Mansoura University
Faculty of Engineering
Mansoura Engineering Journal

A Fog Based Customer Demand Forecasting (CDF) Strategy for Smart Grids Using IoT Technology

Naglaa R. Khalil, Asmaa H. Rabie, Khaled M. Abo-Al-Ez and Ahmed I. Saleh

KEYWORDS:
IOT, Fog cloud, smart grids, demand forecasting, feature selection.

Abstract— In this paper, a novel framework for effective energy management of residential customer is provided to reduce electricity consumption. Advanced Smart Grids (ASGs) can assist a variety of functions thanks to Internet of Things (IoT). These smart devices generate big data, which can be uploaded to the cloud for additional analysis. Fog computing tier operates as a bridge between the IoT devices integrated in Smart Electrical Grid(SEG) and the cloud to overcome cloud issues. Based on the indicated three-tier design, a novel Customer Demand Forecasting (CDF) strategy has been introduced. CDF strategy consists of (i) Feature Selection (FS) stage and (ii) Demand Forecasting (DF) stage. FS stage identifying the most important features that allow the demand forecasting model to produce quick and accurate results. A Hybrid Feature Selection (HFS) approach is used to pick the effective features, which integrates evidence from two feature selectors;(i) Information Gain (IG) as a filter method and (ii) Binary Particle Swarm (BPS)optimization is used as a wrapper method. Then, an Improved KN^{3}B (IKN^{3}B) predictor has been used in DF stage trying to provide accurate demand forecasts based on the selected subset of features from the previous stage. In fact, IKN^{3}B combines both K-Nearest Neighbors (KNN) classifier and Naïve Bayes (NB)classifier and then it improves the characteristics of them to provide the best demand forecasts as possible. Based on experimental results, It is conducted that CDF strategy is demonstrated to have a positive influence on system reliability, resilience, and stability by introducing accurate demand predictions.

I.INTRODUCTION

Advanced Smart Grid (ASGs) is a self-healing, flexible, robust, and sustainable computerized two-way load flow power system with ability to prediction in the face of a variety of suspicion. Smart Grid (SG) systems have been proposed as a cost-effective solution to the global energy crisis due to their inherent connectivity, control, and optimization characteristics, which can result in a real-time equilibrium between power supply and demand [1,2]. Through demand response and energy efficiency technologies, ASG allows distributed generation to shed load demand. Demand response (DR) has become a helpful method for balancing power supply and demand, as well as promoting energy saving as well as emission reduction.

Received: (11 May, 2022) - Revised: (19 October, 2022) - Accepted: (16 November, 2022)

****Corresponding Author:** Naglaa R. Khalil, Master Degree Researcher of Computers and control Department, Faculty of Engineering, Mansoura University, Egypt (E-mail: eng_naglaaragab@yahoo.com).*

Asmaa H. Rabi , Doctor, Computers and control Department, Faculty of Engineering, Mansoura University, Egypt (E-mail: asmaa91hamdy@yahoo.com).

Khaled M. Abo-Al-Ez, Associate professor Cape Peninsula University of Technology (CPUT); Bellville Campus; Symphony way, Bellville 7535 Cape Town, South Africa (E-mail: aboalez40@gmail.com).

Ahmed I. Saleh, Professor, Computers and control Department, Faculty of Engineering, Mansoura University, Egypt (E-mail: aisaleh@yahoo.com)

(Ser. NO. BFEMU-2204-1260)

Nonetheless, employing a diverse variety of alternative resources, such as decentralized generation, is an important. In DR systems, energy storage as well as the regulated load by customers become critical components [3]. DR is essential for improving the reliability and efficiency of electricity networks. It can also help to balance power production and consumption, as well as encourage energy saving and pollution reduction. Analyzing the DR capacity of consumers factors that influence the DR application, customer responsiveness is used to keep track of time the duration of the DR, the preferences of the customers, and the physical characteristics of the customer devices, etc... A thorough comprehension of the DR of customers, potential may assist in determining a customer's profitability and offer customized services as a result.

Currently, DR systems are introduced for industrial and residential customers with DR capital and consumer participation may participate. Additionally, DR systems are introduced to bridge the gap between suppliers and buyers [4]. IoT enables the SGs to monitor and organize smart devices, and also to clear the way for energy management of large-scale system. IoT envisions real-world objects to be integrated, coordinated, communicated, and collaborated for performing everyday tasks more intelligently and efficiently. For energy monitoring, IoT puts emphasis on consumer appliances and analyze their utilization to help comprehend the demand specifications of the customer and the grid before making smart decisions. Home automation services allow multiple members of a household to control the same set of devices [5,6]. Cloud computing provides access to customizable network resources and services, parallel processing, and ubiquitous access to overcome the computational needs of ASG systems [6]. Despite the fact that the cloud computing model is considered effective for ASG, it needs to reach the Quality-of-Experience (QoE)criteria to ASG applications, such as response time, bandwidth, energy exhaustion, and network cost. Hence, fog computing model has been built on cloud computing model by installing localized computing and processing facilities at the network's edge, allowing for location - based services, low response time, and Response time sensual analytics.

Cloud data was initially stored by fog and then transmitted to SG users using high-speed local connections by installing local computing facilities in their residences. In addition, fog is certified as a technical enabler for ASG real-time analytics [7-9]. Many ASG applications are time-sensitive and require rapid data analysis in order to respond rapidly to the different events that occur. In the field of SG, the latest and evolving paradigms of edge and fog computing promise to solve big data storage and analysis. There is a lot of data generated by smart devices and meters called Internet of Things (IoT) smart devices [10,11]. The amount of data collected in SG through smart meters and other IoT sensors is a big data. Huge, complicated, and expanding data sets coming from a variety of reliable sources are referred to as big data. Big data is currently growing rapidly in all sectors of research and engineering thanks to rapid advancements in networking, data storage, and data processing capability [12]. To increase the pliability of the energy system on both the generation and demand sides, new approaches are needed. Using big data analytics in SGs would aid in improving power system [13,14].

Big data analytics approaches are a subset of data mining techniques used to remove, clean, modify, and demonstrate data in order to discover important and useful insights that can aid in determining the best strategy and making decisions for the business in question. For the examination of numerous statistics linked to energy production, distribution mechanization, data transmission, taxation, and customer engagement in SGs, data mining is required. Data mining is vital for converting data into useful information for decision-making reasons, such as demand forecasting. In SG, when data is constantly gathered in actual time, data mining methods are required for energy conservation, stability, and real-time decision-making [15]. Dimensionality reduction is becoming more critical in big data analytics in which low-dimensional structures in high-dimensional data can be identified and exploited by using dimensionality reduction techniques to select the most significant features. However, the energy demand profile is more volatile and does not always follow the same pattern. As a result of the recent implementation of decentralized energy systems, appropriate and relevant feature selection tools and forecasting models for cost-effective and productive consumption modeling are needed.

When data is presented with very high dimensionality, there would be a greater chance of over fitting in forecasting model. The reason is that the training, testing, and validation times for the forecasting model will be increased while the accuracy of the forecasting model will be decreased. Accordingly, Feature Selection (FS) is a critical process before training the demand forecasting model. Filter, wrapper, and hybrid methods are the three types of FS strategies [16-19]. The growth of DR projects in SGs, there is a surge in demand forecasting (DF)for residential customers by forecast individual house energy loads. To supply the DF, anthropologic and structural data from houses have been used [20-22].

This study introduces a new conceptual framework for Customer Demand Forecasting(CDF) in smart grid based on the collected data from IoT devices such as smart meters. CDF consists of two stages, which are; FS and DF. FS aims to select the meaningful features for the next DF stage. This flawless feature selection methodology not only improves the adequacy of model prediction, but also speeds up the prediction process by considering fewer features. Hybrid Feature Selection(HFS) is used in FS to select the most effective features by integrating the manual of two feature selectors, which are; Information Gain(IG) and Binary Particle Swarm (BPS). In DF as the second stage, Improved KN^3B (IKN^3B) predictor that includes K-Nearest Neighbors (KNN) and Naïve Bayes (NB) classifiers is used on filtered data to provide rapid and accurate demand prediction. Results show that the introduced Customer Demand Forecasting (CDF) technology outperforms modern techniques when it comes to prediction accuracy with the lowest time penalty. The structure of this paper is as shown below; Section 2 reviews the previous efforts about the most recent demand forecasting strategies. Section 3 describes the advanced smart grid based on IoT technology and fog computing. Section 4 focuses on the proposed customer demand forecasting strategy. Section 5 depicts the experimental results. Finally, conclusions and future work are presented in section 6.

II. RELATED WORK

The latest research activities on demand forecasting in smart grids will be reviewed in this section. In [23], to predict the use of electrical energy, an efficient model has been created. The introduced model is called CNN-LSTM module, which integrates Convolutional Neural Networks (CNNs) with Long Short-Term Memory Networks (LSTMs) to identify spatial and temporal features that can be used to predict the quantity of energy used in a home. One significant advantage of this method is that it could provide efficient forecast for previously difficult-to-predict electric energy consumption. As provided in [24], a novel Ensemble Forecasting Model based on the Artificial Bee Colony (EFM-ABC) algorithm was proposed and adopted to provide energy demand forecasting. One advantages of EFM-ABC is that it could provide predictive efficiency which can be obtained using the ABC algorithm, which combines the advantages of a local deep search with a global-wide search. Although the benefits of EFM-ABC, it suffers from that it has uncertain time to convergence and dependent sequence of random decisions. In [25], A Deep Learning Model (DLM) was established using a convolutional recurrent neural network with multi-headed attention. The new solution reduced the error by 31.01 percent and increased the prediction performance compared to a state-of-the-art deep learning model.

As introduced in [26], a Hybrid Ensemble Method (HEM) was provided for electricity demand forecasting. HEM consists of three main methods called; Discrete Wavelet Transform (DWT), Particle Swarm Optimization (PSO), and Radial Basis Function Neural Network (RBFNN). In contrast to other models, HEM has a lower average relative error with high forecasting accuracy of 97.5778 %. In [27], the electrical energy consumption of buildings based on a hybrid ARIMA-GBRT model was proposed. This prediction model consists of Autoregressive Integrated Moving Average (ARIMA) and Gradient Boosting Regression Trees (GBRT).The proposed ARIMA-GBRT model provided accurate forecasting performance and lower values of the indices Root Mean Square Error (RMSE)and Mean Absolute Error (MAE).Although the benefits of ARIMA-GBRT, it consumes a long time because GBRT built its trees sequentially. As presented in [28], a Hybrid Demand Forecasting System (HDFS)was proposed. HDFS consists of three methods called Improved Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (ICEEMDAN), Multi-Objective Grey Wolf Optimizer (MOGWO), and Support Vector Machine (SVM). Thus, the framework of this model was represented as data cleaning method-optimizer basic prediction model called ICEEMDAN-MOGWO-SVM. HDFS has a lot of real-world implementations with high prediction accuracy and stability.

III. ADVANCED SMART GRID (ASG)

The traditional smart grid cannot work in real time as it lacks to use IoT technology [29-31]. Actually, IoT produces a real time data that perfectly helps smart grids to perform a variety of tasks [32-34]. In smart grid, the electricity consumers have the ability to manage their energy consumption and use it wisely and efficiently [35]. Accordingly, demand management plays an essential role in the construction of smart grid. At the customer side, demand forecasting is an effective way to optimize the demand scheduling that represents as an advanced automation way of moving a part of the demand from peak to off peak in order to smooth the demand curve [36]. According to traditional smart grids, customer demand forecasting approach is unable to provide real time estimations due to the loss of real time data.

Hence, these traditional grids should be replaced by Advanced Smart Grids (ASGs) which have the ability to cope with real time actions depending on IoT technology and service areas depending on fog computing, as shown in figure1.In figure1, ASG consists of four main levels, which are; IoT, customer, service area, and cloud levels. In IoT level, there are IoT devices such as smart meters which supply the customers (e.g., smart home) with real time data to give them the ability to makefast and accurate decisions. In fact, the communication between the smart meters and the appliances in the system can be created by using wireless networking connections such as Bluetooth, Wi-Fi, etc... While IoT supports the customers with real time data, it cannot formulate this data in appropriate form for analysis. Additionally, the volume of the generated data through IoT will be increased to be larger in size [31-34].

Hence, it is an important to represent the received data at the customer side in a suitable form (e.g., 2-dimensional form) to be analyzed [32]. Data summarization process should be performed on the data in customer level to remove ineffective data and to prevent the data replication before using it to give real time actions. In customer level, the customer can analyze the collected data after performing data representation and data summarization to provide fast and accurate decisions such as demand forecasting. Additionally, every customer will send a copy of data in its cache server to be stored temporarily in the cache server of its closed fog at the nearest service area. Customers can communicate to the nearest fog to store the current amount of data in which fog has a range of network equipment such as smart gates, routers, switches and other similar devices [32,37,38].

According to fog level, fog can support the customer to take quick and effective action on the stored data in its cache. Fog can summarize the data in its cache to be more informative before sending it to the master fog in its service area. In fact, fog can be communicated to customer, another fog in its service area, and master fog of its area through wireless networking connections to support the customers. Additionally, each master fog is able to be connected to another master fog to support the fogs in its service area. The main aim of providing the service area as a middle level between the customers and the cloud level is to give fast and accurate actions coping with real time interactions. Sending all of the data collected by IoT to the cloud will exhaust network bandwidth and will not be viable [32-34]. Cloud is unable to support customers in the customer level to make a real time actions, but it is able to perform other essential tasks in smart grids such as future load prediction. Thus, master fogs have been communicated to the parent cloud to send the collected data from their cache servers to the cloud servers to be permanently stored according to many rules from the cloud.

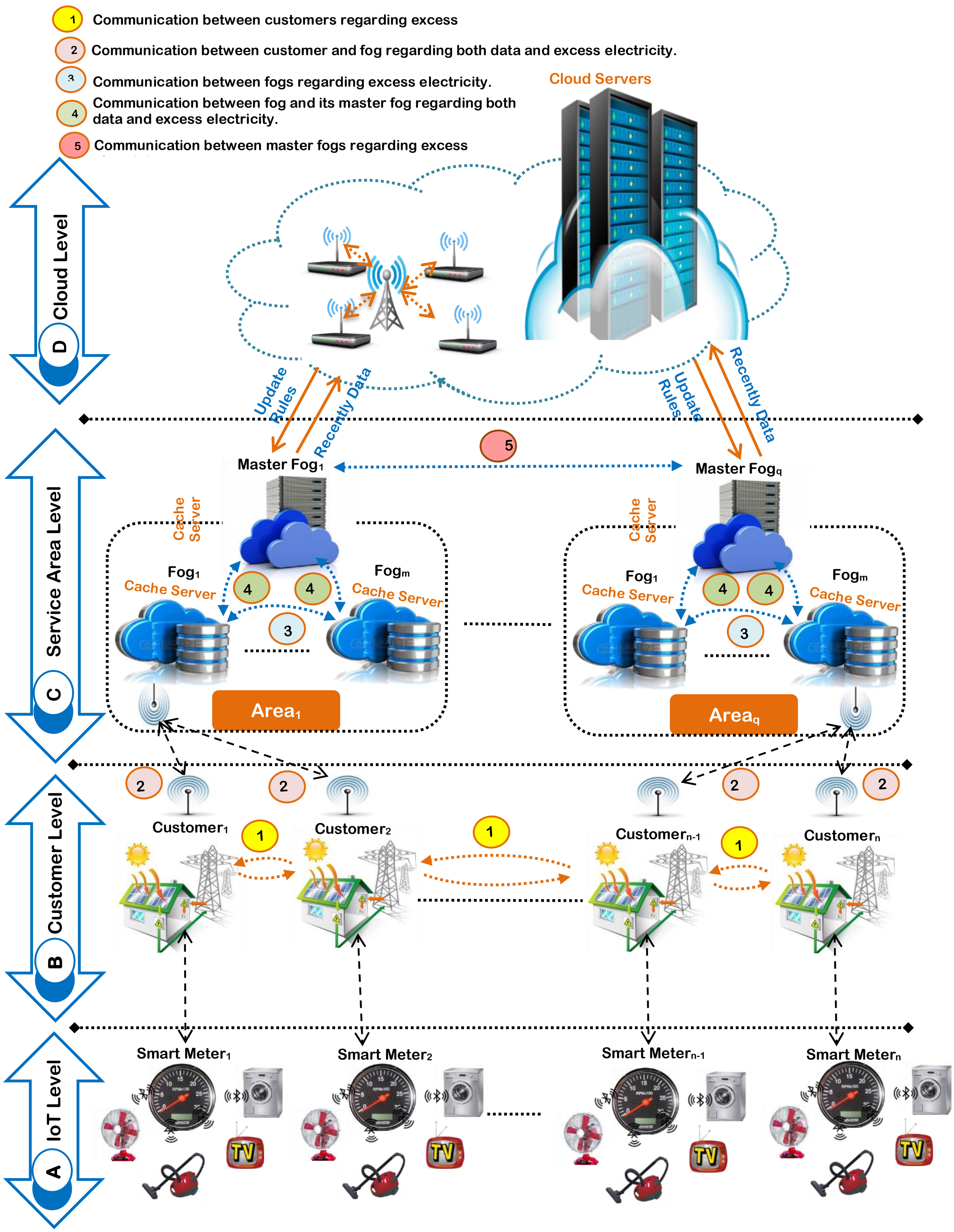

Figure 1. Architecture of the advanced smart grid.

This paper focuses on demand forecasting at the customer side (e.g., smart home). In fact, smart home is an important component in ASG in which it includes the complete IoT infrastructure that composes of both smart devices (e.g., smart meters) and renewable energy sources (e.g., solar, wind etc.) [23,27]. Demand forecasting is an essential task to fulfill the energy demand at the customer. Demand forecasting enables customers to efficiently manage the electricity demand in which they use less energy during the peak time [23-26]. Thus, electricity production expenses will be reduced, and reliable supply will be ensured. Based on the demand forecasting, customer knows if he can self-sufficient with the amount of electricity he has, or he will become a victim customer.

Victim customer begins to communicate with his neighbors in the customer level to buy his need of electricity. If the neighbors cannot support the customer with his need, he starts to ask the closed fog for buying the amount of electricity he needs. In this case, fog either supports the customer with his need or asks other fogs in the fog level to provide this amount of electricity. In the case that these fogs are unable to provide the required electricity, the fog communicates to its master fog to provide its required. Master fog either provides the amount of electricity to the fog and then this fog buys this amount to the customer or asks other master fogs to buy the required electricity. If master fogs are unable to provide the request, then the victim customer should reduce his usage of electricity, otherwise the power will be cut off.

IV. THE PROPOSED CUSTOMER DEMAND FORECASTING (CDF) STRATEGY

The details of CDF strategy will be discussed in this section. CDF strategy aims to provide fast and accurate demand forecasting at the customer side. Actually, fast and accurate demand forecasting helps customers to efficiently manage the electricity demand in which they use less energy during the peak time [23-26]. The input of CDF strategy is an electrical dataset in the form of a set of features that affect electrical demands. Then, irrelevant features should be eliminated by using a suitable feature selection method to select the most informative features that enable the demand forecasting model to perform its tasks well. CDF technique composes of two stages in figure 2, which are; (i) Feature Selection (FS) stage, with (ii) Demand Forecasting (DF) stage. The details of each stage will be discussed in the following subsections. The symbols that used in this paper is presented in table 1.

TABLE 1
NOMINATIONS OF ALL SYMBOLS IN THIS PAPER.

Symbols	Definition
$P_P(P_i)$	The optimal position of each i^{th} particle.
P_i	The particle's current position of i^{th} particle.
P_{pi}	The particle's optimal personal position of i^{th} particle.
$Fitness(P_i)$	The fitness value of the i^{th} particle.
$Fitness(P_{pi})$	The fitness value of the i^{th} particle based on particle's optimal position.
P_G	The optimal particle in swarm.
$Fitness(P_{pi+1})$	The fitness value of the $(i+1)^{th}$ particle.
t	The present iteration
$V_i(t+1)$	The velocity of i^{th} particle during the subsequent iteration.
$V_i(t)$	The velocity of i^{th} particle at the present iteration .
w	The inertia weight; $w\in[0.9\text{-}1.2]$
c_1 and c_2	The cognitive and social acceleration constants; $c_1,c_2\in[2\text{-}4]$
$r_1$and r_2	Random numbers; $r_1,r_2\in[0\text{-}1]$
TD	Training demand data set contents of training samples and its features, TD^2S = (T, F).
T	Training samples.
F	Training samples or testing samples features.
LTS	Labels of training samples, $LTS=L_{ts1};\ldots;L_{tsR}$.
t	No. of samples in training dataset, t=|LTS| or T|.
DT	Demand of training samples, $DT=DT_1;\ldots;DT_t$
ED^2S	Tested demand dataset contents of testing samples and its features, ED^2S = (S, F)
ES	Testing samples.
LES	Labels of testing samples, $LES=LE_{s1};\ldots;LE_{sh}$
h	No. of samples or labels in examining data set, h=| LESI I| or | ES |.
DE	Demand of testing samples, $DE=DE_1;\ldots;DE_s$
m	No of features in training and testing dataset, m=|F|.
RC	Response classes of electrical demand ; RC= c_1; c_2; …; c_R
R	No. of classes in the system, R=|RC|.
FD	Forecasted demand, $FD=fd_1;\ldots;fd_h$.
S_h	Testing samples belongs to ES , $S_h \in$ ES
T_j	Training samples belongs to T, $T_t \in$ T
f_i	Feature of used sample.
W_i	Weight of used sample
C_r	Class belongs to classes of electrical demand, $C_r \in$ RC
Cv_r	Center vector of class of electrical demand, $Cv_r \in C_r$.
W_{itr}	Class C_r probability given the feature value f_i of training samples T_j.
W_{ihr}	Class C_r probability given the feature value f_i of testing sample ES_h.
Ed (ES_h,T_j)	Space between testing sample ES_h and training sample T_j at all features.
FD(ES_h)	Forecasted demand of testing samples ES_h
K	Neighbor's number.
W_{iCvr}	Center vector probability given the feature value f_i of training samples T_j
WS	Weight space of m-dimensions
KNTC	Training samples from T with smallest distance with center vector Cv_r.
a	The feature.
y	The training sample.
r	Class labels.
$P(f_{ayr}\|c_r)$	*The possibility of producing the feature* f_{ayr} given the class c_r.
VH	Very-High demand class.
H	High demand class.
M	Medium demand class.
L	Low demand class.
VL	Very-Low demand class.

(continued on the next page)

(TABLE 1: continued)

Symbols	Definition
C_{VH}	The center of Very-High demand class.
C_H	The center of High demand class.
C_M	The center of Medium demand class.
C_L	The center of Low demand class.
C_{VL}	The center of Very-Low demand class.
C_β	The center of class "β".
α	The number of samples.
$T_{v\beta}$	The training sample that belongs to β class.
C_v	The weighted Center Vector .
Cv_r	The weighted center of class $c_r \in RC$.
Ed	Euclidean distance.
T_{tr}	Sample of the training sample.
$KNTCc_r$	The KNTC according to class c_r.
T_{NTCKr}	The distance among testing sample S_h and the $KNTCc_r$
w_{xhr}	The numerical weighted values for S_h at x^{th} feature in weight space according to class c_r.
W_{xNTCKr}	The numerical weighted values for the weights of t_{ntckr} at x^{th} feature in weight space according to class c_r.
n	The number of classes.
$Ed\ (T_{NTCKr}, S_r)$	The Euclidean Distance between the center of class t_{ntckr} and the testing sample S_r.
$BD_r\ (S_r)$	The smallest Bay Degree (BD) for the testing sample belongs to the corresponding class c_r.
$Ed\ (S_h, T_t)$	The Euclidean Distance between a examined sample S_h a training samples T_t in the m-dimensional weight space of c_r
w_{ihr}	The weights of the i^{th} feature of S_h in the weight space associated to class c_r.
w_{itr}	The weights of the i^{th} feature of T_t in the weight space associated to class c_r.
KAS	K akin samples, $kas=\{t_{1as}, t_{2as}, t_{3as}, \ldots\ldots, t_{kas}\}$.
$FD(S_h)$	The average demand of the identified samples in K Akin Samples (KAS).

A. *Feature Selection (FS) Stage*

Usually, over fitting problem may be caused in smart grids because of the existence of non-informative features in the electrical dataset [29-34]. FS stage in CDF strategy aims to identify the most important features that enable the demand forecasting model to produce quick and reliable results. In fact, extraneous features can reduce the demand forecasting model's accuracy. Thus, it is a necessary operation to eliminate a subset of features that have a little effect on the output before starting to train the demand forecasting model. This step aims to enhance the demand forecasting model's performance, making it more rapid and more cost-effective. Feature selection techniques can be categorized into two main categories, which are; filter and wrapper [29-34]. While filter can quickly select the effective features on the demand forecasting model, it cannot accurately perform its process. On the other hand, wrapper is able to accurately select relevant features but it so slow. Thus, it is an important to utilize the benefits of both types of feature selection methods to rapidly and precisely choose the most effective features on the demands in the smart grid. In this section, a Hybrid Feature Selection (HFS) method as a simple but effective feature selection method is used to choose more important features as shown in figure 3.

HFS consists of two layers called; fast layer and accurate layer. Fast layer aims to quickly select the informative features based on using Information Gain (IG) as a filter method [32]. Accurate layer aims to precisely choose optimum subgroup of the selected features from fast layer. In accurate layer, Binary Particle Swarm (BPS) optimization is used as a wrapper method [39,40]. In fast layer, to implement IG, entropy should be calculated because entropy is a measure of impurity in a training dataset [32]. IG is a symmetrical measure that can reflect extra details about *SS* as well as *KK* that is the degree by which the entropy of SS in which IG can be measured reduces by using (1) [32].

$$IG = H(SS) - H(SS\,|KK) = H(KK) - H(KK\,|SS) \qquad (1)$$

In accurate layer, BPS is implemented based on the subset of features selected from the fast layer. BPS extended the original Particle Swarm Optimization (PSO) by employing a transfer function, called sigmoid function, which converts velocities from the continuous search space into binary space [39,40]. The implementation of BPS passes through many steps as shown in figure 3. Initially, BPS starts by initializing the swarm based on the subgroup of selected features from the fast layer to initialize each particle in swarm. The position of each particle in swarm will be represented in binary space in which every bit will be one or zero value that donated to the selection of feature or not. Then, all particles in the swarm should be evaluated by using a fitness function that indicates the precision of selected classifier such as NB as a standard classifier. Based on fitness values, every particle will update its personal position and also the global position as the best position in swarm will be determined by using (2) and (3) [39,40].

$$P_P(P_i) = \begin{cases} P_i & if(\,Fitness(P_i) > Fitness(P_{pi})) \\ P_{pi} & otherwise \end{cases} \qquad (2)$$

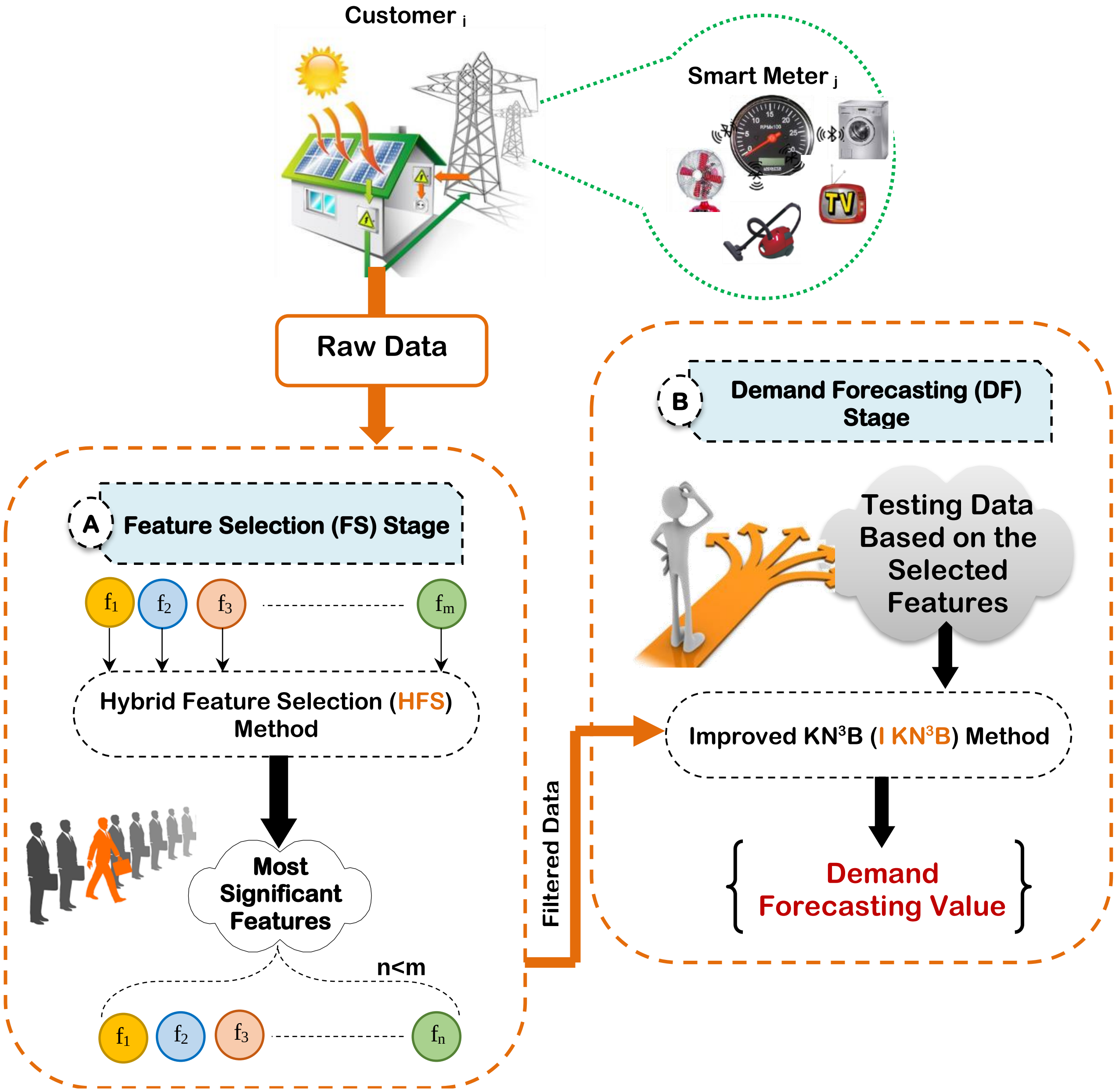

Figure 2. Demand forecasting strategy in smart grid.

$$P_G = \begin{cases} P_{pi} & if\,(Fitness(P_{pi}) > Fitness(P_{pi+1})) \\ P_{pi+1} & otherwise \end{cases} \quad (3)$$

Where $P_P(P_i)$ denotes the optimal position of each i^{th} particle and P_i is the particle's current position of i^{th} particle. P_{pi} also denotes the particle's optimal personal position of i^{th} particle. $Fitness(P_i)$ denotes the fitness value of the i^{th} particle. $Fitness(P_{pi})$ denotes the fitness value of the i^{th} particle based on particle's optimal position. Additionally, P_G is the optimal particle in swarm and $Fitness(P_{pi+1})$ is the fitness value of the $(i+1)^{th}$ particle. Then, the velocity of each particle will be calculated based on the personal and global positions using (4) [39,40].

$$V_i(t+1) = w * V_i(t) + \left(c_1 r_1 \left(P_{pi}(t) - P_i(t)\right)\right) + \left(c_2 r_2 \left(P_G(t) - P_i(t)\right)\right) \quad (4)$$

Where t denotes the present iteration and $V_i(t+1)$ denotes the velocity of i^{th} particle during the subsequent iteration. $V_i(t)$ is the velocity of i^{th} particle at the present iteration. w is the inertia weight; $w \in [0.9\text{-}1.2]$ [39,40]. c_1 and c_2 are the cognitive and social acceleration constants; $c_1, c_2 \in [2\text{-}4]$ and r_1 and r_2 are random numbers; $r_1, r_2 \in [0\text{-}1]$. After the calculation of each particle's velocity, sigmoid function is applied as a transformation function that convert the positions from continuous to binary using (5).

$$P_i^j(t+1) = \begin{cases} 0 & if \quad rand(0,1) \geq sigmoid(V_i^j) = \dfrac{1}{1+e^{-V_i^j}} \\ 1 & otherwise \end{cases} \quad (5)$$

Every particle inside the swarm is assessed using the fitness function depending on its new position. These calculations are repeated until the terminal condition is met. PG is the result and the process comes to an end. All features which have one value in this particle indicate the most important features. Then, the electrical dataset based on the selected features will be passed to the next stage called DF stage to allow the provided methodology to quickly provide accurate results.

B. *Demand Forecasting (Df) Stage*

In this subsection, a new demand forecasting methodology is introduced to quickly offer precise values based on the subset of features selected from the previous stage called HFS stage. The proposed methodology is called Improved KN^{3}B (IKN^{3}B) that aims to improve the features of KNN classier. In fact, IKN^{3}B combines Naïve Bayes (NB) as a weighted method and KNN classifier. Generally, KNN algorithm is easy understanding and implemented in which it is widely used to nonlinear problems solving [41,42]. Additionally, it can reduce the variables influences on the experimental processes [43].

KNN is able to perform effectively in applications where the sample contains many class labels [44]. Although the benefits of KNN, KNN is a lazy learner as all training items are stored and a classifier is not generated till a new item needs to be classified [45]. KNN as a lazy leaner requires more computation time during the classification process [46,47]. For classifying a data item *X*, its K-nearest neighbors are tested and then *X* is allocated to class label to which majority of its neighbors belongs to.

In KNN algorithm, choosing of the hyper parameter k is important. A given single value of K is used to determine the total number of nearest neighbors to classify the class label for unknown item. If k is too small, the algorithm would be more sensitive to outliers. Also, if k is too large, the neighborhood may include too many points from other classes. Thus, the choice of k affects the performance of KNN algorithm [44,48]. In this work, KNN has been enhanced to utilize its advantages and to avoid its mentioned issues. Hence, IKN^{3}B is a modified version of KNN that integrates KNN with NB to provide accurate outcomes. Based on various weighted spaces, NB is used to allocate the class of the nearest training samples to a testing sample in which each space belongs to one of the response [30,49]. Hence, the existence of '*R*' response classes indicates that there are '*R*' of weight spaces. Given every response class, every weight space evaluates the input training data items depending on its feature values. Suppose that there are '*m*' of features and '*R*' of response classes; $RC=\{C_1, C_2, C_3, ..., C_R\}$.

The main aim of using NB is to calculate the conditional probability of a tasted sample in a Feature Space (FS) described by the set; $F=\{f_1, f_2, f_3, ..., f_m\}$ belongs to a response class C_r. Thus, it is an important to use a Weighted Function (WF) which transfers all training sample's feature vector to the relevant weight for every response class. Thus, it is an important to implement WF for all feature of any training sample depending on its response class by implementing NB before using IKN^{3}B classifier [30,32]. Accordingly, *WF:* $f_{ayr} \rightarrow w_{ayr}$, where *a* is the feature, *y* is the training sample, and *r* is class labels. Suppose that the number of weight spaces is '*r'* equals the size of C_r. After that, w_{ayr} can be calculated using (6).

$$W_{ayr} = P(c_r|f_{ayr}) = P(c_r) * P(f_{ayr}|c_r) \quad (6)$$

Where c_r is the possibility of class incidence and $P(f_{ayr}|c_r)$ *is the possibility of producing the feature* f_{ayr} given the class c_r. Thus, to provide the prediction demand of the testing sample; $S(f_1, f_2, ..., f_m) = (f_{1s}, f_{2s}, ..., f_{3s}, f_{ms})$, WF is used as a diversion process to transfer it to each of the available '*r'* weight spaces. For the electrical demand data samples (D^{2}S) divided to a training data sample of '*t*' samples*;* $T=\{T_1, T_2, T_3, ..., T_t\}$ and testing data samples of '*h*' items; $ES=\{S_1, S_2, S_3, ..., S_h\}$. All samples of $T_t \in T$ and $S_h \in ES$ is uttered as an ordered set of '*m*' features; $T_t(f_1, f_2, f_3, ..., f_m)=[f_{1t}, f_{2t}, f_{3t}, ..., f_{mt}]$ and $S_h(f_1, f_2, f_3, ..., f_m)=[f_{1h}, f_{2h}, f_{3h}, ..., f_{mh}]$. Depending on the number of responding classes, the feature space is converted into the respective weight spaces. The conversion of the feature values of T_t to the associated weights in the weight spaces is accomplished using (1). There are *R* weight spaces for response classes (e.g., $RC=\{c_1, c_2, c_3, ..., c_R\}$), For each class $c_r \in$RC uttered in the weight space of c_r by implementing (1). For the space associated to the class c_r, the training sample $T_t(f_1, f_2, f_3, ..., f_m)=[f_{1t}, f_{2t}, f_{3t}, ..., f_{mt}]$, which belongs to class c_r, will be uttered as; $T_t(w_{1r}, w_{2r}, w_{3r}, ..., w_{mr})|_{class=cr}=[w_{1tr}, w_{2tr}, w_{3tr}, ..., w_{mtr}]$.

The feature values of $S_h(f_1, f_2, f_3, ..., f_m)=[f_{1h}, f_{2h}, f_{3h}, ..., f_{mh}]$ adjusted to the associated weights for any class $c_r \in$RC, then uttered as $S_h(w_{1h}, w_{2h}, w_{3h}, ..., w_{mh})|_{class=}c_r = [w_{1hr}, w_{2hr}, w_{3hr}, ..., w_{mhr}]$. Then, for each corresponding weighted KNN class, it is an important to determine the weighted center. The expected demand for the CDF problem might be "Very-High," "High," "Medium," "Low," or "Very-Low," which are indicated as VH, H, M, L, and VL, respectively, and stated by the Response Classes series as; $RC=\{VH, H, M, L, VL\}$. The center of each class shall be specified by (7), represented as; C_{VH}, C_H, C_M, C_L, C_{VL}, respectively whereas C_β is the center of class "β". C_β can be calculated by (7), considering the class 'β' that has α samples represented as; $T_\beta=\{T_{1\beta}, T_{2\beta}, T_{3\beta}, ..., T_{\alpha\beta}\}$.

$$C_\beta = \frac{1}{\alpha}\sum_{v=1}^{\alpha} T_{v\beta} \quad (7)$$

Where α is the number of samples and $T_{v\beta}$ the training sample that belongs to β class. Hence, all data sample $\in D^2S$ are uttered in the m-dimensional weight space. Generally, data samples in the weight space from the same class is logically closed with each other. Then, based on the data samples of each RC, the center of each class is found as a weighted Center Vector (C_v).Considering the class $c_r \in RC$ that has α samples uttered as; $T_r=\{T_{1r}, T_{2r}, T_{3r}, \ldots, T_{\alpha r}\}$ (e.g., $|N_r|=\alpha$). Data sample T_{tr} is uttered as; $T_{tr}(w_{1r}, w_{2r}, w_{3r}, \ldots, w_{mr})|_{class=cr}=[w_{1tr}, w_{2tr}, w_{3tr}, \ldots, w_{mtr}]$while the weighted center of class $c_r \in RC$ is uttered as; Cv_r $(w_1, w_2, w_3, \ldots, w_m)$ $|_{class=cr}$ $=[w_{1Cvr}, w_{2Cvr}, w_{3Cvr}, \ldots, w_{mCvr}]$.Thus, KNN has been improved by transform the data samples from feature space to weight space by using NB to find the relation between features and response classes.

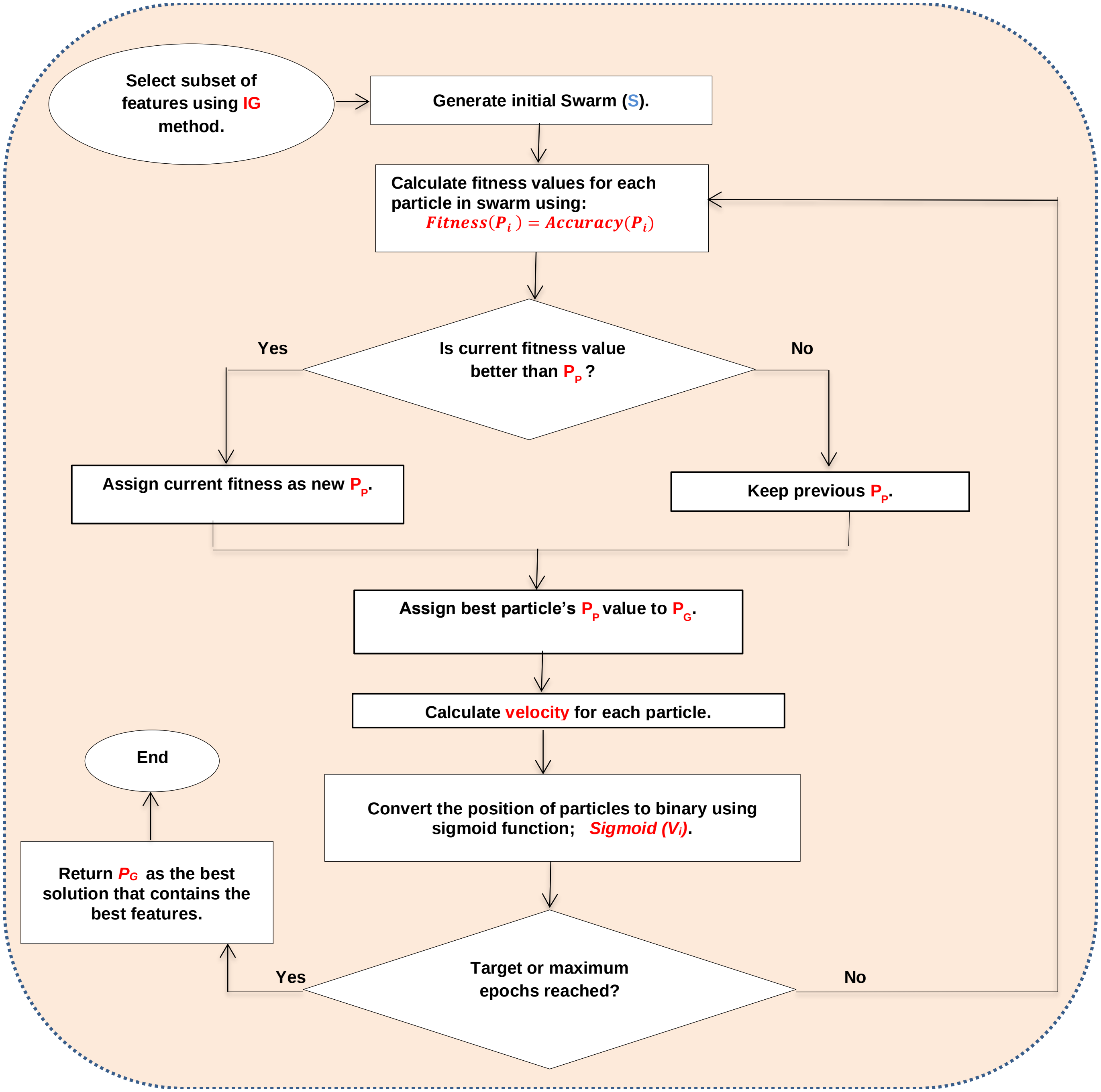

Figure 3. Steps of implementing HFS method.

Additionally, the problem of determining the suitable value of K has been solved by calculating the center of response classes, and then each testing items could be assigned to its relevant class depended solely upon its nearest training samples to the center of their class. To implement IKN^{3}B, Euclidean Distance is evaluated at first to get the distance in between $c_r \in RC$ class's weighted center vector and each sample of the training sample $T_t \in N$ by using (8).

$$Ed\,(Cv_r,T_t)|_{class=Cr} = \sqrt{\sum_{x=1}^{\alpha}(W_{xCvr}-W_{xtr})^2} \quad (8)$$

Where Cv_r is the weighted center vector of class; $Cv_r(w_1, w_2, w_3, ..., w_m)|_{class=cr} = [w_{1Cvr}, w_{2Cvr}, w_{3Cvr}, ..., w_{mCvr}]$. T_{tr} is sample of the training sample; $T_{tr}(w_{1r}, w_{2r}, w_{3r}, ..., w_{mr})|_{class=cr} = [w_{1tr}, w_{2tr}, w_{3tr}, ..., w_{mtr}]$. Corresponding the K-Nearest-To-Center (*KNTC*) of each response classes Cv_r with the smallest Ed $(Cv_r, T_t)|_{class=c_r}$, are defined and represented by the set; $KNTC = \{T_{NTC1}, T_{NTC2}, T_{NTC3}, ..., T_{NTCK}\}$ where *KNTC* is K-nearest training samples to the center in the considered '*r*' weight spaces. The distance between the testing sample S_h and the KNTC of every class is measured using Euclidean distance. Taking the space according to class c_r, the distance among testing sample $S_h(w_{1r}, w_{2r}, w_{3r}, ..., w_{mr})|_{class=cr} = [w_{1hr}, w_{2hr}, w_{3hr}, ..., w_{mhr}]$and the $KNTCc_r$ is uttered as; $T_{NTCKr}(w_{1r}, w_{2r}, w_{3r}, ..., w_{mr})|_{class=cr} = [w_{1NTCKr}, w_{2NTCKr}, w_{3NTCKr}, ..., w_{mNTCKr}]$. The Euclidean distance between each testing sample S_h and the center of class *KNTC* in c_r's m-dimensional weight space $Ed(T_{NTCKr}, S_h)|_{class=c_r}$ can be calculated using (9).

$$Ed\,(T_{NTCKr}, S_h)|_{class=c_r} = \sqrt{\sum_{x=1}^{m}(W_{\mathrm{xNTCKr}} - W_{xhr})^2} \qquad (9)$$

Where$W_{xNTCKr}$and w_{xhr} are the numerical weighted values for the weights of T_{NTCKr} and S_h respectively at x^{th} feature in weight space according to class c_r. The smallest Bay Degree (BD) for the testing sample belongs to the corresponding class c_r can be calculated using (10).

$$\mathrm{BD}_r(S_r) = \frac{1}{n}\sum_{k=0}^{n}\mathrm{Ed}(T_{NTCKr}, S_r) \qquad (10)$$

Where n is the number of classes and $Ed\,(T_{NTCKr}, S_r)$ is the Euclidean Distance between the center of class T_{NTCKr} and the testing sample S_r. For ability to Forecast the Demand (FD) responding to S_h, Euclidean Distance is applied to measure the distance between the examined sample S_h and the training samples For every response space. Taking into account the space relating to class c_r, the distance in between a examined sample $S_h(w_{1h}, w_{2h}, w_{3h}, ..., w_{mh})|_{class=cr} = [w_{1hr}, w_{2hr}, w_{3hr}, ..., w_{mhr}]$and a training samples $T_t(w_{1r}, w_{2r}, w_{3r}, ..., w_{mr})|_{class=cr} = [w_{1tr}, w_{2tr}, w_{3tr}, ..., w_{mtr}]$ in the m-dimensional weight space of c_r using (11).

$$\mathrm{Ed}\,(\mathrm{S_h}, \mathrm{T_t})|_{class=cr} = \sqrt{\sum_{i=1}^{m}(W_{ihr} - W_{itr})^2} \qquad (11)$$

Where w_{ihr} and w_{itr} are the values obtained for the weights of the i^{th} feature of S_h and T_t in the weight space associated to class c_r. To calculate FD associated to S_h, which is the average demand of the identified samples in K Akin Samples (KAS), the training *KAS* are established and described by the set in the considered R weight spaces (R represents response classes); $KAS = \{T_{1AS}, T_{2AS}, T_{3AS}, ..., T_{KAS}\}$ whereas K is number of Akin Samples with the smallest distance, then FD conforming to S_h would be the mean demand of the samples described in KAS, which can be dictated by (12). In algorithm 1, the general steps of the Improved KN^3B (IKN^3B) model are illustrated.

$$FD(S_h) = \Sigma \forall T \subset Sy \in KAS\; demand(T_{yAS}) / k \qquad (12) \qquad (12)$$

To explain the idea, suppose two response classes, "*High*" and "*Low*" respectively, as indicated by; "*H*" and "*L*." Consider of the space of two dimensional features as; $F=\{f1, f2\}$. Figure 4 shows the steps of implementing IKN^3B, suppose K=5.

V. EXPERIMENTAL RESULTS.

The introduced CDF strategy will be implemented in this section. CDF composes of two stages; FS stage using HFS technique and DF stage using IKN^3B techniques. The EU-27 electricity consumption dataset collected from Europe is used in our implementation [50,51,53].EU-27 electricity consumption dataset is a dataset based on the internet that was utilized to corroborate the findings in this study. EU-27 electricity consumption dataset is split into two groups; training and testing sets. The training set is used to learn the demand forecasting model, while the testing set is used to assess the proposed model's accuracy. The implementation of CDF strategy will be passed through two scenarios. In the first scenario, HFS technique will be implemented and compared to the most recent feature selection methods. In the second scenario, the whole strategy called CDF will be implemented and compared to the most recent demand forecasting strategies. To evaluate the model's performance, many metrics based on the confusion matrix will be used [30]. As shown in table 2 These metrics are accuracy, precision, recall, and error [30,31]. Table 3 shows the applied parameters and their corresponding implemented values.

C. *Testing Hybrid Feature Selection (HFS) Method.*

This paper presents HFS technique that is comprised of two layers called; fast layer and accurate layer. While fast layer try to quickly select the informative features based on using IG as a filter method, accurate layer aims to accurately select the best subset of the selected features from fast layer depending on the use of BPS as a wrapper method. To demonstrate the efficiency of the proposed HFS technique, it is compared to many features selection techniques. These techniques are Binary Genetic Algorithm-Gaussian Process Regression (BGA-GPR) [53], Genetic Algorithm (GA) [54], and Least Absolute Shrinkage and Selection Operator (LASSO) [55]. Table 4 shows the applied techniques. NB classifier as a demand forecasting model has been used to evaluate the feature selection techniques. Results are illustrated in figure 5, figure 6, figure 7 and figure 8.

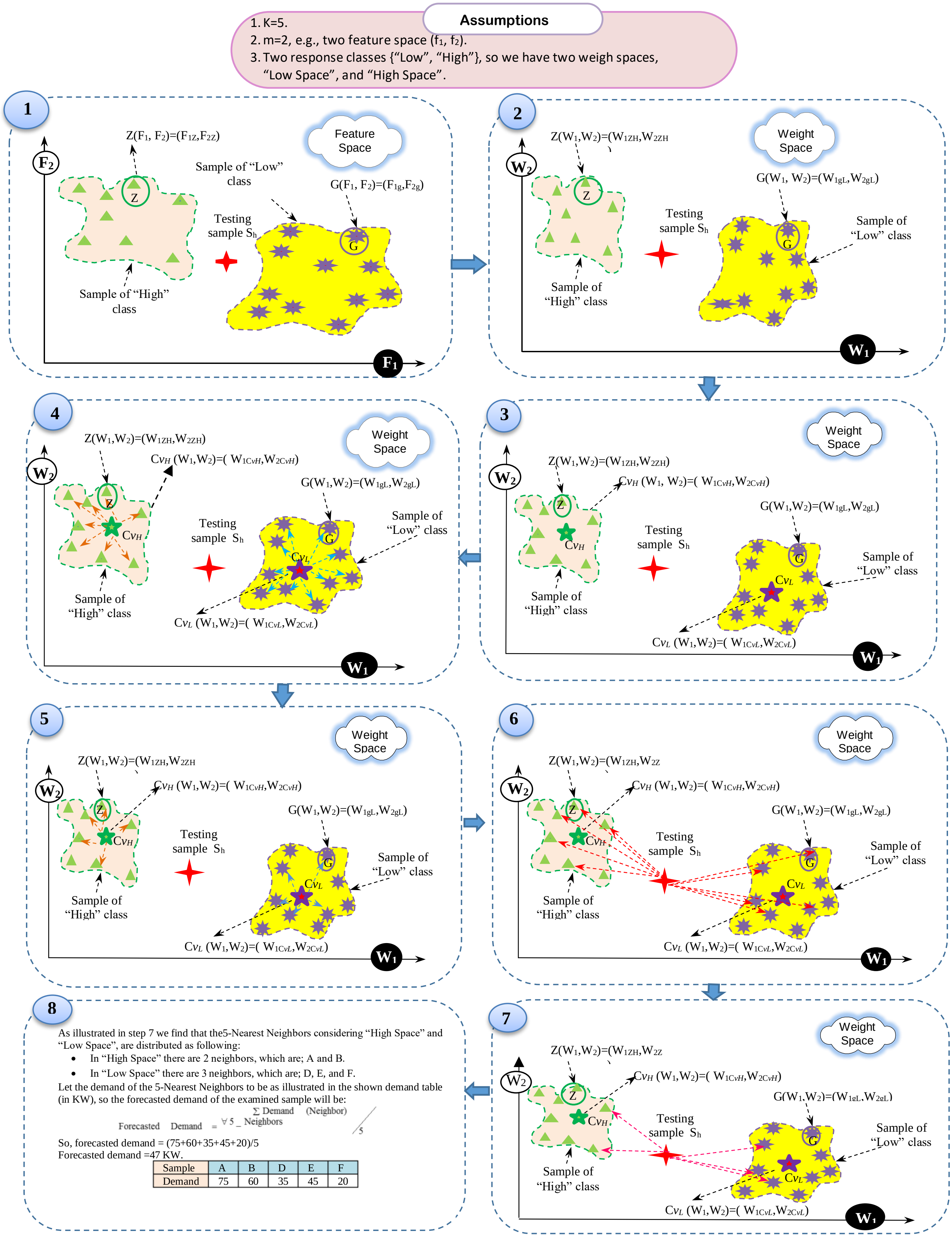

Figure 4. Demand forecasting using IKN³B.

Demand Forecasting Using IKN³B Algorithm

- ***Inputs:***
  - *TD^2S = (T, F); Training demand dataset.*
  - *LTS=L_{sr1};…..;L_{srt} ; Training samples labels.*
  - *t=|LTS| or |T|; No. of samples or labels in training dataset.*
  - *DT=DT_1;….;DT_t, Training samples demand.*
  - *ED^2S= (ES, F); Testing dataset.*
  - *LES=LE_{s1};….;LE_{sh}; Testing samples labels.*
  - *h=|LTE| or |ES|; No. of samples or labels in testing demand dataset.*
  - *DE=DE_1;…;DE_h; Demand of testing samples.*
  - *m=|F|; No. of features in training and testing demand dataset.*
  - *RC=c_1;…..;c_R ; Response classes of electrical demand.*
  - *R=|RC|; No. of classes in the system.*
  - *KNTC= K-Nearest to center samples.*
  - *K=No. of neighbors.*
- ***Output:***
  - *FD=fd_1;…….;fd_h ; forecasted demand .*
- ***Steps:***

/* Training samples weight (probability) calculation of each feature with the class by using Naïve Bayes*/

For every $T_j \in$ T
  For every $f_i \in$ F
    For every $C_r \in$ RC
      Calculate $W_{itr} = P_{Ti,fj}(C_r \mid f_{ij}) = P(C_r)*P(f_{ij}|C_r)$.
    End for
  End for
End for

/* Examining samples weight (probability) calculation of each feature with the class by using Naïve Bayes*/

For every $S_h \in$ ES
  For every $f_i \in$ F
    For every $C_r \in$ RC
      Calculate $W_{ihr} = P_{Sh,fi}(C_r \mid f_{ih}) = P(C_r)*P(f_{ih}|C_r)$.
    End for
  End for
End for

/* Determine the center for each response class at WS*/

For every Cr $\in$ RC
  For every $T_t \in$ T
    For every $W_i \in$ W
      Calculate $C_{Vr} = \frac{1}{\alpha}\sum_{i=1}^{\alpha} T_{tr}$.
      $Cv_r\,(w_1, w_2, w_3, \ldots, w_m)\,|_{class=cr} = [\,w_{1Cvr}, w_{2Cvr}, w_{3Cvr}, \ldots, w_{mCvr}]$
    End for
  End for
End for

/* Space calculation between weight (probability) of training samples and it's corresponding Cv_r*/

For every $Cv_r \in C_r$
  For every $T_j \in$ T
    For every $W_i \in$ W
      Calculate Ed (Cv_r, T_j) = $\sqrt{\sum_{i=1}^{m}(W_{icvr} - W_{ij})^2}$
    End for
  End for
End for

//Corresponding K-nearest to center Cv_r for each response class.

For every $Cv_r \in C_r$
  For every $T_j \in$ T
    Calculate KNTC (Cv_r) =K of training samples with smallest Ed (Cv_r,T_j).
  End for
End for

// Calculate Euclidian distance for each testing sample in ES with KNTC.

For every $K_t \in$ KNTC
  For every $S_h \in$ ES
    Calculate Ed (K_t, S_h) = $\sqrt{\sum_{i=1}^{m}(W_{it} - W_{ih})^2}$
  End for
End for

Algorithm Parameters	
TD	Training demand data set contents of training samples and its features, TD^2S = (T, F).
T	Training samples.
F	Training samples or testing samples features.
LTS	Labels of training samples, LTS=L_{ts1};…..;L_{tsR}.
t	No. of samples in training dataset, t=|LTS| or T|.
DT	Demand of training samples, DT=DT_1;….;DT_t
ED^2S	Tested demand dataset contents of testing samples and its features, ED^2S = (S, F)
ES	Testing samples.
LES	Labels of testing samples, LES=LE_{s1};….;LE_{sh}
h	No. of samples or labels in examining data set, h=| LESI I| or | ES |.
DE	Demand of testing samples, DE=DE_1;…;DE_s
m	No of features in training and testing dataset, m=|F|.
RC	Electrical demand classes; RC= c_1; c_2; …; c_R.
R	No. of classes in the system, R=|RC|.
FD	Forecasted demand, FD=fd_1;……;fd_h.
S_h	Testing samples belongs to ES , $S_h \in$ ES
T_j	Training samples belongs to T, $T_t \in$ T
f_i	Feature of used sample.
W_i	Weight of used sample
C_r	Class belongs to classes of electrical demand, $C_r \in$ RC
Cv_r	Center vector of class of electrical demand, $Cv_r \in C_r$.
W_{itr}	Class C_r probability given the feature value f_i of training samples T_j.
W_{ihr}	Class C_r probability given the feature value f_i of testing sample ES_h.
Ed (ES_h,T_j)	Space between testing sample ES_h and training sample T_j at all features.
FD(ES_h)	Forecasted demand of testing samples ES_h.
K	Neighbor's number.
W_{iCvr}	Center vector probability given the feature value f_i of training samples T_j
WS	Weight space of m-dimensions
W	Training samples or testing samples weight.
KNTC	Training samples from T with smallest distance with center vector Cv_r.

// Calculate Euclidian distance for each testing sample in ES with KNTC.

For every $K_t \in$ KNTC
  For every $S_h \in$ ES
    Calculate Ed (K_t, S_h) = $\sqrt{\sum_{i=1}^{m}(W_{it} - W_{ih})^2}$
End for
End for

// Calculate Euclidian distance for testing sample in ES.

For every $S_h \in$ ES
  For every $T_j \in$ T
    Calculate Ed (S_h, T_j) = $\sqrt{\sum_{i=1}^{m}(W_{ih} - W_{ij})^2}$
  End for
End for

// Determine KNN of training items for each testing sample.

For every $S_h \in$ ES
  For every $T_j \in$ T
    Calculate Neighbors (S_h) =K of training items with smallest Ed (S_h, T_j)
  End for
End for

// Calculate demand forecasting value for each testing sample.

For every $S_h \in$ ES
  FD (S_h) = $\frac{\sum_{i=1}^{k} DT(Nieghbors(Tins))}{K}$
End for

Measure an accuracy of proposed model based on FD and DE values.

Algorithm1: Demand forecasting using IKN³B.

TABLE 2
CONFUSION MATRIX

		Predicted Label	
		Positive	Negative
Known Label	Positive	True Positive (TP)	False Negative (FN)
	Negative	False Positive (FP)	True Negative (TN)

TABLE 3
CONFUSION MATRIX FORMULAS

Measure	Formula	Intuitive Meaning
Precision(P)	TP / (TP + FP)	The percentage of positive classifications that are correct.
Recall / Sensitivity(R)	TP / (TP + FN)	The percentage of positive classifications that are positive.
Accuracy(A)	(TP + TN) / (TP + TN + FP + FN)	The percentage of correct classifications.
Error(E)	1- Accuracy	The percentage of incorrect classifications.

As shown in figure 5, figure 6, figure 7 and figure 8, increasing the number of items in the training dataset improves the performance of all methods. At the maximum number of training items, the highest "Precision," "Recall," and "Accuracy," as well as the lowest "Error," are achieved, at the utmost number of items for training (e.g., 498 items). The reason is that IKN^{3}B depends on KNN that is straightforward classifier and NB as a weighted method to convert the dataset from feature space to weighted space. As a result of weighting the data before learning the forecasting model, the performance of demand forecasting method is enhanced. It should also be observed that the proposed HFS method has the best performance compared to other methods. True Positive (TP) and True Negative (TN) are therefore increased, whereas False Positive (FP) and False Negative (FN) are reduced. This improves the accuracy, precision and recall of the proposed selection method while reducing errors. BGA-GPR, GA, LASSO, and HFS reach to accuracy values equal 0.83,0.81,0.82, and 0.94 respectively at the maximum number of training data (e.g.,498).

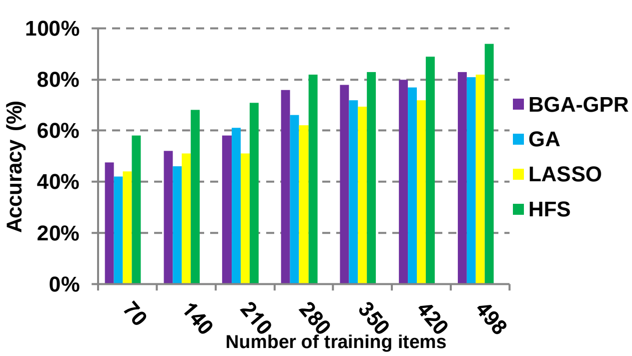

Figure 5. Accuracy of features selection methods using NB.

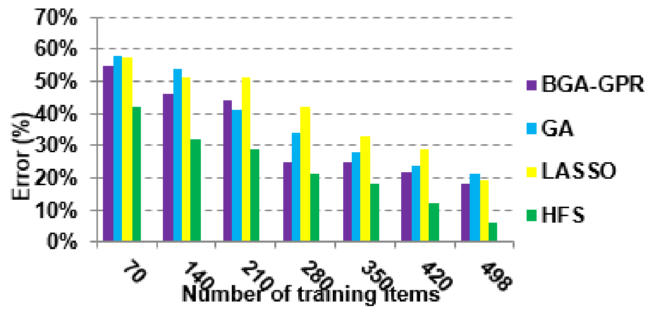

Figure 6. Error of features selection methods using NB.

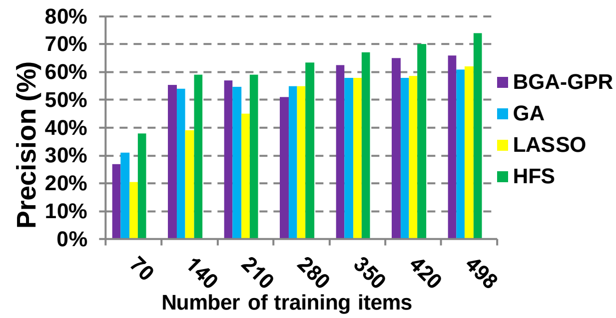

Figure7. Precision of features selection methods using NB.

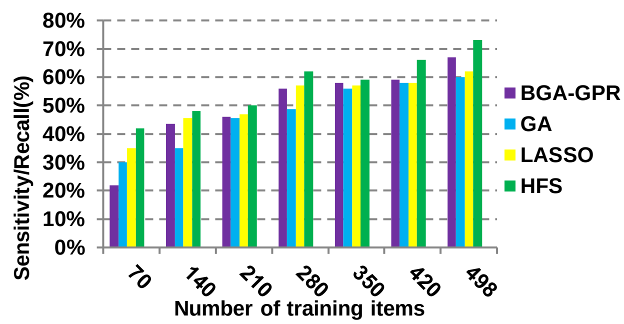

Figure 8. Sensitivity of features selection methods using NB.

HFS achieves the highest accuracy value by using IG as a filter method and then BPS as a wrapper method that improves the performance of the NB classifier as a demand forecasting model by precisely selecting the best features. Accordingly, BGA-GPR, GA, LASSO, and HFS algorithms provide error values reach to 0.18, 0.21, 0.19, and 0.06 respectively. HFS provides precision equals 0.74 while BGA-GPR, GA, and LASSO provide 0.66, 0.61, and 0.62 respectively. While the recall of HFS is 0.73, the recall of BGA-GPR, GA, and LASSO are 0.67, 0.65, 0.60, and 0.62 respectively. Hence, figure 5, figure 6, figure 7 and figure 8 demonstrate that HFS is superior to other current techniques, that are; BGA-GPR, GA, and LASSO because the greatest accuracy and the lowest error are reached by HFS**.**

D. Testing Customer Demand Forecasting (CDF) Strategy.

At the end, it's necessary to turn the proposed CDF **methodology** to the test, ensuring that the used HFS as a feature selection method and IKN^{3}B as a demand forecasting methodology working together well. To measure the performance of CDF methodology, it is examined with some of the most lately utilized demand forecasting methodologies.

TABLE 4:
SOME OF RECENT FEATURE SELECTION METHODS USED FOR EVALUATION.

Features Selection Technique	Description
Binary Genetic Algorithm-Gaussian Process Regression (BGA-GPR)[53]	In [53], BGA-GPR is a new features selection method which has been used to effectively discover the much more important and non-repetitive variables for valid and exact Electricity Demand Forecasting .Binary Genetic Algorithm (BGA) is implemented for the variable selection process and Gaussian Process Regression (GPR) is used to evaluate the fitness score of the variables.
Genetic Algorithm (GA) [54]	In [54], genetic algorithm (GA) is used to enhance feature sets for machine(ML) algorithms which don't include a built-in feature selection technique, Selection, crossover, and mutation are the three key operators in GA, which is a random search optimization process .The GA feature selection increases the efficiency among any algorithms as well as demonstrates that the most important parameters affecting heat demand prediction are historical heat demand, temperature, and "agent schedules," which are generated from major occupancy shifts in the building.
Least Absolute Shrinkage and Selection Operator (LASSO) [55]	In [55], LASSO is a useful way for selecting features aids in the reduction of problem size and the elimination of over fitting. By eliminating redundant variables, it is possible to reduce the number of dependent variables. The number of the absolute values of model variables is constrained by LASSO to be less than a given bound value. Using a regularization method stigmatizes the regression coefficients by decreasing some of them to zeros. Features are chosen after regularization depend on that their coefficient is non-zero. By regularizing and eliminating coefficients to reduce the number of irreverent elements, the LASSO approach can reduce variance without significantly raising bias. In[55] showed that, under the condition of maintaining classification precision, this approach greatly improves feature selection operation performance while reducing modeling and classification time costs.

These recent methods are CNN-LSTM [23], EFM-ABC [24], DLM [25], HEM [26], ARIMA-GBRT [27], and HDFS [28]. All the indicated capabilities are implemented in CDF methodology, therefore HFS is being used to feature selection and IKN^{3}B is being used to demand forecasting. Results are shown in figure 9, figure 10, figure 11 and figure 12.

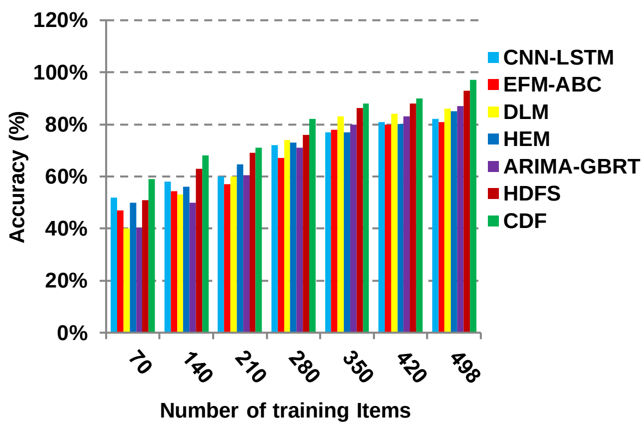

Figure 9. Accuracy of the different demand forecasting methods.

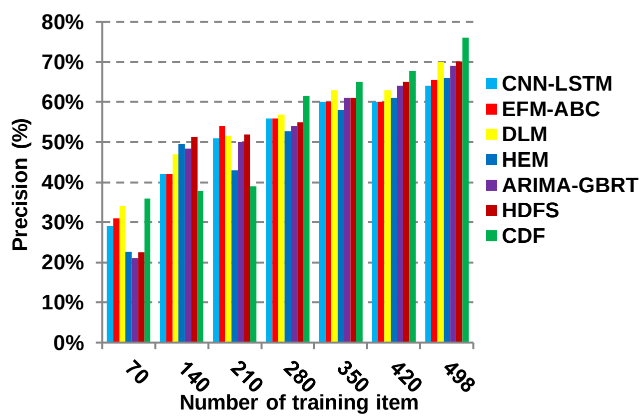

Figure 11. Precision of the different demand forecasting methods.

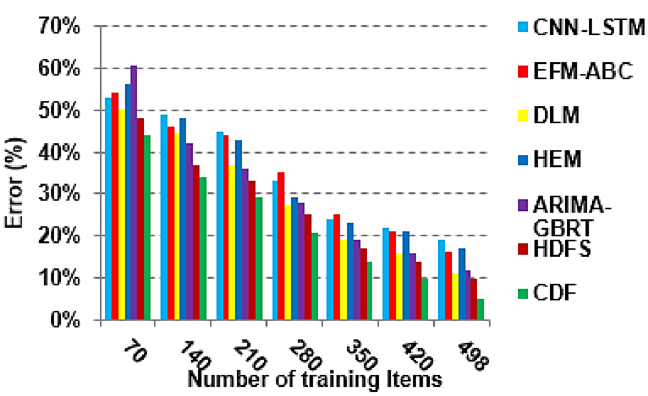

Figure 10. Error of the different demand forecasting methods.

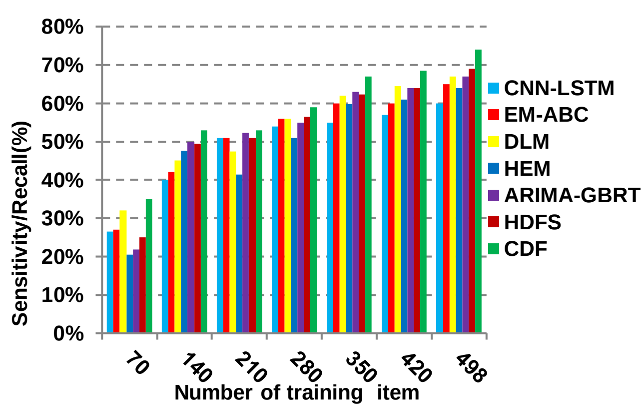

Figure12. Sensitivity of the different demand forecasting methods.

As shown in figure 9, figure 10, figure 11 and figure 12, CNN-LSTM, EFM-ABC, DLM, HEM,ARIMA-GBRT,HDFS, and CDF reach to accuracy values equal 0.82, 0.81, 0.86, 0.85, 0.87, 0.93, and0.97 respectively at the maximum number of training data (e.g.,498). CDF achieves the highest accuracy value by using IKN^{3}B that provides accurate demand forecasts with the shortest possible time penalty based on the most efficient nearest neighbors. Accordingly, CNN-LSTM, EFM-ABC, DLM, HEM, ARIMA-GBRT, HDFS, and CDF algorithms provide error values equals 0.19, 0.16, 0.11, 0.17, 0.12, 0.1, and 0.05 respectively. CDF provides precision value equals 0.76 while CNN-LSTM, EFM-ABC, DLM, HEM, ARIMA-GBRT, and HDFS provide 0.64, 0.66, 0.66, 0.70, 0.69, and 0.70 respectively. while the recall of CDF is 0.74, the recall of CNN-LSTM, EFM-ABC, DLM, HEM, ARIMA-GBRT and HDFS are 0.6, 0.65, 0.64, 0.67, 0.67, and 0.69 respectively. Hence, figure 9, figure 10, figure 11 and figure 12 demonstrate that CDF is superior to other current techniques, that are; CNN-LSTM, EFM-ABC, DLM, HEM, ARIMA-GBRT, and HDFS because the greatest accuracy and the lowest error are reached by CDF.

VI.CONCLUSIONS AND FUTURE WORK.

In this paper, a three-tier architecture of IoT, fog, and cloud has been presented for enhancing the smart electrical grid based on IoT smart devices. In ASGs, CDF is a critical process, thus, it is a vital process to use an effective forecast technique. The results of the literature review reveal that an optimal methodology has yet to be identified. As a result, the appropriate one must be chosen in order to deliver quick and precise forecast values. Hence, an accurate CDF method, which has the potential to bring increased intelligence (smartness) to future smart grids has been provided in this work. CDF strategy in this paper is divided into two main stages, which are; FS and DF stage. HFS as a features selection technique that integrated IG as a filter method and BPS optimization method as a wrapper method has been introduced to eliminate irrelevant features. Then, the filtered data without irrelevant features has been proceeded to DF stage to provide accurate demand forecasts by using Improved KN^{3}B (IKN^{3}B) that combines both KNN and NB classifiers. The provided HFS technique gives more accurate results than existing methods in terms of accuracy, error, precision, and sensitivity/recall. The accuracy, error, precision, and recall values provided by HFS are as follows; 0.94, 0.06, 0.74, and 0.73 respectively. Additionally, experimental results show that the proposed CDF strategy provides better accurate results than the current demand forecasting strategies according to accuracy, precision, and sensitivity. The accuracy, error, precision, and recall values provided by CDF are as follows; 0.97, 0.05, 0.76, and 0.74 respectively.

In the future work, the more superior forecasting model will be investigated. To increase the forecasting model's performance such as bi-directional long short-term memory, many optimization approaches and deep learning methods can be used. The evaluation of the studied techniques' performance on a dataset with higher resolution and longer forecasting horizons could be a future research area of work.

AUTHORS CONTRIBUTION

Naglaa R. Khalil is responsible for conception of the work, software, and drafting the article.

Asmaa H. Rabie is responsible for data collection, data analysis, and interpretation.

Khaled M. Abo-Al-Ez is responsible for supervision and critical revision of the article.

Ahmed I. Saleh is responsible for project administration and final approval of the version to be published.

FUNDING STATEMENT:

The author did not receive any financial support of the research authorship and publication of this article.

DECLARATION OF CONFLICTING INTERESTS STATEMENT:

The author declared that there are no potential conflicts of interest with respect to the research authorship or publication of this article.

REFERENCE

[1] G. Dileep. (2020, February). A survey on smart grid technologies and applications. Renewable Energy. Elsevier. [Online]. 146, PP. 2589–2625. Available: https://doi.org/10.1016/j.renene.2019.08.092

[2] S. Salkuti. (2020, April). Challenges, issues and opportunities for the development of smart grid. *International Journal of Electrical and Computer Engineering (IJECE)*. [Online]. 10(2), PP. 1179-1186. Available: http://doi.org/10.11591/ijece.v10i2.pp1179-1186

[3] M. Hamwi, I. Lizarralde, and J. Legardeur. (2021, February). *Demand response business model canvas:* A tool for flexibility creation in the electricity markets. *Journal of Cleaner Production, Elsevier*. [Online]. 282, PP.1-42. Available: https://doi.org/10.1016/j.jclepro.2020.124539

[4] X. Lu, K. Li, H. Xu, F. Wang, Z. Zhou, et al. (2020, January) Fundamentals and business model for resource aggregator of demand response in electricity markets. *Energy. Elsevier*. [Online]. 204, PP.1-15. Available: https://doi.org/10.1016/j.energy.2020.117885

[5] A. Aman, N. Shaari, and R. Ibrahim. (2021, February). Internet of things energy system: Smart applications, technology advancement, and open issues. *International Journal of Energy Research. Wiley Online Library*. [Online]. 45(6), PP.8389-8419. Available: https://onlinelibrary.wiley.com/doi/full/10.1002/er.6451

[6] S. Hashmi, C. Ali, and S. Zafar. (2020, October). Internet of things and cloud computing-based energy management system for demand side management in smart grid. *International Journal of Energy Research. Wiley Online Library*. [Online]. 45(1), PP.1007-1022. Available: https://onlinelibrary.wiley.com/doi/full/10.1002/er.6141

[7] M. Hussain and M. Beg. (2019, January). Fog computing for internet of things (IoT)-aided smart grid architectures. *Big Data and cognitive computing. Multidisciplinary Digital Publishing Institute*. [Online]. 3(1), PP.1-29. Available: https://doi.org/10.3390/bdcc3010008

[8] M. Forcan and M. Maksimović. (2020, May). Cloud-fog-based approach for smart grid monitoring. *Simulation Modelling Practice and Theory. Elsevier*. [Online]. 101, PP.1-18. Available: https://doi.org/10.1016/j.simpat.2019.101988

[9] Y. Chen, M. Chen, C. Chang, F. Chang, et al. (2021, April). A Smart Home Energy Management System Using Two-Stage Non-Intrusive Appliance Load Monitoring over Fog-Cloud Analytics Based on Tridium's Niagara Framework for Residential Demand-Side Management. *Sensors. Multidisciplinary Digital Publishing Institute*. [Online]. 21(8), PP.1-25. Available: https://doi.org/10.3390/s21082883

[10] D. Avancini, J. Rodrigues, R. Rabêlo, A. Das, et al. (2020, February). A new IoT-based smart energy meter for smart grids. *International Journal of Energy Research. Wiley Online Library*. [Online]. 45(1), PP. 189-202. Available: https://doi.org/10.1002/er.5177

[11] A. Chong, M. Mahmoud, F. Lim, and H. Kasim. (2020, August). A review of Smart Grid Technology, Components, and Implementation. Presented at 2020 8th International Conference on Information Technology and Multimedia (ICIMU). IEEE. [Online]. Available: https://doi.org/10.1109/ICIMU49871.2020.9243430

[12] M. Prajapati and S. Patel. (2020, June). A Review on Big Data with Data Mining. Presented at Data Science and Intelligent Applications. Springer. Singapore. [Online]. Available: https://doi.org/10.1007/978-981-15-4474-3_17

[13] A. Arif, N. Javaid, A. Aldegheishem, and N. Alrajeh. (2021, April). Big data analytics for identifying electricity theft using machine learning approaches in microgrids for smart communities. *Concurrency and Computation: Practice and Experience. Wiley Online Library*. [Online]. 33(17), *PP.1-21*. Available: https://doi.org/10.1002/cpe.6316

[14] V. Chang. (2021, April). An ethical framework for big data and smart cities. *Technological Forecasting and Social Change*. [Online]. 165, PP.1-11. Available: https://doi.org/10.1016/j.techfore.2020.120559

[15] H. K. Mohamed, S. M. El-debeiky, H. M. Mahmoud and K. M. El Destawy. (2006, November). Data Mining for Electrical Load Forecasting In Egyptian Electrical Network. Presented at *2006 International Conference on Computer Engineering and Systems*. [Online]. Available: 10.1109/ICCES.2006.320491

[16] G. Hafeez, K. Alimgeer, A. Qazi, I. Khan, et al. (2020, April). A hybrid approach for energy consumption forecasting with a new feature engineering and optimization framework in smart grid. *IEEE Access. IEEE*. [Online]. 8, PP. 96210-96226. Available: 10.1109/ACCESS.2020.2985732

[17] I. Müller. (2021, June). Feature selection for energy system modeling: Identification of relevant time series information. *Energy and AI. Elsevier*. [Online]. 4, PP.1-14. Available: https://doi.org/10.1016/j.egyai.2021.100057

[18] G. Ruan, H. Zhong, J. Wang, Q. Xia, et al. (2020, April). Neural-network-based Lagrange multiplier selection for distributed demand response in smart grid. *Applied Energy. Elsevier*. [Online]. 264, PP.1-17. Available: https://doi.org/10.1016/j.apenergy.2020.114636

[19] G. Shamim and M. Rihan. (2020, April). Multi-Domain Feature Extraction for Improved Clustering of Smart Meter Data. *Technology and Economics of Smart Grids and Sustainable Energy. Springer*. [Online]. 5(10), PP.1-8. Available: https://doi.org/10.1007/s40866-020-00080-w

[20] S. Singh and N. Kanwar. (2019, December). A Review on Demand Side Management Forecasting Models for Smart Grid. Presented at Intelligent Computing Techniques for Smart Energy Systems. Springer. [Online]. Available: https://doi.org/10.1007/978-981-15-0214-9_96

[21] D. Izidio, P. Neto, L. Barbosa, J. Oliveira, et al. (2021, March). Evolutionary Hybrid System for Energy Consumption Forecasting for Smart Meters. *Energies. Multidisciplinary Digital Publishing Institute*. [Online]. 14(7), PP.1-19. Available: https://doi.org/10.3390/en14071794

[22] D. Syed, H. Abu-rub, A. Ghrayeb, and S. Refaat. (2021, February). Household-level energy forecasting in smart buildings using a novel hybrid deep learning model. *IEEE Access. IEEE*. [Online]. 9, PP.33498-33511. Available: 10.1109/ACCESS.2021.3061370

[23] T. Kim and S. Cho. (2019, September). Predicting residential energy consumption using CNN-LSTM neural networks. *Energy. Elsevier*. [Online]. 182, PP. 72–81. Available: https://doi.org/10.1016/j.energy.2019.05.230

[24] J. Hao, X. Sun, and Q. Feng. (2020, January). A Novel Ensemble Approach for the Forecasting of Energy Demand Based on the Artificial Bee Colony Algorithm. *Energies. Multidisciplinary Digital Publishing Institute*. [Online]. 13(3), PP.1-25. Available: https://doi.org/10.3390/en13030550

[25] S. Bu and S. Cho, (2020, September). Time Series Forecasting with Multi-Headed Attention-Based Deep Learning for Residential Energy Consumption. *Energies. Multidisciplinary Digital Publishing Institute*. [Online]. 13(18), PP.1-17. Available: https://doi.org/10.3390/en13184722

[26] E. Jnr, Y. Ziggah, and S. Relvas. (2021, March). Hybrid ensemble intelligent model based on wavelet transform, swarm intelligence and artificial neural network for electricity demand forecasting. *Sustainable Cities and Society. Elsevier*. [Online]. 66, PP.1-13. Available: https://doi.org/10.1016/j.scs.2020.102679

[27] P. Nie, M. Roccotelli, M. Fanti, Z. Ming, et al. (2021, November). Prediction of home energy consumption based on gradient boosting regression tree. *Energy Reports. Elsevier*. [Online]. 7, PP.1246-1255. Available: https://doi.org/10.1016/j.egyr.2021.02.006

[28] H. Lu, X. Ma, and M. Ma. (2021, March). A hybrid multi-objective optimizer-based model for daily electricity demand prediction considering COVID-19. *Energy. Elsevier*. [Online]. 219, PP. 1-29. Available: https://doi.org/10.1016/j.energy.2020.119568

[29] A. Rabie, A. Saleh, and K. Abo-Al-Ez. (2015, December). ANew Strategy of Load Forecasting Technique for Smart Grids. *International Journal of Modern Trends in Engineering and Research (IJMTER)*. [Online]. 2(12), PP.332-341. Available: https://www.academia.edu/38981085/A_New_Strategy_of_Load_Forecasting_Technique_for_Smart_Grids?from=cover_page

[30] A. Saleh,A. Rabie, and K. Abo-Al-Ezb. (2016, August). A data mining based load forecasting strategy for smart electrical grids. *Advanced Engineering Informatics. Elsevier*. [Online]. 30(3), PP.422- 448. Available: https://doi.org/10.1016/j.aei.2016.05.005

[31] A. Rabie, A. Saleh, and H. Ali. (2020, November). Smart electrical grids based on cloud, IoT, and big data technologies: state of the art. *Journal of Ambient Intelligence and Humanized Computing. Springer*. [Online]. 12(10), PP. 9449-9480. Available:https://link.springer.com/article/10.1007/s12652-020-02685-6

[32] A. Rabie, S. Ali, H. Ali, and A. Saleh. (2018, September). A fog based load forecasting strategy for smart grids using big electrical data. *Cluster Computing. Springer*. [Online]. 22(1), PP.241- 270. Available: https://doi.org/10.1007/s10586-018-2848-x

[33] A. Rabie, S. Ali, A. Saleh and H. Ali. (2019, May). A new outlier rejection methodology for supporting load forecasting in smart grids based on big data. Cluster Computing. Springer. [Online]. 23(2), PP. 509-535. Available: https://doi.org/10.1007/s10586-019-02942-0

[34] A. Rabie, S. Ali, A. Saleh, and H. Ali. (2020, April). A fog based load forecasting strategy based on multi-ensemble classification for smart grids. *Journal of Ambient Intelligence and Humanized Computing. Springer*. [Online]. 11(1), PP. 209-236. Available: https://doi.org/10.1007/s12652-019-01299-x

[35] L. Gelazanskas, and K. Gamage. (2014, February). Demand side management in smart grid: A review and proposals for future direction. *Sustainable Cities and Society. Elsevier*. [Online]. 11, PP. 22-30. Available: https://doi.org/10.1016/j.scs.2013.11.001

[36] Z. MA, J. Xie, H. Li, Q. Sun, et al. (2017, September). The role of data analysis in the development of intelligent energy networks. *IEEE Network. IEEE*. [Online]. 31(5), PP. 88-95. Available: 10.1109/MNET.2017.1600319

[37] M. Jaradat, M. Jarrah, A. Bousselham, Y. Jararweh, and M. Al-Ayyouba. (2015). The Internet of Energy: Smart Sensor Networks and Big Data Management for Smart Grid. *Procedia Computer Science. Elsevier*. [Online]. 56, PP.592-597. Available: https://doi.org/10.1016/j.procs.2015.07.250

[38] Z. Ghanbari, N. Navimipour, M. Hosseinzadeh, and A. Darwesh, (2019, Janury). Resource allocation mechanisms and approaches on the Internet of Things. *Cluster Computing. Springer*. [Online]. 22(4), PP. 1253–1282. Available: https://doi.org/10.1007/s10586-019-02910-8

[39] B. Nguyen, B. Xue, and M. Zhang. (2020, May). A survey on swarm intelligence approaches to feature selection in data mining. *Swarm and Evolutionary Computation. Elsevier*. [Online]. 54, PP. 1-27. Available: https://doi.org/10.1016/j.swevo.2020.100663

[40] N. Kushwaha and M. Pant. (2018, May). Link based BPSO for feature selection in big data text clustering. *Future Generation Computer Systems. Elsevier*. [Online]. 82, PP. 190-199. Available: https://doi.org/10.1016/j.future.2017.12.005

[41] N. Altman. (1992). An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician. Taylor & Francis Group*. [Online]. 46(3), PP. 175–185. Available: https://www.tandfonline.com/doi/abs/10.1080/00031305.1992.10475879

[42] T. Cover, and P. Hart. (1967, January). Nearest Neighbor Pattern Classification. *IEEE Transactions on Information Theory. IEEE*. [Online]. 13(1), PP. 21-27. Available: 10.1109/TIT.1967.1053964

[43] G. Bhattacharya, K. Ghosh, and A. Chowdhury. (2017, June). Granger causality driven AHP for feature weighted KNN. *Pattern Recognition. Elsevier*. [Online].66, PP. 425–436. Available: https://doi.org/10.1016/j.patcog.2017.01.018

[44] H. Bhavsar and A. Ganatra. (2012, September). A Comparative Study of Training Algorithms for Supervised Machine Learning. *International Journal of Soft Computing and Engineering (IJSCE)*. [Online]. 2(4), PP

.74-81. Available:https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.492.6088&rep=rep1&type=pdf

[45] I. Hmeidi, B. Hawashin and E. El-Qawasmeh. (2008, January) Performance of KNN and SVM classifiers on full word Arabic articles. Advanced Engineering Informatics. [Online]. 22 (1), PP. 106–111. Available: https://doi.org/10.1016/j.aei.2007.12.001

[46] C. Nie , and F. Song. (2018, August). Analyzing the stock market based on the structure of kNN network. *Chaos. Solitons & Fractals. Elsevier*. [Online]. 113, PP.148–159. Available: https://doi.org/10.1016/j.chaos.2018.05.018

[47] S. Kotsiantis, I. Zaharakis, and P. Pintelas. (2007, June). Supervised Machine Learning: A Review of Classification Techniques. Presented at *Emerging artificial intelligence applications in computer engineering. Amsterdam.* [Online]. Available: https://dl.acm.org/doi/abs/10.5555/1566770.1566773

[48] M. Soundarya and R. Balakrishnan. (2014). Survey on Classification Techniques in Data mining. *International Journal of Advanced Research in Computer and Communication Engineering*. [Online]. 3(7), PP.268-278. Available: http://www.ijritcc.org

[49] K. Murphy. (2006, October). Naive Bayes Classifiers. *University of British Columbia*. [Online]. 18(60), PP.1-8. Available: https://www.ic.unicamp.br/~rocha/teaching/2011s1/mc906/aulas/naive-bayes.pdf

[50] F. Fernandes, H. Morais, P. Faria, Z. Vale, and C. Ramos. (2013, April). SCADA House Intelligent Management for Energy Efficiency Analysis in Domestic Consumers. Presented at 2013 IEEE PES Conference on Innovative Smart Grid Technologies (ISGT Latin America). IEEE. Sao Paulo. Brazil. [Online]. Available: 10.1109/ISGT-LA.2013.6554494

[51] https://site.ieee.org/pes-iss/data-sets/#fernandes2013

[52] European Environment Agency, "*Final electricity consumption by sector, EU-27*," Available: http://www.eea.europa.eu/data-and-maps

[53] A. Eseye, M. Lehtonen, T. Tukia, S. Uimonen, et al. (2019, September). Efficient Feature Selection Strategy for Accurate Electricity Demand Forecasting. Presented at 2019 IEEE PES Innovative Smart Grid Technologies Europe (ISGT-Europe). IEEE. [Online]. Available: https://doi.org/10.1109/ISGTEurope.2019.8905713

[54] Z. Li, D. Friedrich, and G. Harrison. *(2020, February)*. Demand Forecasting for a Mixed-Use Building Using Agent-Schedule Information with a Data-Driven Model. *Energies. Multidisciplinary Digital Publishing Institute*. [Online]. 13(4), PP.1-20. Available: https://doi.org/10.3390/en13040780

[55] W. Khan, S. Walker, K. Katic, and W. Zeiler. *(*2020, *September)*. A novel framework for autoregressive features selection and stacked ensemble learning for aggregated electricity demand prediction of neighborhoods. Presented at 2020 International Conference on Smart Energy Systems and Technologies (SEST). IEEE. [Online]. Available: https://doi.org/10.1109/SEST48500.2020.9203507

Arabic Title

استراتيجية التنبؤ بطلبات العملاء (CDF) المبنية على الضباب
للشبكات الذكية باستخدام انترنت الأشياء.

Arabic Abstract:

في هذا البحث، يتم توفير إطار عمل جديد لإدارة الطاقة الفعالة للعملاء المقيمين من أجل تقليل استهلاك الكهرباء. حيث يمكن أن تساعد الشبكات الذكية المتقدمة (ASGs) في مجموعة متنوعة من الوظائف بفضل إنترنت الأشياء (IoT).تولد هذه الأجهزة الذكية بيانات ضخمة يمكن تحميلها على السحابة لتحليل إضافي. تعمل طبقة حوسبة الضباب كجسر بين أجهزة إنترنت الأشياء المدمجة في SEG والسحابة للتغلب على مشكلات السحابة. بناءً على التصميم ثلاثي المستويات المشار إليه سابقًا، تم تقديم استراتيجية جديدة لتوقع طلب العملاء .(CDF)تتكون استراتيجية CDF من (1) مرحلة اختيار الميزات (FS)و (2) مرحلة التنبؤ بالطلب .(DF) تحاول مرحلة FS تحديد أهم الميزات التي تسمح لنموذج التنبؤ بالطلب بإنتاج نتائج سريعة ودقيقة. يتم استخدام نهج اختيار الميزات المختلطة (HFS)لاختيار الميزات الفعالة، والتي تدمج الأدلة من اثنين من محددات الميزات؛ (1) اكتساب المعلومات (IG)كطريقة تصفية و (2) يتم استخدام التحسين ثنائي الجسيمات (BPS)باعتباره طريقة المجمع. بعد ذلك، تم استخدام متنبئ KN3B المحسن (IKN3B) في مرحلة تحديد الاتجاه في محاولة لتقديم تنبؤات دقيقة للطلب استنادًا إلى المجموعة الفرعية المحددة من الميزات من المرحلة السابقة. في الواقع، يجمع IKN3B بين مصنف K-Nearest Neighbors (KNN) ومصنف Naïve Bayes (NB) ثم يقوم بتحسين خصائصهما لتوفير أفضل توقعات الطلب قدر الإمكان. بناءً على النتائج التجريبية، يُستنتج أن استراتيجية CDF لها تأثير إيجابي على موثوقية النظام ومرونته واستقراره من خلال تقديم تنبؤات دقيقة للطلب.