



Article An Integrated Model of Deep Learning and Heuristic Algorithm for Load Forecasting in Smart Grid

Hisham Alghamdi ¹, Ghulam Hafeez ^{2,*}, Sajjad Ali ³, Safeer Ullah ⁴, Muhammad Iftikhar Khan ⁵, Sadia Murawwat ⁶ and Lyu-Guang Hua ⁷

- ¹ Electrical Engineering Department, College of Engineering, Najran University, Najran 55461, Saudi Arabia; hg@nu.edu.sa
- ² Department of Electrical Engineering, University of Engineering and Technology, Mardan 23200, Pakistan
- ³ Department of Telecommunication Engineering, University of Engineering and Technology, Mardan 23200, Pakistan; sajjad@uetmardan.edu.pk
- ⁴ Department of Electrical Engineering, Quaid-e-Azam College of Engineering & Technology, Sahiwal 57000, Pakistan; safeer_iiui@yahoo.com
- ⁵ Department of Electrical Engineering, University of Engineering and Technology, Peshawar 25000, Pakistan; miftikhar@uetpeshawar.edu.pk
- ⁶ Department of Electrical Engineering, Lahore College for Women University, Lahore 51000, Pakistan; sadia.murawwat@lcwu.edu.pk
- ⁷ Power China Hua Dong Engineering Corporation Ltd., Hangzhou 311122, China; lv_gh@hdec.com
- * Correspondence: ghulamhafeez393@gmail.com

Abstract: Accurate load forecasting plays a crucial role in the effective energy management of smart cities. However, the smart cities' residents' load profile is nonlinear, having high volatility, uncertainty, and randomness. Forecasting such nonlinear profiles requires accurate and stable prediction models. On this note, a prediction model has been developed by combining feature preprocessing, a multilayer perceptron, and a genetic wind-driven optimization algorithm, namely FPP-MLP-GWDO. The developed hybrid model has three parts: (i) feature preprocessing (FPP), (ii) a multilayer perceptron (MLP), and (iii) a genetic wind-driven optimization (GWDO) algorithm. The MLP is the key part of the developed model, which uses a multivariate autoregressive algorithm and rectified linear unit (ReLU) for network training. The developed hybrid model known as FPP-MLP-GWDO is evaluated using Dayton Ohio grid load data regarding aspects of accuracy (the mean absolute percentage error (MAPE), Theil's inequality coefficient (TIC), and the correlation coefficient (CC)) and convergence speed (computational time (CT) and convergence rate (CR)). The findings endorsed the validity and applicability of the developed model compared to other literature models such as the feature selection-support vector machine-modified enhanced differential evolution (FS-SVM-mEDE) model, the feature selection-artificial neural network (FS-ANN) model, the support vector machine-differential evolution algorithm (SVM-DEA) model, and the autoregressive (AR) model regarding aspects of accuracy and convergence speed. The findings confirm that the developed FPP-MLP-GWDO model achieved an accuracy of 98.9%, thus surpassing benchmark models such as the FS-ANN (96.5%), FS-SVM-mEDE (97.9%), SVM-DEA (97.5%), and AR (95.7%). Furthermore, the FPP-MLP-GWDO significantly reduced the CT (299s) compared to the FS-SVM-mEDE (350s), SVM-DEA (240s), FS-ANN (159s), and AR (132s) models.

Keywords: electric load forecasting; multilayer perceptron; decision making; energy management; deep learning; heuristic optimization algorithm; smart power grid

MSC: 68T20

1. Introduction

Accurate load prediction is indispensable for the effective planning, operation, and energy management of the smart power grid (SPG). It is essential for ensuring the SPG's



Citation: Alghamdi, H.; Hafeez, G.; Ali, S.; Ullah, S.; Khan, M.I.; Murawwat, S.; Hua, L.-G. An Integrated Model of Deep Learning and Heuristic Algorithm for Load Forecasting in Smart Grid. *Mathematics* 2023, *11*, 4561. https:// doi.org/10.3390/math11214561

Academic Editors: Jonathan Blackledge, Jie Wen and Antonio Bolufé-Röhler

Received: 22 August 2023 Revised: 8 October 2023 Accepted: 19 October 2023 Published: 6 November 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). sustainable, secure, and reliable operation, thus benefiting supply and demand-side stakeholders [1–3]. On the supply side, precise load prediction enables effective resource allocation to meet residents' energy demands and optimize resource utilization [4]. Conversely, on the demand side, accurate load prediction is imperative for proactive equipment management, load scheduling, efficient energy utilization, and optimal energy management [5]. However, the accuracy of load forecasting is affected by inherent data uncertainties and randomness. These uncertainties make the task of consistently improving forecast accuracy complex and challenging. Consequently, there is a pressing need to develop models that are capable of enhancing forecast accuracy by effectively addressing the inherent uncertainties in load patterns.

In recent decades, load forecasting techniques are developed by numerous authors like the time series methods: exponential smoothing [6,7], Kalman filters [8], regression methods [9,10], the grey forecasting model (GM) [11], and the autoregressive integrated moving average (ARIMA), as well as ARMAX methods [12–14]. In [15], the authors developed the ARIMA-MPSO model for load forecasting. These prediction methods are capable of forecasting electric load. However, the accuracy improvement is not up to the mark due to the method's inherent shortcomings. For instance, linear regressors are suitable for solving linear problems and have the worst performance while addressing nonlinear problems. Methods such as ARIMA take historical/current records for prediction while ignoring other influencing parameters. GM methods can only cater to exponential growth trends problems. Artificial intelligence (AI) emerged as a smart solution to resolve the issues of traditional methods. For example, these include expert systems [16], radial basis fuzzy logic models [17,18], machine learning models [19], neural networks [20–22], and multilayer perceptron (MLP) models [23]. AI models outperform traditional models in terms of accuracy. However, these methods also suffer from some limitations. For instance, expert systems rely on knowledge acquisition and are challenged when handling uncertainty, radial basis logic models are computationally expensive and have limited generalization capability, and neural network models become trapped in local optima. In [24,25], deep learning models were introduced to resolve the drawbacks of existing models and to improve forecast accuracy. However, these models have high computational complexity. These deep layer models and hybrid methods are superior to intrinsic methods in terms of of accuracy. However, they ignore data preprocessing approaches, which are important for improved accuracy. Considering the limitations of AI methods, hybrid methods have been developed. For example, in [26–28], a hybrid method combining the regression neural network (RGNN) and fruitfly optimization (FFO) algorithm is developed for load forecasting. In [29], an efficient hybrid model using a neural network optimized with the artificial bee colony optimization algorithm is introduced to address the load forecasting issue. The paper in [30] cascaded the support vector machine (SVM) with AI for electric load forecasting in an SPG. Data-driven models have been developed to identify the services needed for load forecasting in smart cities [31]. The authors developed the Mc-SVNN model for sunspot number time series, USD-to-euro currency exchange rate forecasting, daily temperature prediction, and power demand forecasting and wind speed forecasting in Abu Dhabi [32,33]. The proposed model is compared with the literature works in Table 1.

As previously discussed, neural network models have limitations such as interpretability, being trapped into local minima due to limited extrapolation and generalization ability, and being unable to select abstracted features from datasets due shallow layouts. The MLP model employs learning principles (the multivariate autoregressive algorithm (MVARA) and ReLU) and the heuristic optimization algorithm to address these limitations and to minimize the error metric to forecast accuracy enhancement. With this motivation, in this work, the FPP-MLP-GWDO model is developed. First of all, the developed model employs FPP, which uses candidate interaction concepts, redundancy, and relevancy filters to return suitable features to the MLP forecaster. Secondly, the developed model uses MLP as a forecaster by utilizing learning principles, i.e., the MVARA and ReLU, to enhance model generalization and to facilitate accurate prediction. Finally, the developed model employees the genetic wind-drive optimization (GWDO) algorithm [34] as the optimizer due to the powerful search ability for the optimal solution with a faster convergence rate [35]. The GWDO further improves the prediction accuracy by optimizing filter thresholds (irrelevancy and redundancy) weights, as well as biases of the MLP forecaster. This work is a continuation of the earlier work [36] where the FS-FCRBM-GWDO is developed. It is compared with existing models with respect to error metrics and the CC for validation. The novelty and technical contributions are highlighted below:

- A FPP-MLP-GWDO has been developed, where the preprocessing FPP and postprocessing GWDO have been cascaded with the MLP for accuracy improvement.
- Based on existing feature selection techniques [37], FPP has been developed where the feature interaction concept has been introduced, in addition to filters (irrelevancy, redundancy) for the selection of key features.
- The GWDO has been applied to the returned predictions from the MLP to further improve the accuracy by optimizing the filter threshold (irrelevancy and redundancy) weights and biases.
- The developed hybrid model, FPP-MLP-GWDO, is evaluated using Dayton Ohio grid load data regarding aspects of accuracy (the mean absolute percentage error (MAPE), Theil's inequality coefficient (TIC), and the correlation coefficient (CC)) and convergence speed (the CT and CR). The findings endorsed the validity and superiority of the developed model compared to the literature models such as the feature selection–support vector machine–modified enhanced differential evolution (FS-SVM-mEDE) [38], the feature selection–artificial neural network (FS-ANN) [26], the support vector machine–differential evolution algorithm (SVM-DEA) [39], and the autoregressive (AR) model regarding aspects of accuracy and convergence speed.

The remaining sections of this work are organized as follows: Section 2 presents the FPP-MLP-GWDO proposed model. Section 3 presents simulations and discussions. Finally, this work's conclusion is present in Section 4.

Refs.	Methods	Objectives	Performance Metrics	Advantages/Disadvantages	
Time series methods [6–14]	Exponential smoothing, Kalman filters, regression methods, grey model, ARIMA, and ARMAX	Forecast accuracy improvement	MAE, MAPE, RMSE	These prediction methods are capable of forecasting electric load. However, the accuracy improvement is not up to the mark due to the method's inherent shortcomings. For instance, linear regressors are suitable for solving linear problems, but they perform worst while addressing nonlinear problems. Methods such as the ARIMA take historical/current records for prediction while ignoring other influencing parameters. GM methods can only cater to exponential growth trend problems.	
Artificial intelligence methods [16–20]	Expert systems, radial basis fuzzy logic models, machine learning models, and neural networks	Accuracy enhancement and compilation time improvement	MAPE, correlation coefficient	AI models outperform traditional methods when it comes to accuracy. Nevertheless, these advanced techniques have their limitations. For example, expe systems necessitate extensive knowledge acquisition and can need help with handling uncertainty. Radia basis logic models, on the other hand, are computationally expensive and exhibit limited generalization capabilities. Neural network models powerful. However, they easily get stuck in local optima.	

Table 1. Recent related works compared with the proposed model.

Refs.	Methods	Objectives	Performance Metrics	Advantages/Disadvantages	
Deep neural networks and hybrid models [24–28]	MLP, LSTM, RBM, etc.	Accuracy improvement	RMSE, MSE, R, etc.	Deep learning models have been introduced to address the limitations of existing forecasting methods and to enhance prediction accuracy. Nevertheless, these models come with a notable drawback: their high computational complexity. While deep-layer models excel in terms of accuracy compared to intrinsic methods, they often overlook the significance of data preprocessing techniques, which are crucial for achieving improved accuracy.	
Integrated FS-FCRBM- GWDO model [36]	AFC-STLF, MI- mEDE-ANN, FS-ANN, Bi-level.	Accuracy improvement	MAPD, variance, correlation, etc.	A hybrid model is introduced to tackle the constraints associated with current forecasting techniques. The aim is to enhance the prediction accuracy and convergence speed. However, the developed model has comparatively high complexity and a slow convergence.	
This work	FPP-MLP- GWDO	Accuracy, convergence rate, and computational time improvement	TIC, CC, CR, CT, and MAPE	The FPP-MLP-GWDO hybrid model offers several significant advantages over existing models in load forecasting. Its notably improved forecast accuracy stands out, thereby making it a valuable model for precise predictions. Moreover, its ability to converge quickly is ideal for real-time applications, while its adaptability allows it to handle various scenarios and datasets effectively. The model's prowess in capturing nonlinear load patterns ensures accuracy even in complex situations, and its robustness in the face of data variability instills confidence in its reliability. Furthermore, it enhances resource allocation, thus leading to cost savings, and it is designed for scalability to meet the evolving demands of expanding smart city infrastructures. Lastly, the integration of feature preprocessing simplifies data preparation, thereby streamlining the forecasting process. Overall, the FPP-MLP-GWDO model significantly advances load forecasting, thus offering improved accuracy, efficiency, and fast convergence speed.	

Table 1. Cont.

2. Developed Hybrid FPP-MLP-GWDO Model

The developed FPP-MLP-GWDO is a hybrid model with three parts: FPP, an MLP forecaster, and a GWDO optimizer, which is depicted in Figure 1. The developed model goal is to enhance the accuracy and convergence speed simultaneously. The first part, FPP, uses filters (redundancy and irrelevancy) and interaction operations. The FPP takes load data and other influencing parameters such as the dew point, wind speed, humidity, and temperature as input. Through FPP, the received data is first normalized and then fed to filters (redundancy and irrelevancy) and interaction operations. The FPP goal is to return key features and to clean the data (maximum relevancy and feature interactions and minimum redundancy) for the MLP forecaster. The MLP forecaster is trained on the received data to learn future load patterns (forecast) for the Dayton Ohio electric grid. The predicted load pattern returned from the MLP forecaster is fed to the optimizer part based on a GWDO to further improve the accuracy by reducing the error amount. The comprehensive explanation is as follows.



Figure 1. Developed FPP-MLP-GWDO framework for electric load forecasting in SPG.

2.1. Feature Preprocessing

The historical load and exogenous parameters are fed into the FPP part. First, the cleansing operation recovers missing values with earlier day average. Then, the cleaned data are passed through the normalization step to make them within the activation function bounds as follows.

$$Norm = \frac{X - \mu(X)}{std(X),} \tag{1}$$

where *X* represents input data, *std* indicates the standard deviation, and *Norm* shows the normalized data. *X* is the input data, which includes the load demand data denoted by P(hr, d), the temperature illustrated by T(hr, d), the dew point indicated by D(hr, d), and the humidity parameter, which is denoted by (H(hr, d)). The (hr, d) indicates the hour and specific day, respectively. The wind speed, dew point, humidity, and temperature are influencing parameters, which are also known as exogenous parameters. The FPP part includes filters (redundancy and irrelevancy) and feature interaction operations. The FPP aims to discard redundant, irrelevant, and nonconstructive features from the dataset, because redundant information causes execution overhead during the training, and irrelevant features act as outliers. A comprehensive explanation of the FPP filters and feature interaction operations is give below.

2.1.1. Relevant Feature Selection: Relevancy Filter

The relevancy filter in FPP plays a significant role in selecting key features. The relevancy filter selects key features by correlating input features with the target. Relevancy measurements have been made by many techniques [40]. This work uses mutual information (MI) to measure feature relevancy, i.e., how closely *a* and *y* are correlated in the data. The MI observes the *y* under *a*. The MI for *a* and *y* is computed via individual p(a)andp(y) probability distributions (PDs) and the joint PD p(a, y), and it is indicated by I(a; y).

$$S = \{a_1, a_2, \dots, a_M\},$$
 (2)

where *S* is the set having the input variables (a_1, \ldots, a_M) and the target variable *y*. The computation involves checking the information that is common between the input a_i and the target *y* to determine their degree of association. When the common information between the two variables exceeds a certain threshold, it indicates a strong relationship. Moreover, the determination of the significance of the input a_i concerning the target *y* is computed through the following approach.

$$D(a_{i}) = \begin{cases} keep & I(a_{i}; y) > \sigma \\ Discard & I(a_{i}; y) \le \sigma \end{cases}$$
(3)

 $D(a_i)$ represents the measure of relevance between the input and the target variables.

2.1.2. Redundant Feature Elimination: Redundancy Filter

The redundant feature elimination has significant importance with respect to improving the convergence speed. On the other hand, redundant features slow down the convergence speed. On this note, the FPP employed a redundancy filter to find redundant features from the input features set using the MI mechanism. The aim is to rectify the input feature set by discarding redundant features and keeping relevant features. According to the research conducted in [40], it was observed that closely correlated input variables have a negative impact on features selection. The explanation is that two input variables share a significant amount of common information regarding the target variable, but they share very little redundant information. As a result, an input with limited redundant and excluded, even though it could provide essential abstract features for the proposed model. To address these challenges, a redundancy measure called interaction gain (Ig) was introduced in [41], which is defined in Equation (4).

$$RM(a_i, a_s) = Int_g(a_i; a_s; y) = I[(a_i, a_s); y] - I(a_i; a_s) - I(a_s; y),$$
(4)

The redundancy measure, denoted as $RM(a_i, a_s)$, represents the degree of redundancy between the potential inputs a_i and a_s , while y denotes the target variable. The mathematical modeling of the information gain Ig can be expressed in relation to the joint and individual entropy as follows:

$$Int_{g}(a_{i};a_{s};y) = H(a_{i},a_{s}) + H(a_{i},y) + H(a_{s},y) - H(a_{i}) - H(a_{s}) - H(y) - H(a_{i},a_{s},y)$$
(5)

The individual entropy values of a_i , a_s , and y are represented by $H(a_i)$, $H(a_s)$, and H(y), respectively. In contrast, the joint entropy values are denoted by $H(a_i, a_s)$, $H(a_i, y)$, $H(a_s, y)$, and $H(a_i, a_s, y)$.

2.1.3. Feature Interaction

The authors in [37] introduced the concept of filters (irrelevancy and redundancy) with the aim of eliminating redundant/irrelevant features while retaining the relevant ones for subsequent steps. However, a drawback of filter-based methods is that they may discard features that were initially deemed irrelevant, even though such features may be relevant when considered in combination with other features in the set. Building upon this observation, the FPP introduces the notion of interaction, in addition to a filters (irrelevancy and redundancy) approach. If two variables from the input set, a_i and a_s , possess redundant features with respect to the target variable y, the joint MI estimate between a_i and y will be lower than the their combined individual MI estimates. Consequently, this leads to a negative value, as indicated by Equation (4), which signifies the presence of redundant features a_i and a_s for the model. By taking the absolute of the result from Equation (4), we obtain a measure that quantifies the extent of redundancy. In contrast, when the input variables a_i and a_s interact with y (the target variable), their combined MI values with y exceed the sum of their individual MIs. Therefore, a positive value in Equation (4) indicates the presence of interacting features, and its positive/absolute magnitude reflects the extent of the interaction. Therefore, to account for redundancy and interaction, Equation (4) may be expressed with reference to the concept of the interaction gain *Ig*:

$$RM(a_i, a_s) = \begin{cases} \{Int_g(a_i; a_s; y), & \text{if } Int_g(a_i; a_s; y) < 0, \\ 0 & \text{otherwise} \end{cases}$$
(6)

$$In(a_i, a_s) = \begin{cases} Int_g(a_i; a_s; y), & \text{if } Int_g(a_i; a_s; y) > 0, \\ 0 & \text{otherwise} \end{cases}$$
(7)

The equation labeled as Equation (6) is derived by making adjustments to Equation (4) to measure redundancy. Equation (7) calculates the interaction measure. The computation of the interaction measure $IM(a_i)$ for each potential feature is computed below.

$$IM(a_i) = \underset{a_j \in S - \{a_i\}}{Maximize} \{ In(a_i, a_j) \}$$
(8)

2.1.4. FPP Stepwise Procedure

The objective of this adapted technique for feature selection is to optimize the relevance and interaction measures while minimizing the redundancy using a filters-based approach. Unlike existing techniques such as those proposed by [37,41,42], our FPP technique takes into account the interaction between candidates in addition to relevance and redundancy filters. Figure 2 illustrates an FPP technique flow chart [36]. The comprehensive explanation, along with a stepwise explanation, are listed below.

Step 1—Potential inputs: The technique takes the input set consisting of a potential inputs set and the target value *y*.

Step 2—Prefiltering: The prefiltering part of the FPP is illustrated as below:

- The enclosed blocks within the dashed box represent the prefiltering part, during which the relevancy/interaction are computed. The potential inputs are then ranked according to these computed estimates/measures.
- We assess the individual and the gained information to measure the information content. This is done using an adapted form of Equation (4), which is illustrated in the flowchart presented in Figure 2. The function f(,) used in the equation monotonically increases, while the weight factor α balances the relevancy and interaction measures. Depending on the specific forecasting problem, this factor can be adjusted and finely tuned.
- The potential inputs identified in the prefiltering step (*S^p*) are organized in a descending sequence as per their information value.

Step 3—Filtering stage: The filtering stage [36] of the FPP is illustrated in Figure 3 and presented as follows:

• The prefiltering stage output serves as the input for the filtering stage, where the preselected features are divided into selected (*S^s*) and nonselected (*Sⁿ*) features, as illustrated in Figure 2. Redundancy measure is computed using Equation (9), which is modeled below:

$$\mathbf{R}(\stackrel{p}{a_i}) = \underset{\substack{p \\ a_i \in S}}{\text{Minimize}} \left\{ \mathbf{RM}(\stackrel{p}{a_i}, \stackrel{p}{a_j}) \right\},\tag{9}$$

Here, $R(a_i)$ represents the measure of redundancy for every potential input a_i belonging to the set S.

• The assessment of the informational significance of the potential features comprises three metrics: redundancy, relevancy, and interaction. In mathematical terms, this evaluation can be expressed as follows:

$$V(\overset{p}{a_i}) = g\left\{ D(\overset{p}{a_i}), IM(\overset{p}{a_i}), R(\overset{p}{a_i}) \right\}$$

= $D(\overset{p}{a_i}) + \alpha \times IM(\overset{p}{a_i}) + \beta \times R(\overset{p}{a_i}),$ (10)

Here, α , $\beta > 0$, $V(a_i)$ represents the information content, g(,;) denotes a monotonically increasing linear function, and β corresponds to a tuneable parameter.

The determination of the information content is made using the following decision process:

If
$$V(\stackrel{p}{a_i}) \geq R_{th} \rightarrow \stackrel{5}{S} = \stackrel{5}{S} + \{\stackrel{p}{a_i}\}$$

If $V(\stackrel{p}{a_i}) \leq R_{th} \rightarrow \stackrel{n}{S} = \stackrel{n}{S} + \{\stackrel{p}{a_i}\},$
(11)

In this process, information content is compared to the redundancy threshold, which is denoted as R_{th} . If the information value is equal to or greater than the relevancy threshold, it is added to the list of selected features ($\stackrel{s}{S}$). Otherwise, it is included in the list of ($\stackrel{n}{S}$), which includes nonselected features.

• Features, both selected and nonselected, are arranged in descending order based on their information content. Then, a union operator is applied to create a unified set. Subsequently, the postfiltering phase takes both of the sets and their combinations as input [36] as presented in Figure 4.



Figure 2. Overall FPP flow chart, including prefiltering, filtering, and postfiltering.



Figure 3. Filtering part flow chart.

Step 4—Postfiltering: In this phase, adjustments are made to both the selected (\tilde{S}) and nonselected inputs, thus resulting in updates to the V(.) information value. These updated information contents are then reassessed via Equation (11) to determine whether the potential inputs should be included in the selected or nonselected features.

The FPP stops when the nonselected features set \ddot{S} is empty, and no entry remains. During each iteration, the potential input set undergoes prefiltering, filtering, and postfiltering. This ensures that the process avoids becoming stuck in an infinite loop and that it successfully returns the selected features set. Eventually, the selected features are passed into the MLP forecaster.



Figure 4. Postfiltering stage of FPP.

2.2. MLP Forecaster

This part of the proposed model is the MLP forecaster that can be trained to accurately predict future load patterns. The literature review concludes that all the currently available models have the ability to forecast non-linear electrical load. Hence, an MLP has been selected as the preferred intelligent forecaster because it predicts nonlinear load patterns with a satisfactory level of accuracy with earlier converging capability, it possesses a scalable nature, and it exhibits enhanced performance as it scales. The MLP is a variant of ANN comprising multiple layers of interconnected nodes referred to as neurons. Each neuron in one layer is connected to the neurons in the adjacent layers, thus forming a feed-forward architecture. The MLP has the ability to acquire intricate patterns and relationships within load data through a training process, where the weights and biases of neurons are adjusted using input-output pairs. The MLP forecaster employs an MVARA and utilizes ReLU as the learning rule to predict load patterns. The MVARA and ReLU are selected as the learning rules for the MLP forecaster. These rules are chosen because their earlier converging capability ensures a low CT and fast CR, as well as addresses common network challenges such as overfitting and a vanishing gradient. This allows the MLP to make accurate load predictions. The MLP forecaster has a layered architecture comprising the input, hidden, and output layers. Each layer is composed of artificial neurons. The MLP constitutes a feed-forward network comprised of fully connected layers. In this layout, each neuron in a layer is connected to the neurons in the subsequent layer through synaptic weights, as illustrated in Figure 5.

The MLP selects potential inputs from a given dataset and maps the input vector x(t) to the output vector F_t . The MLP's output is represented as follows:

$$F = \sum_{i=1}^{n} W_i f(y_i) + \sum_{j=1}^{m} \beta_j a_j,$$
(12)

In Equation (12), the $f(y_i \text{ is ReLU}, \text{ which is modeled in Equation (13)})$.

$$f(yi) = mx(0, yi)$$

$$f(yi) \begin{cases} 1 & \text{if } yi \ge 0 \\ zero & \text{else.} \end{cases}$$
(13)

The output vector F_t shows the day-ahead forecast results, and it is obtained through a combination of factors. These factors include the weight factor W_i , the linear weight β_j between the input/output nodes, the input elements a_j , and the input to the hidden nodes y_i . The training of the MLP involves utilizing the MVARA and the ReLU transfer functions. The calculation for y_i is expressed in Equation (15).

$$y_i = \sum_{j=1}^3 w_{ij} a_j + b_i,$$
 (14)

In Equation (15), w_{ij} represents the weight between the neurons in the input and hidden layers, and b_i denotes the bias for the hidden layer.



Figure 5. The MLP forecaster with MVARA and ReLU.

The learning process persists until one of the following conditions is satisfied: the iteration maximum limit is reached, the stopping criterion is fulfilled, or the error function is minimized. The error function is modeled below in Equation (15).

$$E = \frac{1}{N} \sum_{t=1}^{N} (R_t - F_t)^2,$$
(15)

In the Equation (15), the actual output of the network pattern is denoted by R_t , while the forecast output is represented by F_t . Additionally, N corresponds to the number of training samples used in the process.

By incorporating ReLU, the MLP forecaster is able to capture nonlinearities and interactions. In the literature, various algorithms have been used to update the weight and bias vectors during the training process, including the Levenberg Marquardt algorithm [41], the MVARA [43], the back-propagation, and the gradient descent [44]. Out of the available training/learning algorithms, the choice of utilizing the MVARA for network training was made because of its ability to converge rapidly and deliver enhanced performance. The selected features from the FPP stage, denoted as S_1, S_2, \ldots, S_n , are inputted into the MLP forecaster stage. In this stage, the network training utilizes data samples from the first three

years, while the testing phase employs data samples from the last year. The ultimate goal is to train the MLP forecaster through this process to accurately predict load patterns. The MLP forecaster generates an error signal, along with weights and biases, that are adjusted using the MVARA [45]. The MAPE serves as the objective function for the optimizer, which aims to enhance the accuracy by adjusting error signal.

2.3. GWDO Optimizer

The MLP forecaster produces a load pattern with a certain level of the MAPE, which is minimized as per the capabilities of the MLP, MVARA, and ReLU. To further reduce the MAPE in the predicted load, the MLP forecaster output is fed into our proposed GWDO optimizer. The GWDO optimizer aims to further decrease the error in the predicted load pattern. Therefore, the optimizer treats the minimization of the error as objective, which is mathematically represented in Equation (16).

$$\begin{array}{l} \underset{I_{th}, R_{th}, C_{i}}{\text{MAPE}} (a) \quad \forall \ a \in \{h, d\}, \end{array}$$
(16)

In Equation (16), I_{th} , R_{th} , and C_i represent the redundancy threshold, irrelevancy threshold, and potential input interaction, respectively. The GWDO optimizer tunes/adjusts these parameters and provides feedback to the FPP. In the FPP stage, the feature selection approach utilizes the optimized I_{th} , C_i , and R_{th} as potential input interactions for the key features selection.

Integrating the GWDO optimizer with the MLP forecaster improves accuracy, albeit at the expense of a degraded convergence rate. Usually, this integration of the optimizer with the forecaster is implemented in applications, where the main emphasis focuses on the accuracy rather than the speed of convergence.

To optimize the forecaster hyperparameters, numerous techniques have been suggested by researchers. These techniques encompass heuristic approaches, as well as quadratic, convex, linear, and nonlinear programming. In this study, linear programming was not utilized because of the nonlinear nature of the problem. Nonlinear programming was excluded because of the extended execution time they entail. Convex optimization was rejected because these processes converge slowly.

Heuristic algorithms such as DE [46], EDE [47], and mEDE [38] were rejected because of challenges such as inadequate precision, sluggish convergence, and the inclination to get stuck in local optima [48]. To overcome the constraints inherent in the current approaches, the GWDO was suggested as a means to effectively optimize the hyperparameters, thereby exhibiting rapid convergence speed. The GWDO is a hybrid approach that combines the key merits of the WDO [35] and the GA [49]. This hybridization proves to be advantageous, as it leverages the fast convergence speed of the WDO while benefiting from the diversity of population provided by the GA.

3. Results and Discussions

To assess effectiveness of the FPP-MLP-GWDO model, MATLAB simulations were performed on a laptop featuring a Core i3, a CPU Intel(R) @2.4GHz, and 8GB of RAM. The performance of the FPP-MLP-GWDO framework was evaluated by comparing it with the literature models: the FS-SVM-mEDE [38], FS-ANN [26], SVM-DEA [39], and AR model. These benchmark frameworks were selected because they share architectural similarities with the developed FPP-MLP-GWDO model.

The FPP-MLP-GWDO model was evaluated using load data from the Dayton Ohio grid. This dataset was obtained from the PJM electricity market, which is publicly available and openly accessible [50] and was also used in a previous study [37]. Figure 6 illustrates the four years of load data from the Dayton Ohio grid, which span from 2014 to 2017. The MLP forecaster uses eighty percent data for training and allocates the remaining twenty percent for testing purposes.



Figure 6. Dayton Ohio grid electricity consumption from 2014 to 2017.

The learning curve assesses the effectiveness of a model by comparing its performance on training and testing data samples over multiple epochs. The objective is to determine whether a model is genuinely learning from the data or merely memorizing it. A poor learning curve is indicative of high variance and bias in the model, thereby suggesting that it is more focused on memorizing the training data than on extracting meaningful patterns. Such a model, characterized by both high variance and bias, typically shows decreased accuracy and a limited ability to generalize effectively. The MLP forecaster learning curve exhibited favorable characteristics for two key reasons. Firstly, there was minimal bias and variance, as evidenced by the small difference between the errors observed during training and testing. Secondly, both the training/testing errors declined as the epochs grew. The MLP forecaster learning curve is illustrated in Figure 7. Initially, when the number of epochs was zero, the MAPE was high, thus indicating that the model was not yet well trained. However, as the number of epochs increased, the MAPE gradually decreased, thus eventually converging to a minimum acceptable value. This point of convergence, known as the saturation point, signifies that the model was effectively trained and had achieved satisfactory performance.



Figure 7. Developed model learning curve on testing and training sets with respect to the MAPE.

During the simulations, we utilized the control parameters listed in Table 2, and their rationale is documented in the literature [36]. These control parameters remained consistent for the FPP-MLP-GWDO and comparative models, thereby ensuring a fair comparative analysis.

$$MAPE = \left(\frac{1}{N}\sum_{t=1}^{N} \frac{|R_t - F_t|}{|R_t|}\right) \times 100,$$
 (17)

$$TIC = \frac{\sqrt{\frac{1}{N}\sum_{t=1}^{N} (R_t - F_t)^2}}{\sqrt{\frac{1}{N}\sum_{t=1}^{N} (R_t)^2} + \sqrt{\frac{1}{N}\sum_{t=1}^{N} (F_t)^2}}$$
(18)

The CC metric is defined in (18).

MAPE and TIC is presented in Equations (17) and (18), respectively.

$$CC = \frac{\mathrm{E}[(R_t - \mu_a) \times (F_t - \mu_F)]}{\sqrt{\left(\sum (R_t - \mu_a)^2\right) \times \left(\sum (F_t - \mu_F)^2\right)}}$$
(19)

In Equation (17), R_t and F_t denote the actual and predicted load values, respectively, while μ_a and μ_F correspond to the mean values of the actual and predicted load, respectively. The accuracy is computed from the error using Equation (20).

$$A = 100 - MAPE(x). \tag{20}$$

where *A* represents the accuracy. The convergence speed is computed using the CT and CR, which are comprehensively presented below:

• The convergence speed encompasses two aspects: the CT and CR. The CT refers to the time it takes for the forecaster to return the predicted load pattern. On the other hand, the CR represents the rate at which the model converges to an iteration returning an optimal result, where the error no longer decreases significantly with increasing iterations. Forecasts with lower CT and CR values (requiring fewer epochs to converge) are considered faster, while higher CT and CR values indicate slower convergence. In this study, the CT is expressed in seconds, while the CR is in aspects of iterations.

Table 2. Forecaster implementation parameters.

Forecaster Parameters	Values
Number of epochs	110
Output layer	1
Number of output neurons	1
Hidden layer	2
Neurons in hidden layer	10
Learning rate	0.0019
Momentum	0.6
Initial weight	0.1
Initial bias	0
Max	0.9
Min	0.1
Feature selection threshold	0.5
Decision variables	2
Number of objectives	0
Population size	24
Delay of weight	0.002

A comprehensive analysis of the performance metrics for the FPP-MLP-GWDO model and existing models is presented below.

3.1. Accuracy Evaluation

The proposed model accuracy was evaluated for both the day- and week-ahead load forecasting below.

3.1.1. Day-Ahead Load Prediction

Figure 8 illustrates the comparison of the day-ahead load predictions for the Dayton Ohio grid between our proposed FPP-MLP-GWDO model and the existing models (the AR, FS-ANN, SVM-DEA, and FS-SVM-mEDE). The results clearly demonstrate the forecasting capability of the proposed model in accurately predicting the day-ahead load for the Dayton Ohio grid. It is evident that all the forecasters, including both the developed and comparative models, have the capability to capture the nonlinear historical load trends. The nonlinear trends capturing the capability of forecasters are due to the activation functions (ReLU, sigmoidal, Tanh, etc.) and learning algorithms (Levenberg Marquardt algorithm, MVARA, back-propagation, gradient descent, etc.).

In this work, the developed FPP-MLP-GWDO model leveraged the MVARA and ReLU to capture nonlinear load trends. In contrast, the comparative models—the FS-ANN, SVM-DEA, and FS-SVM-mEDE—utilized the Levenberg Marquardt algorithm and sigmoidal function to capture nonlinear load trends. The selection of the MVARA and ReLU activation function is based on their advantages, including fast convergence and the ability to address challenges such as overfitting, vanishing gradients, etc. Figure 8 clearly demonstrates that the developed FPP-MLP-GWDO model closely followed the actual pattern, thereby outperforming the benchmark models (the AR, FS-ANN, SVM-DEA, and FS-SVM-mEDE) in terms of load prediction accuracy. The MAPE for the FPP-MLP-GWDO model was recorded at 1.10%, while SVM-DEA achieved 2.5%, the FS-SVM-mEDE achieved 2.1%, the FS-ANN achieved 3.5%, and the AR achieved 4.3%. This comparison is clearly illustrated in Figure 8, thereby affirming the superior accuracy of the FPP-MLP-GWDO model.



Figure 8. The day-ahead load forecasting of the developed model using Dayton Ohio grid data.

3.1.2. Week-Ahead Load Prediction

Figure 9 displays the week-ahead load forecasting with the hour resolution for the proposed FPP-MLP-GWDO model and the existing benchmark frameworks (the AR, FS-ANN, SVM-DEA, and FS-SVM-mEDE) on the Dayton Ohio grid. The FPP-MLP-GWDO model stood out with its accurate, fast, and stable load prediction, thus surpassing the

performance outcomes of the comparative models. Notably, the FPP-MLP-GWDO returned load profile closely aligned with the target load, as clearly depicted in Figure 9.

The proposed hybrid model, FPP-MLP-GWDO, achieved an impressive MAPE accuracy of 1.12%. In contrast, the comparative models such as the AR, FS-ANN, FS-SVM-mEDE, and SVM-DEA exhibited MAPE values of 4.6%, 3.5%, 2.1%, and 2.5%, respectively. The superior performance of the FPP-MLP-GWDO model can be attributed to its unique combination of the MLP with the MVARA, ReLU, and GWDO optimizer. The load forecasting curve generated by the proposed model closely aligned with the target curve, thus further confirming its superior performance compared to the benchmark models.



Figure 9. The week-ahead load forecasting of the developed model using Dayton Ohio grid data.

The developed FPP-MLP-GWDO and the comparative models: FS-ANN, SVM-DEA, and FS-SVM-mEDE were evaluated regarding the commutative distribution function (CDF) of errors, as shown in Figure 10.

The findings reveal that the FPP-MLP-GWDO was superior to the comparative models in terms of the CDF. The utilization of the MLP, with its deep layers designed to capture essential features, enabled reliable prediction, even in highly uncertain situations. Therefore, the proposed model presents an optimal choice for distribution system operators aiming to enhance the efficiency of the SPG.

The statistical evaluation of the accuracy is presented in Figure 11. The MAPE serves as a metric for quantifying the variance between the predicted and the real values. A lower MAPE indicates higher accuracy, while a larger MAPE indicates poorer accuracy. The accuracy analysis, in terms of the MAPE, is depicted in Figure 11. The MAPE values for the proposed model, FS-ANN, SVM-DEA, FS-SVM-mEDE, were 1.10%, 3.5%, 2.5%, and 2.1%, respectively.

The performance evaluations and discussions presented above conclude that the SVM-DEA is superior to the FS-ANN in terms of the MAPE. The improved accuracy of the SVM-DEA model can be attributed to the integration of the DEA optimizer with the SVM forecaster. These modifications devised in the developed model contributed to enhanced accuracy. Additionally, the developed FPP-MLP-GWDO model exhibited better performance in terms of the accuracy than the comparative models (the AR, FS-ANN, SVM-DEA, and FS-SVM-mEDE) illustrated in Figure 11. However, it is noteworthy that the integration of the DEA optimizer led to an increase in the CT.



Figure 10. Evaluation of FPP-MLP-GWDO and comparative models in terms of the CDF using the MAPE.



Figure 11. Proposed model evaluation in terms of the MAPE using Dayton Ohio grid data.

3.2. Convergence Speed Evaluation

The developed model and the comparative models' convergence speed is evaluated using two aspects: the CT and CR, which are comprehensively discussed in the subsequent section.

3.2.1. Convergence Speed in terms of the CT

First, the convergence speed evaluation in terms of the CT is presented in Figure 12. The findings reveal that the FPP-MLP-GWDO had a CT of 299 s. In contrast, the AR, FS-ANN, SVM-DEA, and FS-SVM-mEDE had CT times of 132 s, 159 s, 240 s, and 350 s, respectively. The results demonstrate that the CT increased from 159 s to 240 s when the optimizer was integrated. The intrinsic forecaster models without an optimizer and feature selector had low CT times and vice versa. Thus, adding a preprocessing/feature selector technique or optimizer with an intelligent forecaster increases the CT. The proposed model had a low CT compared to similar nature hybrid forecasting model (where both

the FS and mEDE optimizer were integrated with the forecaster). On the other hand, the AR, FS-ANN, and SVM-DEA models had 132 s, 159 s and 240 s CT times, respectively, which were lower than the proposed model, because, with an ANN, only the feature selector is integrated, and no optimizer is included; with the SVM, only the DEA optimizer is added, and the feature selector is integrated. The findings are illustrated in Figure 12. The proposed FPP-MLP-GWDO model reduced the CT compared to the comparative models for several reasons: the use of the fast-converging GWDO optimizer [34] instead of the EDE or mEDE optimizer [38,46,47], the utilization of the MVARA and ReLU in lieu of the sigmoid function, the adoption of the MLP, which is superior to the ANN, and the introduction of a novel concept for the feature interaction in the FPP for feature selection in addition to filters (irrelevancy and redundancy). In contrast, the comparative models only employ the MI filters approach (irrelevancy and redundancy) or the DE optimizer, which are computationally expensive. It noteworthy that the developed model needs more CT than the FS-ANN. This notable difference in results is due to the absence of an optimizer in the FS-ANN model (see Figure 12). Thus, the discussion concludes that a tradeoff exists between the accuracy and the convergence speed.



Figure 12. Developed model evaluation in comparison with existing models in terms of computational time using Dayton Ohio grid data.

3.2.2. Convergence Speed in Terms of the CR

The performance analysis regarding the convergence speed in terms of the CR for 100 iterations is presented in Figure 13. The figure demonstrates the fast convergence and effective search capability of the FPP-MLP-GWDO compared to the FS-ANN, SVM-DEA, and FS-SVM-mEDE models. Figure 13 illustrates the MAPE of the FPP-MLP-GWDO model and the comparative models (AR, FS-ANN, SVM-DEA, and FS-SVM-mEDE) over more than 110 iterations. The MAPE decreased as the number of iterations increased, which was observed for both the FPP-MLP-GWDO model and the comparative models. Nevertheless, it is noteworthy that the proposed model demonstrated its rapid convergence and effective search capability by reaching convergence approximately by the 18th iteration. On the other hand, the comparative models such as the AR, FS-ANN, SVM-DEA, and FS-SVM-mEDE converged around the 55th, 39th, 35th, and 31th iterations, respectively. This analysis demonstrates that the GWDO is more suitable as an optimizer in hybrid models.



Figure 13. Proposed model evaluation in comparison with existing models' convergence speed values in terns of the CR.

An overall evaluation of the FPP-MLP-GWDO, AR, FS-ANN, SVM-DEA, and FS-SVMmEDE models is summarized in Table 3. This evaluation encompasses various aspects, including the computational complexity, CT, CR, and accuracy.

The simulations, performance analysis, and discussions mentioned above conclude that the hybrid model FPP-MLP-GWDO demonstrates superior performance compared to benchmark models such as the FS-SVM-mEDE, SVM-DEA, and FS-ANN in aspects of the accuracy, CR, CT, complexity, etc.

Table 3. Evaluating the complexity, CT, CR, and accuracy of the suggested model and existing models such as AR, FS-ANN, SVM-DEA, and FS-SVM-mEDE.

Motrico	Load Forecasters						
wietrics	AR	FS-ANN	SVM-DEA	FS-SVM-mEDE	FPP-MLP-GWDO		
Complexity (level)	Low	Low	Moderate	High	Moderate		
CR (epochs)	55th	39th	35th	31th	18th		
CT (s)	132	159	240	350	299		
Accuracy (%)	95.7	96.5	97.5	97.9	98.999		

4. Conclusions

Load forecasting is imperative for decision-making processes in the SPG, thus enabling the efficient utilization of available generation, operational planning, load scheduling, and the assessment of contracts. To address these needs, a novel model called FPP-MLP-GWDO was introduced, where FPP, the MLP forecaster, and the GWDO optimizer cascaded with the aim to achieve accurate load prediction while maintaining an affordable convergence speed. In the FPP-MLP-GWDO model, an innovative approach to feature interaction, in addition to filters (irrelevancy and redundancy), were used in the FPP to find favorable features for the MLP forecaster. Considering the nonlinear and intricate nature of the problem at hand, the GWDO was used as the optimizer to optimize the forecasting results obtained from the MLP forecaster, thereby enhancing accuracy while ensuring reasonable convergence. To asses the developed model performance, experiments were conducted using the Dayton Ohio grid dataset, therein employing metrics such as accuracy (MAPE, TIC, CC) and convergence speed (CT and CR). The findings confirm that the developed FPP-MLP-GWDO model achieved an accuracy of 98.9%, thus surpassing benchmark models such as the AR (95.7%), FS-ANN (96.5%), FS-SVM-mEDE (97.9%), and SVM-DEA (97.5%). Furthermore, the FPP-MLP-GWDO model significantly reduced the CT (299s) compared to the FS-SVM-mEDE (350s), SVM-DEA (240s), FS-ANN (159s), and AR (132s) models. The findings concluded that the FPP-MLP-GWDO model outperformed the comparative models for load forecasting in aspects of accuracy and convergence speed. In future endeavors, there is potential to expand this research for DSM applications in smart cities using the IoTs with data analytics. Additionally, the next promising direction for future research involves incorporating sensors into intelligent models to enhance data analytic applications for energy optimization.

Author Contributions: H.A.: Visualization, Investigation, Writing—review & editing, Formal analysis, Funding acquisition; G.H.: Project administration, Conceptualization, Software, Supervison, Data curation, Methodology, Writing—original draft, Resources, Visualization, Writing—review & editing, Formal analysis, Funding acquisition, Investigation; S.A.: Conceptualization, Data curation, Methodology, Writing—original draft, Resources, Visualization, Investigation, validation; S.U.: Conceptualization, Data curation, Methodology, Writing—original draft, Resources; M.I.K.: Visualization, Investigation, Writing—review & editing, Formal analysis, Funding acquisition; S.M.: Writing—review & editing, Formal analysis, Funding acquisition; L.-G.H.: Visualization, Investigation, Writing—review & editing, Formal analysis, Funding acquisition. All authors have read and agreed to the published version of the manuscript.

Funding: The authors are thankful to the Deanship of Scientific Research at Najran University for funding this work under the Research Priorities and Najran Research funding program grant code (NU/NRP/SERC/12/16).

Data Availability Statement: Data will be made available on the request.

Acknowledgments: The authors are thankful to the Deanship of Scientific Research at Najran University for funding this work under the Research Priorities and Najran Research funding program grant code (NU/NRP/SERC/12/16).

Conflicts of Interest: The authors declare no conflict of interest.

References

- 1. Hafeez, G.; Khan, I.; Jan, S.; Shah, I.A.; Khan, F.A.; Derhab, A. A novel hybrid load forecasting framework with intelligent feature engineering and optimization algorithm in smart grid. *Appl. Energy* **2021**, *299*, 117178. [CrossRef]
- Hashmi, M.H.; Ullah, Z.; Asghar, R.; Shaker, B.; Tariq, M.; Saleem, H. An Overview of the current challenges and Issues in Smart Grid Technologies. In Proceedings of the 2023 International Conference on Emerging Power Technologies (ICEPT), Topi, Pakistan, 6–7 May 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 1–6.
- Asghar, R.; Sulaiman, M.H.; Saeed, S.; Wadood, H.; Mehmand, T.K.; Ullah, Z. Application of linear and nonlinear control schemes for the stability of Smart Grid. In Proceedings of the 2022 International Conference on Emerging Technologies in Electronics, Computing and Communication (ICETECC), Jamshoro, Sindh, Pakistan, 7–9 December 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 1–6.
- 4. Mohan, N.; Soman, K.P.; Kumar, S.S. A data-driven strategy for short-term electric load forecasting using dynamic mode decomposition model. *Appl. Energy* **2018**, 232, 229–244. [CrossRef]
- 5. Glavan, M.; Gradišar, D.; Moscariello, S.; Juričić, Đ.; Vrančić, D. Demand-side improvement of short-term load forecasting using a proactive load management—A supermarket use case. *Energy Build.* **2019**, *186*, 186–194. [CrossRef]
- 6. Billah, B.; King, M.L.; Snyder, R.D.; Koehler, A.B. Exponential smoothing model selection for forecasting. *Int. J. Forecast.* 2006, 22, 239–247. [CrossRef]
- Rendon-Sanchez, J.F.; de Menezes, L.M. Structural combination of seasonal exponential smoothing forecasts applied to load forecasting. *Eur. J. Oper. Res.* 2019, 275, 916–924. [CrossRef]
- 8. Ribeiro, M.I. Kalman and extended kalman filters: Concept, derivation and properties. Inst. Syst. Robot. 2004, 43, 3736–3741.
- 9. Song, K.-B.; Baek, Y.; Hong, D.H.; Jang, G. Short-term load forecasting for the holidays using fuzzy linear regression method. *IEEE Trans. Power Syst.* 2005, 20, 96–101. [CrossRef]
- 10. Wi, Y.-M.; Joo, S.-K.; Song, K.B. Holiday load forecasting using fuzzy polynomial regression with weather feature selection and adjustment. *IEEE Trans. Power Syst.* **2011**, *27*, 596–603. [CrossRef]
- 11. Zhao, H.; Guo, S. An optimized grey model for annual power load forecasting. Energy 2016, 107, 272–286. [CrossRef]
- 12. Huang, S.-J.; Shih, K.R. Short-term load forecasting via ARMA model identification including non-Gaussian process considerations. *IEEE Trans. Power Syst.* 2003, *18*, 673–679. [CrossRef]

- de Andrade, L.C.M.; da Silva, I.N. Very short-term load forecasting based on ARIMA model and intelligent systems. In Proceedings of the 2009 15th International Conference on Intelligent System Applications to Power Systems, Curitiba, Brazil, 8–12 November 2009; IEEE: Piscataway, NJ, USA, 2009; pp. 1–6.
- 14. Bacha, S.A.; Ahmad, G.; Hafeez, G.; Albogamy, F.R.; Murawwat, S. Compensation of Data Loss Using ARMAX Model in State Estimation for Control and Communication Systems Applications. *Energies* **2021**, *14*, 7573. [CrossRef]
- Ma, T.; Wang, F.; Wang, J.; Yao, Y.; Chen, X. A combined model based on seasonal autoregressive integrated moving average and modified particle swarm optimization algorithm for electrical load forecasting. *J. Intell. Fuzzy Syst.* 2017, 32, 3447–3459. [CrossRef]
- Islam, B.U. Comparison of conventional and modern load forecasting techniques based on artificial intelligence and expert systems. Int. J. Comput. Sci. Issues (IJCSI) 2011, 8, 504.
- 17. Gontar, Z.; Hatziargyriou, N. Short term load forecasting with radial basis function network. In Proceedings of the 2001 IEEE Porto Power Tech Proceedings (Cat. No. 01EX502), Porto, Portugal, 10–13 September 2001; IEEE: Piscataway, NJ, USA, 2001; Volume 3.
- 18. Salkuti, S.R. Short-term electrical load forecasting using radial basis function neural networks considering weather factors. *Electr. Eng.* **2018**, *100*, 1985–1995. [CrossRef]
- Chung, W.H.; Gu, Y.H.; Yoo, S.J. District heater load forecasting based on machine learning and parallel CNN-LSTM attention. Energy 2022, 246, 123350. [CrossRef]
- 20. Zambrano-Asanza, S.; Morales, R.E.; Montalvan, J.A.; Franco, J.F. Integrating artificial neural networks and cellular automata model for spatial-temporal load forecasting. *Int. J. Electr. Power Energy Syst.* 2023, 148, 108906. [CrossRef]
- 21. Ullah, S.; Khan, Q.; Mehmood, A.; Kirmani, S.A.M.; Mechali, O. Neuro-adaptive fast integral terminal sliding mode control design with variable gain robust exact differentiator for under-actuated quadcopter UAV. *ISA Trans.* **2022**, *120*, 293–304. [CrossRef]
- 22. Ullah, S.; Khan, Q.; Mehmood, A. Neuro-adaptive fixed-time non-singular fast terminal sliding mode control design for a class of under-actuated nonlinear systems. *Int. J. Control* 2023, *96*, 1529–1542. [CrossRef]
- 23. Yan, Z.; Zhu, X.; Wang, X.; Ye, Z.; Guo, F.; Xie, L.; Zhang, G. A multi-energy load prediction of a building using the multi-layer perceptron neural network method with different optimization algorithms. *Energy Explor. Exploit.* 2023, 41, 273–305. [CrossRef]
- Yazici, I.; Beyca, O.F.; Delen, D. Deep-learning-based short-term electricity load forecasting: A real case application. *Eng. Appl. Artif. Intell.* 2022, 109, 104645. [CrossRef]
- 25. Sekhar, C.; Dahiya, R. Robust framework based on hybrid deep learning approach for short term load forecasting of building electricity demand. *Energy* **2023**, *268*, 126660. [CrossRef]
- 26. Liang, Y.; Niu, D.; Hong, W.-C. Short term load forecasting based on feature extraction and improved general regression neural network model. *Energy* **2019**, *166*, 653–663. [CrossRef]
- 27. Hu, R.; Wen, S.; Zeng, Z.; Huang, T. A short-term power load forecasting model based on the generalized regression neural network with decreasing step fruit fly optimization algorithm. *Neurocomputing* **2017**, 221, 24–31. [CrossRef]
- 28. Li, H.-Z.; Guo, S.; Li, C.-J.; Sun, J.-Q. A hybrid annual power load forecasting model based on generalized regression neural network with fruit fly optimization algorithm. *Knowl.-Based Syst.* **2013**, *37*, 378–387. [CrossRef]
- 29. Awan, S.M.; Aslam, M.; Khan, Z.A.; Saeed, H. An efficient model based on artificial bee colony optimization algorithm with Neural Networks for electric load forecasting. *Neural Comput. Appl.* **2014**, 25, 1967–1978. [CrossRef]
- 30. Dai, Y.; Zhao, P. A hybrid load forecasting model based on support vector machine with intelligent methods for feature selection and parameter optimization. *Appl. Energy* **2020**, *279*, 115332. [CrossRef]
- 31. Massana, J.; Pous, C.; Burgas, L.; Melendez, J.; Colomer, J. Identifying services for short-term load forecasting using data driven models in a Smart City platform. *Sustain. Cities Soc.* **2017**, *28*, 108–117. [CrossRef]
- 32. Saoud, L.S.; Al-Marzouqi, H. Metacognitive sedenion-valued neural network and its learning algorithm. *IEEE Access* 2020, *8*, 144823–144838. [CrossRef]
- 33. Saoud, L.S.; Al-Marzouqi, H.; Deriche, M. Wind speed forecasting using the stationary wavelet transform and quaternion adaptive-gradient methods. *IEEE Access* 2021, *9*, 127356–127367. [CrossRef]
- 34. Nawaz, A.; Hafeez, G.; Khan, I.; Jan, K.U.; Li, H.; Khan, S.A.; Wadud, Z. An intelligent integrated approach for efficient demand side management with forecaster and advanced metering infrastructure frameworks in smart grid. *IEEE Access* 2020, *8*, 132551–132581. [CrossRef]
- 35. Bayraktar, Z.; Komurcu, M.; Bossard, J.A.; Werner, D.H. The wind driven optimization technique and its application in electromagnetics. *IEEE Trans. Antennas Propag.* 2013, *61*, 2745–2757. [CrossRef]
- Hafeez, G.; Alimgeer, K.S.; Wadud, Z.; Shafiq, Z.; Khan, M.U.A.; Khan, I.; Khan, F.A.; Derhab, A. A novel accurate and fast converging deep learning-based model for electrical energy consumption forecasting in a smart grid. *Energies* 2020, 13, 2244. [CrossRef]
- 37. Hafeez, G.; Alimgeer, K.S.; Khan, I. Electric load forecasting based on deep learning and optimized by heuristic algorithm in smart grid. *Appl. Energy* **2020**, *269*, 114915. [CrossRef]
- Hafeez, G.; Alimgeer, K.S.; Qazi, A.B.; Khan, I.; Usman, M.; Khan, F.A.; Wadud, Z. A hybrid approach for energy consumption forecasting with a new feature engineering and optimization framework in smart grid. *IEEE Access* 2020, *8*, 96210–96226. [CrossRef]

- 39. Wang, J.; Li, L.; Niu, D.; Tan, Z. An annual load forecasting model based on support vector regression with differential evolution algorithm. *Appl. Energy* **2012**, *94*, 65–70. [CrossRef]
- 40. Chandrashekar, G.; Sahin, F. A survey on feature selection methods. Comput. Electr. Eng. 2014, 40, 16–28. [CrossRef]
- 41. Amjady, N.; Keynia, F. Day-ahead price forecasting of electricity markets by mutual information technique and cascaded neuro-evolutionary algorithm. *IEEE Trans. Power Syst.* **2008**, *24*, 306–318. [CrossRef]
- Amjady, N.; Keynia, F.; Zareipour, H. Short-term load forecast of microgrids by a new bilevel prediction strategy. *IEEE Trans.* Smart Grid 2010, 1, 286–294. [CrossRef]
- Zhao, C.; Zheng, C.; Zhao, M.; Tu, Y.; Liu, J. Multivariate autoregressive models and kernel learning algorithms for classifying driving mental fatigue based on electroencephalographic. *Expert Syst. Appl.* 2011, 38, 1859–1865. [CrossRef]
- Rehman, M.Z.; Nawi, N.M. Improving the accuracy of gradient descent back propagation algorithm (GDAM) on classification problems. *Int. J. New Comput. Archit. Their Appl.* 2011, 1, 838–847.
- 45. Pavlyuk, D. Short-term traffic forecasting using multivariate autoregressive models. Procedia Eng. 2017, 178, 57–66. [CrossRef]
- Schmidt, M.; Safarani, S.; Gastinger, J.; Jacobs, T.; Nicolas, S.; Schülke, A. On the performance of differential evolution for hyperparameter tuning. In Proceedings of the 2019 international joint conference on neural networks (IJCNN), Budapest, Hungary, 14–19 July 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1–8.
- 47. Albogamy, F.R.; Khan, S.A.; Hafeez, G.; Murawwat, S.; Khan, S.; Haider, S.I.; Basit, A.; Thoben, K.-D. Real-time energy management and load scheduling with renewable energy integration in smart grid. *Sustainability* **1792**, *14*, 1792. [CrossRef]
- 48. Bao, Z.; Zhou, Y.; Li, L.; Ma, M. A hybrid global optimization algorithm based on wind driven optimization and differential evolution. *Math. Probl. Eng.* 2015, 2015, 620–635. [CrossRef]
- 49. Mirjalili, S.; Mirjalili, S. Genetic algorithm. Evol. Algorithms Neural Netw. Theory Appl. 2019, 780, 43–55.
- 50. PJM Electricity Market. Available online: https://www.pjm.com/ (accessed on 23 April 2022).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.