
Running time(s)		*95.249*	*117.09*

The findings presented in Table 8 demonstrate that the CEEMDAN-EWT-LSTM method outperforms the CEEMDAN-RF-LSTM method by achieving superior forecasting accuracy and the lowest error rate. Then, to investigate further the efficiency of our proposed approach in multistep forecasting, several periods were chosen, which are 2 steps (2 hours ahead) to 6 steps (6 hours ahead). These methods have been evaluated within a university dormitory building. Table 9 shows the performance of the proposed approach for different time horizons. The improvement percentage of the proposed approach over CEEMDAN-LSTM and LSTM are summarized in

Table 10 and

Table *11*. It can be seen from Table 9 to Table 11 that forecast accuracy decreases as the time horizon increases. From a general perspective, our experiments show that the hybrid method combining CEEMDAN, EWT, and LSTM is very beneficial for forecasting energy consumption at all time horizons. It has the lowest error compared with the proposed CEEMDAN-RF-LSTM method, CEEMDAN-LSTM, and LSTM.

Table 9 Multistep Forecasting Performance Comparison of Various Methods for the University Dormitory Building

Time Horizons	**Evaluation Metrics**	**LSTM**	**CEEMDAN-LSTM**	**Proposed CEEMDAN-RF-LSTM method**	**Proposed CEEMDAN-EWT-LSTM method**
2 steps (2h ahead)	MAPE (%)	7.202	5.111	4.814	**4.500**
	RMSE	3.688	2.478	1.883	**1.755**
	MAE	2.865	1.995	2.368	**2.230**
3 steps (3h ahead)	MAPE (%)	7.402	5.662	5.434	**5.040**
	RMSE	3.775	2.740	2.124	**1.966**
	MAE	2.923	2.206	2.680	**2.509**
4 steps (4h ahead)	MAPE (%)	7.685	6.083	5.962	**5.813**
	RMSE	3.863	2.991	2.327	**2.268**
	MAE	3.031	2.376	2.958	**2.893**
5 steps (5h ahead)	MAPE (%)	7.611	6.703	6.489	**6.300**
	RMSE	3.876	3.276	2.533	**2.453**
	MAE	3.012	2.605	3.204	**3.122**
6 steps (6h ahead)	MAPE (%)	7.669	6.970	6.773	**6.755**
	RMSE	3.855	3.395	2.624	**2.613**
	MAE	3.020	2.699	3.296	**3.311**

Table 10 The percentage enhancement of Proposed CEEMDAN EWT LSTM compared with other methods for the university dormitory building

Time Horizons	**Improvement Percentage Metrics**	**Proposed CEEMDAN-EWT-LSTM Method vs. LSTM**	**Proposed CEEMDAN-EWT-LSTM Method vs. CEEMDAN LSTM**	**Proposed CEEMDAN-EWT-LSTM Method vs. Proposed CEEMDAN-RF-LSTM**
2 steps (2h ahead)	$\mathbf{P_{MAPE}}$ (%)	37.51%	11.95%	6.51%
	$\mathbf{P_{RMSE}}$ (%)	52.42%	29.19%	6.79%
	$\mathbf{P_{MAE}}$ (%)	22.15%	11.80%	5.82%
3 steps (3h ahead)	$\mathbf{P_{MAPE}}$ (%)	31.91%	10.98%	7.25%
	$\mathbf{P_{RMSE}}$ (%)	47.93%	28.27%	7.45%
	$\mathbf{P_{MAE}}$ (%)	14.15%	13.73%	6.37%
4 steps (4h ahead)	$\mathbf{P_{MAPE}}$ (%)	24.37%	4.44%	2.51%
	$\mathbf{P_{RMSE}}$ (%)	41.29%	24.18%	2.54%
	$\mathbf{P_{MAE}}$ (%)	4.55%	21.75%	2.20%
5 steps (5h ahead)	$\mathbf{P_{MAPE}}$ (%)	17.23%	6.02%	2.91%
	$\mathbf{P_{RMSE}}$ (%)	36.72%	25.12%	3.17%
	$\mathbf{P_{MAE}}$ (%)	3.65%	19.87%	2.55%
6 steps (6h ahead)	$\mathbf{P_{MAPE}}$ (%)	11.92%	3.08%	0.27%
	$\mathbf{P_{RMSE}}$ (%)	32.22%	23.04%	0.41%
	$\mathbf{P_{MAE}}$ (%)	9.64%	22.66%	0.44%

Table 11 The percentage enhancement of Proposed CEEMDAN RF LSTM compared with other methods for the university dormitory building

Time Horizons	**Improvement Percentage Metrics**	**Proposed CEEMDAN-RF-LSTM Method vs. LSTM**	**Proposed CEEMDAN-RF-LSTM Method vs. CEEMDAN LSTM**
2 steps (2h ahead)	$\mathbf{P_{MAPE}}$ (%)	33.16%	5.81%
	$\mathbf{P_{RMSE}}$ (%)	48.95%	24.03%
	$\mathbf{P_{MAE}}$ (%)	17.34%	18.71%
3 steps (3h ahead)	$\mathbf{P_{MAPE}}$ (%)	26.59%	4.02%
	$\mathbf{P_{RMSE}}$ (%)	43.74%	22.50%
	$\mathbf{P_{MAE}}$ (%)	8.31%	21.47%
4 steps (4h ahead)	$\mathbf{P_{MAPE}}$ (%)	22.42%	1.98%
	$\mathbf{P_{RMSE}}$ (%)	39.76%	22.20%
	$\mathbf{P_{MAE}}$ (%)	2.41%	24.49%
5 steps (5h ahead)	$\mathbf{P_{MAPE}}$ (%)	14.74%	3.20%
	$\mathbf{P_{RMSE}}$ (%)	34.64%	22.66%
	$\mathbf{P_{MAE}}$ (%)	6.36%	23.01%
6 steps (6h ahead)	$\mathbf{P_{MAPE}}$ (%)	11.68%	2.82%
	$\mathbf{P_{RMSE}}$ (%)	31.94%	22.72%
	$\mathbf{P_{MAE}}$ (%)	9.15%	22.11%

5.2 Wind Power Dataset

The performance of the proposed approach was tested using wind power datasets from a wind farm located in France [108]. Table 12 compares the proposed hybrid CEEMDAN RF-LSTM method and the proposed hybrid CEEMDAN-EWT-LSTM deep learning method for one-step wind power forecasting. According to the results in Table 12, the proposed CEEMDAN-EWT-LSTM method has the lowest error rate and the best forecasting accuracy compared with the other methods for one-step wind power forecasting.

Table 12 A comparative evaluation of the proposed approach using wind power dataset for one step ahead forecasting

Month	Evaluation Metrics	Proposed CEEMDAN-RF-LSTM method	Proposed CEEMDAN-EWT-LSTM method
Jan	MAPE (%)	1.37	**1.19**
	RMSE	44.54	**35.12**
	MAE	28.08	**24.36**
Running Time (s)		*878.32*	*1805.65*
Feb	MAPE (%)	2.55	**2.54**
	RMSE	82.18	**76.19**
	MAE	52.28	**52.14**
Running Time (s)		*652.32*	*872.34*
Mar	MAPE (%)	1.15	**1.03**
	RMSE	40.65	**34.90**
	MAE	23.68	**21.13**
Running Time (s)		*770.86*	*840.16*
Apr	MAPE (%)	1.39	**1.13**
	RMSE	53.58	**41.24**
	MAE	28.40	**23.17**
Running Time (s)		*657.14*	*716.77*
May	MAPE (%)	1.60	**1.21**
	RMSE	52.56	**38.57**
	MAE	32.77	**24.76**
Running Time (s)		*650.72*	*704.89*
Jun	MAPE (%)	2.49	**2.16**
	RMSE	86.85	**70.94**
	MAE	51.04	**44.37**
Running Time (s)		*488.63*	*589.81*
July	MAPE (%)	1.95	**1.88**
	RMSE	74.01	**60.03**
	MAE	0.07	**38.62**
Running Time (s)		*539.37*	*630.21*
Aug	MAPE (%)	1.03	**0.92**
	RMSE	42.23	**35.10**
	MAE	21.12	**18.86**
Running Time (s)		*523.78*	*605.62*
Sep	MAPE (%)	0.63	**0.57**
	RMSE	28.06	**23.80**
	MAE	12.95	**11.59**
Running Time (s)		*439.20*	*505.71*
Oct	MAPE (%)	1.27	**1.20**
	RMSE	42.46	**40.58**
	MAE	26.11	**24.64**
Running Time (s)		*480.89*	*586.41*
Nov	MAPE (%)	1.53	**1.47**
	RMSE	45.94	**44.07**
	MAE	31.46	**30.04**
Running Time (s)		*4964.13*	*612.31*
Dec	MAPE (%)	2.48	**2.20**
	RMSE	71.16	**64.74**
	MAE	50.77	**45.19**
Running Time (s)		*569.78*	*665.45*

To dig deeper into the effectiveness of our proposed approach for multistep wind power forecasting, we have examined various time intervals ranging from 2 steps (20 minutes ahead) to 6 steps (60 minutes ahead). Table 13 shows the performance of the proposed approach for different time horizons in January. Table 14 and

Table *15* summarize the improvement percentage of the proposed approach over CEEMDAN-LSTM and LSTM in wind power forecasting. As shown in Table 13 to

Table 15, the proposed approach achieves the best accuracy compared to the other methods.

Table 13 Multistep Wind Power Forecasting Performance Comparison using Various Methods in January

Time Horizons	**Evaluation Metrics**	**Long Short-Term Memory**	**CEEMDAN-LSTM**	**Proposed CEEMDAN-RF-LSTM method**	**Proposed CEEMDAN-EWT-LSTM method**
2 steps (20 min ahead)	MAPE (%)	5.110	1.955	1.988	**1.649**
	MAE	153.757	60.872	63.176	**51.931**
	RMSE	104.797	40.101	40.762	**33.806**
3 steps (30 min ahead)	MAPE (%)	5.911	2.264	2.172	**1.817**
	MAE	176.033	71.226	68.535	**56.194**
	RMSE	121.213	46.429	44.536	**37.255**
4 steps (40min ahead)	MAPE (%)	6.547	2.562	2.541	**2.410**
	MAE	192.817	77.060	76.438	**72.796**
	RMSE	134.250	52.535	52.116	**49.423**
5 steps (50min ahead)	MAPE (%)	6.916	2.919	2.897	**2.801**
	MAE	202.781	85.867	83.819	**83.664**
	RMSE	141.830	59.856	59.398	**57.434**
6 steps (60min ahead)	MAPE (%)	7.259	3.171	3.135	**3.115**
	MAE	209.048	94.747	93.735	**93.936**
	RMSE	148.862	65.024	64.298	**63.883**

Table 14 The percentage enhancement of Proposed CEEMDAN EWT LSTM compared with other methods in January for multistep ahead wind power forecasting

Time Horizons	**Improvement Percentage Metrics**	**Proposed CEEMDAN-EWT-LSTM Method vs. LSTM**	**Proposed CEEMDAN-EWT-LSTM Method vs. CEEMDAN LSTM**	**Proposed CEEMDAN-EWT-LSTM Method vs. Proposed CEEMDAN-RF-LSTM**
2 steps (2h ahead)	P_{MAPE} (%)	67.73%	15.65%	17.05%
	P_{RMSE} (%)	66.23%	14.69%	17.80%
	P_{MAE} (%)	67.74%	15.70%	17.06%
3 steps (3h ahead)	P_{MAPE} (%)	69.26%	19.74%	16.34%
	P_{RMSE} (%)	68.08%	21.10%	18.01%
	P_{MAE} (%)	69.26%	19.76%	16.35%
4 steps (4h ahead)	P_{MAPE} (%)	69.26%	19.74%	16.34%
	P_{RMSE} (%)	62.25%	5.53%	4.76%
	P_{MAE} (%)	63.19%	5.92%	5.17%
5 steps (5h ahead)	P_{MAPE} (%)	59.50%	4.04%	3.31%
	P_{RMSE} (%)	58.74%	2.57%	0.18%
	P_{MAE} (%)	59.51%	4.05%	3.31%
6 steps (6h ahead)	P_{MAPE} (%)	57.09%	1.77%	0.64%
	P_{RMSE} (%)	55.06%	0.86%	0.21%
	P_{MAE} (%)	57.09%	1.75%	0.65%

Table 15 The percentage enhancement of Proposed CEEMDAN RF LSTM compared with other methods in January for multistep ahead wind power forecasting

Time Horizons	**Improvement Percentage Metrics**	**Proposed CEEMDAN-RF-LSTM Method vs. LSTM**	**Proposed CEEMDAN-RF-LSTM Method vs. CEEMDAN LSTM**
2 steps (2h ahead)	P_{MAPE} (%)	61.10%	1.69%
	P_{RMSE} (%)	58.91%	3.78%
	P_{MAE} (%)	61.10%	1.65%
3 steps (3h ahead)	P_{MAPE} (%)	63.25%	4.06%
	P_{RMSE} (%)	61.07%	3.78%
	P_{MAE} (%)	63.26%	4.08%
4 steps (4h ahead)	P_{MAPE} (%)	63.25%	4.06%
	P_{RMSE} (%)	60.36%	0.81%
	P_{MAE} (%)	61.18%	0.80%
5 steps (5h ahead)	P_{MAPE} (%)	58.11%	0.75%
	P_{RMSE} (%)	58.67%	2.39%
	P_{MAE} (%)	58.12%	0.77%
6 steps (6h ahead)	P_{MAPE} (%)	56.81%	1.14%
	P_{RMSE} (%)	55.16%	1.07%
	P_{MAE} (%)	56.81%	1.12%

5.3 Discussion

The results of the comparative evaluation show that the proposed approach, namely CEEMDAN-RF-LSTM and CEEMDAN-EWT-LSTM, consistently exhibit superior performance than the other methods in energy building forecasting compared to other methods. This finding aligns with outcomes observed in wind power forecasting, which demonstrates the superiority of proposed approach across diverse forecasting tasks. In contrast to individual forecasting methods, our approaches demonstrate a notable enhancement in the effectiveness of energy forecasting by up to 75%. As the data exhibits complex non-linear and non-stationary patterns, the hybrid models outperform single models owing to their ability to capture and adapt to such complexities. In contrast to other hybrid CEEMDAN methods, specifically CEEMDAN-LSTM, where no special treatment is applied to the highest frequency component, our proposed approaches demonstrate enhanced performance by up to 40%. This underscores the importance of providing specialized treatment to the highest frequency component, resulting in improved forecast accuracy. While the CEEMDAN-EWT-LSTM and CEEMDAN-RF-LSTM architectures achieve superior performance compared to the CEEMDAN-LSTM baseline, the improvement relative to single models like LSTM, ANN, and RF is less pronounced. This discrepancy suggests that even though EWT and RF further refine the data representation, their contributions within the CEEMDAN-LSTM framework exhibit diminishing returns, possibly due to redundancy in successive filtering stages.

Further analysis of the comparative evaluation results indicates that the proposed hybrid CEEMDAN-EWT-LSTM method outperforms the proposed hybrid CEEMDAN-RF-LSTM method in wind power and building energy consumption forecasting. This superiority is primarily attributed to the effectiveness of the EWT-denoising technique in mitigating noise within the highest frequency component, as opposed to approaches employing Random Forest for handling high-frequency components. Notably, the highest frequency component often contains the most noise within the dataset. This component's denoising significantly enhances the forecasting model's accuracy, leading to more precise results.

The proposed approach demonstrates the highest accuracy for multi-step ahead forecasting for both datasets. As seen in Figure 14 to Figure 19, compared to the single forecasting methods, the proposed hybrid can effectively improve the performance of energy forecasting by up to 70%. However, the percentage improvement of the proposed CEEMDAN-EWT-LSTM method over the proposed CEEMDAN-RF-LSTM method decreases as the forecast horizon lengthens. Notably, for 5- and 6-step-ahead forecasting, the proposed CEEMDAN-EWT-LSTM marginally outperforms the proposed CEEMDAN-RF-LSTM. At longer forecasting horizons, the CEEMDAN-EWT-LSTM model may begin to overfit the training data. This is because the model strives to capture both short-term and long-term dependencies in the time series data, which can be challenging with limited training data. Consequently, the CEEMDAN-RF-LSTM model, which is less susceptible to overfitting, may begin to perform comparably to the CEEMDAN-EWT-LSTM model for longer forecasting horizons. Since the Random Forest algorithm utilizes ensemble learning, which combines multiple decision trees to make predictions, it helps to reduce the model's variance and prevent it from overfitting to the training data.

In summary, the outcomes of the comparative evaluation affirm that the CEEMDAN-EWT deep learning method stands out as a promising approach for wind power forecasting and building energy consumption forecasting. It surpasses other state-of-the-art methods in accuracy and effectiveness and exhibits robustness in the presence of noise.

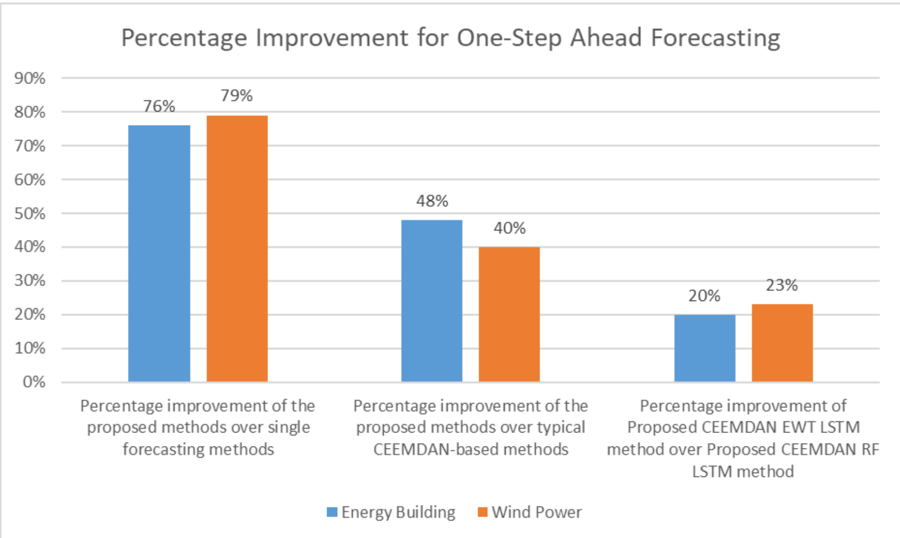

Figure 14 The percentage enhancement of the proposed approach over other methods for one step ahead forecasting

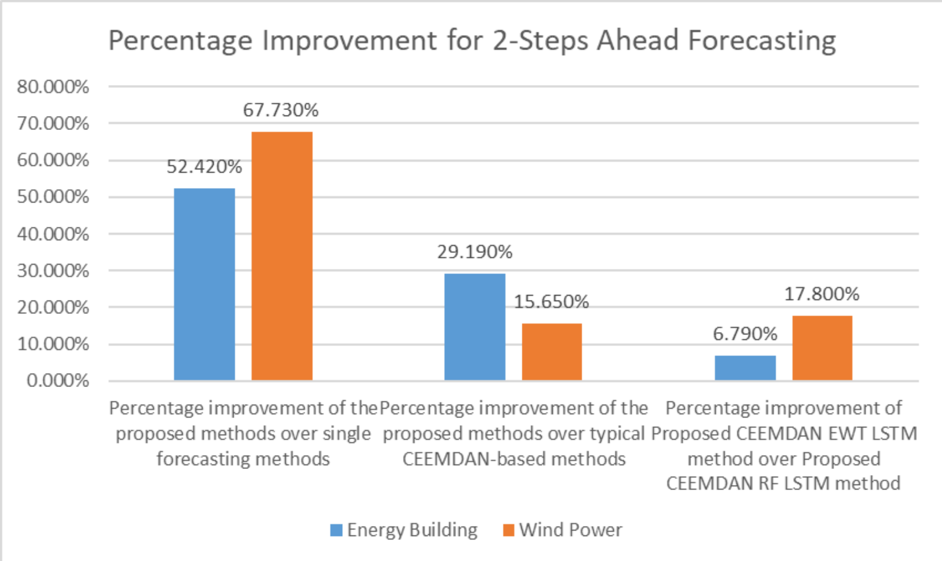

Figure 15 The percentage enhancement of the proposed approach over other methods for 2 steps ahead forecasting

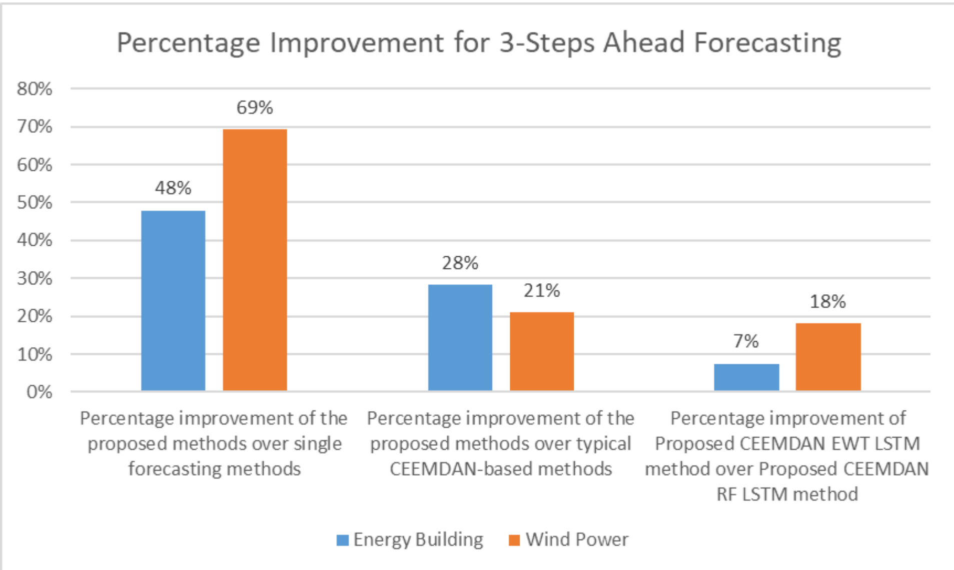

Figure 16 The percentage enhancement of the proposed approach over other methods for 3 steps ahead forecasting

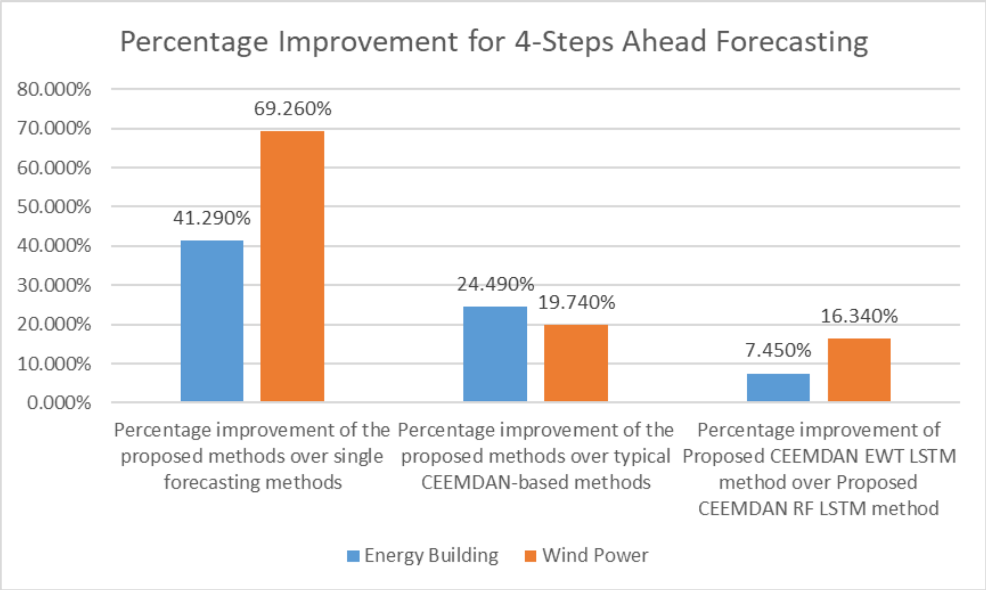

Figure 17 The percentage enhancement of the proposed approach over other methods for 4 steps ahead forecasting

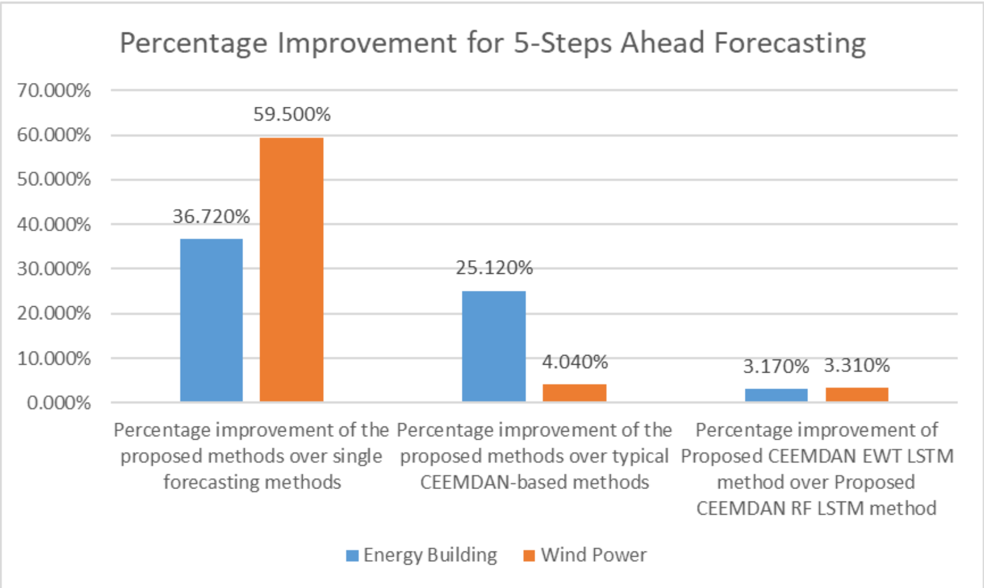

Figure 18 The percentage enhancement of the proposed approach over other methods for 5 steps ahead forecasting

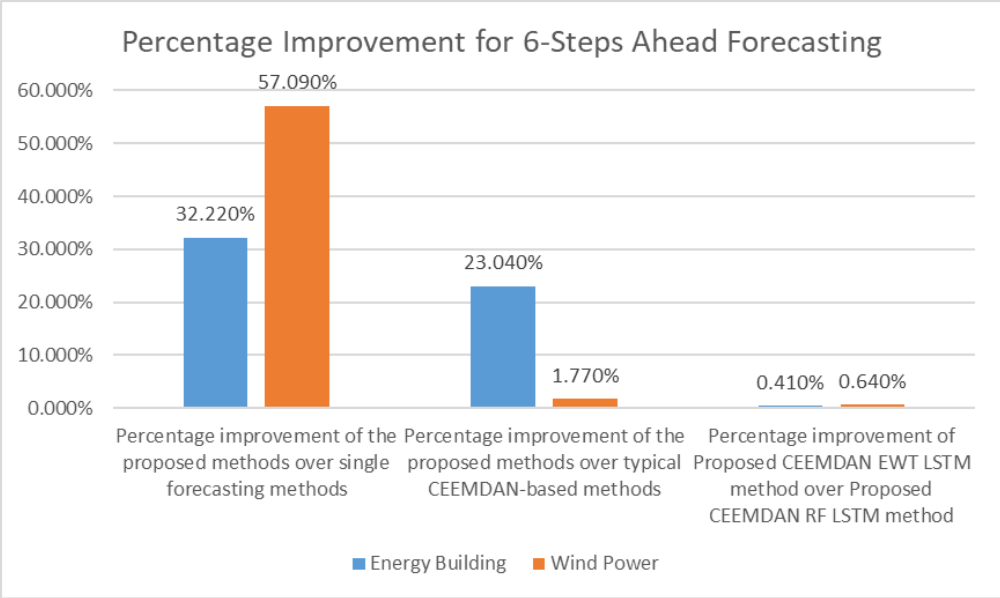

Figure 19 The percentage enhancement of the proposed approach over other methods for 6 steps ahead forecasting

To investigate the generalizability of our proposed CEEMDAN-EWT-LSTM method, we extended its application to the domain of solar forecasting, which inherently represents complex non-stationarity and non-linearity in the data [114]. Utilizing the publicly available global irradiation data retrieved from the National Renewable Energy Laboratory's[115] M2 tower, we assessed the performance of our approach in predicting solar irradiance levels. Our analysis employed hourly data from June and December 2023, as illustrated in the Figure 20.

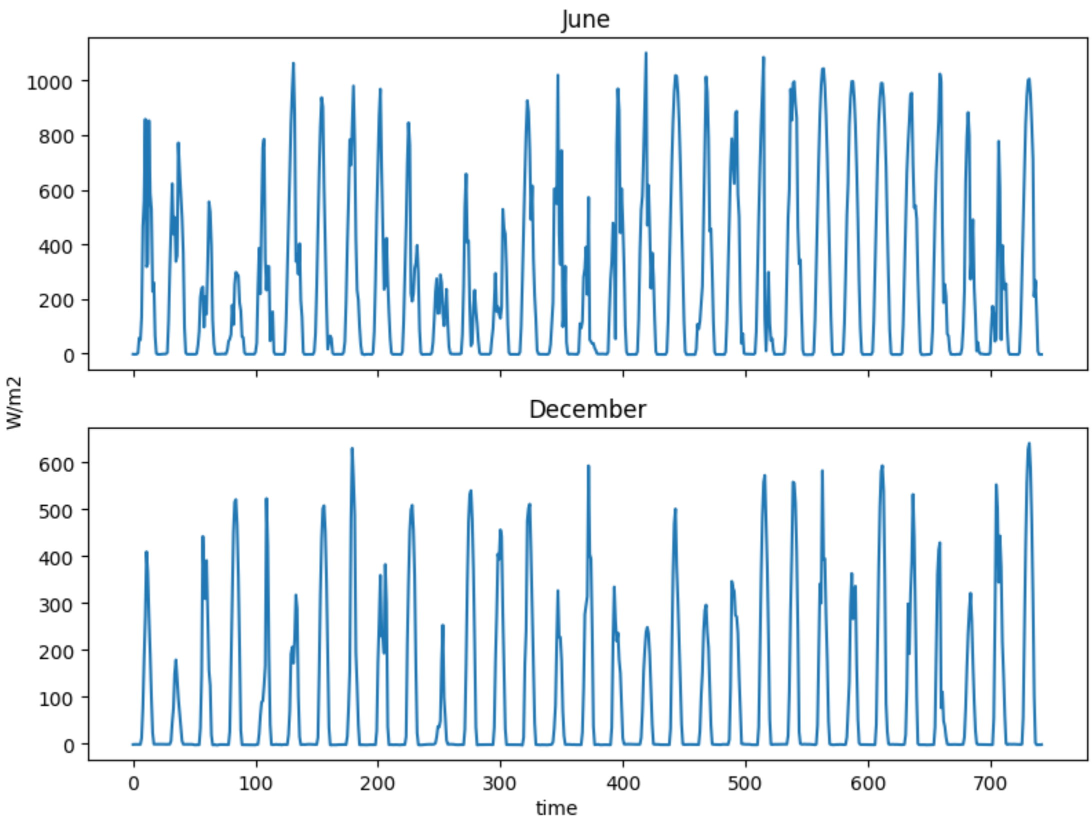

Figure 20 Global Irradiance Data in June and December 2023

We measured the entropy values of the original series and each component after decomposition using CEEMDAN to analyze the data's inherent complexity further. This provided insights into the information distribution within the data and aided in understanding the different components extracted. As shown in Table 16, the original data from June 2023 exhibits a higher entropy value than December 2023. This suggests that the June data possesses greater variability and randomness, likely due to the longer daylight hours and stronger solar fluctuations during summer months. Interestingly, Table 16 also reveals that the first Intrinsic Mode Function (IMF) extracted from the June data possesses a higher entropy value than the first IMF from December. This indicates that the high-frequency information captured by the first IMF is more complex and dynamic in June, further highlighting the influence of seasonal variations. The combined observations from the overall entropy and the first IMF's entropy in Table 16 provide compelling evidence that the solar irradiance data in June 2023 is characterized by significantly higher complexity and variability than December 2023.

Table 16 PE values for Solar Dataset

Month	**Original Series**	**IMF1**	**IMF3**	**IMF4**	**IMF5**	**IMF6**	**IMF7**	**IMF8**
June	0.991	0.921	0.61	0.368	0.26	0.207	0.165	0.146
December	0.967	0.919	0.523	0.358	0.257	0.202	0.181	0.114

Table 17 Forecasting Results in Solar Dataset

Month	**Metrics**	**Forecasting Methods**		
		LSTM	**CEEMDAN LSTM**	**CEEMDAN-EWT-LSTM**
June	MAPE	6.297	4.788	2.684
	RMSE	113.63	83.641	40.763
	MAE	69.325	52.710	29.551
Running time (s)		*9.921*	*244.701*	*208.664*
December	MAPE	4.204	3.278	2.303
	RMSE	54.154	32.726	26.142
	MAE	26.968	21.021	14.765
Running time (s)		*13.075*	*223.15*	*225.10*

Table 18 Enhanced Forecasting Accuracy in Solar Dataset

Month	**Metrics**	**Proposed approach vs LSTM**	**Proposed approach vs CEEMDAN -LSTM**
June	P_{MAPE}	57.373%	43.936%
	P_{RMSE}	64.125%	51.263%
	P_{MAE}	57.373%	43.936%
December	P_{MAPE}	45.247%	29.758%
	P_{RMSE}	51.727%	20.119%
	P_{MAE}	45.247%	29.758%

Our results, as showcased in Table 17 and Table 18 , reveals that:

1. The higher the randomness characteristics of the original data, the greater the accuracy improvement achieved by our proposed CEEMDAN-EWT-LSTM model. This is evident in June's data, where it is outperformed LSTM and CEEMDAN-LSTM by 57.373% and 43.936%, respectively, demonstrating its effectiveness in handling complex data. For December's data, characterized by lower randomness, the improvement offered by our method remains significant. Compared to LSTM and CEEMDAN-LSTM, CEEMDAN-EWT-LSTM achieved percentage improvements of 45.247% and 29.758%, respectively. This showcases its consistent advantage even in conditions with less inherent randomness.
2. Another key observation lies in the connection between the complexity of IMF1 and the performance of our proposed algorithm. When IMF1 exhibits high complexity, as in June's data, the EWT component of our method shines, suppressing the influence of outliers and randomness. This contributes significantly to the accuracy gains observed for June's forecast.

These results suggest that our method should be particularly effective in handling data with high entropy, such as the June data, which could benefit its overall forecasting accuracy. In conclusion, our proposed approach, CEEMDAN-EWT-LSTM, excels at forecasting energy time series data, and below is an analysis of appropriate data for forecasting with CEEMDAN-EWT-LSTM:

1. Non-Stationary Data: CEEMDAN excels at decomposing complex, non-stationary data with varying frequencies and trends into intrinsic mode functions

(IMFs). This allows the LSTM to focus on individual components and their dynamics, leading to more accurate predictions.

2. Time-Series Dependence: This temporal order is crucial for LSTM to exploit past information and make accurate predictions based on past and future data relationships.
3. High randomness: EWT can help reduce noise in the data. The higher the randomness characteristics of the original data, the greater the accuracy improvement our proposed CEEMDAN-EWT-LSTM model achieves.

CHAPTER 6

CONCLUSION AND FUTURE RESEARCH

6.1 Conclusion

This study introduces two hybrid machine learning methods designed for building energy consumption forecasting and wind power forecasting. The first approach proposes a novel hybrid CEEMDAN-RF-LSTM to enhance building energy consumption forecasting accuracy. The second approach presents a hybrid CEEMDAN-EWT-LSTM deep learning method for wind power forecasting. These methods leverage the complementary advantages of various machine learning techniques and advanced data preprocessing to achieve superior forecasting accuracy. Evaluation of real-world datasets demonstrates that our proposed approach outperform state-of-the-art techniques in terms of forecasting accuracy.

Our proposed approach, namely CEEMDAN-RF-LSTM and CEEMDAN-EWT-LSTM, consistently perform better than other methods. This highlights the effectiveness of applying the CEEMDAN method to decompose the nonstationary original data into several relatively stationary components, leading to improved forecasting accuracy. Additionally, applying specialized treatment to the highest frequency component further enhances forecast accuracy.

Our proposed CEEMDAN-EWT-LSTM method outperforms other methods. This is because the EWT-denoising technique effectively reduces noise in the highest frequency component, which is typically the noisiest part of the component. This denoising significantly boosts the accuracy of the forecasting model, leading to more precise forecasts. The proposed approach have the potential to be used to improve the efficiency and reliability of power grids and building energy management systems, respectively. For example, accurate wind power forecasting can help improve the integration of wind power into power grids. An accurate building energy consumption forecasting can help reduce energy consumption and building costs.

6.2 Future Research

There are several directions for future research. One direction is to explore the use of other deep learning models, such as convolutional neural networks (CNNs) and transformers-based models, in the proposed approach. CNNs and transformers are effective for forecasting tasks, and there is potential for them to enhance the efficiency of the proposed approach even more. Another direction for future work is to apply the proposed approach to other forecasting tasks, such as stock price forecasting. In addition, it would be interesting to develop a hybrid forecasting framework that combines the proposed approach with other external factors, such as numerical weather forecasting models. This could lead to even more accurate forecasting results. In forthcoming research, we plan to integrate additional processing steps to assess their potential impact on the overall performance of our methodology and expand the size of the dataset, particularly by incorporating minutely data.

REFERENCES

[1] R. Bhandari, B.E. Arce, V. Sessa, R. Adamou, Sustainability assessment of electricity generation in Niger using a weighted multi-criteria decision approach, Sustainability (Switzerland) 13 (2021). https://doi.org/10.3390/su13010385.

[2] P. Moodley, Sustainable biofuels: opportunities and challenges, Sustainable Biofuels (2021) 1–20. https://doi.org/10.1016/B978-0-12-820297-5.00003-7.

[3] D. AL-Alimi, A.M. AlRassas, M.A.A. Al-qaness, Z. Cai, A.O. Aseeri, M. Abd Elaziz, A.A. Ewees, TLIA: Time-series forecasting model using long short-term memory integrated with artificial neural networks for volatile energy markets, Appl Energy 343 (2023). https://doi.org/10.1016/j.apenergy.2023.121230.

[4] A. Azarpour, O. Mohammadzadeh, N. Rezaei, S. Zendehboudi, Current status and future prospects of renewable and sustainable energy in North America: Progress and challenges, Energy Convers Manag 269 (2022). https://doi.org/10.1016/j.enconman.2022.115945.

[5] D. AL-Alimi, A.M. AlRassas, M.A.A. Al-qaness, Z. Cai, A.O. Aseeri, M. Abd Elaziz, A.A. Ewees, TLIA: Time-series forecasting model using long short-term memory integrated with artificial neural networks for volatile energy markets, Appl Energy 343 (2023). https://doi.org/10.1016/j.apenergy.2023.121230.

[6] D. Kaur, S.N. Islam, M.A. Mahmud, M.E. Haque, Z.Y. Dong, Energy forecasting in smart grid systems: recent advancements in probabilistic deep learning, IET Generation, Transmission and Distribution 16 (2022). https://doi.org/10.1049/gtd2.12603.

[7] M. Zekić-Sušac, S. Mitrović, A. Has, Machine learning based system for managing energy efficiency of public sector as an approach towards smart cities, Int J Inf Manage 58 (2021). https://doi.org/10.1016/j.ijinfomgt.2020.102074.

[8] I. Karijadi, S.Y. Chou, A hybrid RF-LSTM based on CEEMDAN for improving the accuracy of building energy consumption prediction, Energy Build 259 (2022). https://doi.org/10.1016/j.enbuild.2022.111908.

[9] O. Laib, M.T. Khadir, L. Mihaylova, Toward efficient energy systems based on natural gas consumption prediction with LSTM Recurrent Neural Networks, Energy 177 (2019). https://doi.org/10.1016/j.energy.2019.04.075.

[10] M.A.A. Al-Qaness, M.A. Elaziz, A.A. Ewees, Oil Consumption Forecasting Using Optimized Adaptive Neuro-Fuzzy Inference System Based on Sine Cosine Algorithm, IEEE Access 6 (2018). https://doi.org/10.1109/ACCESS.2018.2879965.

[11] Y. Wang, R. Zou, F. Liu, L. Zhang, Q. Liu, A review of wind speed and wind power forecasting with deep neural networks, Appl Energy 304 (2021). https://doi.org/10.1016/j.apenergy.2021.117766.

[12] I. Karijadi, S.-Y. Chou, A. Dewabharata, Wind power forecasting based on hybrid CEEMDAN-EWT deep learning method, Renew Energy 218 (2023) 119357. https://doi.org/10.1016/J.RENENE.2023.119357.

[13] T.Y. Kim, S.B. Cho, Predicting residential energy consumption using CNN-

LSTM neural networks, Energy 182 (2019). https://doi.org/10.1016/j.energy.2019.05.230.

[14] X. Qing, Y. Niu, Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM, Energy 148 (2018). https://doi.org/10.1016/j.energy.2018.01.177.

[15] N.T. Ngo, A.D. Pham, T.T.H. Truong, N.S. Truong, N.T. Huynh, Developing a hybrid time-series artificial intelligence model to forecast energy use in buildings, Sci Rep 12 (2022). https://doi.org/10.1038/s41598-022-19935-6.

[16] C. Fu, M. Quintana, Z. Nagy, C. Miller, Filling time-series gaps using image techniques: Multidimensional context autoencoder approach for building energy data imputation, Appl Therm Eng 236 (2024) 121545. https://doi.org/10.1016/J.APPLTHERMALENG.2023.121545.

[17] W. Khan, S. Walker, W. Zeiler, Improved solar photovoltaic energy generation forecast using deep learning-based ensemble stacking approach, Energy 240 (2022). https://doi.org/10.1016/j.energy.2021.122812.

[18] L. Wang, M. Mao, J. Xie, Z. Liao, H. Zhang, H. Li, Accurate solar PV power prediction interval method based on frequency-domain decomposition and LSTM model, Energy 262 (2023). https://doi.org/10.1016/j.energy.2022.125592.

[19] Y. Song, C. Cai, D. Ma, C. Li, Modelling and Forecasting High-Frequency Data with Jumps Based on a Hybrid Nonparametric Regression and LSTM model, Expert Syst Appl (2023) 121527. https://doi.org/10.1016/J.ESWA.2023.121527.

[20] X. Kong, X. Du, G. Xue, Z. Xu, Multi-step short-term solar radiation prediction based on empirical mode decomposition and gated recurrent unit optimized via an attention mechanism, Energy 282 (2023) 128825. https://doi.org/10.1016/J.ENERGY.2023.128825.

[21] Z. Xing, B. Qu, Y. Liu, Z. Chen, Comparative study of reformed neural network based short-term wind power forecasting models, IET Renewable Power Generation 16 (2022). https://doi.org/10.1049/rpg2.12384.

[22] Y. Ren, P.N. Suganthan, N. Srikanth, A comparative study of empirical mode decomposition-based short-term wind speed forecasting methods, IEEE Trans Sustain Energy 6 (2015). https://doi.org/10.1109/TSTE.2014.2365580.

[23] Q. Zhang, Z. Tang, S. Cao, G. Wang, Wind farm wind power prediction method based on CEEMDAN and de optimized DNN neural network, in: Proceedings - 2019 Chinese Automation Congress, CAC 2019, 2019. https://doi.org/10.1109/CAC48633.2019.8996744.

[24] P. Lv, Y. Shu, J. Xu, Q. Wu, Modal decomposition-based hybrid model for stock index prediction, Expert Syst Appl 202 (2022) 117252. https://doi.org/10.1016/J.ESWA.2022.117252.

[25] Y. Lin, Z. Lin, Y. Liao, Y. Li, J. Xu, Y. Yan, Forecasting the realized volatility of stock price index: A hybrid model integrating CEEMDAN and LSTM, Expert Syst Appl 206 (2022) 117736. https://doi.org/10.1016/J.ESWA.2022.117736.

[26] W. Zeng, Y. Cao, L. Feng, J. Fan, M. Zhong, W. Mo, Z. Tan, Hybrid CEEMDAN-DBN-ELM for online DGA serials and transformer status forecasting, Electric Power Systems Research 217 (2023) 109176. https://doi.org/10.1016/J.EPSR.2023.109176.

[27] N.E. Huang, Z. Shen, S.R. Long, M.C. Wu, H.H. Shih, N. Yen, C.C. Tung, H.H. Liu, The empirical mode decomposition and the Hilbert spectrum for

nonlinear and non-stationary time series analysis, Proceedings of the Royal Society A 454 (1996).
[28] L. Yang, J. Cai, A method to identify wet ball mill's load based on CEEMDAN, RCMDE and SRNN classification, Miner Eng 165 (2021). https://doi.org/10.1016/j.mineng.2021.106852.
[29] K. Barnova, R. Kahankova, R. Jaros, M. Litschmannova, R. Martinek, A comparative study of single-channel signal processing methods in fetal phonocardiography, PLoS One 17 (2022). https://doi.org/10.1371/journal.pone.0269884.
[30] D.O. Afanasyev, E.A. Fedorova, The long-term trends on the electricity markets: Comparison of empirical mode and wavelet decompositions, Energy Econ 56 (2016). https://doi.org/10.1016/j.eneco.2016.04.009.
[31] D. Shu-wen, L. Lu-jun, W. Qing-qu, W. Kang-le, C. Peng-zhan, Fiber optic gyro noise reduction based on hybrid CEEMDAN-LWT method, Measurement 161 (2020). https://doi.org/10.1016/j.measurement.2020.107865.
[32] M.E. Torres, M.A. Colominas, G. Schlotthauer, P. Flandrin, A complete ensemble empirical mode decomposition with adaptive noise, in: ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, 2011. https://doi.org/10.1109/ICASSP.2011.5947265.
[33] O. Singh, R.K. Sunkaria, An empirical wavelet transform based approach for multivariate data processing application to cardiovascular physiological signals, Bio-Algorithms and Med-Systems 14 (2018). https://doi.org/10.1515/bams-2018-0030.
[34] W. Liu, W. Chen, Recent advancements in empirical wavelet transform and its applications, IEEE Access 7 (2019). https://doi.org/10.1109/ACCESS.2019.2930529.
[35] Y. Xu, K. Zhang, C. Ma, Z. Sheng, H. Shen, An Adaptive Spectrum Segmentation Method to Optimize Empirical Wavelet Transform for Rolling Bearings Fault Diagnosis, IEEE Access 7 (2019). https://doi.org/10.1109/ACCESS.2019.2902645.
[36] O. Singh, R.K. Sunkaria, ECG signal denoising via empirical wavelet transform, Australas Phys Eng Sci Med 40 (2017). https://doi.org/10.1007/s13246-016-0510-6.
[37] J. Gilles, Empirical wavelet transform, IEEE Transactions on Signal Processing 61 (2013). https://doi.org/10.1109/TSP.2013.2265222.
[38] L. Breiman, Random forests. Mach Learn, Random Forests (2001).
[39] L. Lin, F. Wang, X. Xie, S. Zhong, Random forests-based extreme learning machine ensemble for multi-regime time series prediction, Expert Syst Appl 83 (2017). https://doi.org/10.1016/j.eswa.2017.04.013.
[40] A. Liaw, M. Wiener, Classification and Regression by randomForest, R News 2 (2002).
[41] Z.H. Zhou, J. Feng, Deep forest: Towards an alternative to deep neural networks, in: IJCAI International Joint Conference on Artificial Intelligence, 2017. https://doi.org/10.24963/ijcai.2017/497.
[42] Y. Guo, S. Liu, Z. Li, X. Shang, BCDForest: A boosting cascade deep forest model towards the classification of cancer subtypes based on gene expression data, BMC Bioinformatics 19 (2018). https://doi.org/10.1186/s12859-018-2095-4.
[43] R. Srivastava, A.N. Tiwari, V.K. Giri, Solar radiation forecasting using MARS, CART, M5, and random forest model: A case study for India, Heliyon 5 (2019).

https://doi.org/10.1016/j.heliyon.2019.e02692.
[44] T. Fischer, C. Krauss, Deep learning with long short-term memory networks for financial market predictions, Eur J Oper Res 270 (2018). https://doi.org/10.1016/j.ejor.2017.11.054.
[45] J. Ren, Z. Yu, G. Gao, G. Yu, J. Yu, A CNN-LSTM-LightGBM based short-term wind power prediction method based on attention mechanism, Energy Reports 8 (2022) 437–443. https://doi.org/10.1016/J.EGYR.2022.02.206.
[46] D. Hládek, J. Staš, S. Ondáš, Comparison of Recurrent Neural Networks for Slovak Punctuation Restoration; Comparison of Recurrent Neural Networks for Slovak Punctuation Restoration, 2019. https://keras.io/.
[47] L. Xiang, J. Liu, X. Yang, A. Hu, H. Su, Ultra-short term wind power prediction applying a novel model named SATCN-LSTM, Energy Convers Manag 252 (2022) 115036. https://doi.org/10.1016/J.ENCONMAN.2021.115036.
[48] J. Wei, X. Wu, T. Yang, R. Jiao, Ultra-short-term forecasting of wind power based on multi-task learning and LSTM, International Journal of Electrical Power and Energy Systems 149 (2023). https://doi.org/10.1016/j.ijepes.2023.109073.
[49] M. Santamouris, Energy Consumption and Environmental Quality of the Building Sector, in: Minimizing Energy Consumption, Energy Poverty and Global and Local Climate Change in the Built Environment: Innovating to Zero, 2019. https://doi.org/10.1016/b978-0-12-811417-9.00002-7.
[50] REN21 Secretariat, 2020 Global Status Report for Buildings and Construction Towards a zero-emissions, efficient and resilient buildings owards a zero-emissions, efficient and resilient buildings and construction sector, 2020.
[51] A. Kalua, Urban residential building energy consumption by end-use in Malawi, Buildings 10 (2020). https://doi.org/10.3390/buildings10020031.
[52] Y. Ye, W. Zuo, G. Wang, A comprehensive review of energy-related data for U.S. commercial buildings, Energy Build 186 (2019). https://doi.org/10.1016/j.enbuild.2019.01.020.
[53] K. Amasyali, N.M. El-Gohary, A review of data-driven building energy consumption prediction studies, Renewable and Sustainable Energy Reviews 81 (2018). https://doi.org/10.1016/j.rser.2017.04.095.
[54] BP, BP Energy Outlook 2019 edition, 2019.
[55] L. Pérez-Lombard, J. Ortiz, C. Pout, A review on buildings energy consumption information, Energy Build 40 (2008). https://doi.org/10.1016/j.enbuild.2007.03.007.
[56] M. Ilbeigi, M. Ghomeishi, A. Dehghanbanadaki, Prediction and optimization of energy consumption in an office building using artificial neural network and a genetic algorithm, Sustain Cities Soc 61 (2020). https://doi.org/10.1016/j.scs.2020.102325.
[57] F. Pallonetto, C. Jin, E. Mangina, Forecast electricity demand in commercial building with machine learning models to enable demand response programs, Energy and AI (2021). https://doi.org/10.1016/j.egyai.2021.100121.
[58] C. Kuster, Y. Rezgui, M. Mourshed, Electrical load forecasting models: A critical systematic review, Sustain Cities Soc 35 (2017). https://doi.org/10.1016/j.scs.2017.08.009.
[59] C. Liu, B. Sun, C. Zhang, F. Li, A hybrid prediction model for residential electricity consumption using holt-winters and extreme learning machine, Appl Energy 275 (2020). https://doi.org/10.1016/j.apenergy.2020.115383.

[60] T. Catalina, V. Iordache, B. Caracaleanu, Multiple regression model for fast prediction of the heating energy demand, Energy Build 57 (2013). https://doi.org/10.1016/j.enbuild.2012.11.010.
[61] Z. Wang, Y. Wang, R. Zeng, R.S. Srinivasan, S. Ahrentzen, Random Forest based hourly building energy prediction, Energy Build 171 (2018). https://doi.org/10.1016/j.enbuild.2018.04.008.
[62] A. Lahouar, J. Ben Hadj Slama, Day-ahead load forecast using random forest and expert input selection, Energy Convers Manag 103 (2015). https://doi.org/10.1016/j.enconman.2015.07.041.
[63] D. Liu, Q. Yang, F. Yang, Predicting Building Energy Consumption by Time Series Model Based on Machine Learning and Empirical Mode Decomposition, in: 2020 5th IEEE International Conference on Big Data Analytics, ICBDA 2020, 2020. https://doi.org/10.1109/ICBDA49040.2020.9101335.
[64] H. Zheng, J. Yuan, L. Chen, Short-Term Load Forecasting Using EMD-LSTM neural networks with a xgboost algorithm for feature importance evaluation, Energies (Basel) 10 (2017). https://doi.org/10.3390/en10081168.
[65] N. An, W. Zhao, J. Wang, D. Shang, E. Zhao, Using multi-output feedforward neural network with empirical mode decomposition based signal filtering for electricity demand forecasting, Energy 49 (2013). https://doi.org/10.1016/j.energy.2012.10.035.
[66] D. Zhang, C. Cai, S. Chen, L. Ling, An improved genetic algorithm for optimizing ensemble empirical mode decomposition method, Systems Science and Control Engineering 7 (2019). https://doi.org/10.1080/21642583.2019.1627598.
[67] Z. Wu, N.E. Huang, Ensemble empirical mode decomposition: A noise-assisted data analysis method, Adv Adapt Data Anal 1 (2009). https://doi.org/10.1142/S1793536909000047.
[68] J. Zheng, J. Cheng, Y. Yang, Partly ensemble empirical mode decomposition: An improved noise-assisted method for eliminating mode mixing, Signal Processing 96 (2014). https://doi.org/10.1016/j.sigpro.2013.09.013.
[69] H. Lin, Q. Sun, S.Q. Chen, Reducing exchange rate risks in international trade: A hybrid forecasting approach of CEEMDAN and multilayer LSTM, Sustainability (Switzerland) 12 (2020). https://doi.org/10.3390/su12062451.
[70] C. Miller, F. Meggers, The Building Data Genome Project: An open, public data set from non-residential building electrical meters, in: Energy Procedia, 2017. https://doi.org/10.1016/j.egypro.2017.07.400.
[71] M. Zekić-Sušac, A. Has, M. Knežević, Predicting energy cost of public buildings by artificial neural networks, CART, and random forest, Neurocomputing 439 (2021). https://doi.org/10.1016/j.neucom.2020.01.124.
[72] J. Xu, K. Xu, Z. Li, F. Meng, T. Tu, L. Xu, Q. Liu, Forecast of dengue cases in 20 chinese cities based on the deep learning method, Int J Environ Res Public Health 17 (2020). https://doi.org/10.3390/ijerph17020453.
[73] S.L. Lin, H.W. Huang, Improving Deep Learning for Forecasting Accuracy in Financial Data, Discrete Dyn Nat Soc 2020 (2020). https://doi.org/10.1155/2020/5803407.
[74] N. Halpern-Wight, M. Konstantinou, A.G. Charalambides, A. Reinders, Training and testing of a single-layer LSTM network for near-future solar forecasting, Applied Sciences (Switzerland) 10 (2020). https://doi.org/10.3390/app10175873.
[75] D. Laszuk, PyEMD Documentation, (2020).

[76] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, É. Duchesnay, Scikit-learn: Machine learning in Python, Journal of Machine Learning Research 12 (2011).
[77] F. Chollet, Keras Documentation, Keras.Io (2015).
[78] A. Lahouar, J. Ben Hadj Slama, Hour-ahead wind power forecast based on random forests, Renew Energy 109 (2017). https://doi.org/10.1016/j.renene.2017.03.064.
[79] D.P. Kingma, J.L. Ba, Adam: A method for stochastic optimization, in: 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings, 2015.
[80] A. Kumar Dubey, A. Kumar, V. García-Díaz, A. Kumar Sharma, K. Kanhaiya, Study and analysis of SARIMA and LSTM in forecasting time series data, Sustainable Energy Technologies and Assessments 47 (2021). https://doi.org/10.1016/j.seta.2021.101474.
[81] A. Kara, Multi-step influenza outbreak forecasting using deep LSTM network and genetic algorithm, Expert Syst Appl 180 (2021). https://doi.org/10.1016/j.eswa.2021.115153.
[82] C. Jörges, C. Berkenbrink, B. Stumpe, Prediction and reconstruction of ocean wave heights based on bathymetric data using LSTM neural networks, Ocean Engineering 232 (2021). https://doi.org/10.1016/j.oceaneng.2021.109046.
[83] I. Kandel, M. Castelli, The effect of batch size on the generalizability of the convolutional neural networks on a histopathology dataset, ICT Express 6 (2020). https://doi.org/10.1016/j.icte.2020.04.010.
[84] Y. Wang, J. Wang, Z. Li, A novel hybrid air quality early-warning system based on phase-space reconstruction and multi-objective optimization: A case study in China, J Clean Prod 260 (2020). https://doi.org/10.1016/j.jclepro.2020.121027.
[85] I.M. Coelho, V.N. Coelho, E.J. da S. Luz, L.S. Ochi, F.G. Guimarães, E. Rios, A GPU deep learning metaheuristic based model for time series forecasting, Appl Energy 201 (2017). https://doi.org/10.1016/j.apenergy.2017.01.003.
[86] I.E.A. IEA, Electricity Market Report, Electricity Market Report (2022).
[87] M.D.A. Al-falahi, S.D.G. Jayasinghe, H. Enshaei, A review on recent size optimization methodologies for standalone solar and wind hybrid renewable energy system, Energy Convers Manag 143 (2017). https://doi.org/10.1016/j.enconman.2017.04.019.
[88] A. Suman, Role of renewable energy technologies in climate change adaptation and mitigation: A brief review from Nepal, Renewable and Sustainable Energy Reviews 151 (2021). https://doi.org/10.1016/j.rser.2021.111524.
[89] R. Lacal Arantegui, A. Jäger-Waldau, Photovoltaics and wind status in the European Union after the Paris Agreement, Renewable and Sustainable Energy Reviews 81 (2018). https://doi.org/10.1016/j.rser.2017.06.052.
[90] J. Li, S. Zhang, Z. Yang, A wind power forecasting method based on optimized decomposition prediction and error correction, Electric Power Systems Research 208 (2022). https://doi.org/10.1016/j.epsr.2022.107886.
[91] S. Teleke, M.E. Baran, S. Bhattacharya, A.Q. Huang, Optimal control of battery energy storage for wind farm dispatching, IEEE Transactions on Energy Conversion 25 (2010). https://doi.org/10.1109/TEC.2010.2041550.
[92] J. Wang, M. Shahidehpour, Z. Li, Security-constrained unit commitment with volatile wind power generation, IEEE Transactions on Power Systems 23

(2008). https://doi.org/10.1109/TPWRS.2008.926719.
[93] Q. Chen, K.A. Folly, Wind Power Forecasting, IFAC-PapersOnLine 51 (2018). https://doi.org/10.1016/j.ifacol.2018.11.738.
[94] S.S. Soman, H. Zareipour, O. Malik, P. Mandal, A review of wind power and wind speed forecasting methods with different time horizons, in: North American Power Symposium 2010, NAPS 2010, 2010. https://doi.org/10.1109/NAPS.2010.5619586.
[95] G. Mu, M. Yang, D. Wang, G. Yan, Y. Qi, Spatial dispersion of wind speeds and its influence on the forecasting error of wind power in a wind farm, Journal of Modern Power Systems and Clean Energy 4 (2016). https://doi.org/10.1007/s40565-015-0151-x.
[96] Z. Peng, S. Peng, L. Fu, B. Lu, J. Tang, K. Wang, W. Li, A novel deep learning ensemble model with data denoising for short-term wind speed forecasting, Energy Convers Manag 207 (2020). https://doi.org/10.1016/j.enconman.2020.112524.
[97] Z.-W. Zheng, Y.-Y. Chen, X.-W. Zhou, M.-M. Huo, B. Zhao, M.-Y. Guo, Short-Term Wind Power Forecasting Using Empirical Mode Decomposition and RBFNN, International Journal of Smart Grid and Clean Energy 2 (2013) 192–199. https://doi.org/10.12720/sgce.2.2.192-199.
[98] J. Wang, W. Zhang, Y. Li, J. Wang, Z. Dang, Forecasting wind speed using empirical mode decomposition and Elman neural network, Applied Soft Computing Journal 23 (2014). https://doi.org/10.1016/j.asoc.2014.06.027.
[99] Y. He, Y. Wang, Short-term wind power prediction based on EEMD–LASSO–QRNN model, Appl Soft Comput 105 (2021). https://doi.org/10.1016/j.asoc.2021.107288.
[100] Y. Ding, Z. Chen, H. Zhang, X. Wang, Y. Guo, A short-term wind power prediction model based on CEEMD and WOA-KELM, Renew Energy 189 (2022). https://doi.org/10.1016/j.renene.2022.02.108.
[101] W. Hu, Y. He, Z. Liu, J. Tan, M. Yang, J. Chen, Toward a digital twin: Time series prediction based on a hybrid ensemble empirical mode decomposition and BO-LSTM neural networks, Journal of Mechanical Design, Transactions of the ASME 143 (2021). https://doi.org/10.1115/1.4048414.
[102] Y. Zhang, Y. Chen, Application of hybrid model based on CEEMDAN, SVD, PSO to wind energy prediction, Environmental Science and Pollution Research 29 (2022). https://doi.org/10.1007/s11356-021-16997-3.
[103] Y. Song, S. Zeng, J. Ma, J. Guo, A fault diagnosis method for roller bearing based on empirical wavelet transform decomposition with adaptive empirical mode segmentation, Measurement (Lond) 117 (2018). https://doi.org/10.1016/j.measurement.2017.12.029.
[104] W. Liu, S. Cao, Y. Chen, Seismic time-frequency analysis via empirical wavelet transform, IEEE Geoscience and Remote Sensing Letters 13 (2016). https://doi.org/10.1109/LGRS.2015.2493198.
[105] J. Hu, J. Wang, K. Ma, A hybrid technique for short-term wind speed prediction, Energy 81 (2015). https://doi.org/10.1016/j.energy.2014.12.074.
[106] J. Hu, J. Wang, Short-term wind speed prediction using empirical wavelet transform and Gaussian process regression, Energy 93 (2015). https://doi.org/10.1016/j.energy.2015.10.041.
[107] J. Hu, J. Wang, L. Xiao, A hybrid approach based on the Gaussian process with t-observation model for short-term wind speed forecasts, Renew Energy 114 (2017). https://doi.org/10.1016/j.renene.2017.05.093.

[108] R. Mandzhieva, R. Subhankulova, Data-driven applications for wind energy analysis and prediction: The case of "La Haute Borne" wind farm, Digital Chemical Engineering 4 (2022). https://doi.org/10.1016/j.dche.2022.100048.
[109] Y. Li, Y. Li, X. Chen, J. Yu, H. Yang, L. Wang, A new underwater acoustic signal denoising technique based on CEEMDAN, mutual information, permutation entropy, and wavelet threshold denoising, Entropy 20 (2018). https://doi.org/10.3390/e20080563.
[110] M. Sun, Q. Li, P. Lin, Short-Term Stock Price Forecasting Based on an SVD-LSTM Model, Intelligent Automation & Soft Computing 28 (2021). https://doi.org/10.32604/iasc.2021.014962.
[111] Y. Shi, L. Zhang, S. Lu, Q. Liu, Short-Term Demand Prediction of Shared Bikes Based on LSTM Network, Electronics (Basel) 12 (2023). https://doi.org/10.3390/electronics12061381.
[112] Y. He, Y. Wang, Short-term wind power prediction based on EEMD–LASSO–QRNN model, Appl Soft Comput 105 (2021). https://doi.org/10.1016/j.asoc.2021.107288.
[113] H. Liu, C. Yu, H. Wu, Z. Duan, G. Yan, A new hybrid ensemble deep reinforcement learning model for wind speed short term forecasting, Energy 202 (2020). https://doi.org/10.1016/j.energy.2020.117794.
[114] J. Tong, L. Xie, S. Fang, W. Yang, K. Zhang, Hourly solar irradiance forecasting based on encoder–decoder model using series decomposition and dynamic error compensation, Energy Convers Manag 270 (2022). https://doi.org/10.1016/j.enconman.2022.116049.
[115] S. Karim, A. Faissal, E.B. Noureddine, National Renewable Energy Laboratory, in: Handbook of Algal Biofuels, 2022. https://doi.org/10.1016/b978-0-12-823764-9.00006-6.