EV fleet charging load forecasting based on multiple decomposition with CEEMDAN and swarm decomposition.

DOKUR, E., ERDOGAN, N. and KUCUKSARI, S.

2022

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 21, 2022, accepted June 7, 2022, date of publication June 13, 2022, date of current version June 16, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3182499

EV Fleet Charging Load Forecasting Based on Multiple Decomposition With CEEMDAN and Swarm Decomposition

EMRAH DOKUR[1,2], NUH ERDOGAN[3], (Member, IEEE), AND SADIK KUCUKSARI[4], (Senior Member, IEEE)

[1]MaREI-SFI Research Centre for Energy, Climate and Marine, University of College Cork, Cork, P43 C573 Ireland
[2]Faculty of Engineering, Bilecik Seyh Edebali University, 11230 Bilecik, Turkey
[3]School of Engineering, Robert Gordon University, Aberdeen AB10 7GJ, U.K.
[4]Department of Applied Engineering and Technical Management, University of Northern Iowa, Cedar Falls, IA 50614, USA

Corresponding author: Emrah Dokur (edokur@ucc.ie)

ABSTRACT As the transition to electric mobility is accelerating, EV fleet charging loads are expected to become increasingly significant for power systems. Hence, EV fleet load forecasting is vital to maintaining the reliability and safe operation of the power system. This paper presents a new multiple decomposition based hybrid forecasting model for EV fleet charging. The proposed approach incorporates the Swarm Decomposition (SWD) into the Complete Ensemble Empirical Mode Decomposition Adaptive Noise (CEEMDAN) method. The multiple decomposition approach offers more stable, stationary, and regular features of the original signals. Each decomposed signal is fed into artificial intelligence based forecasting models including multi-layer perceptron (MLP), long short-term memory (LSTM) and bidirectional LSTM (Bi-LSTM). Real EV fleet charging data sets from the field are used to validate the performance of the models. Various statistical metrics are used to quantify the prediction performance of the proposed model through a comparative analysis of the implemented models. It is demonstrated that the multiple decomposition approach improved the model performance with an R^2 value increasing from 0.8564 to 0.9766 as compared to the models with single decomposition.

INDEX TERMS CEEMDAN, electric vehicle, fleet charging, forecasting, signal decomposition, swarm decomposition.

I. INTRODUCTION

The demand for zero-emission vehicles from the private sector has further accelerated the proliferation of electric vehicles (EVs). Transitioning to electrified fleets has become a global trend. As such, companies on a global and local scale have started the transition to electric mobility by shifting their fleets to EVs. Considering the commitments of only global companies, approximately 5 million vehicles are expected to switch to EVs by 2030 [1]. Accordingly, the number of charging stations has seen rapid growth, reaching almost 10 million private charges in 2020, 25% of which are at workplaces [2]. This equates to more than 15 GW of installed charging capacity at workplaces. While EV fleet charging loads currently account for a small share of global electricity demand, this figure is expected to grow to 670 GW in 2030 [1]. Hence, EV fleet charging loads are estimated to become increasingly significant for power systems, potentially driving increases in peak power generation and transmission capacity. Therefore, EV fleet load forecasting is critical to ensure the smooth operation and security of power systems.

EV forecasting studies can be categorized into two groups: mobility or charging pattern identification and charging load forecasting [3]. Based on probability density functions such as the kernel [4] or Gaussian distributions [5], most of the studies have focused on modeling EV transportation mobility or charging behavior, which are inputs to smart charging algorithms [6]. As such, EV charging loads are better managed from both the grid perspective and EV user convenience. The second group, on the other hand, applies advanced forecasting methods based on artificial intelligence (AI) in order to precisely predict charging loads [3]. Thanks to their adept

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

functionality, AI models are more preferable than probabilistic models. Furthermore, they have the advantage of better representing complex nonlinear problems as in EV charge demand load. The stochastic nature of EV charging demand forces to use of advanced AI models. In [7], the Nearest Neighbor and Modified Pattern Sequence Forecasting methods are applied to predict EV charging powers on an hourly basis. However, the time-series forecasting problem suffer from cross validation. A time-weighted dot product algorithm was proposed to improve accuracy and processing time in the fast demand prediction of the EV charging time series data in [8]. EV charging time series are non-stationary and have unstable characteristics that increase the challenge of forecasting. Hence, statistical time-series forecasting algorithms are combined and their weighted predictions are used to determine EV charging loads in [9]. Convolutional neural networks, long short-term memory (LSTM), and bidirectional LSTM (Bi-LSTM) models based on deep learning were used to forecast EV charging loads [10], [11]. Similarly, Zhu *et al.* [12] obtained better EV load forecasting results with LSTM. While the implemented LSTM model provides adequate forecasting of fast charging demand, it is not validated using conventional AI methods such as Multi-Layer Perceptron (MLP). Based on reinforcement learning, a new forecasting model for EV charging stations is proposed to improve the performance of AI models in [13]. In [14], the parameters of AI based forecasting models were determined by using heuristic optimization algorithms such as genetic algorithm, particle swarm optimization, and ant colony algorithms. However, the optimization based AI models cannot guarantee convergence to the best fit, since they might fall into the local minimum, which in turn leads to premature convergence. Thus, the optimisation-based AI models lead to lower forecasting accuracy. The performance of forecasting model is also affected by the variation and magnitude of the EV charging demand profile. In [3], it was emphasized that forecasting a large-scale EV demand is more difficult compared to small-scale EV demand forecasting. In this regard, short term load forecasting can provide more effective results than those of medium or long term load forecasting. To the best of our knowledge, short term forecasting of EV fleet charging behavior has yet investigated in literature. While the AI based models have been widely used in forecasting problems, there are certain limitations arising due to the stochastic nature of EV charging data. A universal model that represents all charging scenarios of EV load demand has yet to be investigated.

The decomposition of the original series plays a significant role in improving forecasting performance [15]. However, because single decomposition approaches are influenced by many stochastic elements, the high components of the decomposed signal, particularly the first sub-signal, must be thoroughly explored [16]. Therefore, to improve forecasting accuracy, hybrid AI-based models have been the main research focus. While hybrid models use two or more forecasting methods, one approach is to combine a decomposition technique with an AI model [17]. The use of decomposition methods improves the forecasting performance by providing stationary and regular sub-series. In this respect, the complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) has proven to be an effective, and reliable decomposition method that has been widely used in hybrid models. Zhang *et al.* [18] used the CEEMDAN to decompose wind speed data and performed five neural networks with meta-heuristic algorithms for wind speed forecasting. The hybrid approach based on CEEMDAN was shown to outperform single AI models. However, separate weights for each CEEMDAN subcomponent need to be determined. Moreover, non-stationarity of random and irregular data series may not be thoroughly detected by the single decomposition technique. The features of the sub-series have a significant impact on forecasting performance. Therefore, the highest frequency sub-component, in particular, has the most impact on the forecasting performance. To overcome this problem, a secondary decomposition technique was proposed that is a combination of two decomposition approaches [19]. In [20], the two-phase decomposition, namely the complementary ensemble empirical mode decomposition (CEEMD), and variational mode decomposition, were applied for air quality index forecasting. In this approach, the high-frequency components are decomposed into more stationary and regular features using multi-decomposition methods. However, both the proposed methods yield poor noise resistance. In this regard, swarm decomposition (SWD) algorithms based on swarm-prey hunting approach can intelligently decompose the high-frequency components. While the SWD has been used in several recent forecasting studies, including solar [21] and offshore wind [22], the use of SWD as a multi-decomposition tool is worth investigating in order to have more stable sub-series components of the signal. As such, the model performance can be improved.

The main purpose of this study is to develop a charging forecasting tool for EV fleet owners to be used for optimal management of the fleet charging demand with efficient use of the grid assets. Hence, a new multiple decomposition based hybrid model is proposed to forecast EV fleet charging loads. To overcome the impact of the highest signal components on the forecasting performance, a new two stage decomposition technique is introduced to further decompose components. The proposed approach takes advantage of multiple decomposition techniques with CEEMDAN and SWD to produce stationary and regular subseries from the highest frequency signal component. As such, the forecasting performance is improved. In terms of forecasting models, this study implements deep-learning based and conventional forecasting models to evaluate the behaviour of the proposed multiple decomposition approach to their performance. To forecast all subseries components of EV charging demand, Bi-LSTM, LSTM and MLP are implemented. Real EV fleet charging data set from Leeds Council are used to test the performance of the model. Moreover, the model is validated on a public EV charging data with different characteristics. Various statistical

performance metrics commonly used in literature are used to measure the accuracy through a comparative analysis of the implemented models. Finally, the Diebold–Mariano (DM) test results show the superiority of the proposed approach over the implemented models. The remainder of this paper is organized as follows: The proposed approach, along with the decomposition and AI-based forecasting models, is presented in Section 2. Experimental results of EV fleet data with forecasting results are presented in Section 3 along with a detailed performance evaluation. Finally, Section 4 provides concluding remarks.

II. METHODOLOGY

A. COMPLETE ENSEMBLE EMPIRICAL MODE DECOMPOSITION ADAPTIVE NOISE METHOD

The complete ensemble empirical mode decomposition adaptive noise (CEEMDAN) algorithm is an improved version of the ensemble empirical mode decomposition (EEMD) method that decomposes nonlinear and nonstationary time series into multiple stationary components [23]. After applying the CEEMDAN process to the original signal, a residual signal (R_n) and the Intrinsic Mode Functions (IMF) are obtained. Due to adding normally distributed white noise to the original signal using EEMD, the mode mixing effect is reduced. Adding white noise to the EEMD is a disadvantage because it increases the processing time. Furthermore, data loss may occur during the reconstruction process [17]. In order to overcome these problems, the CEEMDAN method was proposed [23].

Where a specific original time series represents $x(t)$, the calculation steps of the CEEMDAN algorithm are given as follows:

Step 1: At time t, white noise of ω^i is added to the original signal, which can be described as:

$$x^i(t) = x(t) + \varepsilon_0 \omega^i(t), \quad i = 1, 2, \ldots, N, \tag{1}$$

where ε_0 is a noise coefficient, and N is the number of realization.

Step 2: The first IMF, IMF_1, is obtained by averaging the components of the EMD [24], as follows:

$$\overline{IMF_1}(t) = \frac{1}{N}\sum_{i=1}^{N} IMF_1^i(t), \tag{2}$$

Then, the residual process is formulated as:

$$r_1(t) = x(t) - \overline{IMF_1}(t). \tag{3}$$

Step 3: Further decomposition for $r_1(t) + \varepsilon_1 EMD_1(\omega^i(t))$ can be performed by using the EMD to calculate the second IMF, and the residual signal can be expressed as follows:

$$\overline{IMF_2}(t) = \frac{1}{N}\sum_{i=1}^{N} EMD_1(r_1(t) + \varepsilon_1 EMD_1(\omega^i(t))), \tag{4}$$

$$r_2(t) = r_1(t) - \overline{IMF_2}(t). \tag{5}$$

Step 4: From steps 2 and 3, the m^{th} residual and $(m+1)^{th}$ IMF component are calculated as follows:

$$r_m(t) = r_{m+1}(t) - \overline{IMF_m}(t), m = 2, \ldots, M, \tag{6}$$

$$\overline{IMF}_{m+1}(t) = \frac{1}{N}\sum_{i=1}^{N} EMD_1(r_m(t) + \varepsilon_m EMD_m(\omega^i(t))), \tag{7}$$

where $\overline{IMF}_{m+1}$ and $EMD_m(.)$ represent the m+1-th IMF mode obtained by the CEEMDAN algorithm and m-th IMF mode is calculated by EMD.

Step 5: Step 4 is repeated until the IMF component and residual reach an insignificant threshold that cannot be decomposed by EMD. Finally, the final decomposed signal $X(t)$ can be calculated as follows:

$$X(t) = \sum_{m=1}^{M} \overline{IMF}_m(t) + R(t), \tag{8}$$

where $R(t)$ represents the final residue. It presents trends in time series.

B. SWARM DECOMPOSITION METHOD

To resolve the non-stationary and multi-component signals, an intelligent swarm decomposition (SWD) algorithm was proposed in [25]. It decomposes the original signal into some oscillating components (OCs) using swarm filtering (SWF) fundamental. This principle has led to the swarm-prey hunting approach. The location information of the preys, the i^{th} member of the swarm at n^{th} step is presented by P_{prey}. The swarm-prey principle has the driving force, $F_{dr}(n, i)$ and the cohesion force, $F^n_{coh,i}$ as follows:

$$F_{dr}(n, i) = P_{prey}(n) - P_i(n-1) \tag{9}$$

$$F^n_{Coh,i} = \frac{1}{M-1}\cdot\sum_{j=1, j\neq i}^{M} f(P_i[n-1] - P_j(n-1)), \tag{10}$$

where, $f(.)$ is the function, which changes with the distance between members, d. The distribution of members in the population is controlled by the critical distance parameter, d_{cr}. The $f(.)$ is given as follows:

$$f(d) = -sgn(d).\ln\left(\frac{|d|}{d_{cr}}\right), \tag{11}$$

where $sgn(.)$ and $ln(.)$ are the sign and logarithmic functions. The swarm updates location and velocity at each iteration because of the tracking of its prey as follows:

$$V_i[n] = V_i[n-1] + \delta.\left(F^n_{Dr,i} + F^n_{Coh,i}\right), \tag{12}$$

$$P_i[n] = P_i[n-1] + \delta.(V_i[n]), \tag{13}$$

where $y[n]$ is the output of the SWF having parameters δ and M in (14). Herein, the δ controls the flexibility of the swarm. The number of swarm is represented by M.

$$y[n] = \beta.\sum_{i=1}^{M} P_i[n], \tag{14}$$

Herein, the weighted factor, β, is preferred to be a small value, i.e., 0.005 [26]. In order to select the optimal values

of vital parameters of SWD, δ and M, the following criterion is followed in (15).The main purpose of the SWF process is to search for the parameters. The relationship of these parameters with each frequency component is given by (16) and (17) [25].

$$\arg_{\delta,M} min \sum_{k} \left\{ \left| Y_{\delta,M}[k] - |S[k]| \right| \right\}^2, \tag{15}$$

$$M(\hat{\omega}) = [33.46\hat{\omega}^{-0.735} - 29.1], \tag{16}$$

$$\delta(\hat{\omega}) = -1.5\hat{\omega}^2 + 3.454\hat{\omega} - 0.01, \tag{17}$$

where $\hat{\omega}$ represents the normalized frequency. The process is performed iteratively to obtain the dominant OCs of the residue. If the residual signal does not consist of any oscillatory mode, the algorithm can be terminated. The detailed information about the SWD can be obtained from [25]. As a result, the SWD method presents subcomponents and residual signals of the input signal.

C. AI BASED FORECASTING METHODS

Many AI-based models such as MLP, LSTM and Bi-LSTM can be used in forecasting studies. Belonging to one of the important neural networks, the Multi-Layer Perceptron (MLP) is a type of supervised learning, and it works on the principle of the feed-forward network. MLP performs the relationship between input and output data using nonlinear activation functions [27]. The output of the MLP is described as:

$$y_p = \varphi_o \left\{ \sum_{j=0}^{N} \omega_{jp}^{o} \left[\varphi_H \left(\sum_{i=0}^{M} \omega_{ij}^{H} x_i \right) \right] \right\}, \tag{18}$$

where ω_{ij}^{H} is input and hidden layer connecting weights and ω_{jp}^{o} is the hidden and output layer connecting weights. The activation functions in the hidden layer and output layer are φ_H and φ_o, respectively. The activation functions have many different forms such as tanh function, linear function, etc. [28]. The learning process of MLP aims to minimize the output error.

Hochreiter and Schmidhuber in [29] proposed the LSTM algorithm based on deep learning. The LSTM has the advantage of solving the gradient disappearance and gradient explosion. A LSTM consists of an input gate (i_t), an output gate (o_t), a forget gate (f_t), and storage elements that transfer information from the previous output to the current output (Fig. 1). The relevant information is processed and stored with the help of these gates. The memory block is the main unit of the hidden layer, as opposed to the other neural networks. [30]. The formulations of the parametric equations of the LSTM gates are given as follows:

$$Inputgate: i_t = \sigma\left(W_i x_t + U_i h_{t-1} + b_i\right) \tag{19}$$

$$Forgetgate: f_t = \sigma\left(W_f x_t + U_f h_{t-1} + b_f\right) \tag{20}$$

$$Outputgate: o_t = \sigma\left(W_o x_t + U_o h_{t-1} + b_o\right) \tag{21}$$

$$Outputs \begin{cases} c_t = f_t \odot c_{t-1} \\ h_t = o_t \odot tanh(c_t), \end{cases} \tag{22}$$

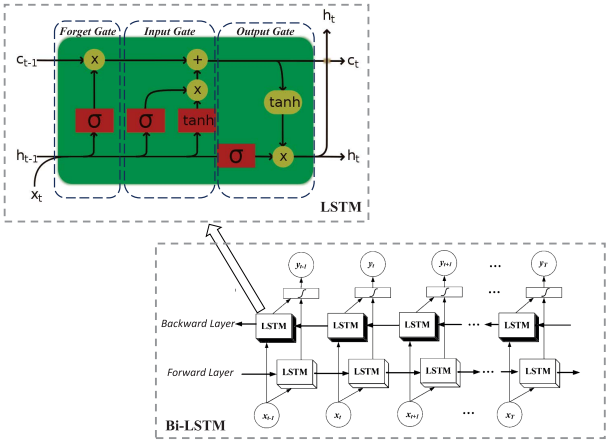

FIGURE 1. Architecture of a single LSTM cell and Bi-LSTM.

where W_i, W_f and W_o are the reccurent weights of the gates. σ indicates sigmoid activation function. The biases are described as b_i, b_f and b_o for each gates. U_i, U_f and U_o provide input weights of three gates. An element-wise multiplication of two vectors is denoted by $\odot$. Bi-LSTM is a modified version of LSTM and the forward and backward computation LSTM networks are trained together in this process (Fig. 1). Compared with the unidirectional LSTM neural network, it has been reported that the BiLSTM neural network has superior performance for forecasting problems [31].

D. PROCEDURES OF PROPOSED HYBRID MODEL

In this section, the proposed hybrid AI model based on multiple decomposition using CEEMDAN and SWD is introduced in detail. Fig. 2 presents the framework of the proposed multiple decomposition based hybrid approach, and the detailed procedures are described as follows:

Step 1. At the first step, the CEEMDAN algorithm is performed to decompose the EV fleet charging load time series data. The implementation process of the CEEMDAN algorithm is presented in Section II-A.

Step 2. The CEEMDAN is employed to decompose the original time series into a set of IMFs with different frequencies. IMF1 is the most unsystematic and chaotic component of the IMFs, with a high frequency [32]. The high-frequency components are more random (non-stationary) and more difficult to forecast as compared with to the low-frequency components. To reduce this effect, IMF1 was re-decomposed by using SWD. The non-stationary decomposed signals are re-decomposed using the SWD algorithm in this step. The SWD algorithm process is given in Section II-B.

Step 3. Each sub-decomposed signal is independently used as an input to Bi-LSTM, MLP and all the benchmark models (MLP, LSTM, Bi-LSTM, CEEMDAN-MLP, CEEMDAN-Bi-LSTM). CEEMDAN-SWD-MLP and CEEMDAN-SWD-Bi-LSTM are then compared. In this case, 1584 data points for EV dataset 1 are configured in such a way that the first 1103 data points are used as a training phase, and the

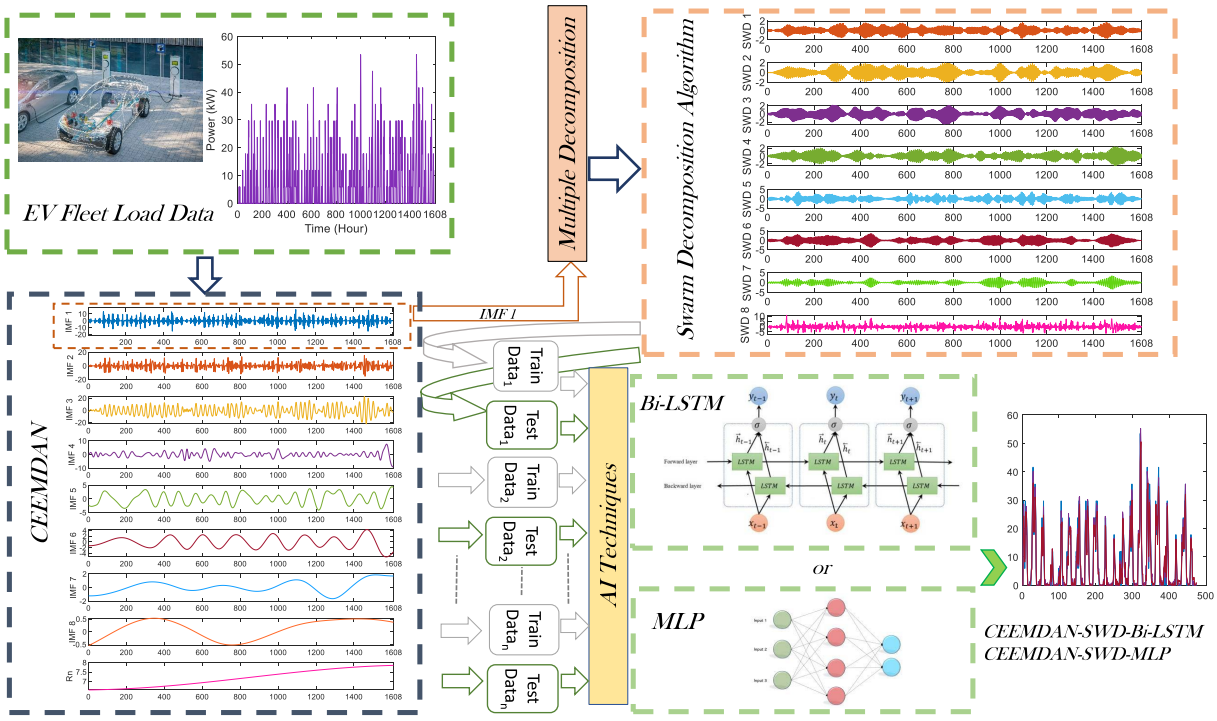

FIGURE 2. The structure of the multiple decomposition forecasting models with CEEMDAN and SWD.

remaining data points are used as a test phase. A similar data selection procedure is applied to the EV dataset 2. It is divided into a training stage of the first 6139 values and a testing stage of the last 2621 values.

Step 4. Finally, aggregation of test and training results of decomposed signals yield the hybrid model results. All final results are compared using performance metrics such as the root mean square error (RMSE), mean square error (MSE), mean absolute error (MAE), and R^2.

III. EXPERIMENTAL RESULTS AND DISCUSSIONS

A. DESCRIPTION OF EV FLEET DATA SET

The EV fleet data set (EV data set 1) used in this study is the Leeds Council EV fleet charging data set provided in [33]. The 724 charging events recorded through the 7.36 kW L2/Mode-3 type charging stations from July 25, 2020 to September 29, 2020 are included in the data set. Each recorded charging event includes information on charge start time, end time, total charging energy, and plug-in duration for each EV, while the time series of charging power and EV information being charged are not provided. It is assumed that all charge events occur on a single day and that there is no following day departure. From the above information, time series charging data for each event with a 1-minute interval is generated by assuming that vehicles are charged with constant power to meet their overall energy demands between the plug-in and plug-off times. The constant charging rate is calculated from the total charging energy and the plug-in duration of the EVs. The final data set includes the sum of the time series of all the events. Since residential, public, and workplace EV charging behaviors are different from each other, the performance of the proposed method is also tested with the public charging data set (EV data set 2) available in [34]. The data set includes 7891 charging events collected from various public stations in 2019. Similar to EV data set 1, the data set provides the charge start, finish, total charging energy provided, and maximum power supplied information. Time series data for all charging events is generated in a similar way to the time series data generation of the EV data set 1. Since the data set includes events for the entire year, a 1-hour time interval is used for the time series data generation.

Fig. 3 presents the EV fleet charging data and the basic architecture of AI-based time series forecasting with a learning step. The original time series data does not have missing data. Also, the noise reduction or smoothing process was not applied to the original data in order not to change the characteristics of the data. The univariate analysis based on only historical EV fleet charging data was investigated. In other words, a fixed number of past values are the configurable inputs of the AI methods. The forecasting of the future data of the time series presents the outputs of the model. This learning phase uses a moving time window known as the sliding window technique. In order to create the best network structure, the width of the input dataset is one of the important factors. Hence, it is examined for measures of association between current and past series values. The autocorrelation

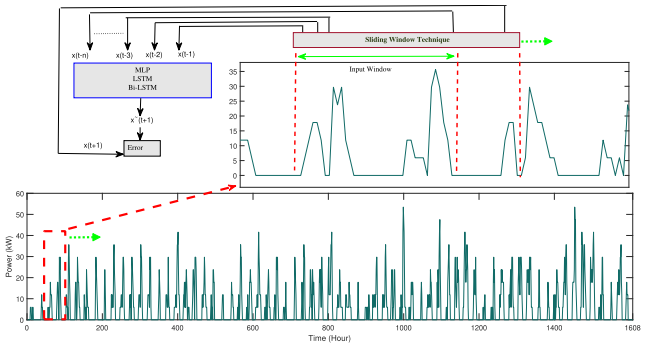

FIGURE 3. Original EV fleet charging data and AI-based time series forecasting procedure.

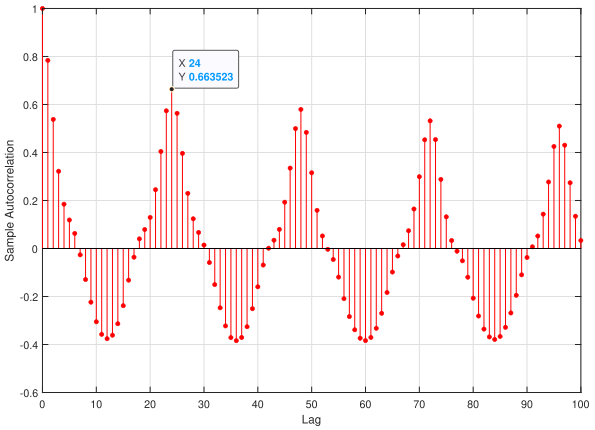

FIGURE 4. Time-series characteristics with autocorrelation function of EV fleet charging load.

of the series (ACF) to determine the window width is shown in Fig.4. As can be seen, the daily pattern is apparent in the ACF, peaking at 24-hr intervals. The sliding window width is determined based on the ACF result.

B. ANALYSIS OF DECOMPOSITION RESULTS

According to the multiple decomposition technique, the CEEMDAN approach is first performed to decompose the EV fleet's original time series. It is obtained that a total of eight IMFs ranged from high frequency to low frequency in Fig. 5. Here, the highest frequency component is represented as IMF1, which has the detailed information of the original series. The final decomposed signal gives the variation trend of the charging data. The AI models are performed to forecast all decomposed IMF signals.

The main challenge of the decomposition-based forecasting step is the prediction of IMF 1, due to its high frequency. In Fig. 6, sample forecasting results for IMF1 and IMF2 are shown. It can be clearly seen that the highest frequency component affects the prediction accuracy more. In order to develop the forecasting performance of this component, in this study, the multiple decomposition approach, or filtering procedure, is included in the estimation process. Since the filtering process might cause data loss, a multiple decomposition method was therefore included in this study. A similar process is performed for the public EV charging load (EV data set 2), and the decomposed signals are obtained. The comparative analysis results in 11 decomposed IMF components and a residual signal. Furthermore, IMF 1 redecomposed 6 separate decomposed signals using SWD for EV dataset 2, similarly.

In this proposed approach to signal decomposition, the SWD method is introduced to conduct the secondary decomposition of IMF1. The IMF1 decomposed eight decompositions in Fig. 7. The final forecast result of IMF1 is found by summing the estimation results of each SWD decomposed component. Thanks to the forecasting of each decomposed signal, which has all the characteristics of IMF1, the model performance is improved. To evaluate the accuracy of the proposed approach, the MLP, LSTM, and Bi-LSTM models were implemented with and without the multiple decomposition method in the forecasting step.

C. FORECASTING RESULTS AND PERFORMANCE EVALUATION

In this section, the forecasting results of all implemented models, namely CEEMDAN-SWD-MLP, the MLP, LSTM, CEEMDAN-MLP, Bi-LSTM,CEEMDAN-Bi-LSTM, CEEMDAN-SWD-Bi-LSTM are discussed in detail. Despite the fact that numerous forecasting techniques have been used in previous research, there is no dominant model in forecasting time series [35]. None of the forecasting techniques outperformed the others in terms of overall performance. As a result, to compare the power of the proposed model, it is compared with some state-of-the-art-models, including LSTM and Bi-LSTM. All experiments are implemented in Matlab R2020b on Windows 10 with a 2.5-GHz Intel Core i5 7200U processor and a 64-bit operating system with 8 GB of RAM. The learning rate, number of epochs, and hidden layers of MLP are chosen by 0.01, 1000, and 3 for all models, respectively. The number of hidden layers in the used LSTM and Bi-LSTM models is designed by 2. For the LSTM-based analysis, a dropout rate of 0.5 and a value of 100 neurons in each layer are chosen. As an optimizer, the Adam algorithm [36] is used. Due to the randomness of the model parameters, all MLP-based techniques were run fifty times to minimize errors. To compare the performance of the models, well-known performance metrics such as the root mean square error (RMSE), mean square error (MSE), mean absolute error (MAE), and R^2 metrics are used as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(y_i - \tilde{y}_i)^2}{N}} \tag{23}$$

$$MSE = \frac{\sum_{i=1}^{N}(y_i - \tilde{y}_i)^2}{N} \tag{24}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \tilde{y}_i| \tag{25}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \tilde{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y}_i)^2} \tag{26}$$

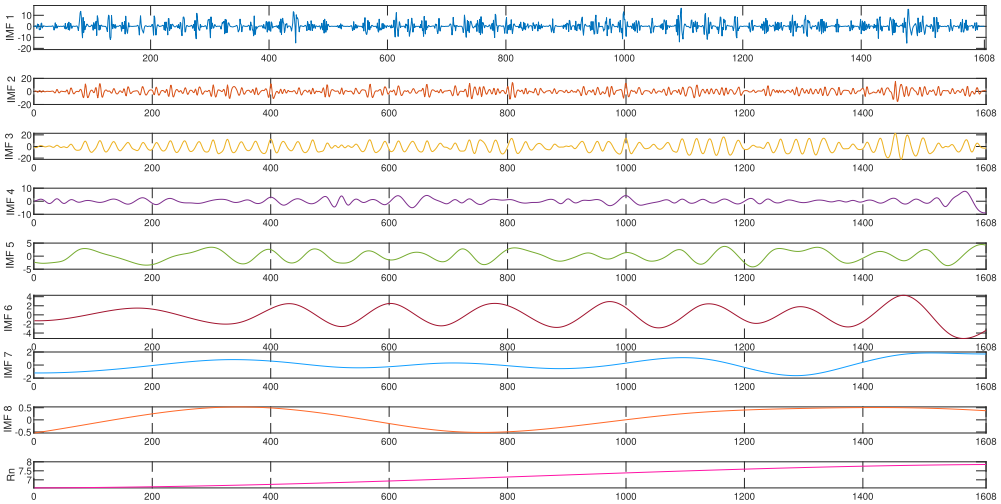

FIGURE 5. CEEMDAN processing outputs: IMFs and residuals of EV fleet charging data.

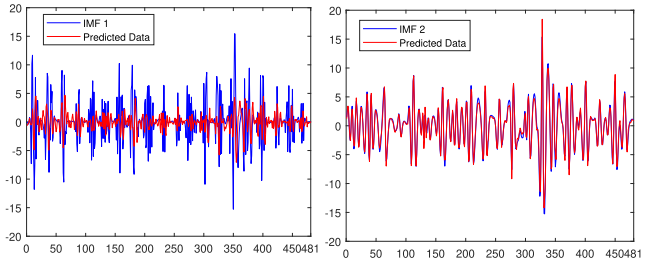

FIGURE 6. Comparison of test forecasting performances of IMF1 and IMF2.

In order to assess the impact of the proposed multiple decomposition technique on the forecasting performance, various AI based models with and without decomposition modules have been compared. The models implemented are standalone models; MLP, LSTM, Bi-LSTM, single decomposition based models; CEEMDAN-Bi-LSTM, CEEMDAN-MLP, and the multiple decomposition based models CEEMDAN-SWD-Bi-LSTM and CEEMDAN-SWD-MLP. Fig. 8 and Fig. 9 present the forecasting test results of the implemented models for EV dataset 1 and EV dataset 2, respectively. It is obtained that the proposed model results (in purple) give the best fit to the real data, while the standalone LSTM model results display the most discrepancies. Traditional metrics have been used for detailed performance comparison. Table 1 reports the comparative analysis results. As shown in the table, the proposed hybrid model achieves the best prediction performance, with the lowest RMSE, MSE, MAE and R^2 values of 1.6858, 2.8421, 1.1166 and 0.9766 for EV dataset 1. In terms of accuracy, the models can be sorted from highest to lowest as CEEMDAN-SWD-Bi-LSTM, CEEMDAN-MLP, CEEMDAN-Bi-LSTM, MLP,Bi-LSTM and LSTM with R^2 values of 0.9501, 0.9028, 0.8564, 0.7344, 0.7327 and 0.7155. It is observed that the LSTM model has poor performance ability to fit the change of signal, particularly when the EV charging characteristic changes rapidly. Although similar results are obtained in EV dataset 2, the only difference is that CEEMDAN-MLP gives better results than CEEMDAN-SWD-Bi-LSTM. While the change in data characteristics is initially anticipated to affect the multiple decomposition performance, the proposed CEEMDAN-SWD based MLP model still yields the best results for EV dataset 2.

In terms of forecasting error measures, Fig. 10 compares the performance of the implemented models. It is shown that the standalone models produce significantly the highest RMSE, MSE, MAE and the lowest R^2 error values. While the signal decomposition reduced the error values, the proposed model achieves the lowest ones, leading to the highest correlation coefficient. This proves that the signal decomposition method can better deal with the noise contained in the EV charging data in order to improve the forecasting accuracy. Fig. 11 presents residual errors in which the variation of errors for the implemented models is presented for each sample. The proposed model has been clearly found to have the lowest residue for each sample. Since the R^2 value is close to 1 in the proposed model, it can be found that there is a strong correlation between the actual and predicted value. Furthermore, Fig. 12 depicts the scatter plots between observed and forecasted values to show the degree of correlation. The more scattering points are around the diagonal line, the better the model's performance is. From Fig. 12, the proposed model CEEMDAN-SWD-MLP is said to be the best correlation. In this regard, it can be seen that the proposed forecasting model is found to be reliable and accurate. The findings confirm that the performance of the well-known MLP model can be improved by the multiple decomposition approach. The proposed model displays better performance for the EV fleet load charging data with higher accuracy rates in terms of RMSE, MSE, MAE and R^2 performance metrics.

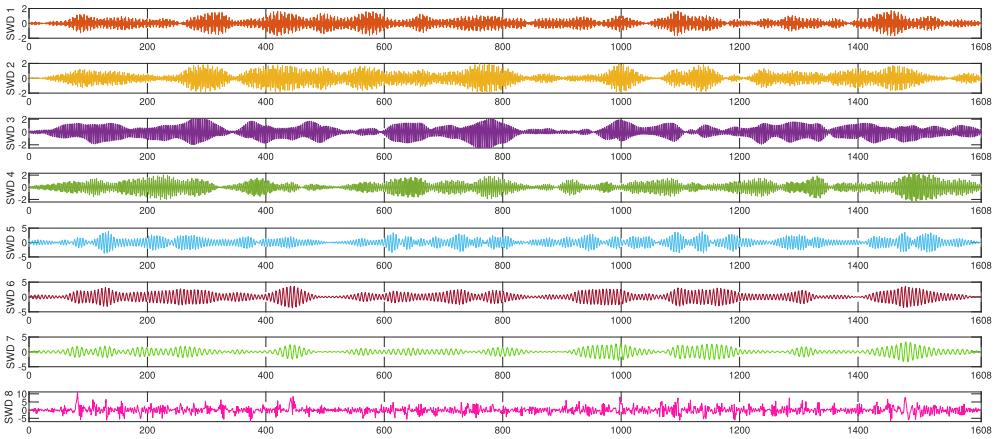

FIGURE 7. Secondary decomposition results of IMF1 by the SWD.

TABLE 1. Comparative Analysis of Forecasting Results for EV Datasets.

	Performance Metrics	MLP	CEEMDAN-MLP	LSTM	Bi-LSTM	CEEMDAN-BiLSTM	CEEMDAN-SWD-Bi-LSTM	Proposed Model
EV Dataset 1	**RMSE**	5.6890	3.4419	5.8880	5.7070	4.1836	2.4647	**1.6858**
	MSE	32.3640	11.8460	34.6690	32.5750	17.5030	6.0750	**2.8421**
	MAE	3.5802	2.1964	3.8344	3.7748	2.5203	1.5835	**1.1166**
	R^2	0.7344	0.9028	0.7155	0.7327	0.8564	0.9501	**0.9766**
EV Dataset 2	**RMSE**	7.8201	4.3465	8.3382	8.3784	5.9720	4.7896	**3.0992**
	MSE	61.1544	18.8921	69.5269	70.2160	35.6651	22.9403	**9.6054**
	MAE	5.0791	2.9334	5.2828	6.2456	4.0439	3.2526	**2.1768**
	R^2	0.7769	0.9309	0.7458	0.7433	0.8696	0.9161	**0.9648**

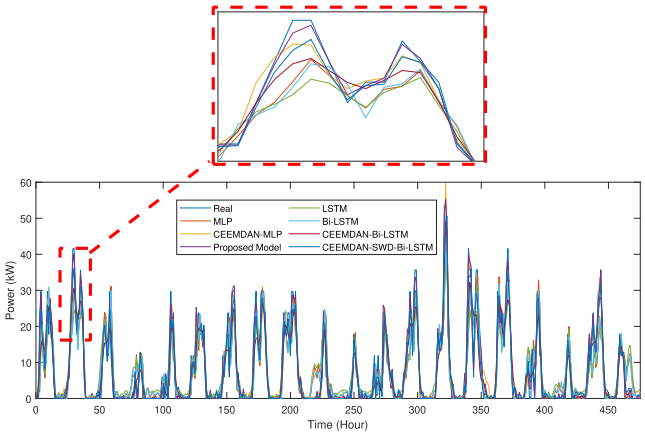

FIGURE 8. Forecasting test results of the EV fleet charging dataset.

The reliability and effectiveness of the proposed algorithm is analyzed using the Taylor diagram as in Fig. 13. Hence, the connection between the correlation coefficient, root mean square deviation (RMSD) and standard deviation is shown in this diagram. In this respect, all the implemented models can be shown on the basis of how well they predict the target data. Herein, the coordinate position of the forecasted value of the CEEMDAN-SWD-MLP model is the closest to the coordinate position of the observed point. The forecasting performance of the proposed model is said to be the best. As shown in the diagram, the proposed method, denoted by the triangular sign in red, has a lower RMSD, a lower standard deviation, and a coefficient value that is closer to 1. It is shown that the correlation between the real and forecasted data is more linear. Considering the standalone forecasting models (e.g., MLP, LSTM and Bi-LSTM), they show lower performance as compared to their single decomposition or multiple decomposition based hybrid counterparts. The simulated results observed that the multiple decomposition based algorithms (e.g. CEEMDAN-SWD-Bi-LSTM and CEEMDAN-SWD-MLP) provide better learning and prediction abilities than single decomposition methods. While multiple decomposition based AI models display competitive forecasting performance, the CEEMDAN-SWD-MLP is preferred thanks to its computational simplicity. All performance analyses have clearly demonstrated that the benefit of the proposed multiple decomposition method has improved the accuracy significantly. A statistical hypothesis test, the Diebold–Mariano (DM), was also performed to discuss the prediction performance of the multiple decomposition based hybrid models. Thus, it is the aim of this paper to examine the effectiveness of the multiple decomposition based hybrid model from a statistical perspective. These test results make the null hypothesis about the difference in accuracy between two forecasting models explicit [37]. If there is no significant difference between the two models, it is defined as the null hypothesis. The other hypothesis can be chosen when there is a significant difference between the two models, on the contrary. The DM test hypothesis can be defined as follows:

$$H_0 : E\left[F\left(e_i^1\right)\right] = E\left[F\left(e_i^2\right)\right] \tag{27}$$

$$H_1 : E\left[F\left(e_i^1\right)\right] \neq E\left[F\left(e_i^2\right)\right], \tag{28}$$

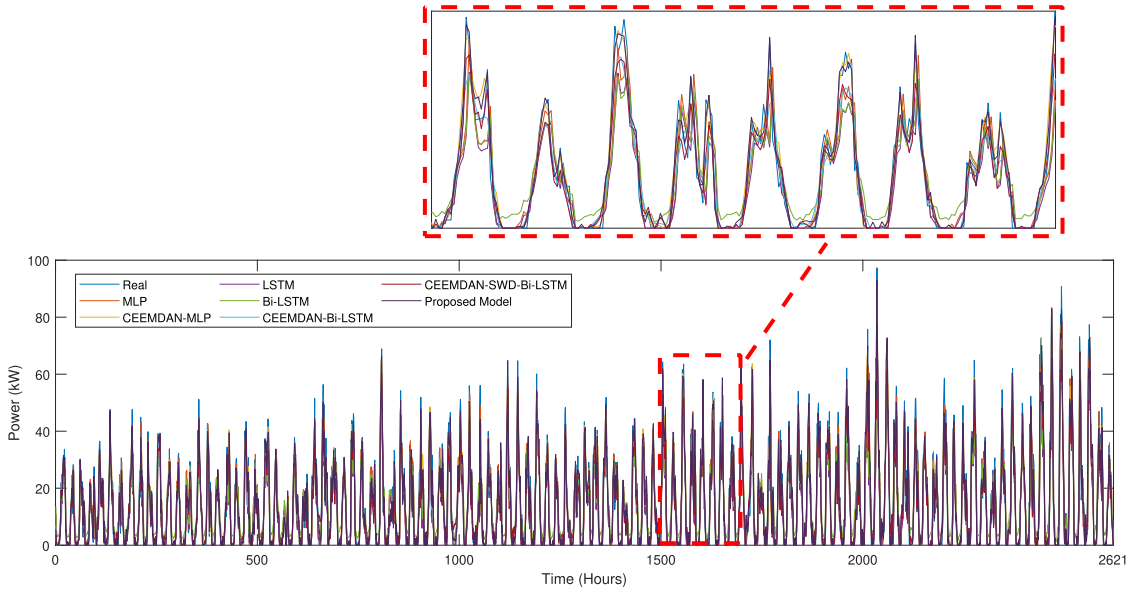

FIGURE 9. Forecasting test results of the EV charging dataset 2.

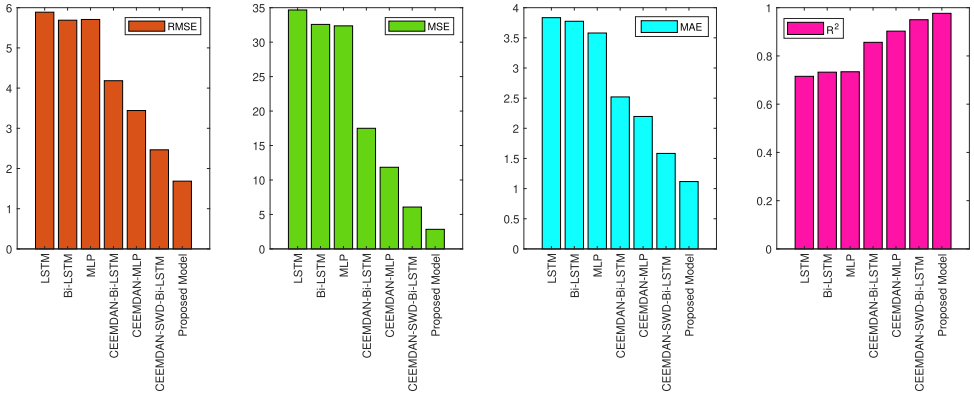

FIGURE 10. The forecasting error performance metrics result from the implemented models.

where F represents the loss function of forecasting errors, and e_i^1 and e_i^2 are the prediction errors between actual values and predicted values of the proposed model and the implemented model, respectively. The statistical value of the DM is given by

$$DM = \frac{\sum_{t=1}^{n} \left(F\left(e_i^1\right) - F\left(e_i^2\right)\right)/n}{\sqrt{S^2/n}} S^2 \tag{29}$$

A comparative analysis of the implemented models in terms of their DM values is reported in Table 2. As reported in Table 2, the multiple decomposition based proposed model is a remarkable difference from implemented models at a 1% significance level for all EV datasets. It was observed that the smallest DM value is 6.2134 for EV Dataset 1. In this respect, the CEEMDAN-SWD-Bi-LSTM shows the closest characteristic features to EV Dataset 1. Thus, the null hypothesis can be rejected at a 1% significance level even with these values. Therefore, the multiple decomposition

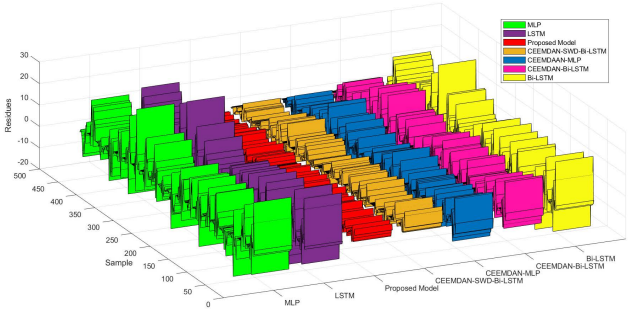

FIGURE 11. Variations of residual value per test sample data.

based hybrid model significantly outperforms all the other models implemented. The results confirm the effectiveness of the proposed model for EV fleet charging load forecasting. Therefore, the hybrid model based on the proposed multiple

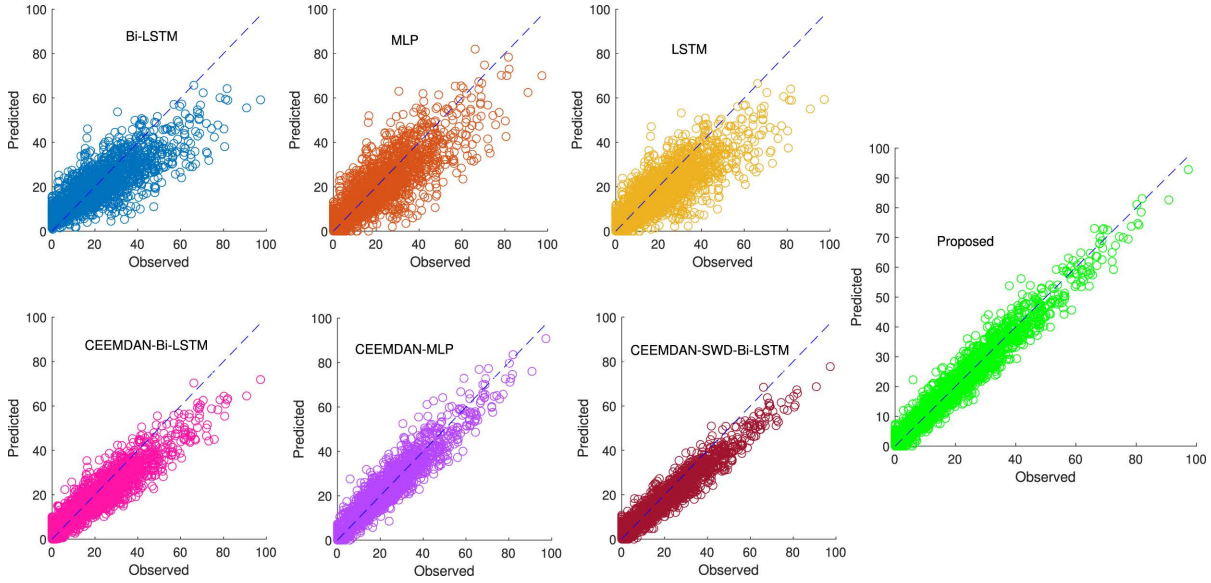

FIGURE 12. Scatter plots between of original and forecasted values for the implemented models for EV dataset 2.

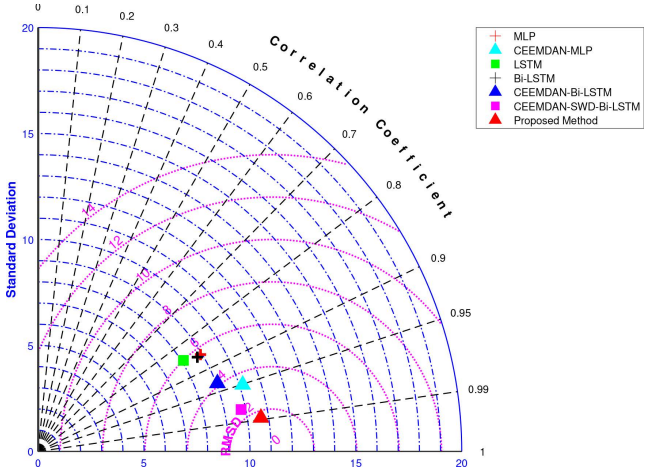

FIGURE 13. The Taylor diagram of the forecasting models of EV fleet charging data.

TABLE 2. Diebold–Mariano (DM) test results of the implemented models for statistical comparison.

Methods	DM Value
MLP	9.0968*
LSTM	9.5066*
Bi-LSTM	9.3429*
CEEMDAN-MLP	8.7867*
CEEMDAN-Bi-LSTM	8.4206*
CEEMDAN-SWD-Bi-LSTM	6.2134*

*1% significance level

decomposition approach significantly outperforms all the other models implemented. The findings confirm the effectiveness of the proposed model in forecasting EV fleet charging load.

IV. CONCLUSION

This study proposed a multiple decomposition-based hybrid forecasting model for EV fleet charging loads. Unlike general decomposition techniques commonly used in the literature, this study developed a two-stage decomposition stage using CEEMDAN and SWD methods to improve the accuracy. To validate the model's performance, two real EV fleet charging datasets with varying characteristics were used.

The decomposition of the highest frequency component with CEEMDAN, which is an additional decomposition stage, improved the model performance. Specifically, the proposed method reduced performance error metrics (e.g., RMSE) by 28.7-51.0% and 48.1-59.7% as compared to the models with a single decomposition method, CEEMDAN-MLP and CEEMDAN-Bi-LSTM, respectively. Furthermore, R^2 increased from 0.9028 to 0.8564 to 0.9766 when compared to CEEMDAN-MLP and CEEMDAN-Bi-LSTM. To further examine the effectiveness of the multiple decomposition approach, a statistical hypothesis test, the Diebold–Mariano (DM), was performed. It has been shown that, in terms of all DM results considered, the proposed hybrid model had a significant difference from the implemented models for EV data sets. As a result, the proposed model outperformed compared to both conventional and hybrid models, MLP, LSTM, and Bi-LSTM, with a single decomposition module.

Future studies of EV charging forecasting will include meta-heuristic approaches with AI methods in order to explore further increases in performance.

REFERENCES

[1] (2021). *Global EV Outlook: Accelerating Ambitions Despite the Pandemic*. Accessed: Mar. 31, 2021. [Online]. Available: https://www.iea.org/reports/global-ev-outlook-2021

[2] N. Erdogan, D. Pamucar, S. Kucuksari, and M. Deveci, "An integrated multi-objective optimization and multi-criteria decision-making model for optimal planning of workplace charging stations," *Appl. Energy*, vol. 304, Dec. 2021, Art. no. 117866. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0306261921011880

[3] A. S. Al-Ogaili, T. J. T. Hashim, N. A. Rahmat, A. K. Ramasamy, M. B. Marsadek, M. Faisal, and M. A. Hannan, "Review on scheduling, clustering, and forecasting strategies for controlling electric vehicle charging: Challenges and recommendations," *IEEE Access*, vol. 7, pp. 128353–128371, 2019.
[4] M. Liang, W. Li, J. Yu, and L. Shi, "Kernel-based electric vehicle charging load modeling with improved Latin hypercube sampling," in *Proc. IEEE Power Energy Soc. Gen. Meeting*, Jul. 2015, pp. 1–5.
[5] N. Erdogan, S. Kucuksari, and U. Cali, "Co-simulation of optimal EVSE and techno-economic system design models for electrified fleets," *IEEE Access*, vol. 10, pp. 18988–18997, 2022.
[6] M. C. Kisacikoglu, F. Erden, and N. Erdogan, "Distributed control of PEV charging based on energy demand forecast," *IEEE Trans. Ind. Informat.*, vol. 14, no. 1, pp. 332–341, Jan. 2018.
[7] M. Majidpour, C. Qiu, P. Chu, H. R. Pota, and R. Gadh, "Forecasting the EV charging load based on customer profile or station measurement?" *Appl. Energy*, vol. 163, pp. 134–141, Feb. 2016.
[8] M. Majidpour, C. Qiu, P. Chu, R. Gadh, and H. R. Pota, "Fast prediction for sparse time series: Demand forecast of EV charging stations for cell phone applications," *IEEE Trans. Ind. Informat.*, vol. 11, no. 1, pp. 242–250, Feb. 2015.
[9] C. Gómez-Quiles, G. Asencio-Cortés, A. Gastalver-Rubio, F. Martínez-Álvarez, A. Troncoso, J. Manresa, J. C. Riquelme, and J. M. Riquelme-Santos, "A novel ensemble method for electric vehicle power consumption forecasting: Application to the Spanish system," *IEEE Access*, vol. 7, pp. 120840–120856, 2019.
[10] X. Zhang, K. W. Chan, H. Li, H. Wang, J. Qiu, and G. Wang, "Deep-learning-based probabilistic forecasting of electric vehicle charging load with a novel queuing model," *IEEE Trans. Cybern.*, vol. 51, no. 6, pp. 3157–3170, Jun. 2021.
[11] M. Chang, S. Bae, G. Cha, and J. Yoo, "Aggregated electric vehicle fast-charging power demand analysis and forecast based on LSTM neural network," *Sustainability*, vol. 13, no. 24, p. 13783, Dec. 2021.
[12] J. Zhu, Z. Yang, M. Mourshed, Y. Guo, Y. Zhou, Y. Chang, Y. Wei, and S. Feng, "Electric vehicle charging load forecasting: A comparative study of deep learning approaches," *Energies*, vol. 12, no. 14, p. 2692, Jul. 2019.
[13] M. Dabbaghjamanesh, A. Moeini, and A. Kavousi-Fard, "Reinforcement learning-based load forecasting of electric vehicle charging station using Q-learning technique," *IEEE Trans. Ind. Informat.*, vol. 17, no. 6, pp. 4229–4237, Jun. 2021.
[14] Y. Li, Y. Huang, and M. Zhang, "Short-term load forecasting for electric vehicle charging station based on niche immunity lion algorithm and convolutional neural network," *Energies*, vol. 11, no. 5, p. 1253, 2018.
[15] S.-X. Lv and L. Wang, "Deep learning combined wind speed forecasting with hybrid time series decomposition and multi-objective parameter optimization," *Appl. Energy*, vol. 311, Apr. 2022, Art. no. 118674.
[16] Z. Qian, Y. Pei, H. Zareipour, and N. Chen, "A review and discussion of decomposition-based hybrid models for wind energy forecasting applications," *Appl. Energy*, vol. 235, pp. 939–953, Feb. 2019.
[17] B. Gao, X. Huang, J. Shi, Y. Tai, and J. Zhang, "Hourly forecasting of solar irradiance based on CEEMDAN and multi-strategy CNN-LSTM neural networks," *Renew. Energy*, vol. 162, pp. 1665–1683, Dec. 2020.
[18] W. Zhang, Z. Qu, K. Zhang, W. Mao, Y. Ma, and X. Fan, "A combined model based on CEEMDAN and modified flower pollination algorithm for wind speed forecasting," *Energy Convers. Manage.*, vol. 136, pp. 439–451, Mar. 2017.
[19] C. Emeksiz and M. Tan, "Multi-step wind speed forecasting and Hurst analysis using novel hybrid secondary decomposition approach," *Energy*, vol. 238, Jan. 2022, Art. no. 121764.
[20] D. Wang, S. Wei, H. Luo, C. Yue, and O. Grunder, "A novel hybrid model for air quality index forecasting based on two-phase decomposition technique and modified extreme learning machine," *Sci. Total Environ.*, vol. 580, pp. 719–733, Feb. 2017.
[21] E. Dokur, "Swarm decomposition technique based hybrid model for very short-term solar PV power generation forecast," *Elektronika ir Elektrotechnika*, vol. 26, no. 3, pp. 79–83, Jun. 2020.
[22] E. Dokur, N. Erdogan, M. E. Salari, C. Karakuzu, and J. Murphy, "Offshore wind speed short-term forecasting based on a hybrid method: Swarm decomposition and meta-extreme learning machine," *Energy*, vol. 248, Jun. 2022, Art. no. 123595.
[23] M. E. Torres, M. A. Colominas, G. Schlotthauer, and P. Flandrin, "A complete ensemble empirical mode decomposition with adaptive noise," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2011, pp. 4144–4147.
[24] N. E. Huang, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N. C. Yen, C. C. Tung, H. H. Liu, and Z. Shen, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. Roy. Soc. London A, Math., Phys. Eng. Sci.*, vol. 454, no. 1971, pp. 903–995, 1998.
[25] G. K. Apostolidis and L. J. Hadjileontiadis, "Swarm decomposition: A novel signal analysis using swarm intelligence," *Signal Process.*, vol. 132, pp. 40–50, Mar. 2017.
[26] S. Wan and B. Peng, "An integrated approach based on swarm decomposition, morphology envelope dispersion entropy, and random forest for multi-fault recognition of rolling bearing," *Entropy*, vol. 21, no. 4, p. 354, Apr. 2019.
[27] Y. Shi, X. Song, and G. Song, "Productivity prediction of a multilateral-well geothermal system based on a long short-term memory and multi-layer perceptron combinational neural network," *Appl. Energy*, vol. 282, Jan. 2021, Art. no. 116046.
[28] S. Bueno and J. L. Salmeron, "Benchmarking main activation functions in fuzzy cognitive maps," *Expert Syst. Appl.*, vol. 36, no. 3, pp. 5221–5229, Apr. 2009.
[29] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
[30] Y. Fan, K. Xu, H. Wu, Y. Zheng, and B. Tao, "Spatiotemporal modeling for nonlinear distributed thermal processes based on KL decomposition, MLP and LSTM network," *IEEE Access*, vol. 8, pp. 25111–25121, 2020.
[31] S. Siami-Namini, N. Tavakoli, and A. S. Namin, "The performance of LSTM and BiLSTM in forecasting time series," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2019, pp. 3285–3292.
[32] Z. Qu, W. Mao, K. Zhang, W. Zhang, and Z. Li, "Multi-step wind speed forecasting based on a hybrid decomposition technique and an improved back-propagation neural network," *Renew. Energy*, vol. 133, pp. 919–929, Apr. 2019.
[33] LC Council. (Sep. 29, 2020). *EV Fleet Chargepoint Use*. [Online]. Available: https://data.gov.U.K./dataset/2279b730-bf4e-40c4-b2de-c82d43ae16d2/ev-fleet-chargepoint-use
[34] ElaadNL. (2021). *Open EV Public Charging Dataset*. [Online]. Available: https://platform.elaad.io/download-data/
[35] H. Abbasimehr, M. Shabani, and M. Yousefi, "An optimized model using LSTM network for demand forecasting," *Comput. Ind. Eng.*, vol. 143, May 2020, Art. no. 106435.
[36] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.
[37] F. Diebold and R. Mariano, "Comparing predictive accuracy," *J. Bus. Econ. Statist.*, vol. 20, no. 1, pp. 134–144, Jan. 2002.