A Multi-Phase Deep Learning Framework for Multi-Step Short-Term Wind Power Forecasting in Presence of Uncertainties

Rashidi, M.; Tara, N.; Mehrasa, M.; Taheri, S.; Vahedi, H.

DOI
10.1109/ACCESS.2025.3631586

Publication date
2025

Document Version
Final published version

Published in
IEEE Access

Citation (APA)
Rashidi, M., Tara, N., Mehrasa, M., Taheri, S., & Vahedi, H. (2025). A Multi-Phase Deep Learning Framework for Multi-Step Short-Term Wind Power Forecasting in Presence of Uncertainties. *IEEE Access*, *13*, 193553-193574. https://doi.org/10.1109/ACCESS.2025.3631586

IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 10 October 2025, accepted 6 November 2025, date of publication 11 November 2025, date of current version 18 November 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3631586

RESEARCH ARTICLE

A Multi-Phase Deep Learning Framework for Multi-Step Short-Term Wind Power Forecasting in Presence of Uncertainties

MOHAMMAD RASHIDI[1], NARGES TARA[1], MAJID MEHRASA[2], (Senior Member, IEEE), SHAMSODIN TAHERI[3], (Senior Member, IEEE), AND HANI VAHEDI[4], (Senior Member, IEEE)

[1]Faculty of Electrical and Computer Engineering, Babol Noshirvani University of Technology, Babol, Mazandaran 4714871167, Iran
[2]Department of Electrical and Computer Engineering, The University of New Orleans, New Orleans, LA 70148, USA
[3]Department of Computer Science and Engineering, University of Quebec in Outaouais (UQO), Gatineau, QC J8X 3X7, Canada
[4]Department of Electrical Sustainable Energy, Delft University of Technology, 2628 CD Delft, The Netherlands

Corresponding author: Mohammad Rashidi (mrashidi@stu.nit.ac.ir)

ABSTRACT With the expanding share of wind energy in power grids, accurate forecasting has become critical for maintaining system stability and operational efficiency. Notwithstanding, forecasting accuracy is compromised by uncertainties from fluctuating wind speeds and meteorological conditions. This paper proposes a novel multi-phase short-term wind power forecasting framework (multi-step ahead forecasting over a 1-hour horizon). Thus, decomposition of the wind power signal and feature extraction are initially implemented using Variational Mode Decomposition (VMD) and Principal Components Analysis (PCA), respectively, aiming to enhance input quality and reduce computational burden. The proposed forecasting model is built on a hybrid DL architecture merging a Convolutional Neural Network (CNN), Attention Mechanism (AM), and Deep Feedforward Neural Network (DFFNN). Given the impact of decomposition levels and extracted PCA components count on forecasting performance, a search-based scheme is developed to explore a pre-defined space (maximum decomposition level and extracted components count) to determine the optimal configuration for each interval. In the next phase, a Fuzzy Decision-Making (FDM) technique is employed to select a balanced and optimal configuration for the proposed model across the year. To demonstrate the proposed architecture's efficacy and generalizability, the model is tested on two real-world data from La Haute Borne wind farm in France and Hill of Towie wind farm in Scotland. Results demonstrate that the proposed architecture with the selected configurations achieves significant accuracy and generalization, with average NRMSEs and NMAEs values of 0.428% and 0.333% for La Haute Borne wind farm and 0.502% and 0.381% for Hill of Towie wind farm.

INDEX TERMS Attention mechanism, fuzzy decision making, hybrid deep learning model, multi-step forecasting, variational mode decomposition, wind power forecasting.

I. INTRODUCTION

Forecasting renewable sources power generation is of paramount significance in energy systems since it affects not only operational costs, but also the optimum planning and management of energy resources. Accurate forecast of wind power generation helps grids maximize energy distribution, remove unwanted interruptions, and reduce the cost of supply of electricity. In this regard, the application of contemporary technologies including Artificial Intelligence (AI) and notably Machine Learning (ML) can be rather effective. Complex electricity usage patterns influenced by several elements including wind speed, wind direction, and temporal condition can be captured and analyzed using ML models. This capability enhances the efficiency of power system and improves the forecasting accuracy.

Numerous studies have been conducted to develop forecasting model and accuracy using different approach. This section reviews recent contributions and key advancements in ML methods, feature extraction, and signal process-

The associate editor coordinating the review of this manuscript and approving it for publication was Mouloud Denai.

ing techniques. In [1], a long short-term memory (LSTM) model along with the empirical Fourier decomposition (EFD) method and the gray wolf optimization algorithm were employed to accurately predict the short-term and single-step wind speeds. In [2], the Bi-LSTM and CNN models along with a graph-based data reconstruction approach were adopted. What makes this approach unique is that it combines these methods for short-term forecasting with multiple steps. The authors of [3] used the Double Attention-based Spatial–Temporal Network (DA-STNet) and the Attentional Graph Network (GAT) to learn how to correlate space and time focusing on the extraction of both spatial and temporal features at the same time as the innovations of this work. In [4], the Generalized Matrix Factorization (GMF) method and Sequential-to-Sequential (Seq2Seq) model were used based on LSTM for ultra-short term and multi-step forecasting. The authors, in [5], combined short-term and multi-step wind power forecasting using LSTM model and SCADA data simultaneously. A model called SSA-VMD-Seq2Seq was suggested in [6] consisting of a CNN and a two-way Gated Recurrent Unit for short-term multi-step forecasting. In [7], CNN-LSTM-AM model was proposed to forecast the power of offshore wind turbines without using meteorological data.

Feature selection and extraction are of utmost importance for wind turbine power generation prediction, as the data related to atmospheric conditions and wind behavior can be complex and volatile. By selecting appropriate features such as wind speed, wind direction, temperature, and air pressure, relevant information can be extracted from raw signals and help forecasting models identify more accurate patterns. Some other research conducted in the field of wind power generation prediction with feature selection or extraction methods are reviewed. In [8], the use of temporal convolutional network (TCN) and hyperparameter optimization with orthogonal representations has brought a certain innovation in automating the extraction of complex features and improving multi-step wind power forecasting. The research conducted in [9] described a GRU-CNN hybrid model that used CNN layers for automatic feature extraction leading to very accurate forecasting in very short amounts of time. The main innovation is the creation of a hybrid model that combines CNN and GRU to improve accuracy and speed. In [10], the authors used both convolutional and recurrent networks together to create a focused hybrid model. This model utilized meteorological data and features extracted from time series data to make accurate medium- and long-term predictions of wind power. The main innovation of this research is the use of three optimization algorithms, Scikit-opt, Optuna, and Hyperopt, for hyperparameters, which optimized the performance of CNN and LSTM by analyzing the effects of investigated methods for tuning hyperparameters to improve forecasting accuracy The research conducted in [11] used a novel approach to decompose eigenstates with the aim of extracting frequency features that were applied to the LSTM model. This innovative approach has significantly increased the accuracy of the model in predicting wind power by reducing the data noise. In [12], two optimization algorithms, Stochastic Fractal Search (SFS) and Particle Swarm Optimization (PSO), were used to fine-tune the parameters of the LSTM model for improving the short-term forecast of wind power resulting in an increase in the accuracy of multi-step forecasts. However, PCA only captures linear latent features as it is a linear technique, it offers the benefits of denoising and reducing redundancy in input dataset [13]. Actually, employing non-linear PCA techniques (e.g., Kernel PCA and Autoencoders) are more computationally expensive. Nonetheless, integrating PCA with non-linear architectures that capture non-linearity contributes to both computational burden reduction and accuracy improvement, according to [14] and [15].

Accurate prediction is fundamentally supported by signal processing, as it facilitates the extraction of meaningful patterns and features from raw data. For the extraction of more particular data features, the authors, in [16], used the discrete wavelet transform (DWT) to break the wind signal into subseries of different frequencies. On the other hand, [17] used both DWT and Euclidean distance concurrently to decompose the wind time series, therefore acting as techniques for dimensionality reduction and data processing. Similarly, in [18], the wavelet transform created much subseries by breaking down the wind speed data. Two techniques were used in [19] and [20] to deconstruct the wind power time series into discrete intrinsic mode functions (IMFs): VMD and Full Empirical Mode Decomposition with Adaptive Noise (CEEMDAN). In [21], VMD was employed, alongside additional techniques such as Singular Spectrum Analysis (SSA) for enhanced signal processing and noise elimination. In [22], the Fourier series method and the Recursive Kalman Filter were employed to process and decompose the requisite signals. In [23], the isolated forest technique was employed to detect outliers in the dataset for processing high-frequency Supervisory Control and Data Acquisition (SCADA) data, being a form of signal processing aimed at mitigating the impact of anomalous data. In [24], the authors additionally utilized the Ensemble Empirical Mode Decomposition (EEMD) technique aiming to address the issue of mode mixing inherent in EMD as an enhanced version of the EMD technique.

In most of the existing research in the field of wind power generation forecasting, simultaneous consideration of signal processing methods and feature selection/extraction has been neglected, which leads to lower accuracy, increased computational burden, increased learning time, and misleading of the learning machine. Therefore, noise in the feature signal will lead to reduced model accuracy and increased computational complexity. However, disregarding the relationships between the input features and the target variable, namely wind power generation, may result in increased computational complexity and extended processing time for forecasting architectures. Proper consideration of these interdependencies is essen-

tial to optimizing model performance and ensuring efficient resource utilization. Although some studies considered both signal processing and feature extraction, they often lacked a systematic approach for optimizing the parameters associated with these methods. In this paper, the wind power forecasting is accomplished using the proposed DL architecture. The contributions of this paper are:

- An integrated framework is proposed for signal processing, feature extraction, and a DL architecture with the aim of determining the optimal configuration of the proposed forecasting framework.
- Implementing and assessing advanced signal processing methods on wind power generation time series with 10-minute steps, in order to find a more effective approach for denoising the wind power signal.
- Proposing an innovative DL architecture that integrates CNN, AM, and DFFNN for multi-step forecasting of farm-level wind power, validated on two different wind farms to demonstrate its generalizability and robustness.
- Proposing an algorithm, called ESA, to examine all feasible configurations of decomposition levels and extracted components.
- Employing an FDM approach to select an optimal framework that minimizes the forecasting errors across the whole year.

The remainder of this paper is structured as follows. Section I provides a description of the La Haute Borne wind farm and wind power penetration in France. Section II presents the design and implementation of the forecasting framework. Section III reviews the state-of-the-art methods used for comparison. Section IV details the experimental setup. Section V discusses the results of the study. Finally, Section VI concludes the paper, emphasizing the key contributions and findings.

II. DESCRIPTION OF LA HAUTE BORNE AND HILL OF TOWIE WIND FARMS

The La Haute Borne wind farm, located near Vaudeville-le-Haut in the Grand Est region of France, has four MM82 turbines, each with a capacity of 2 MW, resulting in a total rated capacity of 8 MW. The onshore turbines were designed by Maia Eolis and managed by ENGIE. The location serves as a research site. The data obtained from the site is freely accessible for scientific research, allowing academics globally to utilize it for their studies in wind energy. The plant is acknowledged as a successful instance of collaboration between industry and academia. By 2023, France will possess an installed capacity of approximately 23 GW for wind turbines. Additionally, France has contributed around 37.9 TWh to the nation's energy consumption, representing 8.3% of the total electricity consumption, derived from wind energy [25]. France, a leading country in Europe's renewable energy sector, aims to enhance its wind energy capacity, targeting an installed capacity of 35 GW to 44 GW by 2030. The historical data pertaining to the wind power plant considered in this study originates from 2016, with a time interval of 10 minutes, provided in [26]. To validate the robustness and generalizability of the proposed framework, a dataset from Hill of Towie wind farm is also adopted. Hill of Towie wind farm is located in Dufftown in Scotland with nominal power of 48.3 MW with 21 SWT-2.3-93 turbines (each with capacity of 2.3 MW). The dataset includes historical wind speed, wind direction, ambient temperature and generated power at 10-minutes step intervals for 2024 [27].

Fig. 1 presents the correlation matrix of La Haute Borne wind farm, illustrating the principal correlations. A significant and deep correlation exists between wind speed and wind turbine apparent power (0.91%). This aligns with the wind power principle ($Power \propto Speed^3$) and indicates that variations in wind speed account for approximately 83% of the fluctuations in output power. The negative correlation of temperature with apparent power (-0.25) and wind speed (-0.22) indicates the effect of reduced air density at high temperatures, therefore reducing the kinetic energy available for the wind turbine. The small correlation of wind direction with other components (0.02 to 0.12) suggests that the yaw control system of the La Haute Borne wind farm efficiently changes the orientation of the turbine in line with the wind. Exploratory data analysis of input signals and target variables is shown in Fig. 2. The scatter plot matrix reveals a strong non-linear positive correlation between wind speed and power generation, with the relationship approaching a saturation plateau at higher speeds. This indicates that wind speed is the primary determinant of power generation. As it is clear from, the distribution of the apparent power is right-skewed, suggesting that most observations correspond to low power levels. In contrast, both temperature and wind direction exhibit much weaker and more indirect influences on power generation. High power outputs are observed across a broad range of temperatures and in all wind directions, implying that these factors act as secondary drivers or that their effects are mitigated by operational control mechanisms, such as the turbine yaw control system. Figs. 3(a)-(d) illustrates the input signals of La Haute Borne wind farm in 2016.

III. DESIGN AND IMPLEMENTATION OF FORECASTING FRAMEWORK

This research proposes a multi-phase framework for farm-scale short-term multi-step wind power forecasting. The forecasting is performed using a multi-step 1-hour ahead approach, continuously predicting for a 24-hour horizon with a single training session, achieving a satisfactory forecasting accuracy. Initially, the historical wind power generation signal is preprocessed using VMD in order to eliminate inherent noise. Subsequently, the resultant features, such as: the decomposed power modes, wind speed, wind direction, and ambient temperature are fed to the feature extraction technique (PCA) to diminishing computational burden and rising learnability of DL models by reducing the dimensionality of dataset. Then, the preprocessed signals are fed to

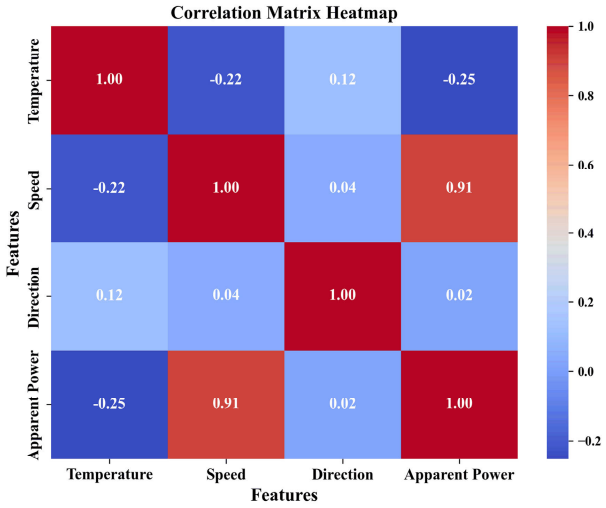


FIGURE 1. Correlation matrix showing the strength and direction of relationships among temperature, speed, direction, and apparent power of dataset from La Haute Borne wind farm.

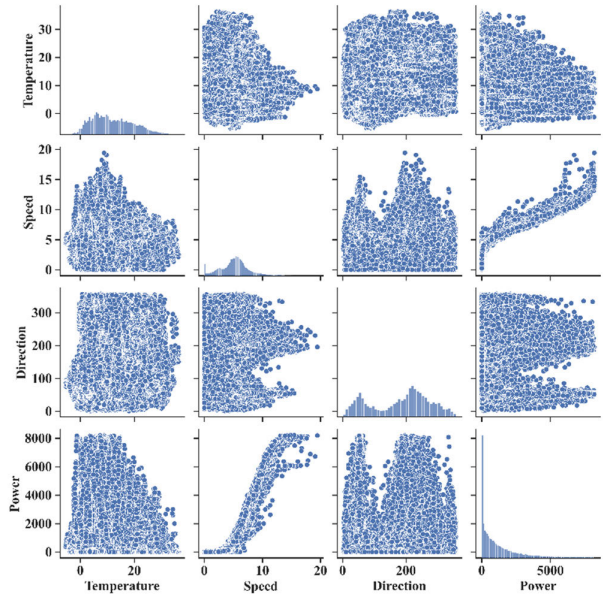


FIGURE 2. Visualization of bivariate correlations and univariate distributions of dataset from La Haute Borne wind farm.

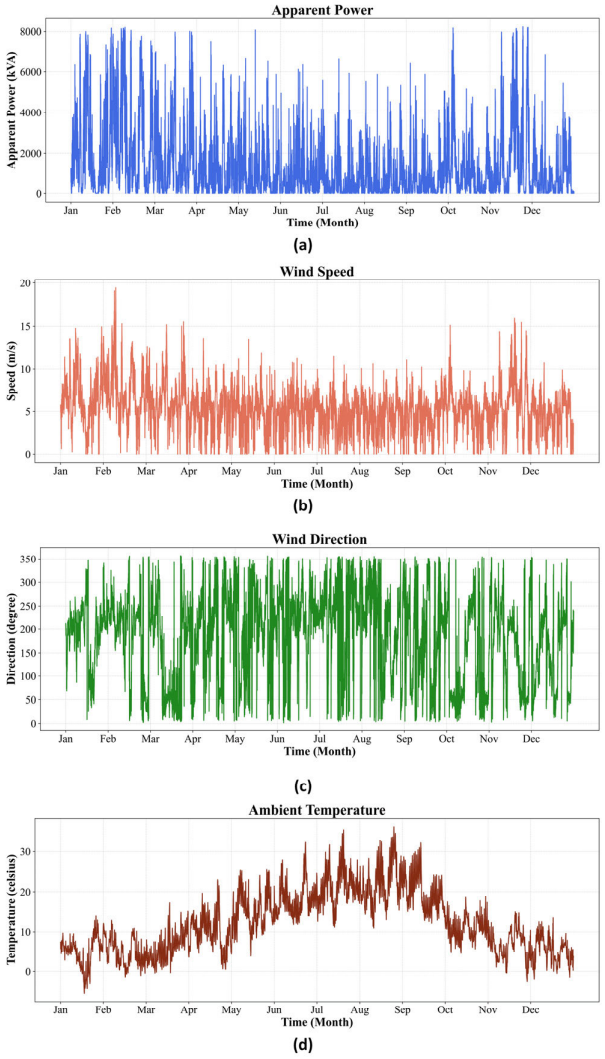


FIGURE 3. Input dataset acquired from measurement sensors from La Haute Borne wind farm in 2016: (a) wind power; (b) wind speed; (c) wind direction; (d) ambient temperature.

the proposed DL architecture (CNN-AM-DFFNN) to learn the inherent wind power generation pattern. Give the impact of the decomposition levels by VMD and the number of extracted components by PCA, and Exhaustive Searching Algorithm is proposed in order to explore into the predefined search space to monitor their impact on the proposed architectures' accuracy. Finally, the FDM technique is employed to select a configuration contributing to the optimal performance and accuracy throughout the year.

A. THE PROPOSED VMD-PCA-CAF MODEL

The forecasting framework outlined in this study is illustrated in Fig. 4. Initially, the aforementioned signal processing techniques are employed to denoise and decompose the wind power signal. Given that the target variable (wind power) is not entirely influenced by all input features, features with higher correlation should be prioritized. Although considered features have non-linear relations and PCA captures linear latent features, this feature extraction technique is selected due to its efficiency, stability, and strong ability to reduce redundancy among correlated features. While nonlinear methods (e.g., kernel PCA, autoencoders) can extract richer patterns, our hybrid model already handles nonlinear dependencies. Thus, PCA serves as a fast and effective preprocessing step that complements the deep model while keeping computational cost manageable.

This study uses a multi-phase forecasting method based on a single training approach to achieve a highly accurate 1-hour forecasting over a 24-hour, with low and acceptable errors. The prediction accuracy is influenced by the

hyperparameters, belonging to the signal processing techniques, the feature extraction technique, and the learning architecture. The objective is to select the architecture that yields the maximum accuracy across all seasons by conducting the training and forecasting process using data from four specific months of the year. To enable the practical implementation of the proposed model by the wind farm operator, an FDM method is employed to determine the optimal configuration comprising the decomposition level and number of extracted components that yields the lowest forecasting error.

1) VARIATIONAL MODE DECOMPOSITION (VMD)

It is a modern and powerful method in signal processing that decomposes non-linear and non-stationary signals into band-limited intrinsic mode functions. The VMD is built on an optimization framework and simultaneously calculates the modes and their corresponding frequencies. This feature makes VMD very effective in analyzing nonlinear, complex, and multiscale data [28], specifically wind power. The goal of VMD is to decompose a signal $f(t)$ into several sub-signals or modes $\mu_k(t)$, each of which covers a distinct frequency band. In this study, the wind power modes are extracted in such a way that their energy is concentrated within specific frequency band. The mathematical structure of VMD is given in (1) [29].

$$f(t) = \sum_{k=1}^{K} \mu_k(t) \tag{1}$$

$\mu_k(t)$ represents the *k-th* wind power signal IMF, while ω_k denotes its corresponding center frequency. $\lambda(t)$ is the Lagrange multiplier enforcing the reconstruction constraint of the original wind power signal, and α is the weighting parameter controlling the balance between data fidelity and spectral compactness in the decomposition. $f(t)$ is the original wind power signal, and $\sigma(t)+j/(\pi.t)$ corresponds to the Hilbert kernel used to construct the analytic signal. The operator $\partial(t)$ denotes the time derivative, and $exp(-j.\omega_k.t)$ is the complex exponential used for frequency shifting. VMD is investigated in this study, due to some advantages. Given the nonlinear and non-stationary nature of wind power signals, VMD is considered an efficient technique, as it provides simultaneous mode extraction, greater stability against noise and unstable signals, and more precise control over mode frequencies through optimization.

$$\begin{aligned} &\arg min_{\mu_k,\omega_k,\lambda} \\ &\alpha \sum_{k=1}^{K} \left\| \partial_t \cdot \left[\left(\sigma(t) + \frac{j}{\pi.t} \right) \cdot \mu_k(t) \right] \cdot \exp(-j \cdot \omega_k \cdot t) \right\|_2^2 \\ &\quad + \left\| f(t) - \sum_{k=1}^{K} \mu_k(t) \right\|_2^2 + \left\langle \lambda(t) \cdot f(t) - \sum_{k=1}^{K} \mu_k(t) \right\rangle \end{aligned} \tag{2}$$

2) PRINCIPAL COMPONENT ANALYSIS (PCA) FOR FEATURE EXTRACTION

Various dataset obtained from measurement sensors and Supervisory Control and Data Acquisition (SCADA) System can be employed in wind power forecasting problem. In this paper, features including wind power, wind speed, wind direction, and temperature are adopted. These signals are often interrelated and may contain noise. Feature extraction converts raw data into a lower-dimension and extracts more significant features that improve the prediction accuracy. Additionally, employing PCA in wind power forecasting problem results in a more meaningful feature set by extracting the components that explain the greatest variance in the data. These components typically capture more significant and impactful information in wind power forecasts. Noise is prevalent in wind power due to unpredictable external factors, including wind turbulence and other meteorological conditions. PCA improves the quality of the data by removing less significant variations and focuses on the main components. PCA is an unsupervised dimensionality reduction method that maps multidimensional data to a lower-dimensional space [30], preserving most of the variance in the data, specifically in seasons with volatile condition (fluctuating wind speed and direction). Steps for PCA implementation are as follows: standardizing the decomposed wind power modes, wind speed, wind direction, and temperature, determining vectors and eigenvalues, and then, creating a covariance matrix. Actually, covariance depicts features interdependence, contributing to the selection of the principal components. Ultimately, the data is converted into a lower-dimension space.

3) CONVOLUTIONAL NEURAL NETWORK (CNN)

Wind power has complex temporal dependencies as it is characterized by nonlinear and nonstationary fluctuations, yielded by varying atmospheric condition (wind speed, and direction). This work integrates a one-dimensional CNNd coupled with an AM layer and a DFFNN for multi-step wind power forecasting. One-dimensional CNNs are a variation of CNNs specifically built for handling sequential or time-series data. Rather than considering two-dimensional inputs like images, these networks extract spatiotemporal patterns from one-dimensional data using convolutional layers. The CNN model contains convolutional, pooling, and activation layers, extracting features from the input dataset together. Each filter is a sliding window with a defined kernel size that identifies information such as recurring patterns or rapid changes through the convolution operation applied to each time interval. The fundamental aspect of the 1-dimensional convolution layer is its capacity to identify local correlations inside the data over a defined time interval. By changing the number of filters, and parameters, including kernel size and stride, the model would be able to capture either long-term or short-term trends in the wind power data. This layer produces a feature map for forecasting that can be fed to following

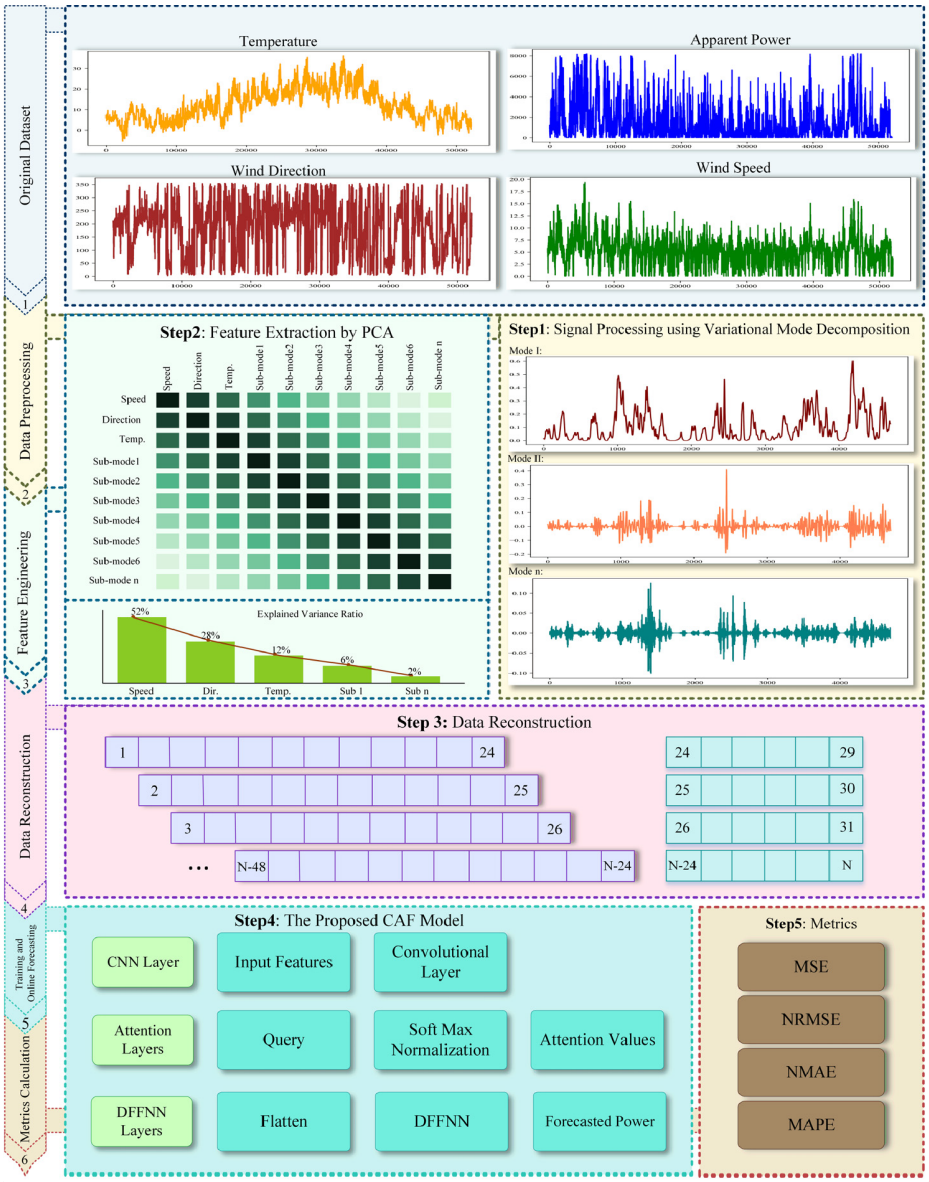


FIGURE 4. The proposed framework.

layers: AM and DFFNN. The mathematical depiction of 1-dimensional convolution layer is described in (3) [31]. Overall, the essential temporal pattern of wind power can be effectively captured by CNN, contributing to solid basis for the following models. $x[i{+}k]$ denotes the input inside the spatial range of the filter, b provides the bias value, y[i] shows the output at index i, k is the kernel number, and S^K is the kernel size.

$$y[i] = \sum_{k=0}^{S^K-1} x[i+k] \cdot \omega[k] + b \tag{3}$$

The 1-dimensional Pooling layer is characterized by two primary components: an aggregation operator applied to each segment, and a sliding window of predefined size that divides the input data into smaller segments. The aggregation operator is often Max Pooling or Average Pooling. This layer is used to reduce the dimensionality of the data and extract key features without losing the original information of the spatiotemporal patterns. The mathematical relationship of Pooling and its variants are presented in (4)-(6). y'[i] denotes the output at position i after applying Pooling. $x[i : i{+}P]$ denotes the part of the input data that falls into the Pooling window. $x[j]$ is the value of the input signal (or input sequence) at index j. P denotes the size of the Pooling window, which determines how many elements of the input data are considered for calculations at each step.

$$\begin{aligned} y'[i] = &\ \text{Pooling}\ (x[i : i+P]) \\ &\text{Max}\ \ \text{Pooling}(x[i : i+P]) \end{aligned} \tag{4}$$

$$= \max(x[i], x[i+1], \ldots, x[i+P-1]) \quad (5)$$

$$\text{Average Pooling}(x[i:i+P]) = \frac{1}{P} \cdot \sum_{j=1}^{i+P-1} x[j] \quad (6)$$

4) ATTENTION MECHANISM (AM)

The AM is an advanced DL approach for wind power forecasting. It enables the model to focus on the most critical parts of the time series, capture long-term dependency, filter the noise, and learn the complex nonlinear interactions among features. This method calculates the dynamic weights at each time point in the input data (including wind speed, wind direction, and temperature) to determine the relative importance of each factor in predicting the output power of wind turbines. Under this scheme, the learnable coefficients first convert the $n \times d$ input data into $n \times d_k$ matrices Q, K, and V. The attention score is then computed by means of the inner product $Q.K^t$ and a scaling factor of $\frac{1}{\sqrt{d_K}}$ is employed to obtain the final attention weights via SoftMax [32]. In essence, this process lets the model detect intricate patterns and nonlinear correlations in the wind data and generate accurate forecasts of the output power. Since this method models the nonlinear interactions and the complex time dependencies in the wind data, it shows a notably greater performance than conventional techniques for time series forecasting.

Equations (7)-(10) represent the AM. α_{ij} is the relative significance of every key vector k_j to the query vector q_i. Inner multiplication of q_i and k_j computes the attention weight α_{ij}. Then, it is normalized using the SoftMax smoothing function. This normalization process yields a single total attention weight for each query vector. Finally, the weight α_{ij} is applied dynamically to the V (value vector) to get the AM layer's output. The larger α_{ij} is, the more influential the corresponding input data point (key) is in the final forecasting.

$$Q = X \cdot W_Q, K = X \cdot W_K, V = X \cdot W_V \quad (7)$$

$$Attention\,(Q, k, V) = \text{soft max}\left(\frac{Q \cdot K^T}{\sqrt{d_K}}\right) \cdot V \quad (8)$$

$$\alpha_{ij} = \frac{\exp\left(\frac{q_i \cdot k_j^T}{\sqrt{d_k}}\right)}{\sum_{j=1}^{n} \exp\left(\frac{q_i \cdot k_j^T}{\sqrt{d_k}}\right)} \quad (9)$$

$$Output = \sum_{j=1}^{n} \alpha_{ij} \cdot v_j \quad (10)$$

5) DEEP FEEDFORWARD NEURAL NETWORK (DFFNN)

A Deep Feedforward Neural Network (DFFNN) is an extension of the standard DFFNN, characterized by increased depth and multiple hidden layers. In wind power forecasting problems, a single DFFNN can be a competing model, because of its low architectural complexity, ease of implementation, and stability across different data. In this architecture, the time-series data flows unidirectionally from the input to the output layer without loops or feedback. Each neuron computes its output using a weighted sum of its inputs followed by a nonlinear activation function, commonly *sigmoid*, tanh, or *ReLU*. The additional hidden layers enable the network to capture intricate patterns and hierarchical features in the data, which is especially valuable for modeling the nonlinear and dynamic nature of wind power generation. A DFFNN can be represented in (11) [33]. $W^{[l]}$ denotes the weight matrix in the l-th layer, $a[l]$ and $a^[l-1]$ denote the outputs of neurons in the l-th and $(l\text{-}1)$-th layers, and $b^{[l]}$ is the bias vector in the l-th layer.

$$a^{[l]} = f\left(W^{[l]} \otimes a^{[l-1]} + b^{[l]}\right) \quad (11)$$

B. THE PROPOSED EXHAUSTIVE SEARCHING ALGORITHM (ESA)

A primary problem in wind power prediction through DL is the effective selection of preprocessing parameters. The decomposition levels, and the number of components extracted through PCA all influence the forecasting accuracy of the learning architectures. Thus, optimizing each component alone does not guarantee the minimum forecasting accuracy for the learning model. The selection of the optimal signal decomposition levels, and the number of extracted of features significantly influence prediction accuracy. Conventional approaches typically employ a predetermined number of modes and extracted components, potentially resulting in the omission of critical information or the incorporation of irrelevant data.

Thus, this study presents an algorithm to address this challenge by carefully assessing all potential combinations of decomposition levels and the number of extracted components through an Exhaustive Searching Algorithm, represented in Algorithm 1. The algorithm explores in a predefined search space and is applied to all specified intervals. Unlike heuristic or stochastic methods (e.g., random or Bayesian search), the ESA systematically evaluates all feasible VMD–PCA configurations within a tractable space, ensuring the true optimum is identified without risk of premature convergence. Its novelty lies in jointly optimizing signal decomposition and feature extraction with the DL model. This method facilitates the identification of the optimal configuration for achieving maximum forecasting accuracy, while simultaneously reducing the necessity for manual tuning (like through trial and error) via an automated framework.

C. FUZZY DECISION MAKING (FDM) FOR FINAL MODEL SELECTION

To comprehensively evaluate and demonstrate the efficacy of the proposed method, the algorithm is applied to four different intervals of the year. These intervals, which include the months of January-February, May-June, July-August, and November-December, are selected as representative of diverse weather conditions in different seasons of the year. This intelligent selection allows evaluating the performance

Algorithm 1 The Proposed ESA Algorithm for Exploring All 546 Configurations

Input: Raw features, S
Input: Target variable, $P^{\text{apparent power}}$
Output: The optimal framework configuration, $C_{optimal}$
Require: Initial feature numbers ($N^{features}$)
Require: Min. and Max. decomposition levels [$N^{Min-VMD}$, $N^{Max-VMD}$]
Require: Min. and Max. extracted components numbers [$N^{Min-PCA}$, $N^{Max-PCA}$]
Require: Sliding window size (W_{length}, S_{step})
Initialize $\mathbf{C_{optimal}} \leftarrow$ NULL, $N^{counter} \leftarrow 1$, $N^{config} \leftarrow 546$
for i =1 to 4 do ▷ Loop through predefined time intervals
  for j from $N^{Min-VMD}$ to $N^{Max-VMD}$ do ▷ Loop through VMD
$D_{power-modes} \leftarrow$ VMD ($P^{apparentpower}$, j)
    ▷ Apply VMD for Decomposition
$D_{new} \leftarrow$ Concatenation ($D_{power-modes}$, S)
    ▷ Concatenating decomposed modes to the dataset
$D_{normalized} \leftarrow$ MinMaxScaler (D_{new})
    ▷ Normalize the resulting data
(X, Y) $\leftarrow$ SlidingWindow ($D_{normalized}$, W_{length}, S_{step})
    ▷ Create feature/target sets
(X_{train}, Y_{train}), (X_{test}, Y_{test}) $\leftarrow$ SplitData (X, Y)
      ▷ Create train and test sets
while $N^{counter} \leq N^{config}$
if ($N^{counter}$ -j-$N^{features}$) ≤ 0 do
    ▷ Loop through PCA
PCA $\leftarrow$ FitPCA (X_{train}) ▷ Apply PCA for feature extraction
$X_{train-pca} \leftarrow$ Transform (PCA, X_{train})
$X_{test-pca} \leftarrow$ Transform (PCA, X_{test})
$M_{trained} \leftarrow$ CAF ($X_{train-pca}$, Y_{train})
    ▷ Train the proposed architecture (CAF)
$Y_{forecasted} \leftarrow$ CAF ($M_{trained}$, $X_{test-pca}$)
    ▷ Forecast with the trained architecture
$E_{recorded} \leftarrow$ Errors ($Y_{forecasted}$, Y_{test})
    ▷ Calculate the NRMSE, NMAE
end if
$N^{counter} = N^{counter}+1$
end while
end for
end for
$\mathbf{C_{optimal}} \leftarrow$ FDM ($E_{recorded}$)
    ▷ Apply the employed FDM to identify the configuration with the lowest Error across all months
return $\mathbf{C_{optimal}}$

of the algorithm under different atmospheric conditions, including temperature changes, and wind patterns in different seasons. The algorithm is applied independently for each time period, and the performance of all possible configurations are recorded for the subsequent analysis.

To address the operational demand for a single and optimal forecasting model, an FDM technique is designed and implemented for the ultimate model selection in this paper. This intelligent system selects a configuration that not only has superior performance but also provides the most stable and reliable results under variable weather conditions. It is achieved by considering the forecasting metrics obtained by the proposed integrated search algorithm in all the intervals.

Decision-making based on fuzzy set theory operates such that each value is assigned a degree of membership in the interval [0, 1], rather than belonging entirely to a single set. In the wind power forecasting problem, choosing the best combination of the decomposition levels (in signal processing step) and the number of components (in feature extraction step) is important, the FDM can help identify the most optimal configuration to be applied to the forecasting model across all months. The NAME and NRMSE are the main metrics used to evaluate the learning models. Based on the membership functions, these errors are categorized into three groups: low error, medium error, and high error. Finally, the fuzzy results are converted into a definite value through the defuzzification method, which indicates the final rank of the model. In this paper, the simple averaging defuzzification process is selected, which is defined using (12) and (13) [34], where N^c denotes the number of configurations.

$$\text{Average NMAE} = \frac{\sum_{c=1}^{N^c} \text{NMAE}_c}{N^c} \tag{12}$$

$$\text{Average NRMSE} = \frac{\sum_{c=1}^{N^c} \text{NRMSE}_c}{N^c} \tag{13}$$

IV. STATE-OF-THE-ART METHODS

A number of DL models and strong signal processing techniques are employed as benchmarks in this study, as described below.

A. SIGNAL PROCESSING

Significant variations and volatility in the power output of wind power facilities can misguide the learning algorithm. Consequently, the signal processing techniques must be adopted for the purification process by eliminating noise from the features, including power output, wind speed, wind direction, and ambient temperature [35].

This paper employs some of the most potent signal processing methods to enhance the data quality through the feature analysis including EMD, and Wavelet Transform. Wavelet transform is one of the effective methods that allows the noise removal and extraction of important features by decomposing the signal into high and low-frequency components. Moreover, these signal processing techniques enhance the quality of the input data and reduce the models' learning time by diminishing the complexity of the patterns within the data. The application of EMD technique to identify the inherent components of the input signals enhances one's

understanding of the system's dynamic behavior. The previous findings demonstrated that employing these signal processing techniques markedly improves the prediction accuracy and bolsters the models stability amongst noisy data.

1) EMPIRICAL MODE DECOMPOSITION (EMD)

In signal processing, the EMD is a potent method for breaking down non-stationary and complicated dataset into a collection of intrinsic modes (IMFs). Analyzing the local maxima and minima of the signal and envelopes of curves help with obtaining such modes [36]. By breaking down input signals into simpler components, EMD helps identify hidden patterns, reduces noise, and increases the accuracy of forecasting models in the field of wind energy forecasting. EMD is different from other signal decomposition techniques, because it can separate nonlinear and non-stationary signals without requiring prior knowledge of basis functions. Hence, EMD is a suitable method for processing wind power signal, as it has non-linear and non-stationary nature. This approach divides the signal into a collection of IMFs, each with a different temporal scale by iteratively screening it. Moreover, EMD allows more exact investigation of the dynamic behavior of complex systems including wind turbines by preserving the physical properties of the original signal during the decomposition process. The mathematical relationship of EMD technique is based on the decomposition of a signal $x(t)$ into a set of intrinsic modes and a residual trend $r(t)$ as expressed in (14) [34], where $c_i(t)$ is the *i-th* intrinsic mode, $r(t)$ is the residual value representing the overall trend of the signal, and n represents the number of intrinsic modes.

$$x(t) = \sum_{i=1}^{n} c_i(t) + r(t) \tag{14}$$

2) WAVELET TRANSFORM (WT)

In signal processing, the wavelet transform is a powerful tool that enables simultaneous analysis of nonlinear and non-stationary signals in both the time and frequency domains [37]. In wind power forecasting, the wavelet transform is very useful, because of its ability to extract time-frequency characteristics of the wind and identify complex patterns. This approach provides information about both high-frequency and low-frequency variations concurrently by separating the original signal into wavelet components at several scales. In wind dynamics analysis, noise reduction, and forecast accuracy enhancement, the WT method proves to be highly efficient. The continuous WT is given (15). $s(t)$ is the original wind power signal, and $W(a,b)$ is the wavelet coefficients that indicates the correlation between the wind power signal and the wavelet at scale a (which is the inverse of the frequency and determines the level of details) and the transfer parameter is shown with b (which is in the time domain). Furthermore, ψ^* is the complex conjugate of the parent wavelet.

$$W(a,b) = \int_{-\infty}^{+\infty} s(t) \cdot \psi^* \left(\frac{t-b}{a} \right) dt \tag{15}$$

In practice, the discrete WT is often used, in which the signal is divided into a set of detail coefficients (high detail) and approximation coefficients (low detail). This process is performed recursively at different levels. The discrete WT is used in wind power forecasting to eliminate noise in signals and pull out the important features. Because, it offers higher computational efficiency, and lower redundancy compared to continuous WT, which is computationally demanding. The discrete WT is expressed in (16), where $D_{j,k}$ represents the wavelet coefficients at scale j and translation parameter k. $s(t)$ is the original wind power signal, and $\psi_{j,k}(t)$ is the scaled and translated wavelet of the parent wavelet.

$$D_{j,k} = \int_{-\infty}^{+\infty} s(t) . \psi_{j,k}(t) \, dt \tag{16}$$

B. OTHER ARCHITECTURES

This paper proposed an architecture, called CAF, that utilizes CNN, AM, and DFFNN. Furthermore, to prove the higher efficiency and accuracy of the proposed architecture in wind power forecasting compared to other powerful ones, some other models (RNN, GRU, LSTM, Transformer-based Network and Temporal Convolutional Network) are also studied and are explained in the following subsection.

1) RECURRENT NETWORKS (RNN, GRU, AND LSTM)

Due to complex nonlinear behavior of wind power and meteorological factors, learning models must be capable of identifying these trends and high-frequency variations. RNNs are widely used DL models for sequence and time series data, due to their ability to capture temporal dependencies. However, vanilla RNNs suffer from issues such as gradient instability and difficulty in learning long-term patterns. To address these limitations, advanced architectures like Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) have been developed. LSTM employs input, output, and forget gates to manage information flow and effectively capture long-term dependencies. GRU, a simplified version of LSTM with update and reset gates, offers lower computational complexity while maintaining competitive performance. These properties make LSTM and GRU well-suited for wind power forecasting, where data are nonlinear, volatile, and require models with both accuracy and adaptability. Given their inherent capability for capturing complex trends and fluctuation, these models were selected for this present analysis.

The mathematical formulation of RNN shown in (17) and (18) illustrates how the hidden and output layers are calculated at each time step t. In this network, the hidden state h_t combines the information about the current input x_t and the previous hidden state h_{t-1} and is updated using the weights W_h and U_h and the bias vector b_h. This operation applies an activation function σ such as $ReLU$, and the tanh

function. It also applies weights W_y, bias b_y, and corresponding activation function ϕ to calculate the network y_t from the new hidden state h_t. This operation enables the RNN model to absorb information from previous time steps and learn temporal relationships for the purpose of processing or predicting sequential input.

$$h_t = \sigma\left(W_h \cdot x_t + U_h \cdot h_{t-1} + b_h\right) \tag{17}$$
$$y_t = \phi\left(W_y.h_t + b_y\right) \tag{18}$$

The update and reset gates are updated with the new hidden state at each time step according to (19) and (20) [38]. In GRU, there are several parameters that play a role in the updating process and control the information flow. Two update gates z_t and reset r_t define major parameters. Every gate updates the different weights and biases. The weights consist of W_z to update gates and also W_r and U_r focus on resetting gates. Apart from every gate having a calculated value of the gate, bias vectors b_z and b_r help calibrate the models for each gate. The sigmoid activation function then appends a nonlinear function to the computed gate value and to the inputs. These settings enable the GRU network to maintain valuable knowledge in past states without including the new data in the hidden state.

$$z_t = \sigma\left(W_z \cdot x_t + U_z \cdot h_{t-1} + b_z\right) \tag{19}$$
$$r_t = \sigma\left(W_r \cdot x_t + U_r \cdot h_{t-1} + b_r\right) \tag{20}$$

Equations (21)-(26) describe the LSTM model. The LSTM network consists of a set of parameters that are used to learn temporal dependencies and accurately predict data. The input at each time step x_t and the hidden state of the previous step h_{t-1} along with the previous cell state c_{t-1} are used as input data for the calculations. The weights consist of matrices W_f, W_i, and W_c for the inputs of forget gate, input gate, and candidate cell state, respectively. U_f, U_i, *and* U_c are weight matrices that handle information from previous hidden states h_{t-1} and apply to the respective components of the forget, input, and candidate cell gates. Each gate is also provided with bias vectors b. Such values enable the LSTM to retain core information for valid forecasting and retrieve long and short-term dependencies. c_t is the cell state of the LSTM network at time step t. O_t is the output gate value. It is clear from the formulation that unnecessary information can be filtered out during learning, improving the model's ability to identify patterns and key meteorological data.

$$f_t = \sigma\left(W_f \cdot x_t + U_f \cdot h_{t-1} + b_f\right) \tag{21}$$
$$i_t = \sigma\left(W_i \cdot x_t + U_i \cdot h_{t-1} + b_i\right) \tag{22}$$
$$c'_t = \tanh\left(W_c \cdot x_t + U_c \cdot h_{t-1} + b_c\right) \tag{23}$$
$$c_t = f_t \odot c_{t-1} + i_t \odot c'_t \tag{24}$$
$$O_t = \sigma\left(W_O \cdot x_t + U_O \cdot h_{t-1} + b_O\right) \tag{25}$$
$$h_t = O_t \odot \tanh(c_t) \tag{26}$$

2) TRANSFORMER-BASED NETWORK

The Transformer is DL model that is primarily based on self-attention [39]. The Transformer architecture has been recently utilized for wind power forecasting as an alternative to recurrent and convolutional models. The Transformer's self-attention mechanism adaptively captures both local variations and long-range dependencies in wind power time series, while positional encodings preserve sequential order. The Transformer is a scalable and efficient framework that improves forecasting accuracy for not stationary or stable renewable energy dataset. It involves multi-head attention, position-wise feed-forward networks, and residual normalization.

3) TEMPORAL CONVOLUTIONAL NETWORK (TCN)

Temporal Convolutional Networks (TCNs) are a type of DL model that uses causal and dilated convolutions to model sequential data. It enables capturing long-range temporal dependencies without using recurrent connections. Unlike traditional recurrent networks, TCNs process the whole input sequence at the same time. This framework contributes to more efficiency and avoids vanishing gradients. Actually, the receptive field grows exponentially with depth with dilated convolutions. It allows network to learn both short-term changes and long-term trends. TCNs are effective at capturing the nonlinear temporal patterns of power generation, which improves accuracy of short-term and medium-term forecasting. TCNs are a great tool for forecasting wind energy as they are able to capture complex changes with high training efficiency. The details and structure of TCN are described in [40].

V. EXPERIMENTAL SETUP

A. FORECASTING PERFORMANCE METRICS

In the field of wind power forecasting, it is essential to use multiple evaluation criteria to accurately evaluate the performance of forecasting models. The power generation rate in wind power plants fluctuates greatly due to changes in wind speed, and atmospheric conditions. MSE (mean square of error) is more sensitive to sharp changes in the forecast of generated power and is useful for identifying large forecasting errors. NMAE, by normalizing the mean absolute magnitude of the errors, allows for a comparison of the performance of models in forecasting wind power. Another metric termed NRMSE (normalized root mean squared error), provides a suitable benchmark for evaluating both long-term and short-term wind power forecasts. The simultaneous use of these metrics helps with holistic evaluation of the wind power forecasting architectures performance from multiple perspectives and enhances the reliability of forecasts for generation planning and grid management. The metrics used in this paper are shown in (27)-(30) [41]. The value of the forecasted target variable and its actual value are represented by y'', and y_i, respectively. y_c denotes the capacity of the wind

TABLE 1. Architectures parameters.

Models	Hyperparameters	
DFFNN	Hidden Layer 1	Units = 256 Activation = `ReLU` Drop Out 1 = 0.3
	Hidden Layer 2	Units = 32 Activation = `ReLU` Drop Out 2 = 0.2
	DFFNN Output Layer	Units = 6
RNN-DFFNN	RNN Layer 1	Units = 256
	RNN Layer 2	Units = 128
	DFFNN Layer 1	Units = 64 Activation = `ReLU`
	DFFNN Output Layer	Units = 6
GRU-DFFNN	GRU Layer 1	Units = 128
	GRU Layer 2	Units = 64
	DFFNN Layer 1	Units = 64
	DFFNN Output Layer	Units = 6
LSTM-DFFNN	LSTM Layer 1	Units = 256
	LSTM Layer 2	Units = 128
	DFFNN Layer 1	Units = 64 Activation = `ReLU`
	DFFNN Output Layer	Units = 6
DeepTCN-Net	TCN	Filters = 64 Kernel Size = 3 Dilation Rates = [1, 2, 4, 8] Drop Out = 0.1
	DFFNN Layer 1	Units = 64
	DFFNN Output Layer	Units = 6
CTrans-Net	Convolution Layer 1	Filters = 128 Kernel Size = 3 Padding = Valid Strides = 3
	Convolution Layer 2	Filters = 128 Kernel Size = 1 Padding = Valid Strides = 1
	Multi-Head Attention Layer	Attention Units = 128 Attention Heads = 4 Activation = `tanh`
	DFFNN Layer 1	Units = 64 Activation = `ReLU`
	DFFNN Layer 2	Units = 6 Activation = Linear
CNN-AM-DFFNN	Convolution Layer 1	Filters = 128 Kernel Size = 3 Padding = Valid Strides = 3
	Convolution Layer 2	Filters = 128 Kernel Size = 1 Padding = Valid Strides = 1
	Attention Layer	Attention Units = 128
	DFFNN Layer 1	Units = 64 Activation = `ReLU`
	DFFNN Layer 2	Units = 6

farm, and n denotes the number of samples.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(\widehat{y_i} - y_i)^2 \tag{27}$$

$$\text{NMAE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\widehat{y_i} - y_i}{y_c}\right| \times 100 \tag{28}$$

$$\text{NRMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\frac{(\widehat{y_i} - y_i)^2}{y_c}} \times 100 \tag{29}$$

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\widehat{y_i} - y_i}{y_i}\right| \times 100 \tag{30}$$

B. DATA RECONSTRUCTION AND STANDARDIZATION (DATA SCALING)

In multi-step forecasting, especially for wind power forecasting, it is necessary to create a time series using the sliding window approach to leverage past information for predicting the future values. This method allows the model to capture patterns and temporal relationships between the data and provides more accurate forecasts for different time steps. For this purpose, the first 10 minutes of data, including features such as wind speed, wind direction, temperature, and generated power, must be converted into time-series input. A fixed time window (1 hours, 6 samples) of the past data is slid forward continuously. For each window, the inputs and forecast variable for the next time steps (wind power forecast 1 hour later) are defined. After this step, the data should be divided into training, validation, and test sets. Furthermore, to improve the performance of the model, data normalization is performed using the Min-Max Scaling method so that features with different scales do not negatively affect the performance of the model. Data standardization using the Min-Max scaling method is depicted in (31), where X^{scaled} represents the standardized value of the data, X represents the input signal's value, and X^{max} and X^{min} represent the upper and lower limits of the data, respectively.

$$X^{scaled} = \frac{X - X^{min}}{X^{max} - X^{min}} \tag{31}$$

C. MODEL CONFIGURATIONS

The architecture and configuration specifications of the key components used in the proposed model are comprehensively represented in Fig. 5 and TABLE 1, respectively.

To track temporal interdependence in wind power data, the proposed CAF architecture is meticulously designed. The first convolutional layer (configured with 128 filters, a kernel size of 3, and stride of 3) extracts short-term temporal variations with simultaneous reduction of redundancy through down-sampling. This attribute enables the model to capture rapid variation in wind power without overwhelming computational burden. The features extracted from the first layer are employed in the second convolutional layer (with 128 filters, a kernel size of 1 and stride of 1). The convolutional layer acts as a feature projection mechanism and preserves expressive capacity. These integrated convolutional

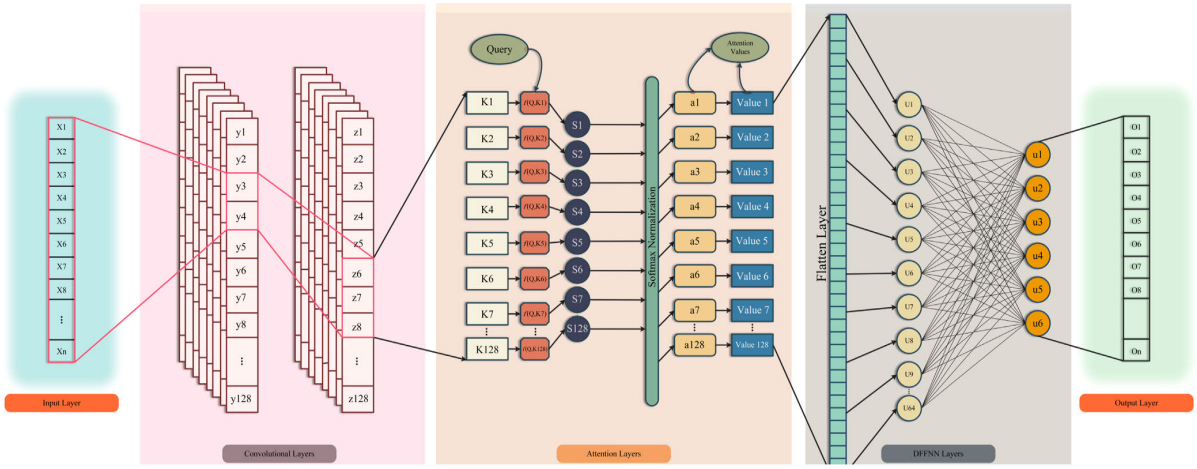


FIGURE 5. Graphical representation of the proposed CAF architecture.

layers create a comprehensive hierarchical representation that strikes a balance between scalability ands accuracy.

To further enhance forecasting accuracy, an attention mechanism with 128 units is used, which provides the ability to assign importance weights to extracted features from convolutional layers. It enhances the model's potential to prioritize informative time steps and suppresses irrelevant fluctuations. This improves both interpretability and robustness against noisy input features. Subsequently, the dense feed-forward layers integrate the existent features into a nonlinear latent space. The first dense layer, with 64 neurons and activation, captures higher-order nonlinear dependencies while ensuring efficient training and avoiding vanishing gradient issues, whereas the final dense layer outputs the six-step-ahead (1-hour ahead) forecast. The parameter choices (number of filters, kernel size, stride, and units) were decided to balance expressive capacity with generalization capability, ensuring that the model remains computationally efficient while delivering accurate and stable multi-step forecast.

TABLE 2. Case studies for ablation study.

Case Studies	VMD	PCA	Architecture		
			CNN	AM	DFFNN
Case 1	×	×	✓	×	✓
Case 2	×	✓	✓	×	✓
Case 3	✓	×	✓	×	✓
Case 4	✓	✓	✓	×	✓
Case 5	×	×	×	✓	✓
Case 6	×	✓	×	✓	✓
Case 7	✓	×	×	✓	✓
Case 8	✓	✓	×	✓	✓
Case 9	×	×	✓	✓	✓
Case 10	×	✓	✓	✓	✓
Case 11	✓	×	✓	✓	✓
Case 12	✓	✓	✓	✓	✓

D. COMPUTATIONAL ENVIRONMENT

The experiments were performed on a computer with Intel(R) Core (TM) i5-7500 CPU @ 3.40 GHz and 16 GB RAM with Windows 10 64-bit operating system with Python language and Tensorflow, and Scikit-Learn libraries.

VI. RESULTS AND DISCUSSION

This section evaluates the proposed VPCAF data-driven forecasting architecture and its accuracy in forecasting wind power generation in three phases. To evaluate and validate the generalizability the proposed architecture, two datasets from La Haute Borne in France and Hill of Towie in Scotland are adopted. For the first wind farm, training periods include [January 5 to February 5], [July 4 to August 4], [May 13 to June 13], and [November 24 to December 24]. Forecasts are then conducted for the day following each period: February 6, June 14, August 5, and December 25. The training periods for the second wind farm are [December 5 to January 5], [April 12 to May 12] and [July 27 to August 27]. The forecasts are performed for January 6, May 13 and August 28, respectively. This approach is designed to evaluate the accuracy of the forecast under different meteorological and temporal conditions.

In this study, a set of advanced signal processing techniques, including EMD (with decomposition into 10 intrinsic modes), WT (with decomposition into 3 levels that include an approximation coefficient, and 3 detail coefficients), and VMD (with decomposition into 6 modes), are used to denoise and decompose input apparent power signal and forecast wind power in a 1-hour time horizon (through the multi-step scheme). To validate the superiority of VMD for denoising and decomposition of wind power signal, its performance is compared with EMD and WT in the first phase. For this purpose, after training the proposed CNN-AM-DFFNN model with data from July 4 to

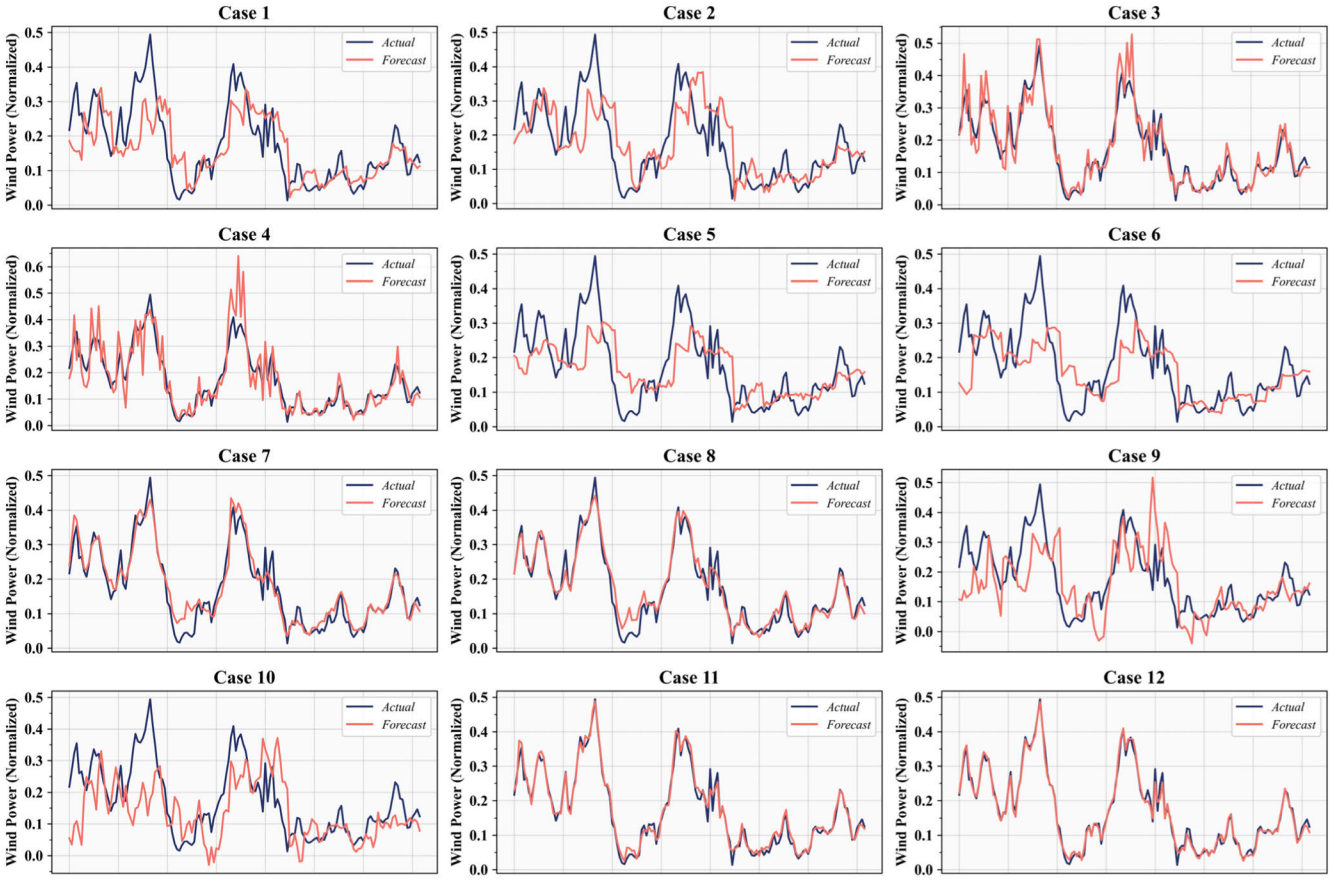


FIGURE 6. Ablation Study for 12 Case Studies: Forecasting performance analysis under different model variations.

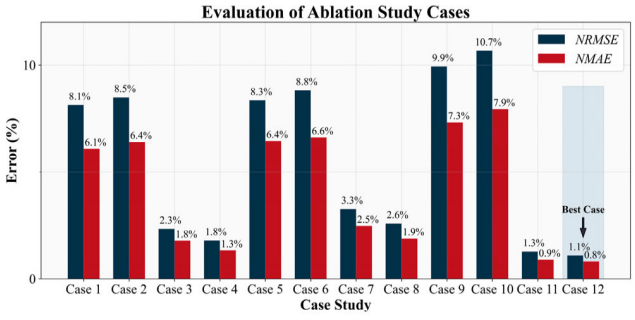


FIGURE 7. NRMSE and NMAE evaluation of ablation study cases.

August 4, the forecast for wind power generation on the next day (August 5) is performed. In addition to the proposed model, several benchmark models were also evaluated for comparison.

In the second phase of the study, all configurations are independently assessed across all time intervals using the proposed ESA. Specifically, for each period, the DL model is trained and conducts forecasting with all feasible configurations of decomposition and feature extraction for each month, aiming to identify and select the best configuration matched with the trend of that specific month.

In the third phase of the study, considering the operational demand of the wind farm to use a single and optimal forecasting architecture throughout the year, an FDM scheme based on simple averaging is proposed in this study to select the optimal configuration that can provide acceptable forecast accuracy for each month. This approach is designed to ensure the accuracy and efficiency of the model under different temporal and atmospheric conditions. The aforementioned phases utilize the dataset from La Haute Borne wind farm. For validation, the dataset from Hill of Towie wind farm is employed.

A. PHASE 1: SIGNAL PROCESSING, TRAINING, AND FORECASTING USING THE PREDICTIVE MODELS

In the initial phase, the ablation study of the proposed model and efficacy of DL-based models (DFFNN, RNN, GRU, LSTM, DeepTCN-Net, CTrans-Net and the proposed architecture) combined with different signal processing techniques are investigated. In the next part, the results for forecasts using the proposed and baseline architectures are presented. Finally, computational complexity of the proposed model in

TABLE 3. Results of forecasting based on architectures and signal processing techniques.

Models	VMD			EMD			WT		
	MSE	NMAE (%)	NRMSE (%)	MSE	NMAE (%)	NRMSE (%)	MSE	NMAE (%)	NRMSE (%)
DFFNN	0.1119	2.5608	3.3449	0.7058	6.5759	8.4014	0.2036	3.4101	4.5119
RNN-DFFNN	0.0197	1.0493	1.4034	0.4943	5.3628	7.0306	0.1639	3.2405	4.0487
GRU-DFFNN	0.0204	1.0353	1.4276	0.6201	5.7280	7.8744	0.1254	2.7214	3.5413
LSTM-DFFNN	0.0251	1.1275	1.5836	0.6120	5.9949	7.8227	0.1549	3.0202	3.9354
DeepTCN-Net	0.0597	1.8865	2.4434	0.6623	6.3990	8.1384	0.1509	3.0994	3.8856
CTrans-Net	0.1095	2.6517	3.3086	0.3471	1.3291	5.8919	0.3098	4.2173	5.5663
CAF	**0.0166**	**0.9474**	**1.2899**	0.4859	5.4275	6.9705	0.2122	3.4362	4.6065

terms of training time, inference time and number of trainable hyperparameters is discussed.

1) ABLATION STUDY

In order to demonstrate the efficacy of the proposed method and each element's contribution to the enhancement of the forecasting accuracy, an ablation study is implemented in 12 case studies shown in TABLE 2.

To demonstrate the superiority and effectiveness and of the proposed DL architecture (case 12), a comparative analysis of ablation study is carried out. The results for forecasts and corresponding error values are illustrated in Figs. 6 and 7. Architectures without VMD or PCA (e.g., cases 1, 2, 5, 6, 9, and 10) present relatively higher errors, with NRMSE and NMAE values exceeding 8% and 6%, respectively, which are the sign of their limited forecasting capability. The adoption of VMD and PCA integrated with the proposed CAF model significantly improves performance, as observed in cases 3, 4, 7, and 8, with NRMSE values between 1.7% and 3.3%. As it is demonstrated, the proposed model (CAF) in case 12 outperforms other cases with the lowest NMRSE and NMAE values (1.087% and 0.812%). Nonetheless, the synergistic use of VMD and PCA for signal decomposition and feature extraction, and CNN and AM for temporal feature learning, and DFFNN for prediction leads to robust and superior accuracy, proving the advantage of the proposed architecture.

2) PERFORMANCE COMPARISON WITH BASELINE ARCHITECTURES

The results of the forecasting criterion obtained from the investigated combinations are displayed in TABLE 3. Figs. 8(a)-(g) illustrate the forecasting results. The forecasting results obtained using VMD technique clearly demonstrate that the proposed CAF model significantly outperforms other methods, achieving the lowest errors across all evaluation metrics, with an MSE of 0.0166, an NMAE of 0.9474%, and an NRMSE of 1.2899%. Other hybrid architectures outperformed the RNN-DFFNN model, which achieved an MSE of 0.0197, an NMAE of 1.0493%, and an NRMSE of 1.4034%. The GRU-DFFNN model ranked third, with an MSE of 0.0204, an NMAE of 1.0353%, and an NRMSE of 1.4276%. Despite the intricate architecture of LSTM-DFFNN, it exhibited suboptimal performance with an MSE of 0.0251, an NMAE of 1.1275%, and a NRMSE of 1.5836%. The DeepTCN-Net model recorded higher errors compared with these hybrid architectures, with an MSE of 0.0597, an NMAE of 1.8865%, and an NRMSE of 2.4434%. The CTrans-Net model exhibited even weaker performance, with an MSE of 0.1095, an NMAE of 2.6517%, and an NRMSE of 3.3086%. The DFFNN model had the weakest performance among all models, with an MSE values of 0.1119, an NMAE of 2.5608%, and an NRMSE of 3.3449%, indicating a substantial performance gap. The results clearly indicate the superiority of the proposed model (CAF) in the first phase of this study over hybrid architectures, due to its strong VMD-based feature processing capacity.

The implementation of the EMD technique, combined with the proposed CAF model, exhibited superior performance relative to the other analyzed models, with an MSE of 0.4859, an NMAE of 5.4275%, and an NRMSE of 6.9705%. The RNN-DFFNN model ranks second, exhibiting a marginal performance difference from the CAF, with an MSE of 0.4943, an NMAE of 5.3628%, and an NRMSE of 7.0306%. The CTrans-Net model recorded the next best performance, with an MSE of 0.5471, an NMAE of 5.9591%, and an NRMSE of 7.8919%. In contrast,

DeepTCN-Net produced much higher errors, with an MSE of 0.6623, an NMAE of 6.3990%, and an NRMSE of 8.1384%. The LSTM-DFFNN and GRU-DFFNN models exhibited comparable performance, with an MSEs of 0.612 and 0.6201, respectively. The DFFNN model has the weakest performance among all examined architectures, yielding the greatest errors, including an MSE of 0.7058, an NMAE of 6.5759%, and an NRMSE of 8.4014%. The forecasting results demonstrates that integrating EMD with CAF yields high forecasting accuracy. While the relatively close performance of RNN-DFFNN model highlights positive contribution of EMD to signal decomposition, the CAF

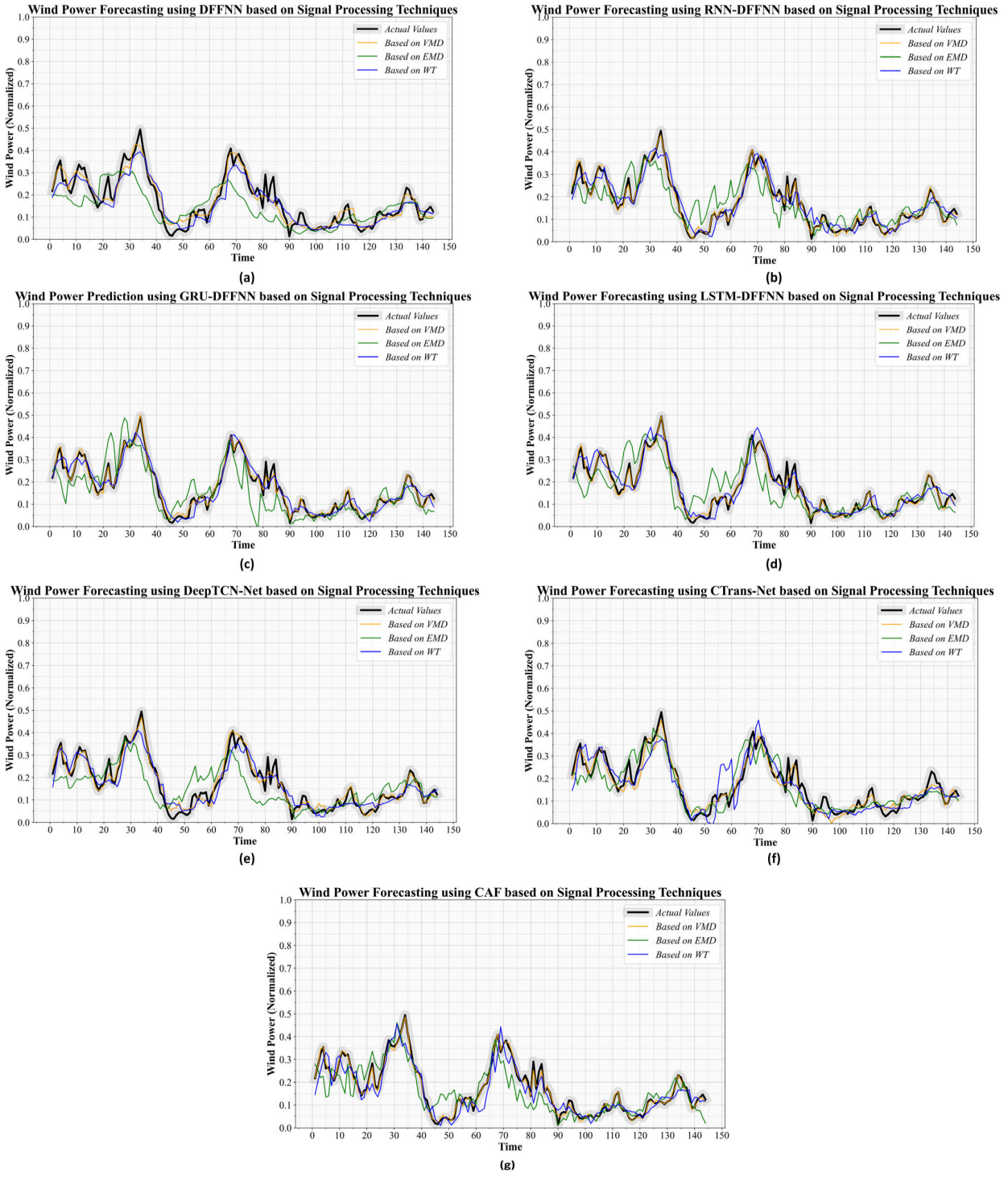


FIGURE 8. Results for multi-step forecasting (August 5) using different models and signal processing techniques: (a) DFFNN; (b) RNN-DFFNN; (c) GRU-DFFNN; (d) LSTM-DFFNN; (e) DeepTCN-Net; (f) CTrans-Net; (g) the proposed model (CNN-AM-DFFNN).

model still outperforms all others, confirming its superior forecasting capability.

The WT technique is employed for signal processing of the apparent power. The GRU-DFFNN model demonstrated exceptional performance, achieving the lowest error metrics, with an MSE of 0.1254, an NMAE of 2.7214%, and an NRMSE of 3.5413%. After that, the LSTM-DFFNN model ranks next with an MSE of 0.1549, an NMAE of 3.0202% and an NRMSE of 3.9354%. The DeepTCN-Net model followed closely with an MSE of 0.1509, an NMAE of 3.0994%,

TABLE 4. Model complexity and computational cost of the proposed and baseline architectures.

Models	Training Time per Epoch (ms)	Inference Time per Forecast (ms)	Total Trainable Hyperparameters
DFFNN	169	50	65126
RNN-DFFNN	980	54	126022
GRU-DFFNN	1510	58	95174
LSTM-DFFNN	3590	70	478150
DeepTCN-Net	626	71	93446
CTrans-Net	880	59	185158
CAF	568	60	118982

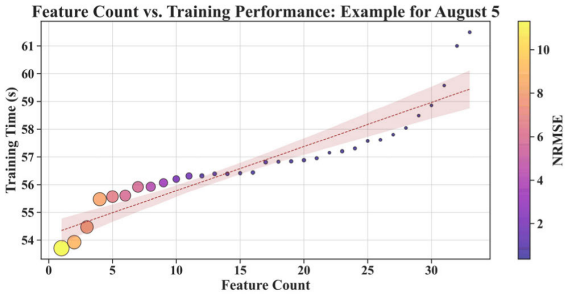


FIGURE 9. Scalability Analysis of the propose architecture: Effect of Feature Count on Training Time and NRMSE for August 5.

and an NRMSE of 3.8856%. The RNN-DFFNN model also showed acceptable performance by recording an MSE of 0.1639, an NMAE of 3.2405% and an NRMSE of 4.0487%. The CTrans-Net model recorded larger errors compared with these, with an MSE of 0.3098, an NMAE of 4.2173%, and an NRMSE of 5.5663%. Using the VMD signal processing technique and the forecasting architecture proposed in this paper has led to superior outcomes relative to other combination of signal processing techniques and forecasting models. This combination is able to extract complex features and distinct frequency modes from the input data with high accuracy, which in turn leads to improved performance of the forecasting model and a significant reduction in forecasting errors. Compared with other signal processing techniques and other forecasting models, the proposed architecture in this paper has been able to achieve the most accurate results in various evaluations, including MSE, NMAE, and NRMSE, and shows significant superiority. These findings indicate the high effectiveness of combining VMD with the proposed CAF model in processing and analyzing complex data. Signal processing-based forecasting with the proposed forecasting model using the VMD method has provided significantly better results than the results obtained by using the empirical mode decomposition and wavelet transform methods, which can be attributed to several reasons. First, VMD, using the adaptive decomposition method, allows for the extraction

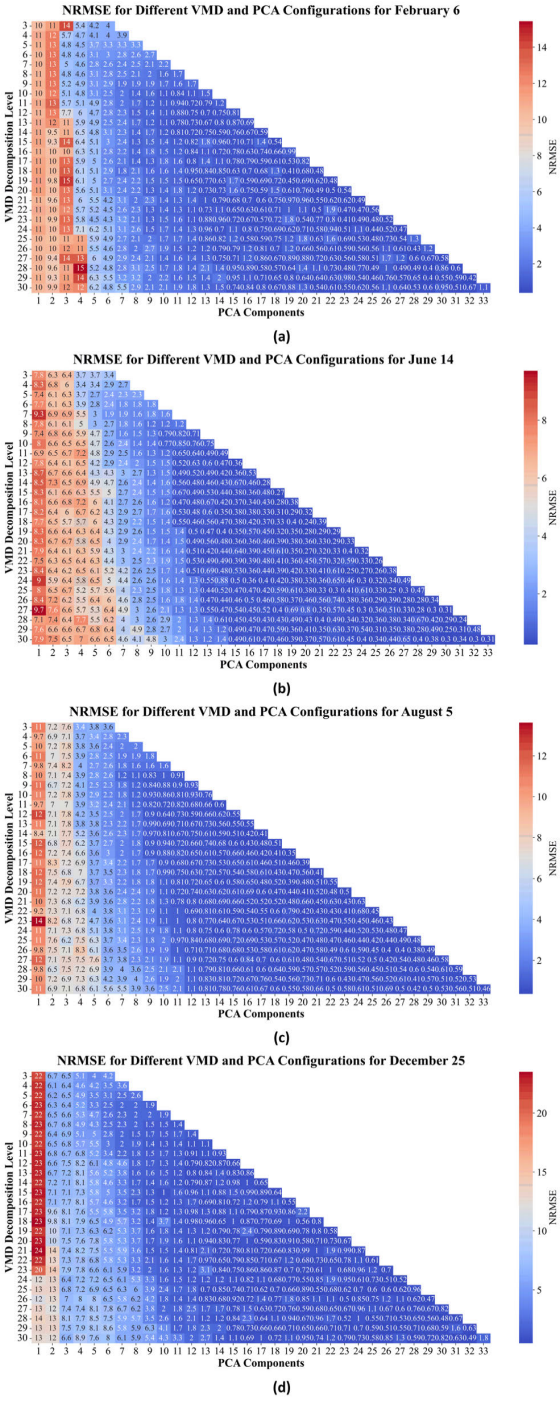


FIGURE 10. Results for NRMSE based on different VMD and PCA configuration for four intervals: (a) February 6; (b) June 14; (c) August 5; (d) December 25.

of signal frequency components, which tend to overlap less with each other. This feature leads to better separation of dynamic patterns in the data. On the other hand, the empirical mode decomposition and wavelet transform methods may not get key signal information correctly, as they are sensitive to noise. The second thing is that VMD made the forecasting model work better by keeping time-frequency coherence.

TABLE 5. Forecast wind power using the FDM-recommended configuration (N_{level} =26 and $N_{components}$ =28).

Time (h)	February 5		June14		August 5		December 25	
	Forecasted (kVA)	Actual (kVA)	Forecasted (kVA)	Actual (kVA)	Forecasted (kVA)	Actual (kVA)	Forecasted (kVA)	Actual (kVA)
1	1899.81	1940.87	1438.46	1443.74	1754.05	1756.79	1894.57	1875.81
2	2093.79	2105.95	1486.27	1497.33	3532.98	3505.60	1535.69	1514.44
3	2047.52	2048.84	2132.51	2123.80	1258.98	1244.72	1867.25	1846.89
4	1606.60	1553.45	1084.25	1117.07	1416.15	1421.18	2134.23	2102.07
5	1708.10	1669.48	1519.84	1523.13	1003.38	1006.90	3907.33	3886.01
6	2243.43	2202.17	2456.97	2454.11	1539.09	1558.65	2562.80	2523.37
7	1513.23	1483.88	1847.07	1892.26	3147.88	3144.13	3007.81	2993.07
8	2914.02	2921.54	480.26	482.68	815.31	778.32	2247.65	2217.55
9	2737.46	2720.03	300.60	268.89	1438.26	1418.89	3824.82	3819.29
10	2859.00	2878.23	658.48	645.92	670.55	679.04	2482.47	2445.24
11	2294.40	2281.99	950.46	923.75	576.17	583.46	3002.21	2976.25
12	2295.61	2270.65	2480.54	2499.53	942.33	936.71	3521.74	3502.81
13	2341.92	2281.52	2190.21	2198.00	797.13	799.32	2575.89	2576.36
14	2588.62	2542.37	1341.24	1330.55	1761.19	1731.59	2696.87	2715.81
15	2615.99	2586.02	1010.66	1029.20	767.39	747.60	3301.32	3325.67
16	2862.97	2783.99	409.81	382.52	1491.53	1475.27	2243.43	2255.41
17	2321.69	2271.80	377.81	364.02	1002.96	990.59	3512.95	3531.13
18	2350.23	2258.72	284.88	266.15	643.44	638.64	6032.07	5951.86
19	2634.59	2679.33	513.52	540.03	1286.06	1274.66	5520.78	5497.63
20	3115.10	3157.02	507.16	512.55	699.27	704.72	5900.56	5890.20
21	2968.36	2908.51	305.00	283.04	943.89	946.77	6723.22	6693.02
22	2400.33	2275.51	763.92	775.95	1194.59	1185.15	7905.38	7910.68
23	2361.22	2274.44	1248.17	1249.10	816.45	823.98	7855.22	7821.54
24	1621.29	1582.68	577.20	605.05	1370.60	1353.46	8033.71	7996.75

This makes it easier to look at complicated data like wind power generation. According to the results obtained from comparisons, the combined VMD-PCA-CAF approach is selected as the optimal architecture proposed in this paper.

3) COMPUTATIONAL COMPLEXITY

TABLE 4 compared the proposed and baseline architectures in term of complexity and computational cost. Model complexity is assessed based on training time per epoch, inference time per forecast and total number of trainable hyperparameters.

Results show that DFFNN has the lowest complexity due to its shallow structure, whereas LSTM-DFFNN indicates the highest complexity that has yielded significantly longer training time (3590 ms/epoch) and a much larger trainable parameter set (∼478k). Given the trainable hyperparameter counts of DeepTCN-Net and CTrans-Net (∼93k and ∼185k), the training times reached to 628 ms/epoch and 880 ms/epoch. Despite the superior performance of CAF compared to LSTM-DFFNN, it has moderately lower complexity and less trainable parameters (∼119k) achieving training time and inference time of 568 ms/epoch and 60 ms/forecast, respectively.

The results indicate that the CAF framework achieves a favorable balance between complexity and performance, being more computationally efficient than recurrent (LSTM, GRU, and RNN) and temporal architectures (DeepTCN-Net and CTrans-Net) while achieving superior forecasting accuracy.

Fig. 9 shows the effect of the number of input features on the training performance and forecasting error of CAF architecture. It indicates that the increasing feature count

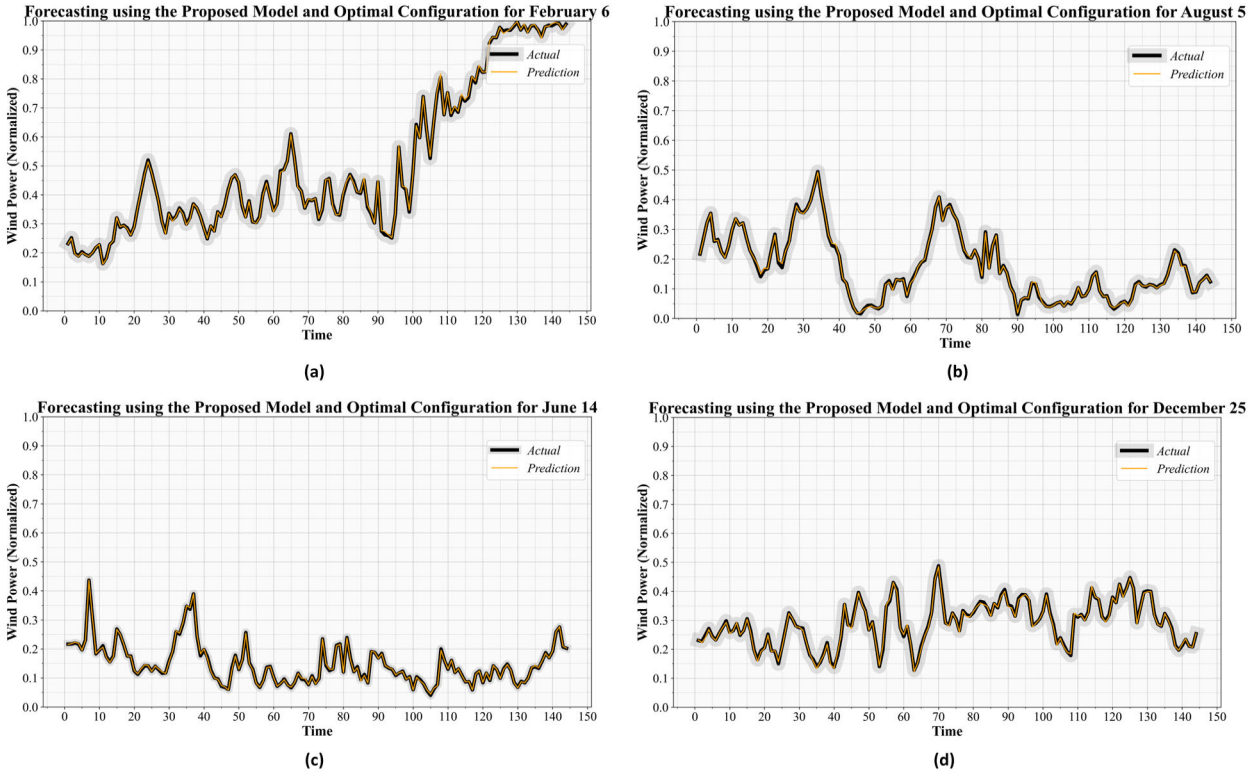


FIGURE 11. Results for multi-step forecasting (next 1 hour) using the proposed architecture (CAF) with optimal decomposition level and extracted components: (a) February 6; (b) August 5; (c) June 14; (d) December 25.

contributes to higher training time as computational burden increases. However, NRMSE does not consistently decrease with more input features, suggesting that additional features raise computational burden without guaranteeing higher accuracy. This underlines the need to carefully choose the right set of features so the model achieves a good balance between accuracy and efficiency, making it both practical and scalable for real-world applications.

B. PHASE 2: APPLYING THE PROPOSED ESA

As proven in the previous section, the combination of VMD and the CAF leads to the highest forecasting accuracy. Algorithm 1 illustrates the design of the proposed ESA, which aims to evaluate all configurations of decomposition levels and extracted components in a predefined space. The ESA systematically explores all possible decomposition levels and number of extracted components using VMD and PCA for each month independently. Then, this configuration is applied to the preprocessing step of the input signals. Following preprocessing, the new dataset is fed into the CAF architecture. Next, the desired forecasting metrics are calculated after training and forecasting. Results for NRMSE Based on Different VMD and PCA Configurations are shown in Figs. 10(a)-(d).

In this work, the effect of the model's dependency to the decomposition levels and number of extracted components are extensively studied. The forecast for February 5 shows that higher accuracy can be achieved with more extracted components (like 29 or 32) and more decomposition levels (like 29 or 28), as shown in Fig. 10(a). Notably, the lowest error is achieved with the combination of 29 levels for decomposition and 29 components after feature extraction, resulting in an NRMSE of 0.399%. The results revealed that increasing the number of decomposition level and extracted components leads to improved forecasting accuracy. Due to the nature of wind power pattern in February, using a more of decomposition levels helps the model overcome fluctuations. In addition, the large number of features extracted by PCA caused more information to be retained from the original data. Then, the model has a greater ability to learn the complex behavior of the signal. At the same time, the combination with 18 decomposition levels and 19 components also demonstrated relatively good performance (an NRMSE of 0.412%), implying that for data with less details, it is acceptable to choose a lower number of decomposition levels and components. However, selecting a decomposition level and number of extracted components close to 28 and 29, respectively, generally yields desirable performance. These results indicate the necessity for a more comprehensive study to extract additional information from the signal, particularly in months such as February that exhibit more intricate patterns. This elucidates the rationale for employing

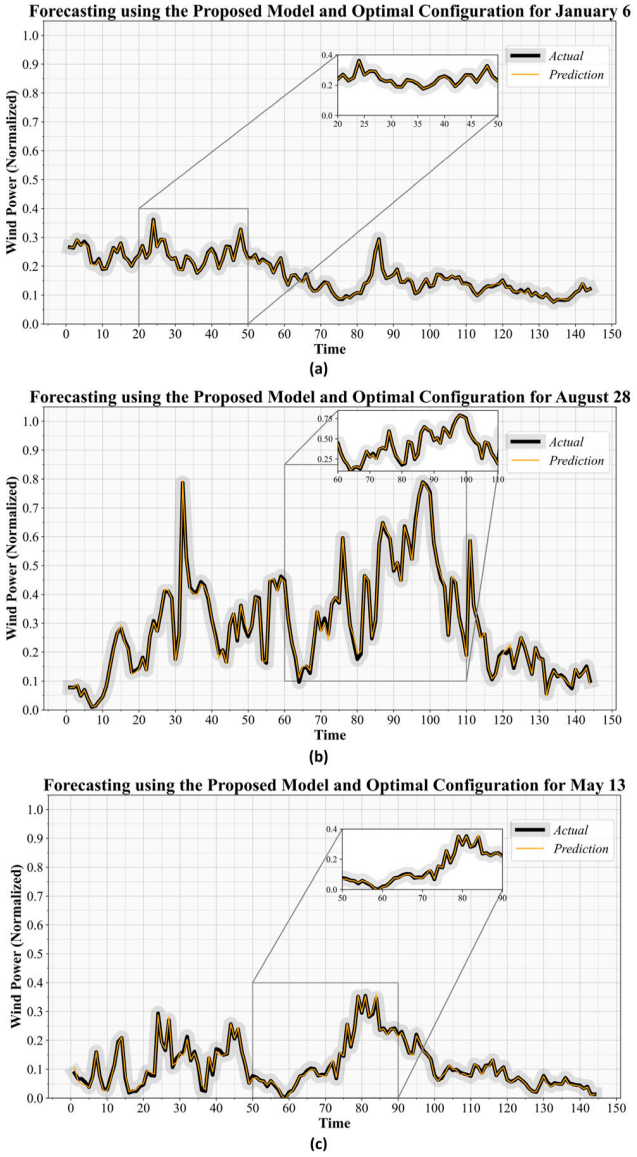


FIGURE 12. Forecasting results for Hill of Towie wind farm using the proposed architecture (CAF) and optimally identified configuration: (a) January 6; (b) August 28; (c) May 13.

a higher number of decomposition levels and extracted components.

The forecast results for June 14 (illustrated in Fig. 10(b)) indicate that employing 28 decomposition levels and 31 components derived from PCA yields optimal model performance, achieving the minimum NRMSE of 0.24%. These results emphasize how crucial careful feature selection and accurate signal decomposition are to raising model accuracy. June is typically marked by gradual changes and the emergence of more stable trends. With 28 decomposition level, VMD enables a finer-grained signal decomposition, facilitating the model to distinguish subtle changes. Furthermore, selecting 31 decomposition levels help the model to keep its complex structure and learn important information in the raw data. However, combinations such as N_{level} =27 and Ncomponents=30 also perform well (an NRMSE of 0.265%). Nonetheless the configuration with 28 decomposition levels and 31 final components also provides the most accurate forecast.

As it is shown in Fig. 10(c), the forecast for August 5 shows that the configuration with 27 decomposition levels and 26 PCA-extracted features achieves the lowest NRMSE, which is 0.380%. The model demonstrated superior performance because of its capacity to learn and adapt to the complex variations and existent noise in the data for this day. It is deduced that better results arise from decomposition levels ranging from 25 to 30 and number of extraction components between 24 and 29. The observed benefits stem from the inherent characteristics of summer wind patterns, which can include localized thermal effects, leading to significant short-term fluctuations in wind speed.

The December 25 forecast shows that the best model with 26 decomposition levels and 29 PCA-extracted features achieves the lowest NRMSE, which is 0.467%, as it is depicted in Fig. 10(d). The model demonstrated superior performance because of its capacity to learn and adapt to the complex variations and noise present in the data for this month. It is deduced that better results usually arise from decomposition levels ranging from 24 to 28 and extraction components between 27 and 31. The observed benefits stem from the inherent fluctuations in winter winds during December, which leads to variation in wind speed and erratic patterns.

C. PHASE 3: IMPLEMENTATION OF THE FDM FOR IDENTIFYING THE OPTIMAL AND FINAL FORECASTING FRAMEWORK

This study employed the simple averaging FDM method to identify the optimal configuration from all possible combinations. This scheme analyzes the normalized values of performance indices (NRMSE) from four distinct time intervals. Each interval contains 546 NRMSE values, collectively forming a 2184-dimensional decision-making space. The goal of this proposed system is to comprehensively analyze the performance of all models in this multidimensional space and select the configuration that not only has the best overall performance, but also shows the most stable behavior under variable weather conditions.

The analysis of the results proves that the configuration N_{level}=*26* and $N_{components}$=*28* is identified as the optimal configuration for the preprocessing step. The forecasts with the selected configuration for four months are shown in Figs. 11(a)-(d). Also, the denormalized values of the forecasts are presented in TABLE 5. This selection is based on the mean value of the performance index (NRMSE), which is calculated as 0.428. The NRMSE values for periods are 0.43% on 5 February, 0.281% on 14 June, 0.383% on 5 August, and 0.617% on 25 December, respectively. These values indicate that the selected framework has an acceptable performance in all time periods. Similar to the outcome obtained from

considering the NRMSE metric, applying FDM to the NMAE values also selects the configuration with $N_{level}=26$ and $N_{components}=28$. The average NMAE value for this configuration is 0.333%, comprising 0.333 on January 5, 0.214% on May 13, 0.291% on July 4, and 0.494% on November 24.

These results demonstrates that the selected combination of decomposition levels and number of final components delivers superior performance across all time periods, achieving a very low error is some cases like on 13 May. Although a relative increase in the error rate is observed on 24 November, the model generally shows good stability across different periods. Therefore, the configuration $N_{level}=26$ and $N_{components}=28$ is identified as the optimal and most stable combination.

D. PERFORMANCE VALIDATION ON THE SECOND WIND FARM

To further evaluate the generalization capability and robustness of the proposed framework, comprising VPCAF with the proposed ESA and employed FDM technique, the second dataset from Hill of Towie wind farm in Scotland is adopted for validation. The second dataset is fed into the ESA to conduct training and forecasting using the proposed CAF architecture across all combinations of decomposition level and extracted features counts. Finally, the FDM technique is applied to identify the most optimal combination over the investigated intervals, presenting each season.

The results for forecasts are shown in Figs. 12(a)-(c). The configuration $N_{level}=28$ and $N_{components}=31$ is determined as the optimal combination. With such decomposition levels and extracted feature count, CAF architecture performs forecasts with the average NRMSE of 0.502% (0.585% for May 13, 0.692% for August 28 and 0.229% for January 6) and NMAE of 0.381% (0.434% for May 13, 0.53% for August 28 and 0.178% for January 6). It demonstrates that the trained model successfully tracks variations and forecasts the wind power generation with substantially high accuracy and precision throughout all seasons.

VII. CONCLUSION

This research presents an innovative deep learning framework, termed VPCAF-FDM, for multi-step short-term (1)-hour) wind power forecasting. The benchmark systems adopted in this study are derived from La-Haute-Borne and Hill of Towie wind power plants, which provide reliable and representative real-world datasets. A multi-phase framework is designed to tackle the substantial issues of forecast accuracy caused by fluctuating wind speeds and uncertain weather conditions. The framework begins with advanced data preprocessing analysis, in which VMD is proved to surpass other techniques in phase 1. The VMD technique is utilized to decompose the historical wind power signal, followed by PCA for feature extraction and noise removal. The core of the proposed forecasting methodology is the DL architecture, termed CAF, which integrates CNN, AM, and DFFNN and has demonstrated the superior accuracy. This work proposed an integrated algorithm during the second phase of this study, called ESA, to examines different key hyperparameters of VMD and PCA, specifically the decomposition levels and the number of extracted components. The model's generalizability and robustness were evaluated using data from four distinct months, which encompassed all four seasons of the year. In the final phase, the FDM technique was implemented to determine a configuration that provides the most balanced and appropriate forecasting efficacy across all seasons. Using the dataset from La-Haute-Borne wind farm, the results confirm the effectiveness accuracy of the proposed framework by identifying and then applying the optimal configuration of $N_{level}=26$ and $N_{components}=28$ to the signal decomposition and feature extraction steps. The identified configuration exhibits extraordinary accuracy, with average NRMSE and NMAE values of 0.428% and 0.333%, respectively, for the four studied intervals. The forecasts using CAF model with the identified configuration for the dataset from Hill of Towie wind farm ($N_{level}=28$ and $N_{components}=31$) provides the average NRMSE and NMAE of 0502% and 0.381%, respectively.

Future work will focus on applying Deep Reinforcement Learning (DRL) for probabilistic forecasting of wind power generation with deep risk assessments and prediction drift analysis.

REFERENCES

[1] B. Kumar, N. Yadav, and Sunil, "A novel hybrid algorithm based on empirical Fourier decomposition and deep learning for wind speed forecasting," *Energy Convers. Manage.*, vol. 300, Jan. 2024, Art. no. 117891, doi: 10.1016/j.enconman.2023.117891.

[2] K. Wang, X.-Y. Tang, and S. Zhao, "Robust multi-step wind speed forecasting based on a graph-based data reconstruction deep learning method," *Expert Syst. Appl.*, vol. 238, Mar. 2024, Art. no. 121886, doi: 10.1016/j.eswa.2023.121886.

[3] B. Zhao, X. He, S. Ran, Y. Zhang, and C. Cheng, "Spatial correlation learning based on graph neural network for medium-term wind power forecasting," *Energy*, vol. 296, Jun. 2024, Art. no. 131164, doi: 10.1016/j.energy.2024.131164.

[4] Y. Li, H. Wang, J. Yan, C. Ge, S. Han, and Y. Liu, "Ultra-short-term wind power forecasting based on the strategy of 'dynamic matching and online modeling,'" *IEEE Trans. Sustain. Energy*, vol. 16, no. 1, pp. 107–123, Jan. 2025, doi: 10.1109/TSTE.2024.3424932.

[5] Z.-H. Liu, C.-T. Wang, H.-L. Wei, B. Zeng, M. Li, and X.-P. Song, "A wavelet-LSTM model for short-term wind power forecasting using wind farm SCADA data," *Expert Syst. Appl.*, vol. 247, Aug. 2024, Art. no. 123237, doi: 10.1016/j.eswa.2024.123237.

[6] W. Bai, M. Jin, W. Li, J. Zhao, B. Feng, T. Xie, S. Li, and H. Li, "Multi-step prediction of wind power based on hybrid model with improved variational mode decomposition and sequence-to-sequence network," *Processes*, vol. 12, no. 1, p. 191, Jan. 2024, doi: 10.3390/pr12010191.

[7] Y. Sun, Q. Zhou, L. Sun, L. Sun, J. Kang, and H. Li, "CNN–LSTM–AM: A power prediction model for offshore wind turbines," *Ocean Eng.*, vol. 301, Jun. 2024, Art. no. 117598, doi: 10.1016/j.oceaneng.2024.117598.

[8] R. Meka, A. Alaeddini, and K. Bhaganagar, "A robust deep learning framework for short-term wind power forecast of a full-scale wind farm using atmospheric variables," *Energy*, vol. 221, Apr. 2021, Art. no. 119759, doi: 10.1016/j.energy.2021.119759.

[9] M. A. Hossain, R. K. Chakrabortty, S. Elsawah, and M. J. Ryan, "Very short-term forecasting of wind power generation using hybrid deep learning model," *J. Cleaner Prod.*, vol. 296, May 2021, Art. no. 126564, doi: 10.1016/j.jclepro.2021.126564.

[10] S. Hanifi, A. Cammarono, and H. Zare-Behtash, "Advanced hyper-parameter optimization of deep learning models for wind power prediction," *Renew. Energy*, vol. 221, Feb. 2024, Art. no. 119700, doi: 10.1016/j.renene.2023.119700.

[11] M. Balci, E. Dokur, U. Yuzgec, and N. Erdogan, "Multiple decomposition-aided long short-term memory network for enhanced short-term wind power forecasting," *IET Renew. Power Gener.*, vol. 18, no. 3, pp. 331–347, Feb. 2024, doi: 10.1049/rpg2.12919.

[12] Z. Tarek, M. Y. Shams, A. M. Elshewey, E.-S. M. El-kenawy, A. Ibrahim, A. A. Abdelhamid, and M. A. El-dosuky, "Wind power prediction based on machine learning and deep learning models," *Comput., Mater. Continua*, vol. 74, no. 1, pp. 715–732, 2023, doi: 10.32604/cmc.2023.032533.

[13] D. He and R. Liu, "Ultra-short-term wind power prediction using ANN ensemble based on PCA," in *Proc. 7th Int. Power Electron. Motion Control Conf.*, vol. 3, Harbin, China, Jun. 2012, pp. 2108–2112, doi: 10.1109/IPEMC.2012.6259170.

[14] Z. Wang, Y. Ying, L. Kou, W. Ke, J. Wan, Z. Yu, H. Liu, and F. Zhang, "Ultra-short-term offshore wind power prediction based on PCA-SSA-VMD and BiLSTM," *Sensors*, vol. 24, no. 2, p. 444, Jan. 2024, doi: 10.3390/s24020444.

[15] D. Wang, X. Cui, and D. Niu, "Wind power forecasting based on LSTM improved by EMD-PCA-RF," *Sustainability*, vol. 14, no. 12, p. 7307, Jun. 2022, doi: 10.3390/su14127307.

[16] F. Rodríguez, S. Alonso-Pérez, I. Sánchez-Guardamino, and A. Galarza, "Ensemble forecaster based on the combination of time-frequency analysis and machine learning strategies for very short-term wind speed prediction," *Electric Power Syst. Res.*, vol. 214, Jan. 2023, Art. no. 108863, doi: 10.1016/j.epsr.2022.108863.

[17] K. Dhibi, M. Mansouri, M. Trabelsi, K. Abodayeh, K. Bouzrara, H. Nounou, and M. Nounou, "Enhanced PSO-based NN for failures detection in uncertain wind energy systems," *IEEE Access*, vol. 11, pp. 15763–15771, 2023, doi: 10.1109/ACCESS.2023.3244838.

[18] G. An, Z. Jiang, X. Cao, Y. Liang, Y. Zhao, Z. Li, W. Dong, and H. Sun, "Short-term wind power prediction based on particle swarm optimization-extreme learning machine model combined with AdaBoost algorithm," *IEEE Access*, vol. 9, pp. 94040–94052, 2021, doi: 10.1109/ACCESS.2021.3093646.

[19] B. Qu, Z. Xing, Y. Liu, and L. Chen, "Research on short-term output power forecast model of wind farm based on neural network combination algorithm," *Wind Energy*, vol. 25, no. 10, pp. 1710–1734, Oct. 2022, doi: 10.1002/we.2763.

[20] X. Zhou, C. Liu, Y. Luo, B. Wu, N. Dong, T. Xiao, and H. Zhu, "Wind power forecast based on variational mode decomposition and long short term memory attention network," *Energy Rep.*, vol. 8, pp. 922–931, Nov. 2022, doi: 10.1016/j.egyr.2022.08.159.

[21] R. Yu, J. Gao, M. Yu, W. Lu, T. Xu, M. Zhao, J. Zhang, R. Zhang, and Z. Zhang, "LSTM-EFG for wind power forecasting based on sequential correlation features," *Future Gener. Comput. Syst.*, vol. 93, pp. 33–42, Apr. 2019, doi: 10.1016/j.future.2018.09.054.

[22] H. H. H. Aly, "A novel deep learning intelligent clustered hybrid models for wind speed and power forecasting," *Energy*, vol. 213, Dec. 2020, Art. no. 118773, doi: 10.1016/j.energy.2020.118773.

[23] A. Sonawat, S.-J. Kim, H.-M. Yang, Y.-S. Choi, K.-M. Kim, Y.-K. Lee, and J.-H. Kim, "Positive displacement turbine—A novel solution to the pressure differential control valve failure problem and energy utilization," *Energy*, vol. 190, Jan. 2020, Art. no. 116400, doi: 10.1016/j.energy.2019.116400.

[24] S. Wang, "Differences between energy consumption and regional economic growth under the energy environment," *Energy Rep.*, vol. 8, pp. 10017–10024, Nov. 2022, doi: 10.1016/j.egyr.2022.07.065.

[25] K. EL Kinani, S. Meunier, L. Vido, and S. L. Ballois, "Interdisciplinary analysis of wind energy—A focus on France," *Sustain. Energy Technol. Assessments*, vol. 55, Feb. 2023, Art. no. 102944, doi: 10.1016/j.seta.2022.102944.

[26] R. Mandzhieva and R. Subhankulova, "Data-driven applications for wind energy analysis and prediction: The case of 'La haute Borne' wind farm," *Digit. Chem. Eng.*, vol. 4, Sep. 2022, Art. no. 100048, doi: 10.1016/j.dche.2022.100048.

[27] A. Clerc and E. Lingkan, "Hill of Towie wind farm open dataset," V1.0.0, Renew. Energy Syst., U.K., Mar. 2025. [Online]. Available: https://doi.org/10.5281/zenodo.14870023

[28] H. Liu and C. Chen, "Data processing strategies in wind energy forecasting models and applications: A comprehensive review," *Appl. Energy*, vol. 249, pp. 392–408, Sep. 2019, doi: 10.1016/j.apenergy.2019.04.188.

[29] K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 531–544, Feb. 2014, doi: 10.1109/TSP.2013.2288675.

[30] C. Zhang, Z. Tao, J. Xiong, S. Qian, Y. Fu, J. Ji, M. S. Nazir, and T. Peng, "Research and application of a novel weight-based evolutionary ensemble model using principal component analysis for wind power prediction," *Renew. Energy*, vol. 232, Oct. 2024, Art. no. 121085, doi: 10.1016/j.renene.2024.121085.

[31] M. Krichen, "Convolutional neural networks: A survey," *Computers*, vol. 12, no. 8, p. 151, Jul. 2023, doi: 10.3390/computers12080151.

[32] R. Albano, L. Giusti, E. Maiorana, and P. Campisi, "Explainable vision transformers for vein biometric recognition," *IEEE Access*, vol. 12, pp. 60436–60446, 2024, doi: 10.1109/ACCESS.2024.3393558.

[33] M. M. Taye, "Understanding of machine learning with deep learning: Architectures, workflow, applications and future directions," *Computers*, vol. 12, no. 5, p. 91, Apr. 2023, doi: 10.3390/computers12050091.

[34] M. Hu, S. Zhang, W. Dong, F. Xu, and H. Liu, "Adaptive denoising algorithm using peak statistics-based thresholding and novel adaptive complementary ensemble empirical mode decomposition," *Inf. Sci.*, vol. 563, pp. 269–289, Jul. 2021, doi: 10.1016/j.ins.2021.02.040.

[35] Y. Liu, Y. Wang, Q. Wang, K. Zhang, W. Qiang, and Q. H. Wen, "Recent advances in data-driven prediction for wind power," *Frontiers Energy Res.*, vol. 11, Jun. 2023, Art. no. 1204343, doi: 10.3389/fenrg.2023.1204343.

[36] B. K. Pradhan, B. C. Neelappu, J. Sivaraman, D. Kim, and K. Pal, "A review on the applications of time-frequency methods in ECG analysis," *J. Healthcare Eng.*, vol. 2023, no. 1, Jan. 2023, Art. no. 3145483, doi: 10.1155/2023/3145483.

[37] T. Guo, T. Zhang, E. Lim, M. Lopez-Benitez, F. Ma, and L. Yu, "A review of wavelet analysis and its applications: Challenges and opportunities," *IEEE Access*, vol. 10, pp. 58869–58903, 2022, doi: 10.1109/ACCESS.2022.3179517.

[38] S. Zhang, E. Robinson, and M. Basu, "Wind power forecasting based on a novel gated recurrent neural network model," *Wind Energy Eng. Res.*, vol. 1, Aug. 2024, Art. no. 100004, doi: 10.1016/j.weer.2024.100004.

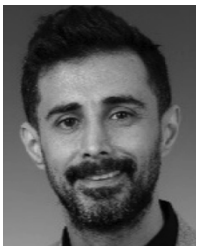
[39] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2022, pp. 5998–6008.

[40] Y. Sun, J. Yang, X. Zhang, K. Hou, J. Hu, and G. Yao, "An ultra-short-term wind power forecasting model based on EMD-EncoderForest-TCN," *IEEE Access*, vol. 12, pp. 60058–60069, 2024, doi: 10.1109/ACCESS.2024.3373798.

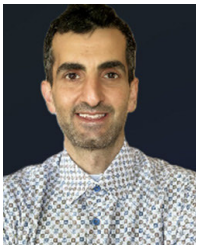
[41] D. Song, X. Tan, Q. Huang, L. Wang, M. Dong, J. Yang, and S. Evgeny, "Review of AI-based wind prediction within recent three years: 2021–2023," *Energies*, vol. 17, no. 6, p. 1270, Mar. 2024, doi: 10.3390/en17061270.

MOHAMMAD RASHIDI received the M.Sc. degree in power systems engineering from Babol Noshirvani University of Technology, Babol, Iran, in 2024. Since 2024, he has been with Regional Electricity Company. He is also a Researcher with the Faculty of Electrical and Computer Engineering, Babol Noshirvani University of Technology. His research interests include power systems, the application of AI in power systems, deep learning, deep reinforcement learning, power system resilience, renewable energy, energy management, and electric vehicles.

NARGES TARA received the B.S. degree in electrical power engineering from Lorestan University, Khorramabad, Iran, in 2022, and the M.Sc. degree in electrical power engineering from Babol Noshirvani University of Technology, Babol, Iran, in 2025. Her research interests include artificial intelligence, power systems, deep learning, wind power forecasting, and renewable energy.



MAJID MEHRASA (Senior Member, IEEE) received the Ph.D. degree from the University of Beira Interior (UBI), Covilha, Portugal. During last 16 years, he persistently worked on different projects in power electronics, power electronic applications, and power systems in various countries, including Iran, Portugal, Italy, France, and the USA. As a Postdoctoral Research Fellow, he was involved with the industry-based projects in different universities, such as the University of Trieste, Italy; G2Elab, University of Grenoble Alpes, France; EERC, University of North Dakota, USA; San Diego State University, USA; and California State University, Sacramento, USA. He is currently an Assistant Professor with The University of New Orleans (UNO), New Orleans, USA. He has authored more than 120 articles, including scholarly book chapters, refereed-journal articles, and conference papers. His research interests include power electronics (PEs) with its applications and related control systems, developing artificial intelligent (AI) for the cybersecurity and stability enhancement of PEs-based low-inertia power systems, and electric vehicles.



SHAMSODIN TAHERI (Senior Member, IEEE) received the B.Sc. degree from the University of Mazandaran, Babolsar, Iran, in 2006, the master's degree from Iran University of Science and Technology (IUST), Tehran, Iran, in 2009, and the Ph.D. degree from the University of Quebec in Chicoutimi (UQAC), Chicoutimi, Canada, in 2013, all in electrical engineering. From 2013 to 2014, he was with the Technical Services and Research Department, SaskPower, Saskatchewan, Canada. Since 2014, he has been a Professor with the University of Quebec in Outaouais, Gatineau, Canada. He is the founder of LAR3E, a CFI-funded laboratory on energy efficiency in power networks. His research focuses on integrating renewable energy sources into power grids under cold climate conditions, with expertise in photovoltaic systems.

HANI VAHEDI (Senior Member, IEEE) received the Ph.D. degree (Hons.) from the École de Technologie Supérieure (ÉTS), Université du Québec, Montreal, in 2016.

After seven years in the industry, as a Power Electronics Designer and later as a Chief Scientific Officer, he joined Delft University of Technology, as an Assistant Professor to advance electrification for the clean energy transition. He leads TU Delft's 24/7 Energy Hub at The Green Village, an operational microgrid that integrates renewables, green hydrogen production, and hybrid energy storage systems. He has authored over 100 IEEE conferences and journals, a Springer Nature book, and a book chapter published by Elsevier. He is an inventor of the five-level Packed U-Cell (PUC5) converter, holds several U.S. and international patents, and transferred this technology to industry, assisting in the development of a bidirectional EV charger based on his invention. His research interests include multilevel power converter topologies, modulation and control, and applications in smart grids; renewable energy conversion; EV charging; green hydrogen production through electrolysis; and fuel-cell systems.

Dr. Vahedi has been an active member of the IEEE Industrial Electronics Society (IES), since 2012. He received the ÉTS's Best Ph.D. Thesis Award (2016–2017). He has serving on the local organizing committee for IECON12 in Montreal. He is acting as a students and young professionals chair for several IES-sponsored conferences from 2016 to 2024. He has serving as the Technical Program Chair (GreenTech Cluster) for IECON25, Madrid; and the Special Sessions Chair for IECON26, Doha. He currently manages the IES Chapters and Joint Chapters Program. He is an Associate Editor of IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS, IEEE OPEN JOURNAL OF THE INDUSTRIAL ELECTRONICS SOCIETY, and IEEE OPEN JOURNAL OF POWER ELECTRONICS.