

Received February 21, 2020, accepted March 14, 2020, date of publication April 6, 2020, date of current version June 4, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2985732

A Hybrid Approach for Energy Consumption Forecasting With a New Feature Engineering and Optimization Framework in Smart Grid

GHULAM HAFEEZ^{1,2}, **KHURRAM SALEEM ALIMGEER**¹,
ABDUL BASEER QAZI³, (Member, IEEE), **IMRAN KHAN**², (Senior Member, IEEE),
MUHAMMAD USMAN⁴, (Senior Member, IEEE),
FARRUKH ASLAM KHAN⁵, (Senior Member, IEEE), AND **ZAHID WADUD**⁶

¹Department of Electrical and Computer Engineering, COMSATS University Islamabad, Islamabad 44000, Pakistan

²Department of Electrical Engineering, University of Engineering and Technology, Mardan 23200, Pakistan

³Department of Software Engineering, Bahria University, Islamabad 44000, Pakistan

⁴Department of Software Engineering, University of Engineering and Technology, Mardan 23200, Pakistan

⁵Center of Excellence in Information Assurance, King Saud University, Riyadh 11653, Saudi Arabia

⁶Department of Computer Systems Engineering, University of Engineering and Technology at Peshawar, Peshawar 25000, Pakistan

Corresponding author: Farrukh Aslam Khan (fakhan@ksu.edu.sa)

This work was supported by the Deanship of Scientific Research at King Saud University, Saudi Arabia, through the Research Group Project No. RGP-214.

ABSTRACT Electric energy consumption forecasting enables distribution system operators to perform efficient energy management by flexibly engaging energy consumers under the intelligent demand-response program in the smart grid (SG). With this motivation, in this paper, a fast and accurate hybrid electrical energy forecasting (FA-HELFF) framework is developed. The proposed framework integrates two modules with support vector machine (SVM) based forecaster. These modules are data pre-processing and feature engineering, and modified enhanced differential evolution (mEDE) based optimizer. First, feature selection algorithms like random forests and relief-F are combined to devise a hybrid feature selection algorithm to alleviate redundancy. Secondly, for feature extraction, a radial basis Kernel-based principal component analysis algorithm is employed to eliminate the dimensionality reduction problem. Finally, to conduct accurate and fast electrical energy consumption forecasting, the mEDE based optimizer is integrated with the SVM based forecaster. The resulting FA-HELFF framework is tested on publicly available independent system operator New England (ISO-NE) control area hourly load data. The results demonstrate that the FA-HELFF framework is robust and shows significant improvements when compared to other benchmark frameworks in terms of accuracy and convergence speed.

INDEX TERMS Electrical energy consumption forecasting, energy management, smart grid, grey correlation analysis, differential evolution, radial basis kernel-based principal component analysis, support vector machine.

I. INTRODUCTION

The electrical power system is a network that includes two types of agents: power agents (generation, transmission, and distribution) and consumer agents, where both the agents attempt to maximize benefits and minimize expenses. The emergence of advanced metering infrastructure (AMI) has renovated the legacy power grid with smart grids

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

(SGs) [1]–[3]. This renovation has created a platform for both agents to forecast electrical energy consumption for decision making in the SG. Thus, electrical energy consumption forecasting is an indispensable task in planning, operation, and energy management of the SG. Generically, electrical energy consumption forecasting has four types based on time horizon: (i) very short-term forecasting, (ii) short-term forecasting, (iii) medium-term forecasting, and (iv) long-term forecasting. The forecast time horizon for these classes ranges from one minute up to one hour, more than an hour up to a

week, more than a week up to a year, and more than a year, respectively.

The first three-term electrical energy consumption predictions are favorable for the operation and energy management of the SG. However, long-term forecasting is beneficial for the infrastructure planning and development of the SG. Thus, it is essential to acquire accurate load forecasting because incorrect forecasting may result in improper operation, planning, and energy management, which leads to imprudent reservoirs, excessive generators, and additional operating costs. Moreover, optimistic electrical energy consumption forecast may cause risk to the power grid or lead to unnecessary energy purchases from operators who have forecasted energy consumption accurately by paying high costs. A few recent papers have highlighted the problem of electrical energy consumption forecasting in [4]–[6]. However, some authors focused on net electrical energy consumption forecasting and very few authors focused on sub-metered electrical energy consumption forecasting. Therefore, electric energy consumption forecasting issues remain open for solutions. Realizing the importance of the problem, authors have developed a variety of electrical energy forecasting models during the past few years. Recently, a model based on enhanced deep neural network (EDNN) is developed to predict week and year ahead electrical energy consumption [7]. In [8], a framework based on a multi-task regressor is proposed to predict electrical energy consumption based on the recorded energy consumption data by the smart meters. Sideratos *et al.* [9] proposed a load prediction model based on a hybrid DNN (HDNN). The proposed HDNN based model utilizes key parameters of ANN and deep learning models to resolve the forecasting issues. Though these papers are a good start for studying electrical energy consumption forecasting models, the frameworks used in literature are condemned for issues like impotent learning, handcrafted features, inaccurate appraisal, insufficient guiding significance, and limited learning capacity.

Thus, a fast and accurate hybrid electrical energy consumption forecasting (FA-HELFF) framework is developed in this work to solve issues related to electrical energy consumption forecasting in real-time for efficient energy management. The proposed FA-HELFF framework is a hybrid model having three modules: (i) feature engineering, (ii) forecaster, and (iii) optimizer. The main features and contributions of this research are described as follows:

- 1) A non-linear integrated forecasting framework is designed that is capable of handling the complex real situation and is implementable in the energy management of the SG.
- 2) To develop a novel feature engineering framework, a hybrid feature selector technique is proposed by fusing two algorithms: relief-F and random forests techniques, to monitor and control the feature selection process. Then, a radial basis Kernel-based principal component analysis (KPCA) is proposed for

feature extraction to resolve the dimensionality reduction issue. Finally, the optimizer module based on modified enhanced differential evolution (mEDE) algorithm is integrated with the forecaster module based on SVM to improve the accuracy by optimizing the hyperparameters.

- 3) To develop an optimization framework, a novel algorithm namely mEDE is proposed by devising modifications in the EDE algorithm to enhance forecast accuracy and convergence speed for efficient energy management.
- 4) The proposed FA-HELFF framework is tested on hourly load data of the ISO New England (ISO-NE) control area. The experiments validate that the proposed FA-HELFF framework outperforms other existing frameworks.

The remainder of the paper is structured as follows: First, the recent and relevant background study of electric load forecasting is discussed in section II. In section III, the proposed FA-HELFF framework is introduced and explained. The simulation results and discussion are presented in section IV. Finally, in section V, the paper is concluded along with future research directions.

II. RELATED LITERATURE STUDY

In this section, the existing recent and relevant research on electrical energy consumption forecasting for efficient energy management in the SG is discussed. Various electrical energy consumption forecasting strategies like statistical methods, machine learning methods, time series models, dynamic systems, and hybrid forecasting models are presented. Some of them are explained below.

A hybrid model is presented in [10] for load forecasting, which employs the combination of extreme learning machine (ELM), wavelet transform (WT), and least square regression (LSR), to improve the forecast accuracy. An ensemble strategy based on WT is proposed to avoid the trivial process of utilization of the complementary information of wavelet parameters and wavelet parameter selection to forecast accuracy. This model is also extended to electricity price forecasting by altering the input variables. However, the criterion to select decomposition levels and mother wavelet is not precisely defined. Besides this, the objectives are achieved at the cost of increased computation time. In [11], a framework for time-series load forecasting based on random forests and ensemble of ELM is proposed. The training sample sets are generated through a bootstrap sampling technique for random forests-based ELM ensemble model to handle multi-regime time series forecasting. However, accurate, stable, and efficient performance is achieved at the cost of high execution time. Authors developed a subsampled support vector regression ensemble (SSVRE) framework in [12] to solve the prediction problem. The accuracy is improved by support vector regression (SVR) and the learning process is enhanced by a subset of small sized subsamples. The proposed

framework is validated on publicly available datasets of New South Wales and Jiangxi Province. However, some forecasters in the ensemble are more accurate and the final output should be the average of the single outputs. An integrated framework of wavelet neural network (WNN), empirical mode decomposition (EMD), autoregressive integrated moving average (ARIMA), and fruit fly optimization algorithm (FFOA) is developed in [13] to predict future load. However, the accurate load forecasting is influenced by social and natural parameters that make the prediction process cumbersome, time-consuming, and slow. An enhanced Gaussian process mixture (EGPM) is presented in [14] to solve the load forecasting problem in the most effective manner. The efficacy of the model is evaluated by comparing with benchmark schemes like Gaussian process mixture (GPM), radial basis function neural network, and SVM classifier, in terms of forecast accuracy. The simulation results indicate that EGPM outperforms existing methods in accuracy. However, the accuracy is enhanced at the cost of increased compilation time. In [15], a hybrid model of EMD, general regression neural network, and FFOA are proposed for load forecasting having minimal redundancy and maximum relevancy. However, it is cumbersome to forecast due to various influencing parameters that lead to high volatility and computational complexity. A load forecasting model is presented in [16] for efficient energy management. The hybrid model is an integrated framework of least square SVM, auto-correlation, and grey wolf optimization algorithm. However, efficient energy management is performed at the cost of high system complexity.

In [17], a hybrid short-term electric load forecasting framework based on ANN is proposed. The forecasting is achieved by training the multi-layer neural network using the genetic algorithm (GA), general regression network, and Elman neural network. Accurate results are achieved by validating the system model on real data of Victoria and New South Wales states of Australia. In [18], authors focused on features selection based on genetic strategy and forecasting on the ELM. The forecasting model is validated on the load dataset of the Australian electricity market operator (AEMO). The simulation results validated that the forecasting model reduced the error and improved forecast accuracy. However, the ELM degree of freedom reduces with the increase in network complexity that leads to the problem of model overfitting. The work in [19] is based on multi-scale modeling for accurate demand forecasting ranging from short-term to medium-term time horizon. The emphasis of the authors is more on data analysis. Moreover, both seasonal and non-seasonal cycles are detected using autoregressive and moving average methods. The proposed model's accuracy is assessed based on the Akaike-Bayesian criterion. The accuracy is, however, enhanced while the rate of convergence is compromised.

A deep learning approach is presented in [20] for the household load forecasting. The uncertainty in the household is learned using deep belief recurrent neural network (RNN)

model for accurate load forecasting. However, the increasing number of layers in the deep model, data, and volume diversity may sometimes lead to the problem of overfitting. In [21], a data-driven framework based on copula model and DNN is proposed for load forecasting. The data is pre-processed by Box-Cox transformation and the parameters are tuned using the copula model for accurate load forecasting. The proposed model is tested on the power load data of both Texas and Arkansas in the United States. The proposed model performs better than classical neural networks and ELM based models.

A load forecasting framework based on artificial intelligence (AI) and DNN is described in [22] to improve the electric load prediction process for the South African distribution network. The proposed model outperforms both optimally tuned ELM and adaptive neuro-fuzzy inference system (ANFIS) based models. However, the accuracy of the forecasting process for the distribution power system is increased by including the effect of temperature, which in turn results in slow convergence. In [23], an intelligent hybrid model based on Kalman filtering (KF), WNN, and ANN using clustering techniques is presented to forecast the day and week ahead load of the commercial sector of Egypt and Canada. However, the accuracy is achieved at the cost of increased model complexity.

It can be confidently concluded from the literature that tremendous progress has been made in the field of electric load forecasting for energy management. However, the existing methods are weak in processing large data and it is hard to tune control parameters, which results in high computational complexity and inability to quickly converge because the redundancy, irrelevancy, and dimensional reduction are not averted. Moreover, the aforementioned literature does not cater for both forecast accuracy and convergence rate simultaneously. To solve such problems, a fast and accurate framework is the need of the day. Thus, in [24], SVM and gradient descent algorithm-based framework is proposed. However, this framework introduces much computational complexity and is unable to converge. Some authors focus on feature selection algorithms and traditional classifier decision trees and artificial neural networks [25]. However, decision trees face the problem of overfitting, which means that a decision tree performs well in training but poorly in prediction, and the artificial neural network has limited generalization capability and has difficulty in controlling its convergence. In [26], the authors proposed hybrid feature selection, extraction and classification-based framework for load forecasting. However, this method has high system complexity and is unable to converge.

In this context, a novel hybrid forecasting FA-HELFF framework is designed in this study to forecast the convergence rate and accuracy simultaneously. The objective of this framework is to provide fast and accurate forecasting for efficient energy management to fulfill energy needs of the society. The recent and relevant work discussed in this section is summarized in Table 1.

TABLE 1. A summary of the recent and relevant work in terms of methodologies, proposed approach, forecasting horizon, datasets, objectives, and limitations.

Refs.	Methodologies	Time horizon	Datasets	Objectives	Limitations
[10]	ELM, WT, and LSR	Hour and day ahead	ISO-NE	Forecast accuracy improvement by alleviating overtraining and uncertainty problems	Difficult to select mother wavelet and number of decomposition levels, and increased computational complexity
[11]	Random forest and ensemble of ELM based strategy	One and multi step ahead	NASA	Accurate, stable, and efficient forecasting	Time-consuming, difficult parameters identification, and high execution time
[12]	SSVRE framework	Half hour ahead	Jiangxi Province, and New South Wales	Computational accuracy and efficiency improvement	Several types of subsampled models
[13]	EMD, ARIMA, WNN, and FFOA	Hours, day, week, and month ahead	AEMO	Forecast accuracy improvement	Forecasting process difficult, time-consuming, and slow
[14]	An EGPM based load forecasting	Week ahead	ISO-NE	Forecast accuracy improvement	Time-consuming and slow
[15]	EMD, general regression neural network, and FFOA	Day ahead	Langfang China	Forecast accuracy improvement	Parameter settings and statistical properties leads much computational complexity
[16]	DNN model based on FCRBM and CRBM	15 minutes, hour, and day ahead	Single residential customer	Building energy consumption estimation	High computational complexity
[17]	A multi-layer neural network using GA, Elman neural network, and general regression network Genetic strategy	Day, and two day ahead with half hour resolution	AEMO	Accuracy improvement for various power system applications	High complexity and time-consuming
[18]	Genetic strategy based feature selection and ELM based forecasting	Year ahead with half hour resolution	AEMO	Effective subset of features selection for accurate and robust load forecasting	Difficult acquisition and overfitting
[19]	Autoregressive and moving average methods base multi-scale model	Hour ahead	PJM market	Accuracy improvement for efficient energy management	Difficult to train and time-consuming
[20]	Pooling based deep belief RNN	Hour ahead load forecasting	IRISH	House load forecast accuracy improvement	Overfitting, time-consuming, and computational complexity
[21]	A data driven framework based on copula model and deep neural network	Hour and week ahead	PJM market	Forecast accuracy improvement	Difficult parameters identification
[22]	A hybrid framework of artificial intelligence (AI) and DNN	Hour and week ahead	South African distribution networks	Forecast accuracy improvement	Difficult to train and time-consuming
[23]	A hybrid intelligent model of KF, WNN, and ANN using clustering approach	Two days and week ahead	Commercial dataset of Egypt and Canada	MAPE reduction	Increased model complexity

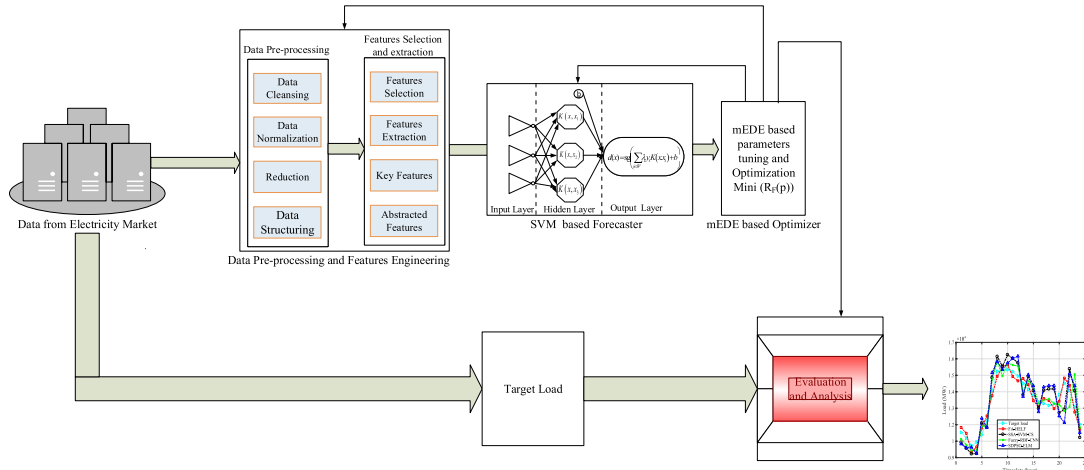


FIGURE 1. Proposed schematic framework for fast and accurate electrical energy consumption forecasting with new feature engineering and optimization modules.

III. PROPOSED HYBRID FRAMEWORK WITH NEW FEATURE ENGINEERING AND OPTIMIZATION MODULES

The proposed FA-HELFF framework has three modules: data pre-processing and feature engineering module, SVM based forecaster, and mEDE based optimizer, as depicted in Figure 1. This work is an extension of our earlier work published in a conference [27]. The feature engineering part is based on Grey correlation analysis (GCA), and radial basis kernel principal component analysis (KPCA). The forecaster module is based on SVM. The optimizer module is based on the proposed mEDE algorithm. The overall step by step procedure of the proposed framework is depicted in Figure 2, and a brief discussion is given below:

A. DATA PRE-PROCESSING AND FEATURE ENGINEERING MODULE

Data from the ISO-NE control area is fed to the data pre-processing and feature engineering module, which is composed of two phases: (i) pre-processing, and (ii) feature engineering, as shown in Figure 2. The brief description of (i) and (ii) is as follows:

1) DATA PRE-PROCESSING

The first phase is the data pre-processing. In the pre-processing phase, data cleansing action is conducted on the dataset to recover defected, erroneous, and missing data by taking the average of the previous day’s load. Then, the cleansed data is forwarded to the normalization phase to pre-prepare the data within limits of the activation function. After normalization, the normalized data is structured in descending order via a structuring action. Finally, the obtained results are normalized to move towards desired energy consumption forecasting. The prepared and cleansed data is fed into the feature engineering phase, as depicted in Figure 2.

2) FEATURE ENGINEERING

The second phase is feature engineering. In this phase, the abstractive and key features are selected and extracted from the prepared data, and redundant and irrelevant features are discarded, as depicted in Figure 2. The desired features from the dataset are selected and extracted through GCA and radial basis KPCA, respectively. To control feature selection, the feature selector is based on GCA, which combines relief-F and random forests algorithms to calculate the importance of the features. Furthermore, the feature selector decides based on feature importance whether to reserve or discard a feature. The radial basis KPCA based feature extractor uses the Kernel function to deal with high dimensional non-linear data because the PCA is not suitable for this dilemma. The purpose of feature extraction is to reduce redundant features. A brief demonstration of the feature engineering phase is given below:

1) Feature selector:

The feature selection process is based on GCA, which is developed by combining relief-F and random forests and is controlled by combined controlling threshold σ . The GCA roughly selects a feature space where the most relevant and desired features are kept and irrelevant features are discarded based on feature importance and feature selection controlling threshold σ .

The electric load data matrix is represented by D and is defined as follows:

$$D = \begin{bmatrix} d_{11} & d_{12} & d_{13} & \dots & d_{1n} \\ d_{21} & d_{22} & d_{23} & \dots & d_{23} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ d_{m1} & d_{m2} & d_{13} & \dots & d_{mn} \end{bmatrix} \quad (1)$$

The columns denote the feature index and rows represent the time-stamps. Furthermore, d_{mn} is the m th component of data, which is n th hour ahead of electrical

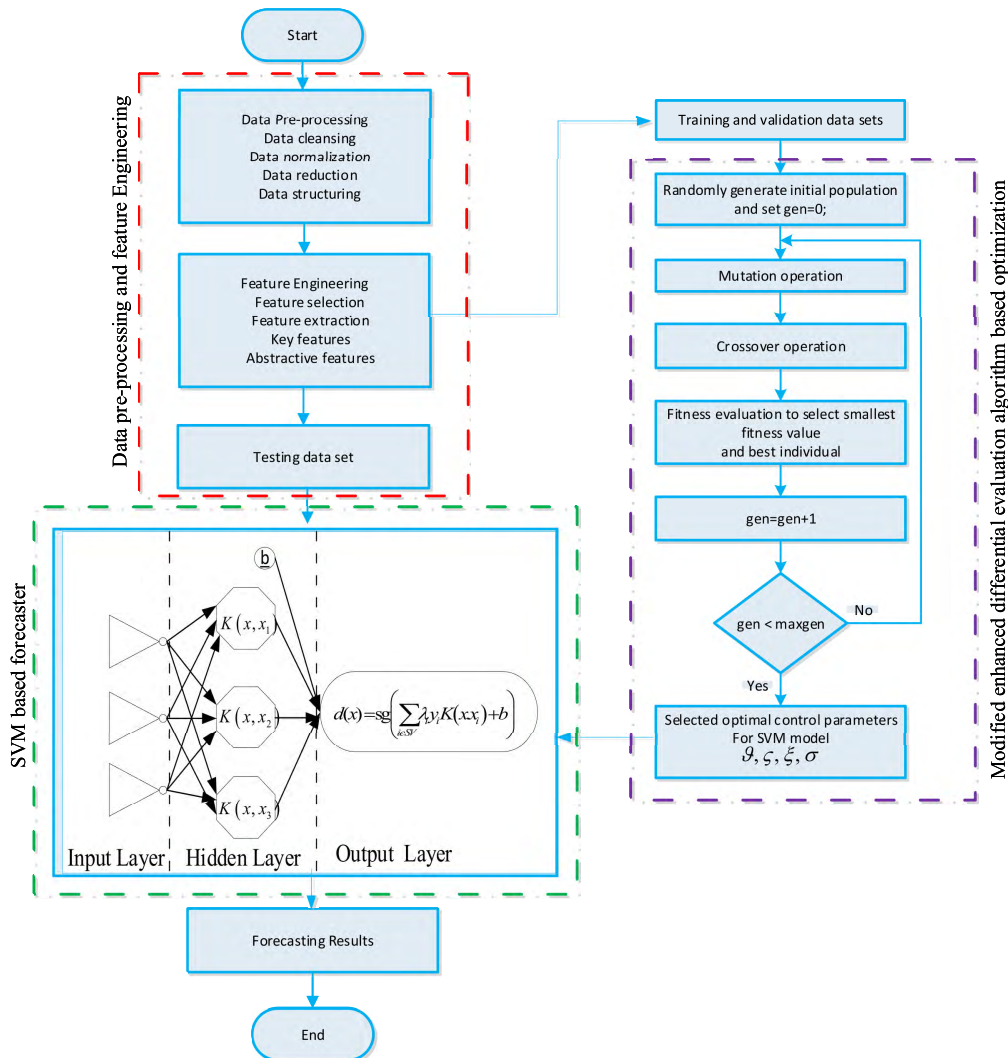


FIGURE 2. Step by step working flow chart of the proposed schematic framework for fast and accurate electrical energy consumption forecasting. Red dotted box shows data pre-processing and feature engineering module, green dotted box shows SVM based forecaster module, and the blue dotted box represents mEDE algorithm-based optimization module.

energy consumption pattern that is to be forecasted. Equation 1 can also be written in the time sequence form as:

$$D = \begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_n \end{bmatrix}, \quad (2)$$

where,

$$t_j = [d_{j1} \ d_{j2} \ d_{j3} \ \dots \ d_{jn}] \quad j \in [1, m]. \quad (3)$$

The different features have various degrees of impact on forecasted electrical energy consumption patterns, and the GCA calculates the importance of each feature and its influence on electrical energy consumption forecasting. The GCA determines the correlation between each feature and the final electrical energy

consumption pattern to effectively control the feature selection process. The GCA via correlation determines the closeness between data signals. Closer the two data signals, greater is the correlation and vice versa. Thus, GCA measures the closeness between two data signals. Since each feature has different physical meaning and different dimension in a framework, when the GCA is carried out, non-dimensional data is normalized either via their average value or the maximum value. The original data sequence is normalized as follows:

$$\chi_i^*(j) = \frac{\chi_i(j) - \min \chi_i(j)}{\max \chi_i(j) - \min \chi_i(j)} \quad (4)$$

where $\chi_0(j)$, $i \in [1, n]$ and $\chi_i(j) = \alpha_{ij}$, $i \in [1, m]$ are the two sequences, n is the count of features, and m is the time sequence length. The Grey correlation

coefficient, after normalization, is calculated as:

$$C_g(\chi_0^*(j), \chi_i^*(j)) = \frac{\Delta_{\min} + \rho \Delta_{\max}}{\Delta_{oi}(j) + \rho \Delta_{\max}}, \quad \rho \in (0, 1)$$

$$\Delta_{oi}(j) = |\chi_0^*(j) - \chi_i^*(j)|,$$

$$\Delta_{\max} = \max_{i,j} |\chi_0^*(j) - \chi_i^*(j)|,$$

$$\Delta_{\min} = \min_{i,j} |\chi_0^*(j) - \chi_i^*(j)|, \quad (5)$$

where ρ represents a distinguishing coefficient, which is set as 0.50 by [28]. The grade of Grey correlation is determined as follows:

$$G_i(\chi_0^*, \chi_i^*) = \frac{\Delta_{\min} + \min_{i,j} |\chi_0^*(j) - \chi_i^*(j)|}{\sum_{j=1}^m C_g(\chi_0^*(j), \chi_i^*(j)) + \min_{i,j} |\chi_0^*(j) - \chi_i^*(j)|} \quad (6)$$

The low correlated features are dropped and the remaining selected features are sorted in descending order and the time sequence becomes as follows:

$$t_i = [d_{j1} \ d_{j2} \ d_{j3} \ \dots \ d_{jn-\delta}], \quad (7)$$

where δ is the dropped features and t_i is the time sequence.

After GCA, the data sequence is passed through feature selectors. At this stage, the data sequence is processed by two evaluators: relief-F evaluator κ , and random forests evaluator γ , respectively. The purpose of evaluators is to evaluate the importance of each feature to select desired features. The feature evaluation and selection processes are shown in Algorithm 1. The first evaluator is the random forests evaluator that moves on boot-strap bagging samples [29]. The bagged samples are divided into training samples and out of bag samples. For the first evaluator γ , all weights are initialized to zero, and the training of random forests is initiated. Then, the feature's importance is determined by the out of bag data with noise. For the second evaluator κ , weights are updated on the concept of distance among near hits and misses. The two evaluators (γ and κ) forward the determined feature importance to the selector to perform feature selection based on the controlling threshold.

In Algorithm 1, $r[n]$ represents the random forests having n number of decision trees. The $\omega^F[\tau_j]$ and $\omega^r[\tau_j]$ represent feature importance that is calculated by relief-F and random forests, respectively. The parameters are updated as follows:

$$\omega^F[\tau_j] = \omega^F[\tau_j] - \frac{\sum_{i=1}^j \text{diff}(D, d^*, H_i)}{m * j} + \frac{\sum_{C \neq \text{class}(\chi)} \text{diff}(D, d^*, M_i(C))}{m * j}, \quad (8)$$

where in class C_i , d^* represents randomly selected item, and function $\text{diff}(D, r_1, r_2)$ calculates difference D between r_1 and r_2 . Mathematical definition of

the difference function is as follows:

$$\text{diff} = \begin{cases} 0 & \text{values are different} \\ 1 & \text{values are same} \\ \text{Difference between attributes is} & \\ \text{normalized to be within [0 1]} & \end{cases} \quad (9)$$

The features selection process is based on importance ω^F and ω^r , respectively. Both are normalized as follows:

$$\hat{\omega}^r = \frac{\omega^r}{\max(\omega^r)}$$

$$\hat{\omega}^F = \frac{\omega^F}{\max(\omega^F)} \quad (10)$$

The features that have combined importance value greater than σ are considered as key features, while those features that have combined importance value less than σ are considered as irrelevant features. The key features are kept and irrelevant features are discarded. This process is mathematically modelled by Equation 11.

$$\tau_j = \begin{cases} \text{keep} & \omega^r[\tau_j] + \omega^F[\tau_j] > \sigma \\ \text{discard} & \omega^r[\tau_j] + \omega^F[\tau_j] \leq \sigma \end{cases} \quad (11)$$

The selected features are sent to the feature extraction phase based on a radial basis KPCA to remove redundancy among the selected features.

2) Feature extractor:

In the second phase, the feature extraction operation is performed, which is based on radial basis KPCA. The purpose of this operation is to remove the redundant information to resolve dimensionality reduction issue. The output of the feature selector is given as an input to the radial basis KPCA based feature extractor in order to generate dimensionally reduced matrix having desired and most relevant features, which can be modelled as:

$$S = (s_1, s_2, s_3, \dots, s_N)^T, \quad (12)$$

where s_i is the i^{th} variable related to the electric load. The correlation of features and eigenvalues is computed as:

$$\lambda_e = V^{f*} e, \quad \lambda \geq 0 \ \& \ e \in f^*, \quad (13)$$

where λ represents the eigenvalue, V denotes covariance matrix of S , and f^* denotes the feature space. Moreover, $V^{f*} e$ is calculated through Equation 13:

$$V^{f*} e = \frac{1}{N} \sum_{i=1}^N \langle \varphi(s_i), e \rangle \varphi(s_i),$$

$$\sum_{k=1}^N \varphi(s_k) = 0, \quad (14)$$

where φ denotes input data mapping and feature space, and $\langle s, y \rangle$ represents the product of s and y . Equation 13

becomes Equation 15 by devising the above modifications:

$$\lambda \langle \varphi(s_k), e \rangle = \langle \varphi(s_k), V^{f*} e \rangle, \quad (15)$$

where e for $\lambda = 0$ can be determined as:

$$e = \sum_{i=1}^N \beta_i \varphi(s_i), \quad (16)$$

where β_i represents coefficients that correspond to s_i . Now, the Kernel function defined in [30] is used as:

$$K_{ij} = \langle \varphi(s_i), \varphi(s_j) \rangle \quad \forall i, j \in [1, N] \quad (17)$$

Equations 15 and 16 are combined, and the combined form is defined as follows:

$$\lambda \sum_{i=1}^N \beta_i K_i = \frac{1}{N} \sum_{i=1}^N \beta_i \sum_{j=1}^N K_{ki} K_{ij}, \quad (18)$$

where β_i represents coefficients that correspond to s_i and $\beta = [\beta_1, \beta_2, \dots, \beta_N]^T$, then Equation 15 can be rewritten as:

$$\lambda NK\beta = K^2\beta \quad (19)$$

The eigenvectors β and λ are selected to perform dimensionality reduction via normalization. Therefore, we have:

$$\langle e_i, e_j \rangle = 1 \quad \forall i, j \in [1, N]. \quad (20)$$

The resultant Equation 21 can be obtained by substituting Equation 16 in Equation 20, which is as follows:

$$\begin{aligned} & \left\langle \sum_{i=1}^N \beta_i^k \varphi(s_i), \sum_{j=1}^N \beta_j^k \varphi(s_j) \right\rangle \\ &= 1 \\ &= \sum_{i=1}^N \sum_{j=1}^N \beta_i^k \beta_j^k \langle \varphi(s_i), \varphi(s_j) \rangle \\ &= \sum_{i=1}^N \sum_{j=1}^N \beta_i^k \beta_j^k K_{ij} \\ &= \langle \beta_k, K\beta_k \rangle \\ &= \lambda_k \langle \beta_k, \beta_k \rangle \end{aligned} \quad (21)$$

The principal component extraction can be determined in the following manner:

$$p_k = \langle e_k, \varphi(s) \rangle = \sum_{i=1}^N \beta_i^k \langle \varphi(s_i), \varphi(s) \rangle, \quad (22)$$

where p shows the principal component. The Kernel functions have the following generic forms:

Linear form : $K(s, y) = \langle s, y \rangle$

Activation basis form : $K(s, y) = \tanh(\alpha_0 \langle s, y \rangle^d + \alpha_1)$

Radial basis form :

$$K(s, y) = \exp(-\theta \|s - y\|^2) \quad (23)$$

The feature extraction phase is shown in Algorithm 1. After the feature engineering phase, the selected and extracted features matrix is fed as an input to the forecaster module based on SVM to forecast electrical energy consumption pattern.

B. SVM BASED FORECASTER

The data is cleaned after the data pre-processing and feature engineering phases, and has no redundant and irrelevant features anymore. This module achieves desired electrical energy consumption forecasting through the cleaned and processed data. Various machine learning techniques exist in the literature for electrical energy consumption forecasting. The SVM, among the machine learning models, is chosen as a forecaster for electrical energy consumption forecasting due to its robust and efficient performance to produce accurate results with less computational time. In this section, we formulate and investigate the classification problem. The SVM based forecaster is depicted in Figure 2 and the detailed description of the problem is as follows:

1) PROBLEM FORMULATION

The classification problem is mathematically modeled as:

$$f(s, p) = \sum_{i=1}^D p_i \chi_i(s) + \partial, \quad (24)$$

where $p_i^\infty (i = 1, 2, 3, \dots)$ are the parameters of the forecaster to be determined, D is the dimensional space, and ∂ depends on data distribution and parameters of the classifier. The objective of SVM is to define a hyperplane in D -dimensional feature space that differentiates the data points. In this work, the hyperplane is defined in Equation 24. After that, the regularized risk function is defined as:

$$R_F(p) = \frac{\sum_{i=1}^D |L_i^a - f(s, p)|_\zeta + \sigma p^2}{D}, \quad (25)$$

where σ represents feature selection controlling threshold, ζ denotes insensitive loss function parameter, and L_i^a represents the target electrical energy consumption pattern. The minimization of this regularized risk function is required to obtain the parameter p . Robust error function can be computed as follows:

$$s = \begin{cases} 0 & \text{if } |L_i^a - f(s, p)| < \zeta \\ |L_i^a - f(s, p)| & \text{otherwise.} \end{cases} \quad (26)$$

In Equation 26, a function is used to minimize Equation 25 and can be modeled as:

$$f(s, \alpha, \alpha^*) = \sum_{i=1}^N (\alpha^* - \alpha) K^*(s, s_i) + \partial, \quad (27)$$

where $\alpha^* \geq 0$ for all values of i . $K^*(s, y)$ is the Kernel function for SVM that shows dot product in the feature space f^* of radial basis KPCA as:

$$K^*(s, y) = \sum_{i=1}^D \chi_i(s) \chi_i(y) \quad (28)$$

The Kernel function makes χ_i feature not needed to be calculated in an infinite feature space. The α and α^* can be achieved by maximizing the quadratic form as:

$$R(\alpha^*, \alpha) = -\zeta \sum_{i=1}^N (\alpha_i^* + \alpha_i) + \sum_{i=1}^N L_i^a (\alpha_i^* - \alpha_i) - \frac{1}{2} \sum_{i,j=1}^N (\alpha_i^* + \alpha_i) (\alpha_i^* - \alpha_i) K^*(s_i, y_j). \quad (29)$$

The SVM based forecasting process is indicated in Algorithm 1. The forecasted energy consumption pattern is fed into the optimizer module to improve the accuracy by further minimizing the error.

C. MODIFIED ENHANCED DIFFERENTIAL EVOLUTION ALGORITHM BASED OPTIMIZER

The goal of this module is to further improve the forecast accuracy by minimizing regularized risk function. Since the returned value of the regularized risk function from SVM based forecaster is minimum as per its capabilities, therefore, the optimizer module is integrated with the forecaster based on SVM to further minimize the regularized risk function. Thus, the optimization module takes the regularized risk function minimization as an objective function. But this function is related to the hyper parameters like Kernel parameters ξ , cost penalty ϑ , and insensitive loss function parameters ζ . However, optimizing these hyperparameters for fast, accurate, and efficient load forecasting is still a crucial issue. In this view, scholars have used various methods like gradient descent algorithm, cross validation, and back-propagation algorithm to optimize hyperparameters [31]. However, these methods have high computational complexity and are unable to converge.

Therefore, DE among the optimization algorithms is chosen due to two reasons: (i) it avoids premature convergence, and (ii) it has optimal search capability. Authors in [32] used an improved version of DE (EDE), which is proposed in [33]. The work done in [32] is enhanced in terms of accuracy of the trial vector generation and the convergence rate. Therefore, mEDE is used with SVM model to optimally tune and select control parameters, as shown in Figure 2. The SVM with mEDE is depicted in Figure 3. A brief discussion is given below:

In [32], the trial vector V for i^{th} individual in t generation is represented as follows:

$$V_t(i, j) = \begin{cases} u_t(i, j) & \text{if } rand(j) \leq FF(U_t(i)) \\ x_t(i, j) & \text{if } rand(j) > FF(U_t(i)), \end{cases} \quad (30)$$

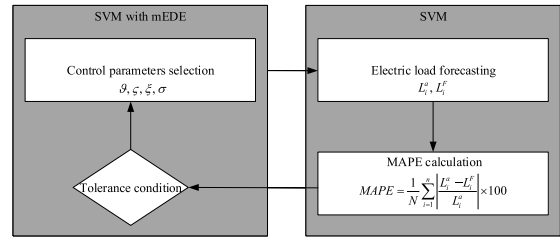


FIGURE 3. Support vector machine with modified enhanced differential evolution algorithm.

where $u_t(i, j)$ is the mutant vector, and $x_t(i, j)$ is the parent vector. In Equation 30, $FF()$ represents the fitness function ranging between 0 and 1, and $rand()$ represents the random number that lies between 0 and 1. Based on $X_t(i)$ and $Y_t(i)$, the next generation $X_{t+1}(i)$ offspring is generated as:

$$V_{t+1}(i, j) = \begin{cases} V_t(i, j) & \text{if } R_F(V_t(i)) \leq FF(x_t(i)) \\ x_t(i, j) & \text{otherwise.} \end{cases} \quad (31)$$

From the above Equations 30 and 31, it is obvious that the selection of next generation $t + 1$ offspring depends on trial vector of the previous generation that relies on $rand()$ and $FF()$ functions. The EDE algorithm in [32] updates load values by comparing random number $rand()$ with fitness $FF()$. This random updating of load is a big puzzle. Thus, this issue is resolved by removing the dependence of offspring selection on the randomly generated number. The process of updating load values is devised by comparing the fitness function of the candidate load value with the previous load value. Thus, the new load values will become optimal and will contribute to the improvement of the forecast accuracy. The devised modifications in Equation 30 are as follows:

$$V_t(i, j) = \begin{cases} u_t(i, j) & \text{if } \frac{X_t(i)}{X_t(i_{max})} \leq FF(U_t(i)) \\ x_t(i, j) & \text{if } \frac{X_t(i)}{X_t(i_{max})} > FF(U_t(i)) \end{cases} \quad (32)$$

In this view, the fitness function of parent and mutant vectors is defined in [32] as:

$$FF(U_t(i)) = \frac{1}{\frac{1}{R_F(U_t(i))} + \frac{1}{R_F(X_t(i))}} \quad (33)$$

$$FF(X_t(i)) = \frac{1}{\frac{1}{R_F(X_t(i))} + \frac{1}{R_F(U_t(i))}} \quad (34)$$

In fitness functions of Equations 33 and 34, it is assumed that each mathematical operator, i.e., division and addition, requires 1 unit of time to execute. Since Equations 33 and 34 will take 5 units of time in each iteration to execute, therefore, according to [32], the total iterations for EDE algorithm are 100. During each iteration, one fitness function is calculated by EDE in 500 units of time and two fitness functions are computed in 1000 units of time. Thus, to alleviate the execution time and improve the convergence speed, the modifications in the fitness functions in Equations 33 and 34 are as

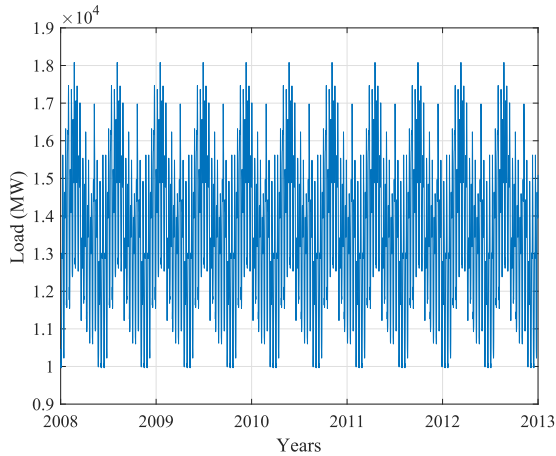


FIGURE 4. Five-year historical hourly load data of independent system operator New England control area with year index.

follows:

$$FF(U_t(i)) = \frac{R_F(X_t(i))}{R_F(U_t(i)) + R_F(X_t(i))} \quad (35)$$

$$FF(X_t(i)) = \frac{R_F(U_t(i))}{R_F(X_t(i)) + R_F(U_t(i))} \quad (36)$$

Using Equations 35 and 36, the algorithm takes 400 units of time to calculate two fitness functions in 100 iterations. In this way, the convergence speed of the EDE algorithm used in [32] is enhanced. The pseudo-code of the proposed framework is presented in detail in Algorithm 1.

IV. SIMULATION RESULTS AND DISCUSSION

To evaluate the validity and applicability of the proposed FA-HELFF framework, extensive simulations are performed using MATLAB. For this evaluation, the hourly load data of the ISO-NE control area is used. The used dataset includes 5-year (from 2008 to 2013) historical load data with hour resolution, and is publicly available at [34]. The hourly load data is split into three datasets: training, testing, and validation, as depicted in Figure 4. The 80% load data is for training, and the remaining 20% is kept for testing and validation purposes, respectively. For validation, the FA-HELFF framework is compared with benchmark frameworks like F-RBF-CNN [9], SDPSO-ELM [35], and SSA-SVM-CS [36] in terms of convergence speed and accuracy. These frameworks are selected because of their architectural resemblance with the proposed FA-HELFF framework, which is needed for a fair comparison. The simulation parameters are listed in Table 2 and are kept the same for the proposed and benchmark schemes. The detailed description of the simulation results is presented as follows:

In feature engineering, the first GCA is applied to the selected abstractive features from the hourly load data of the ISO-NE control area during 01-01-2008 to 31-12-2013 time horizon by calculating the correlation between features and the target load. The purpose is to remove irrelevant features. As the correlation threshold grade (σ) is set to 0.5 by [28],

TABLE 2. Parameters used in the simulations for the proposed FA-HELFF framework and benchmark frameworks like SSA-SVM-CS F-RBF-CNN, and SDPSO-ELM.

Parameters	Values
Population	24
Maximum upper population bound	0.9
Minimum lower population bound	0.1
Decision variables	2
Maximum epochs	100
Learning rate	0.002
Weight decay	0.0002
Initial value of weights	0.1
Initial value of bias	0
Momentum	0.5
Number of objectives	2
Features selection threshold	0.5

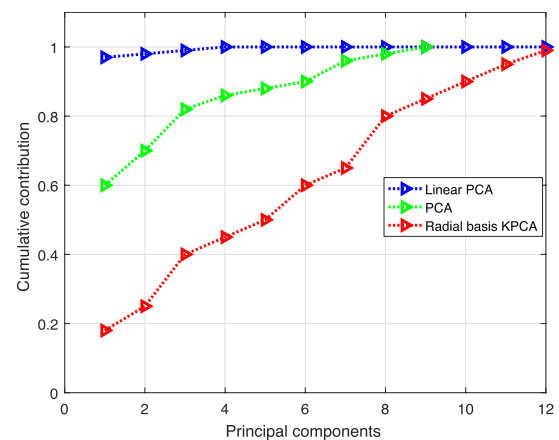


FIGURE 5. Performance evaluation of KPCA with different Kernel functions and PCA.

the features having a value less than this threshold σ are discarded and the features having a value greater than σ are reserved. The importance value of the reserved features is evaluated through two evaluators, γ and κ , as described in Algorithm 1. We observed that with the increase of feature selection threshold σ , large features are dropped, which leads to the best training speed and worst forecast accuracy.

Then, the radial basis KPCA in feature engineering is used to extract principal components and eliminate redundant information within the reserved features. The comparison of KPCA, PCA, and different Kernels in terms of cumulative contribution is depicted in Figure 5. The radial basis KPCA can extract principal components and the cumulative contribution approaches 95%, as depicted in Figure 6. Thus, among the different Kernels, radial basis function is chosen as the Kernel for KPCA because radial basis KPCA distributes data points along coordinate axes to extract principal components, which contributes in the accurate load forecasting.

A learning curve of error vs the number of epochs for SVM based forecaster is depicted in Figure 7, which enables us to evaluate whether the selected model is learning or

Algorithm 1 Pseudo-Code of the Proposed FA-HELFF Framework for Energy Consumption Forecasting

Importinputs : $\omega^r[\tau_j] = 0$, $\omega^F[\tau_j] = 0$, $D[]$, *and* $r[n]$ *Output* : $L_i^F = L_1^F, L_2^F, L_3^F, \dots, L_n^F$

begin

Restore the defective and missing values by data cleansing phase

Normalize the data w.r.t. its maximum value by data normalization phase

Change the data structure by data structuring phase

Parameters initialization: set all weights equal to zero and read data from $D = []$

First evaluator γ : Evaluation based on first evaluator γ

begin

for $k = 1 : m$ **do do**

for $i = 1 : n$ **do do**

 Determine $errOOB1_i$ using OOB dataset of decision tree [i]

 Add noise randomly to OOB dataset on feature τ_k

 Determine $errOOB2_i$ using OOB dataset of decision tree [i]

end

 Determine the importance of the feature $\omega^r[\tau_j] \leftarrow \frac{\sum_{i=1}^n errOOB2_i - errOOB1_i}{n}$

end

end

Second evaluator κ : Evaluation based on second evaluator κ

begin

for $i = 1 : m$ **do do**

 Randomly choose an item in class C_i

 Findout k nearest hits item $H_j(C_i)$

 For each class $C_j \neq C_i$

 Findout k nearest hits item $M_j(C_i)$

end

for $i = 1 : m$ **do do**

 Update $\omega^F[\tau_i]$ using Equation 8

end

 Feature selector: Select features based on ω^F and ω^r

begin

 Normalize ω^F and ω^r using Equation 10

 Perform feature selection using Equation 11

end

end

Feature extractor

begin

 Perform feature extraction based on KPCA using Equations 11-23

end

Forecaster

begin

 Perform electrical energy consumption forecasting using SVM based forecaster using Equations 24-29

end

Optimizer

begin

 Perform optimization on regularization risk function based on mEDE to optimally tune parameters using Equations 30-36

end

Return fast and accurate results

end

memorizing the data. At the start, when the number of epochs is zero, the error is maximum, indicating that the SVM based forecaster is not well trained. When the number of epochs increases, the error decreases, indicating that the SVM based

forecaster is training. During this training course, a point is reached where the error is not decreased any further with the increase in epochs. That point is known as saturation point and the SVM based forecaster is well trained by that

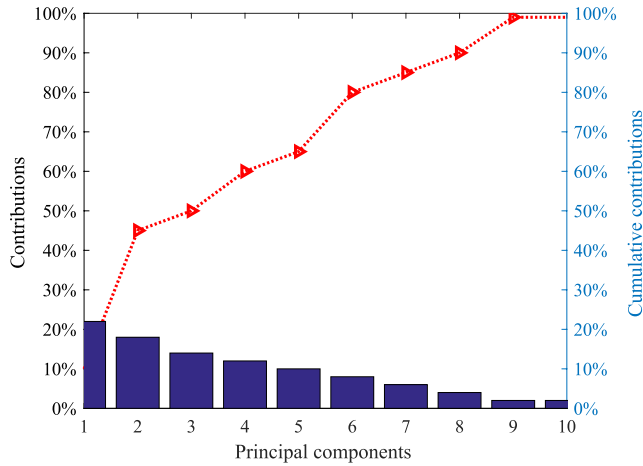


FIGURE 6. Radial basis KPCA in terms of cumulative contribution.

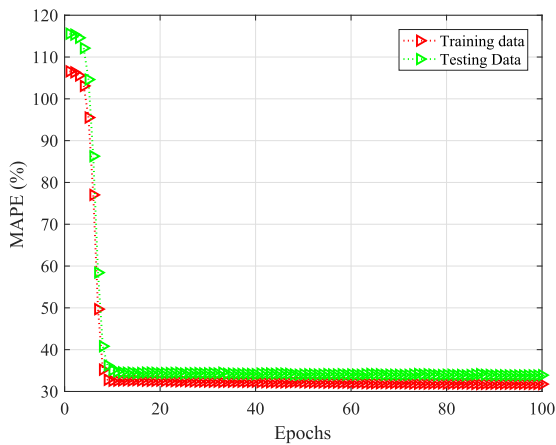


FIGURE 7. Proposed model learning curve on training and testing data samples for 100 number of iterations.

time. The simulation results of the well-trained network for hourly electrical energy consumption forecasting, weekly electrical energy consumption forecasting, and monthly electrical energy consumption forecasting are discussed at the same time.

The performance evaluation of FA-HELFF framework and the benchmark schemes such as F-RBF-CNN [9], SDPSO-ELM [35], and SSA-SVM-CS [36] in comparison with actual electrical energy consumption pattern with an hour time horizon is shown in Figure 8. The statistical analysis of the proposed FA-HELFF framework and benchmark frameworks in terms of mean absolute percentage error (MAPE) is listed in Figure 3. The proposed FA-HELFF framework based forecasted electrical energy consumption pattern is closely related to the target electrical energy consumption pattern, which demonstrates that significant improvement in forecast accuracy is obtained when the optimization module and feature engineering module are integrated with the forecaster module based on SVM. The SSA-SVM-CS framework outperforms both F-RBF-CNN and SDPSO-ELM in terms of MAPE. Moreover, F-RBF-CNN outperforms SDPSO-ELM in terms of MAPE. The superior performance of the proposed

TABLE 3. Evaluation of the proposed FA-HELFF framework and the benchmark frameworks like SSA-SVM-CS, F-RBF-CNN, and SDPSO-ELM in terms of MAPE (%) for the 28 May 2013 with hour resolution.

Proposed and benchmark forecasting frameworks				
Hours	SDPSO-ELM MAPE (%)	F-RBF-CNN MAPE (%)	SSA-SVM-CS MAPE (%)	FA-HELFF MAPE (%)
00	2.25	1.28	1.65	0.42
01	1.73	1.60	1.25	0.34
02	1.69	1.55	1.23	0.42
03	1.80	1.95	1.32	0.43
04	1.94	1.65	1.05	0.33
05	2.45	1.92	1.32	0.56
06	3.72	2.25	2.02	0.52
07	3.14	2.45	1.75	0.53
08	1.60	1.53	1.25	0.35
09	1.44	1.40	1.18	0.34
10	1.85	1.46	1.35	0.45
11	2.19	1.68	1.45	0.44
12	2.30	1.72	1.23	0.43
13	2.15	1.55	1.50	0.42
14	1.65	1.65	1.10	0.33
15	1.78	1.73	1.39	0.41
16	1.90	1.65	1.32	0.43
17	1.84	1.42	1.40	0.32
18	2.80	1.95	1.52	0.52
19	3.40	2.10	2.12	0.51
20	3.25	2.40	1.90	0.56
21	1.72	1.33	1.10	0.31
22	1.64	1.55	1.28	0.31
23	1.95	1.52	1.20	0.42
Avg.	2.15	1.82	1.36	0.42

FA-HELFF framework is due to the integration of feature engineering and optimization module with the forecaster module based on SVM. Feature engineering avoids irrelevancy and redundancy, and the optimization module helps to minimize the error by optimizing hyperparameters.

Table 4 illustrates statistical evaluation of forecasted electrical energy consumption and target electrical energy consumption in terms of accuracy, where FA-HELFF and benchmark frameworks: F-RBF-CNN [9], SDPSO-ELM [35], and SSA-SVM-CS [36], are chosen for analysis. The table shows that the average MAPE for a specific day (28 May 2013) is significantly reduced in case of the proposed FA-HELFF framework. The MAPE of the proposed FA-HELFF framework, SDPSO-ELM, F-RBF-CNN, and SSA-SVM-CS, is 0.410%, 1.655%, 0.988%, and 0.899%, respectively. Hence, FA-HELFF has improved accuracy of 75.22%, 58.50%, and 54.39% in comparison to the benchmark frameworks SDPSO-ELM, F-RBF-CNN, and SSA-SVM-CS, respectively.

Figure 9 illustrates the evaluation of target electrical energy consumption vs forecasted electrical energy consumption for the proposed FA-HELFF framework and benchmark

TABLE 4. Evaluation of accuracy in terms of MAPE of the proposed framework FA-HELFF and benchmark frameworks like SSA-SVM-CS, F-RBF-CNN, and SDPSO-ELM by taking target electrical energy consumption as a reference.

Forecasting frameworks	MAPE (%)	Forecast accuracy improvement (%)
F-RBF-CNN	0.988%	FA-HELFF to F-RBF-CNN (58.50%)
SSA-SVM-CS	0.899%	FA-HELFF to SSA-SVM-CS (54.39%)
SDPSO-ELM	1.655%	FA-HELFF to SDPSO-ELM (75.22%)
FA-HELFF	0.410%	

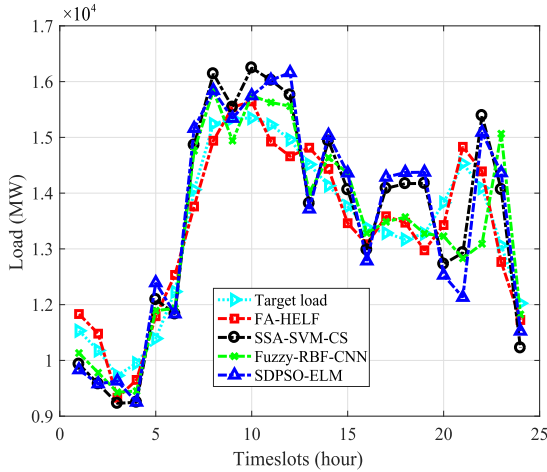


FIGURE 8. Hourly electrical energy consumption forecasting on May 28, 2013 of ISO-NE control area. The cyan-line represents the target energy consumption pattern, the red, blue, black, and green, represent FA-HELFF, SDPSO-ELM, SSA-SVM-CS, and F-RBF-CNN frameworks based forecasting, respectively.

frameworks for the week time horizon of 12/22/2013 to 12/29/2013. The week ahead results provide electrical energy consumption forecasting for different days including working days, holidays, and weekends. Results demonstrate that the proposed FA-HELFF framework enhanced the forecast accuracy by 76.80%, 58.56%, and 55.39% as compared to SDPSO-ELM, F-RBF-CNN, and SSA-SVM-CS, respectively.

The forecasted electrical energy consumption vs target electrical energy consumption evaluation for the month ahead time horizon (11/01/2013 to 11/30/2013) is illustrated in Figure 10. The monthly forecasted results demonstrate the significant enhancement in forecast accuracy. This improvement in forecast accuracy is due to the integration of both prior feature engineering and post mEDE algorithm based modules with SVM based forecaster. The simulation results show that the FA-HELFF framework improved the forecast accuracy by 77.65%, 58.66%, and 57.39% as compared to SDPSO-ELM, F-RBF-CNN, and SSA-SVM-CS, respectively.

The accuracy results in terms of MAPE of the proposed FA-HELFF framework and benchmark models like F-RBF-CNN, SSA-SVM-CS, and SDPSO-ELM are for different months of the year 2013, as listed in Table 5. The simulation results in Table 5 demonstrate that the proposed FA-HELFF framework and benchmark models like SDPSO-ELM,

TABLE 5. Evaluation of the proposed FA-HELFF framework and the benchmark frameworks like F-RBF-CNN, SSA-SVM-CS, and SDPSO-ELM, in terms of MAPE (%) for the year 2013 with monthly time horizon.

Months	Proposed and benchmark forecasting frameworks			
	SDPSO-ELM MAPE (%)	F-RBF-CNN MAPE (%)	SSA-SVM-CS MAPE (%)	FA-HELFF MAPE (%)
Jan	2.22	1.67	1.55	0.41
Feb	1.68	1.50	1.15	0.39
Mar	1.67	1.51	1.33	0.43
Apr	1.60	1.29	1.35	0.42
May	1.74	1.62	1.09	0.36
Jun	2.40	1.81	1.52	0.54
Jul	3.82	3.22	2.02	0.53
Aug	3.44	2.64	1.82	0.54
Sept	1.62	1.49	1.20	0.33
Oct	1.54	1.40	1.28	0.39
Nov	1.81	1.42	1.30	0.47
Dec	2.12	1.78	1.46	0.41
Avg.	2.12	1.79	1.44	0.41

F-RBF-CNN, and SSA-SVM-CS have average MAPE of 0.41, 2.12, 1.79, and 1.44, respectively.

The performance analysis of FA-HELFF and benchmark frameworks like F-RBF-CNN [9], SDPSO-ELM [35], and SSA-SVM-CS [36] in terms of computational time is depicted in Figure 11. The individual models, ELM, CNN, and SVM, without the integration of both feature engineering and optimization modules have low computational time and worst error performance. However, when both feature engineering and optimization modules are integrated with these individual models, the computational time is increased and the error is reduced due to the tradeoff between accuracy and convergence rate. The statistical evaluations of the proposed FA-HELFF framework and benchmark frameworks in terms of forecast errors and computational time are listed in Table 6. The average values of computational speed and forecast error for both individual (ELM, CNN, and SVM) models and hybrid (F-RBF-CNN, SDPSO-ELM, SSA-SVM-CS, and FA-HELFF) models for daily, weekly, and monthly time horizon are listed in Table 6. The first row, second row, and third-row show daily, weekly, and monthly evaluations of models (both individual and hybrid), respectively, in terms of computational speed and forecast error. The computational speed of individual models without the integration of both feature

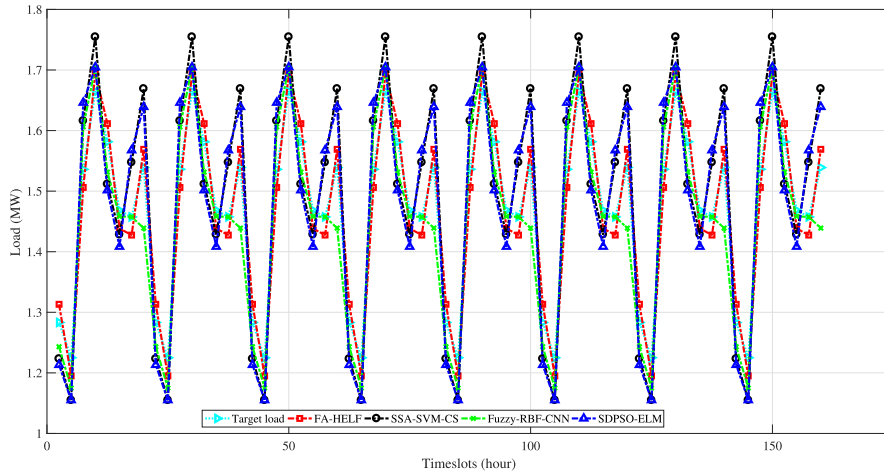


FIGURE 9. Week ahead (12/22/2013 to 12/29/2013) electrical energy consumption forecasting of the ISO-NE control area. The cyan-line shows the target energy consumption profile, the red, blue, black, and green, represent FA-HELF, SDPSO-ELM, SSA-SVM-CS, and F-RBF-CNN frameworks based forecasting, respectively.

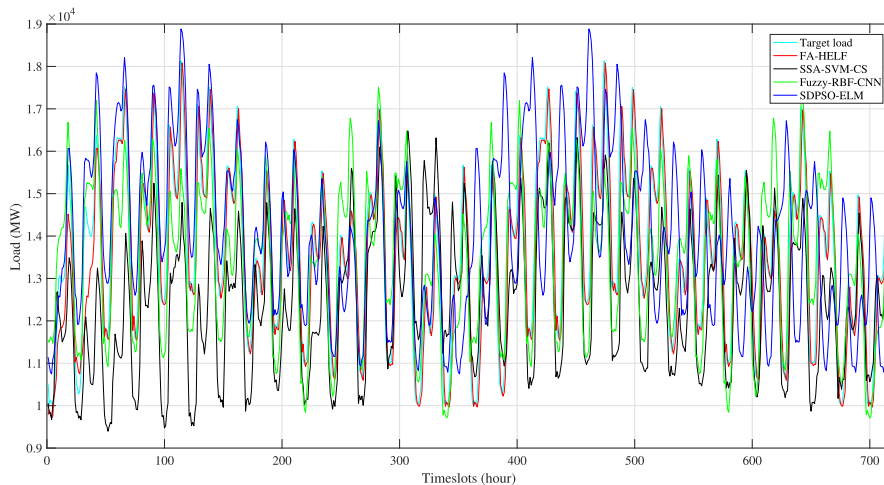


FIGURE 10. Month ahead (11/01/2013 to 11/30/2013) electric energy consumption forecasting on the ISO-NE control area. The cyan-line shows the target electrical energy consumption profile, the red, blue, black, and green, represent FA-HELF, SDPSO-ELM, SSA-SVM-CS, and F-RBF-CNN frameworks based forecasting, respectively.

engineering and optimization modules (ELM, CNN, and SVM) and hybrid models with the integration of both feature engineering and optimization modules (SSA-SVM-CS, F-RBF-CNN, SDPSO-ELM, and FA-HELF) is computed on a Core(TM)-i3-3110M, CPU@2.40 GHz with 8 GB RAM system. The individual models, ELM, CNN, and SVM, have the lowest computational time of 170 sec, 185 sec, and 115 sec, respectively, while the hybrid models have increased computational time of 242, 512, 355, and 285 secs for SDPSO-ELM, F-RBF-CNN, SSA-SVM-CS, and FA-HELF frameworks, respectively. The increase in computational time is encountered as the optimizer module or the feature engineering module, or both modules are integrated with the individual forecaster models. Moreover, this increment in time is due to the trade-off between convergence speed and accuracy, thus achieving more accuracy at the cost of

surplus computational time. The proposed FA-HELF framework reduces the computational time by devising modifications in the EDE optimization algorithm (see Equations: 35-36).

Thus, the feature engineering module rectifies the feature space by removing redundant and irrelevant features and the optimization module based on the mEDE algorithm tunes the control parameters of the SVM, which ensures accurate electrical energy consumption forecasting.

Robustness evaluation of the proposed FA-HELF framework and benchmark frameworks like SDPSO-ELM, F-RBF-CNN, and SSA-SVM-CS is depicted in Figure 12. The evaluation is conducted by adding error (noise) to each feature and observing the accuracy of each scheme. The proposed FA-HELF framework is more robust than the benchmarks frameworks like SSA-SVM-CS, F-RBF-CNN,

TABLE 6. Evaluation of actual and forecasted electrical energy consumption in terms of MAPE and computational time of individual frameworks (ELM, CNN, and SVM) and hybrid frameworks like SSA-SVM-CS, SDPSO-ELM, F-RBF-CNN, and FA-HELF for day ahead, week ahead, and month ahead time horizon.

Frameworks without and with feature engineering and optimization modules													
ELM		CNN		SVM		SDPSO-ELM		F-RBF-CNN		SSA-SVM-CS		FA-HELF	
CT (s)	MAPE (%)	CT (s)	MAPE (%)	CT (s)	MAPE (%)	CT (s)	MAPE (%)	CT (s)	MAPE (%)	CT (s)	MAPE (%)	CT (s)	MAPE (%)
170	2.25	185	2.20	115	2.98	242	1.65	512	0.99	355	0.89	285	0.42
235	2.25	265	2.20	195	2.98	310	1.65	588	0.99	415	0.89	365	0.42
305	2.25	335	2.20	275	2.98	385	1.65	695	0.99	495	0.89	435	0.42

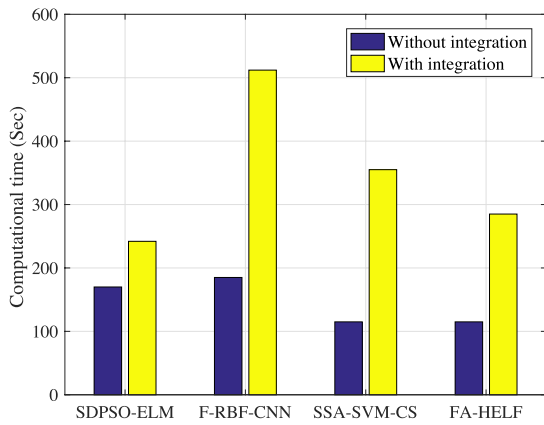


FIGURE 11. Comparative evaluation of the proposed and benchmark frameworks with and without integration of feature engineering and optimization modules.

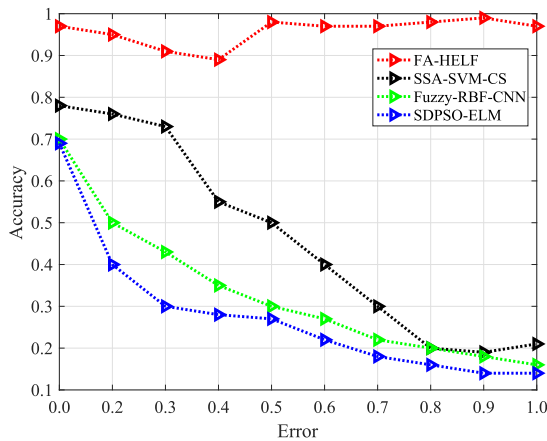


FIGURE 12. Comparative evaluation of the proposed and benchmark frameworks in terms of robustness.

and SDPSO-ELM as illustrated in Figure 12, because the noise within features has little influence on accuracy, and therefore, less important and irrelevant features are dropped during the feature engineering phase. Thus, the proposed FA-HELF framework is also robust against noise in the features.

V. CONCLUSION AND FUTURE WORK

In this work, a non-linear framework (FA-HELF) is proposed to perform efficient and accurate energy consumption

forecasting for energy management in smart grid. The proposed framework combines random forests and relief-F with GCA to choose the most appropriate features, and uses radial basis KPCA for the extraction of features to overcome the dimensionality reduction problem. The selected and extracted features provide the most relevant information about electrical energy consumption to train SVM based forecaster. In addition, the mEDE optimization algorithm is used in the proposed framework to optimize the hyperparameters for improving the forecast accuracy along with minimizing the computational time.

To analyze the validity of FA-HELF framework, load data with hour resolution of the ISO-NE control area is used. We can draw the following conclusions based on our results and evaluation: (i) GCA and radial basis KPCA feeds the most relevant and desired features to the SVM based forecaster and removes the redundant and irrelevant features from the input data to improve the forecasting performance; (ii) although feature engineering assists in the improvement of forecast accuracy, mEDE optimization algorithm contributes in the improvement of both convergence rate and forecast accuracy; (iii) the proposed FA-HELF framework is better as compared to the benchmark frameworks like SDPSO-ELM, F-RBF-CNN, and SSA-SVM-CS. Simulation results show that the developed framework is robust, fast, and powerful to forecast electrical energy consumption for efficient energy management. Thus, it is believed that the proposed FA-HELF framework is scalable and reliable and can be applied in real-life for efficient energy management.

In future, some other advanced heuristic techniques for suitable parameter selection can be integrated with SVM based forecaster for fast and accurate electrical energy consumption forecasting. Moreover, the advanced heuristic techniques can be integrated with advanced deep learning models and the work can be extended to medium-term and long-term electrical energy consumption forecasting.

REFERENCES

[1] N. Javid, G. Hafeez, S. Iqbal, N. Alrajeh, M. S. Alabed, and M. Guizani, "Energy efficient integration of renewable energy sources in the smart grid for demand side management," *IEEE Access*, vol. 6, pp. 77077–77096, 2018.

- [2] G. Hafeez, N. Javaid, S. Iqbal, and F. Khan, "Optimal residential load scheduling under utility and rooftop photovoltaic units," *Energies*, vol. 11, no. 3, p. 611, 2018.
- [3] G. Hafeez, N. Islam, A. Ali, S. Ahmad, and M. U. A. K. S. Alimgeer, "A modular framework for optimal load scheduling under price-based demand response scheme in smart grid," *Processes*, vol. 7, no. 8, p. 499, 2019.
- [4] G. Hafeez, N. Javaid, M. Riaz, A. Ali, K. Umar, and Z. Iqbal, "Day ahead electric load forecasting by an intelligent hybrid model based on deep learning for smart grid," in *Conference on Complex, Intelligent, and Software Intensive Systems*. Cham, Switzerland: Springer, 2019, pp. 36–49.
- [5] G. Hafeez, N. Javaid, M. Riaz, K. Umar, Z. Iqbal, and A. Ali, "An innovative model based on FCRBM for load forecasting in the smart grid," in *Proc. Int. Conf. Innov. Mobile Internet Services Ubiquitous Comput.*, Cham, Switzerland: Springer, 2019, pp. 49–62.
- [6] G. Hafeez, N. Javaid, S. Ullah, and Z. Iqbal, "Short term load forecasting based on deep learning for smart grid applications," in *Proc. Int. Conf. Innov. Mobile Internet Services Ubiquitous Comput.* Cham, Switzerland: Springer, 2018, pp. 276–288.
- [7] L. Han, Y. Peng, Y. Li, B. Yong, Q. Zhou, and L. Shu, "Enhanced deep networks for short-term and medium-term load forecasting," *IEEE Access*, vol. 7, pp. 4045–4055, 2019.
- [8] J. Fiot and F. Dinuzzo, "Electricity demand forecasting by multi-task learning," in *Proc. IEEE Manchester PowerTech*, Jun. 2017, p. 1.
- [9] G. Sideratos, A. Ikononopoulos, and N. D. Hatzigiorgiou, "A novel fuzzy-based ensemble model for load forecasting using hybrid deep neural networks," *Electr. Power Syst. Res.*, vol. 178, Jan. 2020, Art. no. 106025.
- [10] S. Li, L. Goel, and P. Wang, "An ensemble approach for short-term load forecasting by extreme learning machine," *Appl. Energy*, vol. 170, pp. 22–29, May 2016.
- [11] L. Lin, F. Wang, X. Xie, and S. Zhong, "Random forests-based extreme learning machine ensemble for multi-regime time series prediction," *Expert Syst. Appl.*, vol. 83, pp. 164–176, Oct. 2017.
- [12] Y. Li, J. Che, and Y. Yang, "Subsampled support vector regression ensemble for short term electric load forecasting," *Energy*, vol. 164, pp. 160–170, Dec. 2018.
- [13] J. Zhang, Y.-M. Wei, D. Li, Z. Tan, and J. Zhou, "Short term electricity load forecasting using a hybrid model," *Energy*, vol. 158, pp. 774–781, Sep. 2018.
- [14] L.-L. Li, J. Sun, C.-H. Wang, Y.-T. Zhou, and K.-P. Lin, "Enhanced Gaussian process mixture model for short-term electric load forecasting," *Inf. Sci.*, vol. 477, pp. 386–398, Mar. 2019.
- [15] Y. Liang, D. Niu, and W.-C. Hong, "Short term load forecasting based on feature extraction and improved general regression neural network model," *Energy*, vol. 166, pp. 653–663, Jan. 2019.
- [16] E. Mocanu, P. H. Nguyen, M. Gibescu, and W. L. Kling, "Deep learning for estimating building energy consumption," *Sustain. Energy, Grids Netw.*, vol. 6, pp. 91–99, Jun. 2016.
- [17] L. Xiao, J. Wang, X. Yang, and L. Xiao, "A hybrid model based on data preprocessing for electrical power forecasting," *Int. J. Electr. Power Energy Syst.*, vol. 64, pp. 311–327, Jan. 2015.
- [18] P. Jiang, F. Liu, and Y. Song, "A hybrid forecasting model based on date-framework strategy and improved feature selection technology for short-term load forecasting," *Energy*, vol. 119, pp. 694–709, Jan. 2017.
- [19] K. G. Boroojeni, M. H. Amini, S. Bahrami, S. S. Iyengar, A. I. Sarwat, and O. Karabasoglu, "A novel multi-time-scale modeling for electric power demand forecasting: From short-term to medium-term horizon," *Electr. Power Syst. Res.*, vol. 142, pp. 58–73, Jan. 2017.
- [20] H. Shi, M. Xu, and R. Li, "Deep learning for household load forecasting—A novel pooling deep RNN," *IEEE Trans. Smart Grid*, vol. 9, no. 5, pp. 5271–5280, Sep. 2018.
- [21] T. Ouyang, Y. He, H. Li, Z. Sun, and S. Baek, "Modeling and forecasting short-term power load with copula model and deep belief network," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 3, no. 2, pp. 127–136, Apr. 2019.
- [22] S. Motepe, A. N. Hasan, and R. Stopforth, "Improving load forecasting process for a power distribution network using hybrid AI and deep learning algorithms," *IEEE Access*, vol. 7, pp. 82584–82598, 2019.
- [23] H. H. H. Aly, "A proposed intelligent short-term load forecasting hybrid models of ANN, WNN and KF based on clustering techniques for smart grid," *Electric Power Syst. Res.*, vol. 182, May 2020, Art. no. 106191.
- [24] A. Shiri, M. Afshar, A. Rahimi-Kian, and B. Maham, "Electricity price forecasting using support vector machines by considering oil and natural gas price impacts," in *Proc. IEEE Int. Conf. Smart Energy Grid Eng. (SEGE)*, Aug. 2015, pp. 1–5.
- [25] H. Jiang, K. Wang, Y. Wang, M. Gao, and Y. Zhang, "Energy big data: A survey," *IEEE Access*, vol. 4, pp. 3844–3861, 2016.
- [26] K. Wang, C. Xu, Y. Zhang, S. Guo, and A. Y. Zomaya, "Robust big data analytics for electricity price forecasting in the smart grid," *IEEE Trans. Big Data*, vol. 5, no. 1, pp. 34–45, Mar. 2019.
- [27] G. Hafeez, I. Khan, M. Usman, K. Aurangzeb, and A. Ullah, "Fast and accurate hybrid electric load forecasting with novel feature engineering and optimization framework in smart grid," in *Proc. 6th Conf. Data Sci. Mach. Learn. Appl. (CDMA)*, Mar. 2020, pp. 31–36.
- [28] D. Julong, "Introduction to grey system theory," *J. Grey Syst.*, vol. 1, no. 1, pp. 1–24, 1989.
- [29] L. Breiman, "Bagging predictors," *Mach. Learn.*, vol. 24, no. 2, pp. 123–140, Aug. 1996.
- [30] Z. Du, X. Wang, L. Zheng, and Z. Zheng, "Nonlinear system modeling based on KPCA and MKSVM," in *Proc. ISECS Int. Colloq. Comput., Commun., Control, Manage.*, Aug. 2009, pp. 61–64.
- [31] Anamika and N. Kumar, "Market clearing price prediction using ANN in indian electricity markets," in *Proc. Int. Conf. Energy Efficient Technol. Sustainability (ICEETS)*, Apr. 2016, pp. 454–458.
- [32] N. Amjadi, F. Keynia, and H. Zareipour, "Short-term load forecast of microgrids by a new bilevel prediction strategy," *IEEE Trans. Smart Grid*, vol. 1, no. 3, pp. 286–294, Dec. 2010.
- [33] R. Storn and K. Price, "Differential evolution—a simple and efficient heuristic for global optimization over continuous spaces," *J. Global Optim.*, vol. 11, no. 4, pp. 341–359, 1997.
- [34] *ISO New England Data*. [Online]. Available: <http://www.iso-ne.com>
- [35] N. Zeng, H. Zhang, W. Liu, J. Liang, and F. E. Alsaadi, "A switching delayed PSO optimized extreme learning machine for short-term load forecasting," *Neurocomputing*, vol. 240, pp. 175–182, May 2017.
- [36] X. Zhang, J. Wang, and K. Zhang, "Short-term electric load forecasting based on singular spectrum analysis and support vector machine optimized by cuckoo search algorithm," *Electr. Power Syst. Res.*, vol. 146, pp. 270–285, May 2017.



GHULAM HAFEEZ received the B.Sc. degree in electrical engineering from the University of Engineering and Technology at Peshawar, Pakistan, and the M.S. degree in electrical engineering from COMSATS University Islamabad, Islamabad, Pakistan, where he is currently pursuing the Ph.D. degree. He is a lifetime Chartered Engineer of the Pakistan Engineering Council. He is working as a Lecturer with the Department of Electrical Engineering, University of Engineering and Technology, Mardan. He has authored or coauthored over 15 peer-reviewed research articles in reputed international journals and conferences. His research interests include optimization, planning, energy management, and machine learning applications in smart/micro grids.



KHURRAM SALEEM ALIMGEER received the bachelors's degree in IT, in 2002, the M.S. degree in telecommunications (gold medal), in 2006, and the Ph.D. degree in electrical engineering with a specialization in antenna design, in 2014. He is currently an Assistant Professor with COMSATS University Islamabad, Pakistan, where he is working as a Researcher with the RF-Lab. He has published more than 80 research articles in reputed journals and conferences in the fields of antenna design, wave propagation, mathematical modeling, wireless communications, image processing, and energy management in smart/micro grids. He has been serving as an editor and a reviewer of few reputed journals, since 2008.



ABDUL BASEER QAZI (Member, IEEE) received the M.S. degree in information and communication systems from the University of Technology, Hamburg, Germany, and the Ph.D. degree from the University of Maastricht (UNU-MERIT), The Netherlands. He also worked as an Assistant Professor with CECOS University, Peshawar, Pakistan. His industrial experience includes working with four Fortune 500 companies such as IBM, Siemens, Philips Medical Systems, and NXP

Semiconductors in Germany and The Netherlands. He is currently a Senior Assistant Professor with the Department of Software Engineering, Bahria University, Islamabad, Pakistan. Prior to this, he was an Assistant Professor with the Capital University of Science and Technology, Islamabad. His research interests include wireless networks, antennas, technology policy, innovation, and entrepreneurship.



IMRAN KHAN (Senior Member, IEEE) received the B.Sc. degree in electrical engineering from the NWFP University of Engineering and Technology, Peshawar, Pakistan, in 2003, the M.Sc. degree in telecommunication engineering from the Asian Institute of Technology, Thailand, in 2007, and the Ph.D. degree in telecommunications from the School of Engineering and Technology, Asian Institute of Technology, in 2010. He is currently working as a Professor with the Department of

Electrical Engineering, University of Engineering and Technology, Mardan. His research interests include the performance analysis of wireless communication systems, OFDM, OFDMA, MIMO, cooperative networks, cognitive radio systems, and energy management in smart grids.



MUHAMMAD USMAN (Senior Member, IEEE) received the B.Sc. degree in computer information systems engineering from the University of Engineering and Technology (UET) at Peshawar, Pakistan, in 2004, the M.Sc. degree in computer engineering from the Center of Advanced Studies in Engineering (CASE), Islamabad, Pakistan, in 2007, and the Ph.D. degree from the University of Ulsan, South Korea, in 2016. He is associated with the UET, Mardan, as an Associate Professor.

His research interests include security and energy efficiency in wireless networks, next-generation networks, and wireless sensor networks. In 2015, he received the Best SCI(E) Paper Award from the Korean Government through the BK21+ Project. He is the Reviewer of various reputable international journals such as the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, IEEE ACCESS, the IEEE TRANSACTIONS ON COGNITIVE COMMUNICATION AND NETWORKING, the IEEE SENSORS JOURNAL, the IEEE SYSTEMS JOURNAL, *Applied Mathematics & Information Sciences*, and *The Computer Journal*.



FARRUKH ASLAM KHAN (Senior Member, IEEE) received the M.S. degree in computer system engineering from the GIK Institute of Engineering Sciences and Technology, Pakistan, in 2003, and the Ph.D. degree in computer engineering from Jeju National University, South Korea, in 2007. He has supervised/co-supervised five Ph.D. students and 17 M.S. thesis students. Several M.S. and Ph.D. degrees students are currently working under his supervision. He received professional

trainings at the Massachusetts Institute of Technology, New York University, IBM, and other institutions. He is currently working as a Professor with the Center of Excellence in Information Assurance, King Saud University, Riyadh, Saudi Arabia. He is also the Founding Director of the Wireless Networking and Security (WiNGS) Research Group, National University of Computer and Emerging Sciences, Islamabad, Pakistan. He has published more than 100 research articles in refereed international journals and conferences. His research interests include cybersecurity, wireless sensor networks and e-health, bio-inspired and evolutionary computation, and the Internet of Things. He has co-organized several international conferences and workshops. He is a Fellow of the British Computer Society (BCS). He serves as an Associate Editor for prestigious international journals, including IEEE ACCESS, *PLOS One*, *Neurocomputing* (Elsevier), *Ad Hoc and Sensor Wireless Networks*, *KSII Transactions on Internet and Information Systems*, *Human-Centric Computing and Information Sciences* (Springer), and *Complex & Intelligent Systems* (Springer). He serves on the panel of reviewers of over 50 reputed international journals and numerous international conferences.



ZAHID WADUD received the B.Sc. and master's degrees in electrical engineering from the University of Engineering and Technology (UET) at Peshawar, Pakistan, in 1999 and 2003, respectively, and the Ph.D. degree from the Capital University of Science and Technology, Islamabad, Pakistan, with a thesis entitled Energy Balancing With Sink Mobility in the Design of Underwater Routing Protocols. He is currently working as an Assistant Professor with the Department of

Computer Systems Engineering, at UET, Peshawar. He published over a dozen state-of-the-art publications in the renowned international journals. His research interests include wireless sensor networks, energy efficient networks and subsystems, the mathematical modeling of wireless channels, embedded systems, and sensors interfaces.

...