IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received March 6, 2022, accepted March 23, 2022, date of publication April 29, 2022, date of current version June 2, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3171229

Non-Technical Losses Detection Using Autoencoder and Bidirectional Gated Recurrent Unit to Secure Smart Grids

PAMIR[1], NADEEM JAVAID[1,2], (Senior Member, IEEE), UMAR QASIM[3], ADAMU SANI YAHAYA[1], EMAN H. ALKHAMMASH[4], AND MYRIAM HADJOUNI[5]

[1]Department of Computer Science, COMSATS University Islamabad, Islamabad 44000, Pakistan
[2]School of Computer Science, University of Technology Sydney, Ultimo, NSW 2007, Australia
[3]Department of Computer Science, University of Engineering and Technology at Lahore (New Campus), Lahore 54000, Pakistan
[4]Department of Computer Science, College of Computers and Information Technology, Taif University, P.O. Box 11099, Taif 21944, Saudi Arabia
[5]Department of Computer Sciences, College of Computer and Information Science, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia

Corresponding author: Nadeem Javaid (nadeemjavaidqau@gmail.com)

This work is supported by Taif University Researchers Supporting Project number (TURSP-2020/292) Taif University, Taif, Saudi Arabia. This work is also supported by Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2022R193), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

ABSTRACT Electricity theft is considered one of the most significant reasons of the non technical losses (NTL). It negatively influences the utilities in terms of the power supply quality, grid's safety, and economic loss. Therefore, it is necessary to effectively deal with the electricity theft problem. For detecting electricity theft in smart grids (SGs), an efficient and state-of-the-art approach is designed in the underlying work based on autoencoder and bidirectional gated recurrent unit (AE-BiGRU). The proposed approach consists of six components: (1) data collection, (2) data preparation, (3) data balancing, (4) feature extraction, (5) classification and (6) performance evaluation. Moreover, bidirectional gated recurrent unit (BiGRU) is used for the identification of the anomalies in electricity consumption (EC) patterns caused due to factors like family formation changes, holidays, parties, and so on, which are referred as non-theft factors. The proposed autoencoder-bidirectional gated recurrent unit (AE-BiGRU) model employs the EC data acquired from state grid corporation of China (SGCC) for simulations. Furthermore, it is visualized from the simulation results that 90.1% accuracy and 10.2% false positive rate (FPR) are obtained by the proposed model. The results are better than different existing classifiers, i.e., logistic regression (LR), decision tree (DT), extreme gradient boosting (XGBoost), gated recurrent unit (GRU), etc.

INDEX TERMS Autoencoder, bidirectional gated recurrent unit, deep learning algorithms, smart grid, electricity theft detection, machine learning algorithms, synthetic theft attacks.

I. INTRODUCTION

Electricity is a great and significant favor of science to human beings, which makes their life easier [1], [2]. It is used in various sectors such as medical, agriculture, transportation, industrial, commercial and residential. However, the electric utility companies suffer from several challenges due to losses in electricity transmission and distribution [3], being partitioned into non technical losses (NTL) and technical losses (TL). TL cover a small and unavoidable amount of the losses incurred to the electric utility, e.g., transmission line and transformer losses [4]. Whereas, the NTL cover a large amount of the losses. The reasons of the NTL include interfering with the meter so that it can report less electricity consumption (EC) value instead of the real EC to the utility, bypassing the meter, meter readers reporting the false EC values in return of some bribe, etc., [5]. The economic loss faced due to the electricity fraud is greater than $25 billion per year throughout the world. The USA suffers around $1.6 billion loss annually while Fujian (China) suffers more than ¥100 million loss per annum [6].

The manual meter inspection method is currently used for detecting electricity theft, which is costly in terms of human resource, money, and time. With the help of energy data acquisition system, electricity theft detection (ETD) is directed by some novel methods, which are distinguished

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

into network, data and hybrid methods [7]. In the first type of methods, ETD is performed by learning the historical EC data. With the exponential growth in the development and usage of the artificial intelligence (AI), the data based methods are well-suited options for ETD in smart grids (SGs) [8] and [9]. These methods are further grouped into supervised learning methods and unsupervised learning methods. In the supervised learning methods, labeled data is employed for training the algorithms [10]. The supervised data based methods include wide-and-deep convolutional neural networks (WDCNNs) [11], hybrid deep neural networks [12], support vector machine (SVM) [13], etc. These methods are trained to perform binary classification for ETD in SGs, and they give the best performance. However, the labeled data are needed to train these models, which are rarely available. If available, the data imbalance issue exists. The imbalance is between the normal data samples and the abnormal data samples. The imbalance is due to easy availability of the normal users' EC data in real environment as compared to abnormal users' EC data. One-sided, biased, and skewed decisions towards the class with higher number of samples, i.e., honest class, are made due to data imbalance issue. It also causes the model to have high false positive rate (FPR) value.

The authors in [13] proposed an SVM based consumption pattern based ETD (CPBETD) model for binary classification. However, the curse of dimensionality issue is ignored, which results in a high FPR that leads to increase in the cost associated with on-site inspections. Moreover, using the transformer meters and k-means clustering technique, the CPBETD is made strong against the identification of the EC pattern irregularities because of the non-theft reasons (drift). However, the CPBETD is still inappropriate in identification of the anomaly due to drift and that is why it achieves high FPR value. Furthermore, as discussed previously, the authors in [11] and [12] proposed WDCNNs and hybrid deep neural network for efficient ETD in SG, respectively. However, the concept of drift is not considered in these articles. This is another issue due to which ML models generate high FPR value. Drift means the irregularities in the EC patterns due to the non theft factors such as holidays, changes in the number of residents in a family, and changes in the number of electric appliances.

There is a special need of developing a modern approach to efficiently deal with the above challenges. An efficient and state-of-the-art approach is designed for ETD in SG based on the autoencoder and bidirectional gated recurrent unit (AE-BiGRU). The major contributions made in the underlying work are listed below.

- An AE-BiGRU theft detector is developed for efficient and effective ETD in SGs. For dealing with the curse of dimensionality issue, the autoencoder (AE) algorithm is used.
- We implement six synthetic theft attacks to tackle the data imbalance or inadequacy of the minority class data.
- We employ the BiGRU model to deal with drifts. The BiGRU has long term memory, which makes it able to learn long term temporal correlation and identify the drift easily.
- The dropout regularization method is added in BiGRU to avoid the proposed AE-BiGRU model from overfitting issue.
- To stop the proposed model from trapping into the local optima, Adam optimization method is employed.
- For dealing with the missing values, simple imputer (SI) is used. While for tackling outliers, three sigma rule (TSR) of thumb is used. Whereas, for handling the unscaled data present in the selected dataset, min-max scaler is employed.

The organization of the remaining manuscript is done in the following manner. Section 2 analyzes the related work. Sections 3 and 4 describe the problem statement and the system model proposed in the underlying study, respectively. The results obtained from simulations are elaborated in Section 5. At the end, the concluding remarks and future work of the paper are given in Section 6.

II. RELATED WORK

The work done in the field of ETD in SG is discussed in this section. The work is categorized into hardware, game theoretic and artificially intelligent literature. In hardware based category [14]–[16], some hardware instruments are employed to deal with the electricity theft. For instance, in [14], the authors proposed a hardware based solution for ETD in SG. They use some sensors possessed by the smart meter (SM) to indicate the electricity theft. However, the installation and maintenance of these additional hardware devices need additional cost.

ETD is taken to be a game played between electric utility company and the electricity thief in game theory methods [17]–[19]. The key objective of these game theory based approaches is to achieve the Nash equilibrium for the game between the electric company and abnormal consumer. As no hardware equipment is involved in these ETD approaches, so no additional cost is required, hence, game theory based methods are cheaper than the hardware based ETD methods. However, the game theory based ETD solutions are also not the suitable approaches to deal with the electricity theft problem due to the reason that it is very difficult to find an optimal and satisfactory equilibrium between the malicious users and utility.

In AI based literature [20]–[33], the authors employed AI related techniques in order to detect the theft of electricity using EC data provided by the SM. In [20], genetic algorithm (GA) is employed for adjusting the hyperparameters' values. However, synthetic minority oversampling (SMOTE) is used for data balancing that causes the proposed model to overfit. The authors in [21] employed some ensemble boosting and bagging techniques for ETD in SG. The SMOTE is used to tackle the data imbalance problem. The results show that high true positive rate (TPR) and low FPR are achieved

by the bagging techniques, i.e., extra trees and random forest (RF). However, data balancing is performed via SMOTE, where, the models tend to overfit. Moreover, an improved SMOTE, i.e., k-means clustering SMOTE (K-SMOTE) based data balancing and improved RF based electricity theft classification is done in [22]. The proposed method provides accurate and reliable locations for manual on-site inspection, so that NTL is reduced and the power system's stability and reliability are improved. The computational overhead of the proposed technique is also reduced as the decision trees (DTs) in RF are working in parallel. In [23], convolutional neural network (CNN) is used in combination with long short term memory (LSTM). For extracting important features, CNN is utilized. While the LSTM detects electricity theft. In addition, the unbalanced class issue is resolved by generating the synthetic data of the theft class using SMOTE. However, SMOTE leads the proposed model to the overfitting problem.

In [24], the authors proposed a model that performs conditional wasserstein generative adversarial network based data balancing to deal with unbalanced data problem, stacked convolution denoising autoencoder based feature extraction to deal with the curse of dimensionality issue, and light gradient boosting machine based detection of electricity theft. It is abbreviated as (CSL) classifier. In [25], the authors proposed a conditional variational autoencoder (CVAE) model for the augmentation of the data points to solve the data imbalance issue. Moreover, feature extraction is done by the encoder with the convolution layers to finally achieve dimensionality reduction. Finally, the CNN classifier is employed for ETD in SG. Furthermore, the authors in [26] proposed a siamese network for electricity theft classification in SG. Moreover, the imbalanced data issue is resolved by the adaptive synthesis technique. In addition, in order to make the proposed deep siamese model more generalized, two dimensional EC data along with one dimensional EC data are considered. The two dimensional data (i.e., weekly EC data) are handled by the CNN, whereas, the one dimensional data are handled by the LSTM.

In [27], a categorical boosting (CBoost) technique is proposed for ETD. For handling the missing values in the dataset, the k nearest neighbors (KNN) algorithm is employed. Moreover, in order to deal with the issue of imbalanced data, the SMOTETomek method is used. For selection and extraction of the important features from the dataset, the features extraction and selection based on the scalable hypothesis (FRESH) technique is used. Finally, the CBoost classifier is used to classify the electricity theft in SG. Furthermore, a modern ETD algorithm, i.e., text CNN is proposed in [28]. The datasets of both residential and industrial type users are considered. The residential EC dataset of the Irish users and the industrial dataset of the Chinese users are utilized for ETD. Moreover, the augmentation of the minority class data is done using a newly proposed data augmentation technique. The authors in [29] proposed a new model that is the combination of the existing DT, KNN, and SVM models, termed as DT-KSVM, for ETD in SG. The data augmentation is done using WGAN to solve the imbalanced data problem. A two level detection of theft is performed in this paper. The primary or first level theft detection is done utilizing a popular similarity measure method, whereas, the proposed DT-KSVM is responsible for secondary level ETD in SGs. Futhermore, in [30], a modified ensemble algorithm is designed for detection of the abnormal EC behavior, called dynamic ETD algorithm. Moreover, the imbalanced data issue is tackled using the hyperparameter of the XGBoost, i.e., scale-pos-weight.

In [31], the authors considered the imbalance data problem, high dimensionality issue, and classification. SMOTE and near miss (NM) are used in combination to resolve the data imbalance issue. The combined model is termed as SMOTE-NM. Moreover, to solve high dimensional data problem, the residual network model is utilized. The suitable adjustment of the hyperparameters' values of the adopted classifiers is done using the bayesian optimization algorithm. RF, Adaboost, and DT are employed for ETD in SGs. Results show that RF outperforms other employed algorithms in terms of FPR and false negative rate (FNR). Furthermore, another state of the art solution for ETD is designed in [32]. The theft detection is done using the XGBoost model. The Irish EC dataset is employed for ETD. The proposed classifier outperforms the other existing classifiers in terms of accuracy and FPR. In [33], the authors designed another state of the art technique that is capable of efficiently dealing with fault and privacy disclosure issues. So a new fault and privacy maintained ETD is proposed, termed as FPETD. Moreover, principal component analysis (PCA) is employed for tackling the curse of dimensionality issue. The superiority of the proposed model over the existing models with respect to accuracy is shown via simulations. However, the computational overhead is high. Furthermore, the authors in [34] presented a novel deep learning (DL) algorithm, i.e., gated recurrent unit (GRU) for theft classification in SGs. Moreover, the unbalanced data problem is tackled using the application of the synthetic six theft attacks. The state grid corporation of China (SGCC) utility corporation's dataset is employed for model training and testing, which is one of the high dimensional datasets in engineering applications that consists of 1035 features. The results show that GRU achieved better results in comparison with the benchmark SVM model with respect to accuracy, FPR, AUC score, F1 score, recall, and precision. However, curse of dimensionality problem is neglected.

III. PROBLEM STATEMENT

In [11] and [12], a CNN with the wide-and-deep learning ability and a model that combines MLP with LSTM are proposed for detecting the theft of electricity in SGs. Nevertheless, data imbalance problem is ignored. Generally, the normal class EC data are available in a large amount in real environment, whereas, the theft data are rarely available. So a ML model learns the normal class samples to a great degree in comparison with the minority class samples. Finally, the model looks skewed towards the majority class

TABLE 1. Details of dataset.

Attribute	Value
Timeframe of data collection	01-Jan-2014 to 31-Oct-2016
Number of abnormal users	3615
Number of normal consumers	38757
Total consumers	42372

and generates false results in terms of the minority class samples. It means that the model is biased that leads to high FPR value, where, high FPR results in high on-site inspection cost. Moreover, the drift is also neglected by these articles. Ignoring the drift causes high FPR value that is directly proportional to the on-site inspection cost. Moreover, only one layer is involved in the wide component of the model proposed in [11] for 1D data analysis, which leads the model to local optimal trapping problem.

In [13], an SVM based CPBETD approach is proposed for detecting electricity theft in SG. However, the curse of dimensionality problem is not considered. Curse of dimensionality issue refers to a concept that occurs when dealing with the high dimensional data; the classification error increases if the number of features increase. Neglecting the curse of dimensionality issue results in model's overfitting that leads to high FPR value, which simply implies the high on-site inspection cost. Moreover, the proposed model deals with drift by using the k-means clustering technique and transformer meters. However, the mechanism employed by the proposed model for drift identification is still not suitable and it results in a high FPR value (11%).

In [20] and [21], SMOTE data balancing technique is employed to deal with the imbalance data problem. However, employing SMOTE for balancing the data leads to the model's overfitting issue. In addition, SMOTE is not an appropriate choice for the time series sequential data. Furthermore, the selection of the suitable performance parameters is necessary for a model's effectiveness evaluation. However, authors in [6] and [13] ignore the selection of appropriate, suitable, and reliable performance measures.

IV. PROPOSED MODEL

The system proposed in this work for ETD is presented in Figure 1, which consists of six different modules: data collection, data preparation, data balancing, feature extraction, classification module, and performance evaluation. These six modules are elaborated in the current section.

A. DATA COLLECTION

SGCC provides the EC data that is being used in the underlying work [35] for simulations' purpose. The data are labeled. A limited number of theft records are there in the dataset. The dataset has the consumer number column, daily basis EC data (1034 columns of EC data), and the flag or label column that consists of 0 and 1, where, 0 represents the honest user and 1 represents the theft consumer. More details related to the dataset are given in Table 1.

It is noteworthy that the SGCC dataset contains outliers, missing values, and unscaled data. Therefore, data preparation is performed to deal with these issues.

B. DATA PREPARATION

In many cases, an EC dataset consists of some unscaled data, missing values, and outliers due to multiple reasons, i.e., meter's fault, storage problem, etc., [36]. To efficiently deal with the unscaled data, outliers, and the missing values, min-max scaler, TSR, and SI techniques are utilized, respectively. The main working flow of the SI method is taken from [11] and is shown in Equation 1,

$$f(x_{i,s}) = \begin{cases} \dfrac{x_{i,s-1} + x_{i,s+1}}{2} & \text{if } x_{i,s} \in NaN,\ x_{i,s-1},\ x_{i,s+1} \notin NaN \\ 0 & \text{if } x_{i,s} \in NaN,\ x_{i,s-1} \text{ or } x_{i,s+1} \in NaN \\ x_{i,s} & \text{Otherwise,} \end{cases} \quad (1)$$

where s and i show a specific slot, i.e., day and a specific electricity user, respectively. $x_{i,s}$ is the EC data of the user i for the day s. The next day's EC data is given by $x_{i,s+1}$ while the previous day's EC data is given by $x_{i,s-1}$. Not a number (NaN) shows the missing data while $f(x_{i,s})$ is the result of the SI implementation, which provides us with an imputed dataset. Moreover, the missing values can be tackled through other statistical techniques, such as moving average (MA), double MA and exponential MA [37]. However, the selection of SI method is made because the missing value is filled using the previous and the next values' average. Furthermore, SI adds diversity in the dataset where statistical techniques lack. The reason is that the mentioned statistical techniques fill missing values with the duplicate values suggested by the moving window, which lead to the overfitting problem.

More oftenly, in EC datasets, some outliers can be found, which affect the ML models' detection accuracy. So we use an outlier removal technique, i.e., TSR in the proposed model. The mathematical formulation of TSR of thumb is borrowed from [11] and is presented in Equation 2,

$$f(x_{i,s}) = \begin{cases} avg(X) + 2.\sigma(X) & \text{if } x_{i,s} > avg(X) + 2.\sigma(X), \\ x_{i,s} & \text{Otherwise.} \end{cases} \quad (2)$$

In the equation, multiple values of $x_{i,s}$ form the vector X. The standard deviation is given by $\sigma(X)$ while the average value of X is given by $avg(X)$. Finally, $f(x_{i,s})$ presents the output of TSR implementation. It means that $f(x_{i,s})$ represents the non-outlier EC data of a consumer i at day s. The reason for using the TSR technique over other outlier detection techniques (e.g., isolation forest (IF) and local outlier factor (LOF) [38]) is that it follows the principle of normal distribution where an outlier is the one that deviates from the distribution; while the normal value does not. This process makes the outlier detection process accurate and efficient. Whereas, IF first isolates a feature from the given dataset and then declares it as an anomaly or not. By doing this, the computational overhead

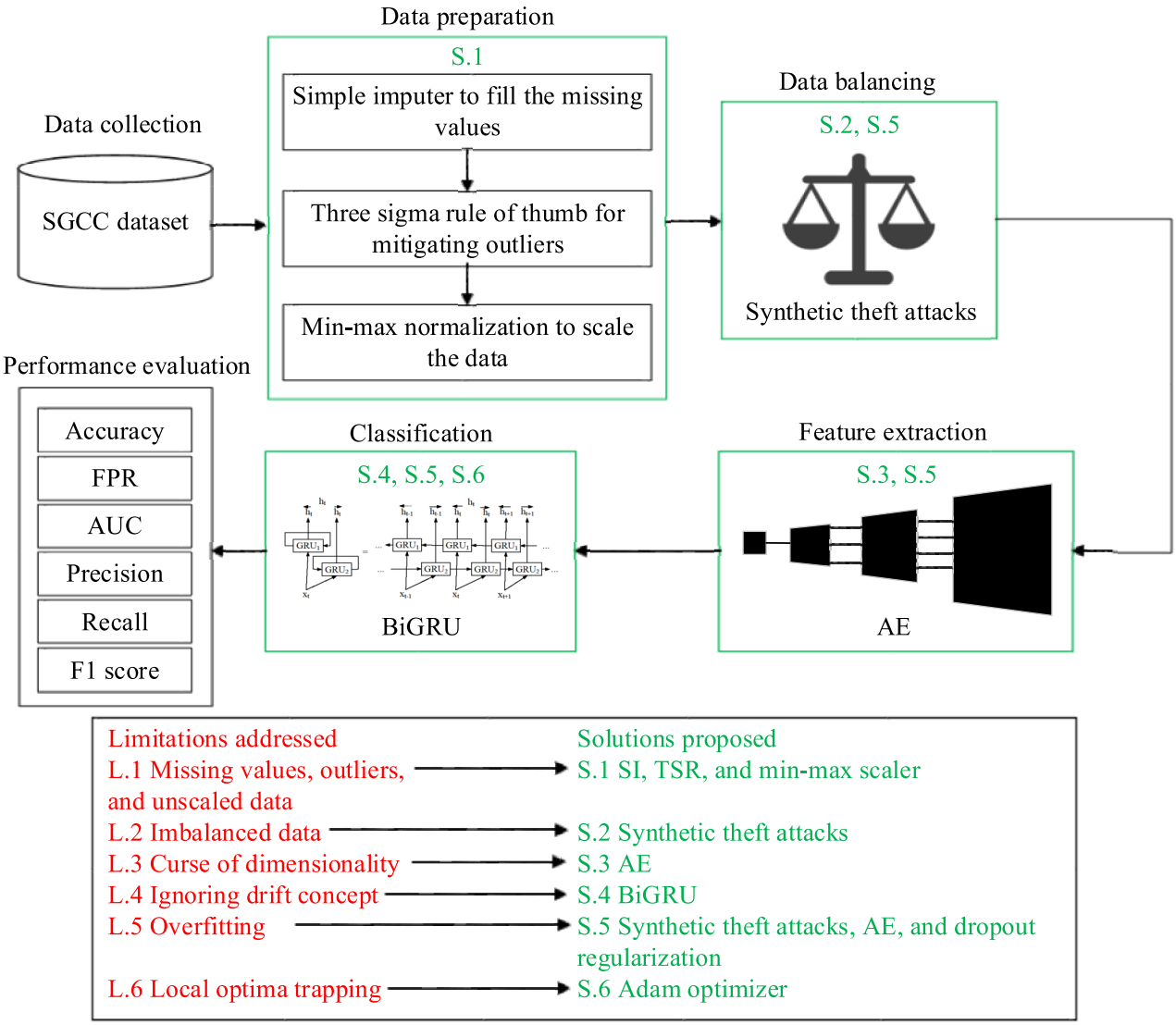


FIGURE 1. Overview of the model proposed for detecting NLT in SGs.

is increased to a greater extent. Similarly, LOF identifies anomalies by checking the deviation of an observation from its neighbors, which is not efficient for high-dimensional data. Keeping these concerns in view, TSR is used, which is a time and cost-efficient technique to remove outliers.

Furthermore, the min-max scaler technique is employed to deal with the unscaled data to normalize it or bring it in a particular range because the DL techniques are sensitive to diverse data. The formula used for the normalization technique is taken from [11] and is given in Equation 3,

$$f(x_{i,s}) = \frac{x_{i,s} - \min(X)}{\max(X) - \min(X)}, \tag{3}$$

where the lowest possible value of vector X is given by $\min(X)$ while the highest possible value of vector X is given by $\max(X)$. The term $f(x_{i,s}$ is the normalized output of the min-max scalar. Moreover, several other normalization techniques are available such as robust scalar and Z-score [39]. However, we have found the best performance results using min-max scalar because it scales values between 0 and 1. Whereas, the above mentioned techniques normalize values in different ranges, which are not efficient for the accurate training of the classification model. That is the reason of using min-max normalization in this work.

C. DATA BALANCING

A dataset is considered imbalanced if there is a huge difference in the number of observations of the normal and abnormal classes. Due to imbalanced data, in testing phase, the model will be biased towards the normal class, which minimizes the abnormal or theft detection accuracy. So to achieve good and satisfactory theft detection accuracy, we need to balance our data first and then train the ML algorithm. Many data balancing techniques are used in the existing literature such as SMOTE [20], [21], CVAE [25],

ADASYN [26], etc. SMOTE is most widely used by many researchers in the literature for data balancing. However, SMOTE creates model overfitting issue. Therefore, in this research, synthetic theft attacks [40] are employed for dealing with the issues of data imbalance and model overfitting. In addition, theft attacks based synthetic data generation is most suitable for the time series data. These attacks successfully maintain the non-linearity in the EC patterns.

The dataset (SGCC) we considered for the analysis is of imbalanced nature, where the imbalance nature is dealt with using the theft attacks. For denoting the electricity user's EC data, e_t is used where ($t \in [0, 1034]$). In the dataset, we have EC data of 1035 days. The mathematical representations for the synthetic theft attacks are taken from [40] and are given in Equations 4-9,

$$t_1(x_{i,s}) = x_{i,s} * random(0.1, 0.9), \tag{4}$$
$$t_2(x_{i,s}) = x_{i,s} * r_s, r_s = random(0.1, 1), \tag{5}$$
$$t_3(x_{i,s}) = x_{i,s} * random[0, 1], \tag{6}$$
$$t_4(x_{i,s}) = mean(V) * random(0.1, 1), \tag{7}$$
$$t_5(x_{i,s}) = mean(V), \tag{8}$$
$$t_6(x_{i,s}) = V_{1034-s}, \tag{9}$$

where $V = \{x_{1,1}, x_{1,2}, x_{1,3}, \ldots, x_{1,1034}\}$ represents the overall EC values of a consumer. The six attacks mentioned above are applied to the normal consumers' EC data in order to balance the normal and abnormal electricity consumers' data of SGCC dataset. The considered SGCC dataset is originally imbalanced. As the authors in [13] considered the benign consumers dataset and synthetically generated the theft data using theft attacks to balance the data. Therefore, we also assumed that we have the dataset of only benign consumers and synthetically generated the theft data using theft attacks for data balancing. To avoid higher time complexity, only 12000 data instances of 38757 benign consumers' EC data are selected for analysis purpose. The theft attacks are applied for data balancing. Theft attack 1, theft attack 2, theft attack 3, theft attack 4, theft attack 5, and theft attack 6 are applied on benign data ranges 0-1000, 1000-2000, 2000-3000, 3000-4000, 4000-5000, and 5000-6000, respectively. So the first 6000 records will become theft consumers' data, whereas, the remaining 6000 records (starting from 6000 to 12000 records) are the normal data. In this way, the dataset is balanced. Now, this dataset is forwarded to the AE for extracting the necessary features to achieve higher ETD accuracy and decrease FPR. Figure 2 displays the selected normal patterns while Figure 3 shows the selected theft attack patterns. In the figures, the EC patterns of a synthetic theft attack and a normal consumer are depicted for 30 days (numbered from day 0 to day 29).

D. FEATURE EXTRACTION

To achieve an efficient solution for ETD in SG with higher theft classification accuracy and lower FPR value, the feature extraction process is needed. However, the curse of dimensionality problem is ignored by the existing studies, which leads to the classifier's overfitting issue that maximizes the FPR value and minimizes the theft detection accuracy. Therefore, we employed a powerful method for the extraction of the key features, i.e., AE [6]. Feature extraction is a sub type of feature engineering that is used for dimensionality reduction, which is performed by deleting the redundant and irrelevant data [40]–[42]. Feature extraction results in lower computational and storage overhead as well as higher classification performance.

AE [6], [44] is one of the popular feature extraction techniques that consists of two modules: encoder and decoder. Encoder transforms the original input data into the lower dimensions (compressed representation) while the decoder seeks to recreate the original input from the compressed representation provided by the encoder module. When training the encoder and decoder modules, the target is to recreate the original input data with minimum reconstruction loss. After model training, only encoder module is employed for feature extraction [44]. Furthermore, it is important to mention that AE efficiently compresses the original representation of EC data into a low dimensional feature space and then learns non-linear patterns from it. This unique property of AE makes it superior over other deep

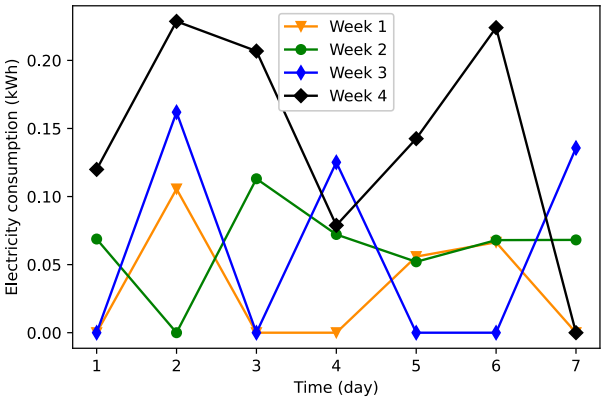


FIGURE 2. Electricity consumption pattern of a normal consumer.

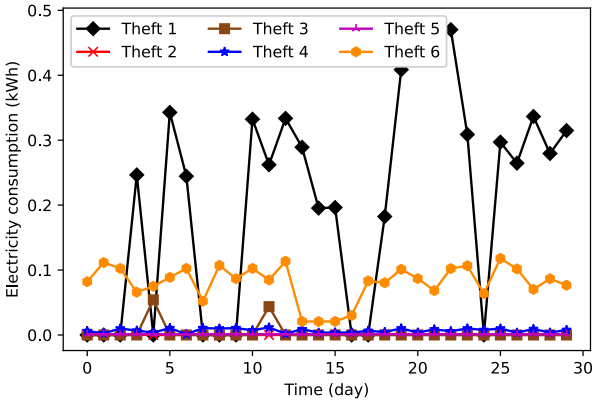


FIGURE 3. Electricity consumption patterns of a consumer after applying synthetic theft attacks.

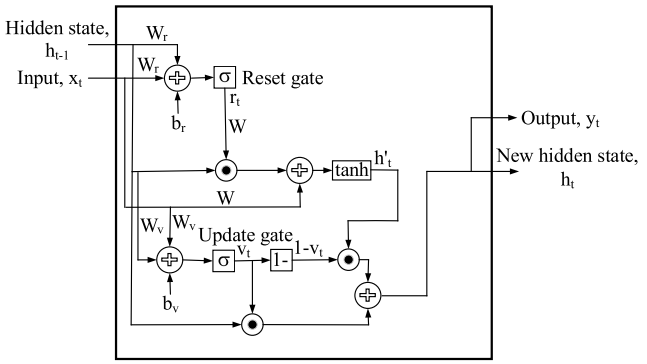


FIGURE 4. Generic framework of GRU.

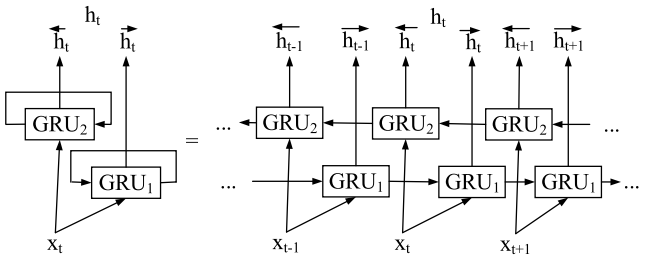


FIGURE 5. Generic framework of BiGRU.

feature extractors. This is the reason of using AE as a feature extractor in this work. Moreover, the encoding process of AE is taken from [6] and is presented in Equation 10,

$$f = a(w_x * x_t + b_x), \tag{10}$$

where, $a(.)$ represents the activation function, f shows the extracted feature vector, w_x represents the input weight and b_x represents the bias term of the input value. Moreover, in AE feature extractor, we use four Dense layers having 800, 400, 200, and 100 neurons, respectively, to compress (encode) the original EC data into lower dimensions. For each Dense layer, we employ ReLU as an activation function.

E. ELECTRICITY THEFT CLASSIFICATION

To classify the electricity theft, the BiGRU model [43] is employed. In the primary DL networks and their subsequent models such as CNN, weights' updation is conducted during the backpropagation process. Due to which the exploding and vanishing gradient issues occur. Therefore, recurrent neural networks (RNN) were designed to deal with these issues, such as LSTM and GRU. On the other hand, LSTM is also not a good and suitable model as it has high time complexity and has numerous parameters in comparison with the GRU. The GRU is a more suitable classification technique as compared to conventional DL approaches, i.e., MLP, CNN, and LSTM. Furthermore, GRU extracts the features that are beneficial to detect the presence of energy theft. A generic architecture of GRU is depicted in Figure 4 [44]. The GRU has update gate and reset gate. The former gate assists the GRU algorithm to decide about the historical data that is to be moved ahead. Whereas, the latter gate is used in the model to determine that how much of the previous knowledge to forget. The complete computations of the update gate, reset gate, candidate hidden state and the new hidden state are given in Equations 11-14 [44], respectively. The process flow of GRU from the input to the output is given in Figure 4.

$$v_t = \{sigmoid * (W_v x_t + W_v h_{t-1} + b_v)\}, \tag{11}$$
$$r_t = \{sigmoid * (W_r x_t + W_r h_{t-1} + b_r)\}, \tag{12}$$
$$h'_t = \{\tanh *(W x_t + W r_t \odot h_{t-1})\}, \tag{13}$$
$$h_t = \left\{v_t \odot h_{t-1} + (1 - v_t) \odot h'_t\right\}. \tag{14}$$

In the above equations, v_t, r_t, h'_t, and h_t represent the update gate, reset gate, candidate hidden state, and the new hidden state, respectively. W, h_{t-1}, and b show weight, hidden state at previous time step, and the bias term, respectively. Moreover, the terms *tanh* represents the hyperbolic activation function. The symbol $\odot$ represents Hadamard product.

Generally, the bidirectional models have a special property of learning information from the previous as well as the subsequent (future) data in order to predict the current data [45]. BiGRU is an enhanced version of the GRU. It is widely utilized in various domains like optical communication [45], network security [46], structural damage recognition [47], natural language processing [48], etc. However, it is narrowly used in engineering applications, specially in ETD [43]. Therefore, in this work, we employed the BiGRU for ETD in SG for further investigation. The generic architecture of BiGRU is shown in Figure 5 [45].

The BiGRU technique is the combination of two GRUs that are unidirectional in completely different directions [45]. One GRU, i.e., GRU_1 (also called forward GRU) moves from left to right side and the second GRU, i.e, GRU_2 (also called backward GRU) moves from right to left side. Consequently, the information traverses from right to left and then in the opposite direction during the final prediction, where it allows to predict the current data and observations using the information from the past and future states. The basic working mechanism of BiGRU is that the input data sequence is initially passed through backward and forward neural networks. Then, the output result of both are connected and combined at the same output layer. For theft classification, we use five layers: two BiGRU layers, a flatten layer, a dropout layer, and a dense layer. We employ 50 neurons in both the BiGRU layers, and use 0.2 dropout probability value for Dropout layer. Finally, we employ 1 neuron and sigmoid activation function for Dense layer to obtain the output. Furthermore, an important property of BiGRU is that it has a long term memory, which makes it able to learn and retain the long term temporal dependencies between the features. It helps BiGRU to detect electricity variance in different consumption circumstances such as holidays or parties, and easily differentiate between the real abnormal pattern and drift. The BiGRU model's process flow is modeled mathematically using the Equations 15-17 [45].

$$\overrightarrow{h_t} = GRU_1(x_t, \overrightarrow{h_{t-1}}), \tag{15}$$

$$\overleftarrow{h_t} = GRU_2(x_t, \overleftarrow{h_{t+1}}), \tag{16}$$
$$h_{ut} = \overrightarrow{h_t} \oplus \overleftarrow{h_t}, \tag{17}$$

where $\overrightarrow{h_{ut}}$, $\overleftarrow{h_t}$, $\overrightarrow{h_{t-1}}$ and $\overleftarrow{h_{t+1}}$ show the newly updated states of GRU_1 and GRU_2. The symbol $\oplus$ denotes the concatenation of two vectors.

F. PERFORMANCE EVALUATION METRICS

ETD is a binary classification task, which is performed to classify an EC pattern as either theft or honest. The most efficient and reliable performance evaluation metrics for this task are accuracy, FPR, AUC, recall, precision, and the F1 score. The results can be generated in terms of the confusion matrix that consists of four sub parameters: TPR, FPR, FNR, and true negative rate (TNR). These parameters are defined below.

- FPR: when an actually honest EC user is predicted as fraudulent via the classification technique.
- TPR: when an actually fraudulent EC user is predicted as fraudulent via the classification technique.
- TNR: when an actually honest EC user is predicted as honest via the classification technique.
- FNR: when an actually fraudulent EC user is predicted honest via the classification technique.

TNR and TPR show the correct classification, whereas, the FNR and FPR show the misclassification. The fundamental goal of ETD is to correctly detect the theft values, which ultimately reduces the on-site inspection cost [3]. The suitable performance parameters utilized for evaluation of the proposed technique are described in the subsections below.

1) ACCURACY

The accuracy is not known as a suitable performance evaluation parameter if the dataset is imbalanced [3]. The reason is that even if a technique shows bad performance, it still may have good accuracy value. It is due to the model's tendency towards the normal class. However, in our scenario, we initially performed the data balancing using the theft attacks. That is why, accuracy in our case is the most suitable performance metric to perform efficient performance evaluation. The formula used for calculating accuracy is given in Equation 18 [49].

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}. \tag{18}$$

2) AREA UNDER THE CURVE SCORE

AUC score is one of the important metrics used for performance evaluation. It can also be called as the separability measure. It shows the tradeoff between TPR and FPR. High AUC score means high TPR and high TPR shows model's better classification capability. The AUC's calculation formula is taken from [11] and is given in Equation 19.

$$AUC = \frac{\sum_{i \in positiveclass} RANK_i - \frac{M(M+1)}{2}}{M \times N}, \tag{19}$$

where data sample i's rank is given by $RANK_i$, negative samples are given by N and positive samples are given by M.

3) PRECISION

Precision is another very necessary and suitable performance metric needed for our model's performance evaluation. Precision shows the classifier's performance with respect to the FPR. The higher value of precision means low FPR that is widely recommended. The calculation of precision is done via Equation 20 [3].

$$Precision = \frac{TP}{TP + FP}. \tag{20}$$

4) RECALL

The second name of recall is sensitivity. It is calculated to show the proportion that how many of the actual abnormal electricity users are predicted abnormal by the algorithm. Recall shows the classification algorithm's performance with respect to the FNR. The high value of recall means low FNR that leads to high misclassification rate. Recall is calculated using Equation 21 [3].

$$Recall = \frac{TP}{TP + FN}. \tag{21}$$

5) F1 SCORE

It is another performance metric, which is used for model's evaluation. It is calculated using the precision and recall values. The high value of F1 score means that our model is correctly classifying the abnormal as well as the normal electricity users. The F1 score is calculated using the mathematical formula given in Equation 22, being taken from [50].

$$F1\ score = 2 * \frac{Precision * Recall}{Precision + Recall}. \tag{22}$$

6) FALSE POSITIVE RATE

The high value of FPR leads to high misclassification rate that results in increase in the cost of on-site inspections. The cost is incurred in order to confirm that the users are really honest that are predicted as dishonest by the model. Therefore, low FPR is highly recommended in theft users' classification to avoid high on-site inspection cost. FPR is computed using Equation 23 [50].

$$FPR = \frac{FP}{FP + TN}, \tag{23}$$

where TP denotes true positive, TN denotes true negative, FP denotes false postive and FN denotes false negative.

V. DISCUSSION OF SIMULATION RESULTS

The results obtained after performing extensive simulations are discussed in this section. AE-BiGRU's performance is evaluated via performing simulations in Google Colaboratory using Python programming language. SGCC dataset is used for training the proposed and existing models. The dataset

details are provided in Table 1. The initial preprocessing of the dataset is done using TSR, SI, and min-max scaler methods. After that, data balancing is performed using the synthetic theft attacks. Then, the splitting of dataset in training set and testing set is done in the ratio of 80% and 20%, respectively, during training process. The short introduction to these performance measures is already given in the previous subsections. Furthermore, the details of the existing techniques are given in the subsequent subsections.

A. SELECTED BENCHMARK TECHNIQUES

Before discussing simulation results, this section includes the details of the existing approaches being implemented in our study for comparison purpose. The compared methods include LR, SVM, DT, RF, XGBoost, CNN, GRU, and BiGRU.

1) LOGISTIC REGRESSION

LR is a supervised learning algorithm used for the classification tasks. It has three types: binary, ordinal, and multinomial. We employed the binary LR that is used to output only two possible values, 0 and 1. In our scenario, 0 means that the consumer is honest while 1 means that the consumer is dishonest or theft. This technique is also used by many researchers in the literature for ETD, which is a binary classification problem [12], [31], [43], etc.

2) SUPPORT VECTOR MACHINE

It is a widely used classifier that receives large popularity in the literature for performing different classification tasks. It is also employed by many researchers for solving the ETD problem [3], [12], [13], [31]. To classify the theft and non theft consumers, the SVM draws a hyperplane with a huge margin between the support vectors. Furthermore, the hyperparameters' (kernel, γ, and C) values are adjusted. The radial basis function (RBF) kernel is employed in the SVM. Moreover, the default values for γ and C hyperparameters are used.

3) DECISION TREE

DT is a supervised learning technique employed for both the regression and classification tasks [44]. It is widely used in the literature for the binary classification tasks such as ETD [29], [31], [51].

4) RANDOM FOREST

RF is an ensemble model that is used for classification tasks. It is the collection of the DTs used to make accurate and reliable predictions [12]. RF is widely utilized for binary classification problems like ETD [12], [21], [22], [43].

5) EXTREME GRADIENT BOOSTING

XGBoost is one of the popular supervised learning techniques. This algorithm is extensively used for ETD in SG [21], [30], [32]. In XGBoost, trees are combined to perform accurate ETD.

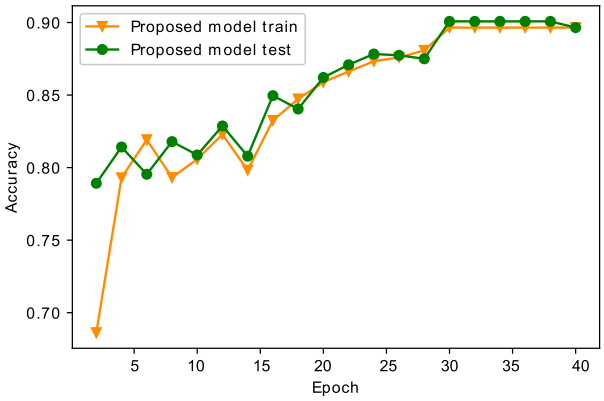


FIGURE 6. Training and testing accuracy of AE-BiGRU.

6) CONVOLUTIONAL NEURAL NETWORK

CNN is a DL model used for performing the classification tasks. It consists of multiple layers such as convolution layers, pooling layers, fully connected layer, and softmax or logistic layer. The convolution, pooling, and fully connected layers are used to extract high level features, to decrease the spatial size of those high level features, and to connect the neurons of one layer to the another layer's neurons, respectively. Finally, the softmax layer is used for multi class classification while the logistic layer is used for binary class classification [44]. CNN is also used by many researchers in the recent literature for ETD in SGs [12], [23], [43], [52].

B. PERFORMANCE COMPARISON OF THE PROPOSED AE-BIGRU MODEL WITH THE BENCHMARK TECHNIQUES

To check the effectiveness of our proposed method, we run our model for 40 iterations of training and testing. Figures 6 and 7 represent the convergence plots for accuracy and loss of AE-BiGRU, respectively. The Figure 6 shows that the training accuracy value, at the 40th epoch, is 0.896 and the testing accuracy value is 0.901, which are the maximum values. The dataset we used for analysis contains some zero values. At the 6th iteration, the proposed AE-BiGRU model is trained on a batch that contains zero values, which causes the overfitting problem. The final accuracy of the proposed AE-BiGRU model is 90.1%. Moreover, Figure 7 presents the AE-BiGRU model's training and testing loss for 40 epochs. The training and testing loss finally reaches value of 0.245 and 0.237, respectively, which are the minimum values. As the batch in the 6th epoch contains zero values, therefore, the overfitting issue takes place at the 6th iteration only. Figures 6 and 7 clearly indicate and prove AE-BiGRU model's excellent performance on training as well as testing data, which means that overfitting is successfully avoided.

Figures 8 and 9 represent the GRU's accuracy and loss convergence, respectively. From Figure 8, it is observed that, the final training accuracy and testing accuracy values, at the 40th epoch, are 0.880 and 0.791, respectively. As no feature extraction and no proper dropout regularization is done in GRU model, therefore, after the 14th iteration,

TABLE 2. Comparison of proposed AE-BiGRU model with benchmark techniques.

Classification techniques	P1	P2	P3	P4	P5	P6
LR	$6.80*10^{-1}$	$6.76*10^{-1}$	$5.40*10^{-1}$	$7.30*10^{-1}$	$6.21*10^{-1}$	$3.48*10^{-1}$
SVM	$7.13*10^{-1}$	$7.14*10^{-1}$	$7.46*10^{-1}$	$6.88*10^{-1}$	$7.16*10^{-1}$	$2.58*10^{-1}$
DT	$8.34*10^{-1}$	$8.35*10^{-1}$	$8.47*10^{-1}$	$8.18*10^{-1}$	$8.32*10^{-1}$	$1.60*10^{-1}$
RF	$7.65*10^{-1}$	$7.67*10^{-1}$	$8.38*10^{-1}$	$7.22*10^{-1}$	$7.76*10^{-1}$	$1.79*10^{-1}$
XGBoost	$8.20*10^{-1}$	$8.21*10^{-1}$	$8.36*10^{-1}$	$8.02*10^{-1}$	$8.19*10^{-1}$	$1.16*10^{-1}$
CNN	$7.77*10^{-1}$	$7.78*10^{-1}$	$7.99*10^{-1}$	$7.55*10^{-1}$	$7.77*10^{-1}$	$2.16*10^{-1}$
GRU	$7.91*10^{-1}$	$8.22*10^{-1}$	$7.90*10^{-1}$	$8.36*10^{-1}$	$8.13*10^{-1}$	$1.76*10^{-1}$
BiGRU	$8.60*10^{-1}$	$8.61*10^{-1}$	$8.82*10^{-1}$	$8.38*10^{-1}$	$8.59*10^{-1}$	$1.17*10^{-1}$
Proposed model	$9.01*10^{-1}$	$9.01*10^{-1}$	$9.13*10^{-1}$	$8.86*10^{-1}$	$8.99*10^{-1}$	$1.02*10^{-1}$

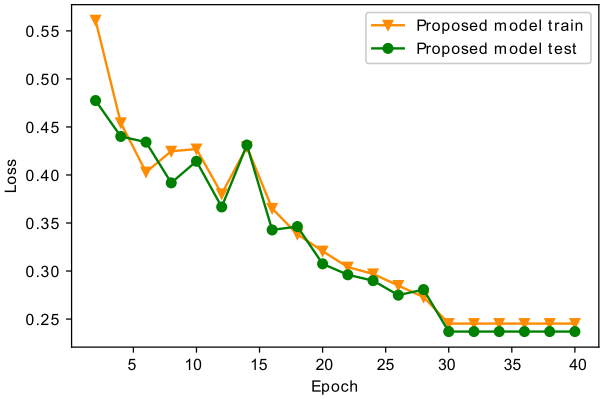


FIGURE 7. Training and testing losses of AE-BiGRU.

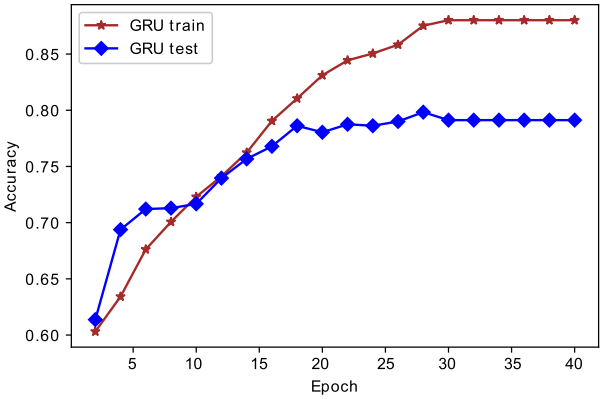


FIGURE 8. Training and testing accuracy of GRU.

GRU starts overfitting, which continues till the last iteration. The GRU's final testing accuracy (at the 40th iteration) is 0.791. Moreover, Figure 9 represents GRU's training and testing loss. The minimization of training and testing loss continues till the 8th iteration. After 8th epoch, overfitting starts, which goes till the 40th iteration due to the improper dropout regularization values and no feature extraction. Finally, the Figures 8 and 9 very clearly depict that the benchmark GRU model performs well on training data and shows lower performance on the testing data, which means that the overfitting issue has occurred. There are three reasons in which the proposed AE-BiGRU model helps in avoiding overfitting issue, which include synthetic theft attacks based abnormal data augmentation, employing AE as a feature extractor or dimensionality reduction algorithm, and using of the dropout regularization strategy. The dropout probability value selected for our proposed AE-BiGRU model is 0.2. Finally, we can say that the most significant reason out of the above three reasons is AE based feature extraction, that plays a vital role and proves the proposed AE-BiGRU model to be superior in comparison with the benchmark GRU in terms of overfitting prevention.

The summary of the results of the selected benchmark models and the proposed model with respect to different performance measures is given in Figure 10 and Table 2. In Table 2, P1, P2, P3, P4, P5 and P6 represent accuracy, AUC score, precision, recall, F1 score and FPR, respectively. The results show the inferiority of RF as compared to other

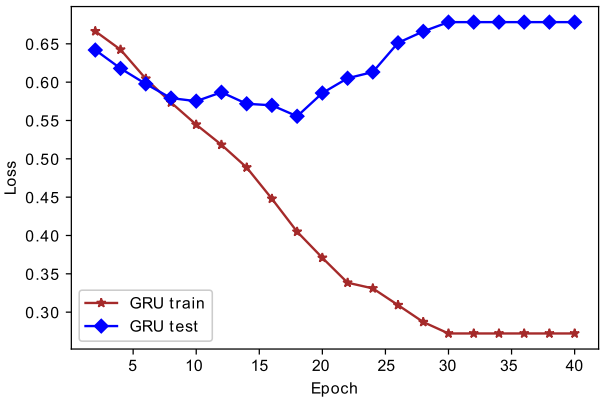


FIGURE 9. Training and testing losses of GRU.

models. RF provides good results when it is trained using relatively small number of samples and will rapidly reach to a point where adding more data samples will not improve the classification accuracy. On the other hand, CNN, GRU, BiGRU, etc., require more data samples to provide the same accuracy, however, the deep models benefit from the large data, and continuously enhance the classification accuracy of the model. Hence, CNN, GRU and BiGRU, provide better results as compared to RF, LR, and SVM, which is seen in Table 2 and Figure 10. RF, LR, and SVM do not perform well while dealing with the large datasets such as SGCC because they are prone to the overfitting problem. Conversely, the proposed AE-BiGRU model provides better results in

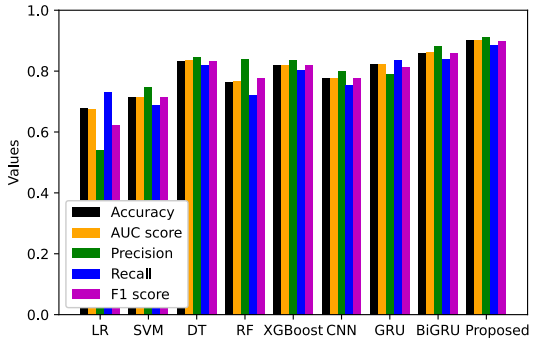


FIGURE 10. Performance comparison of different classifiers.

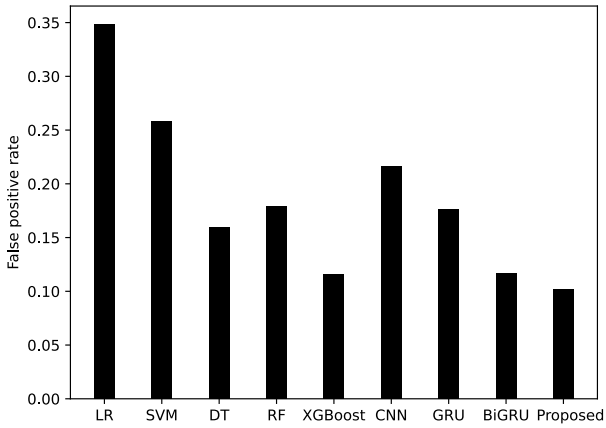


FIGURE 11. Performance comparison using FPR.

comparison with all the existing benchmark schemes with regard to all the performance measures.

The proposed AE-BiGRU model outperforms all the existing models with respect to all performance measures due to many strong reasons. Firstly, it can efficiently handle the data imbalanced problem using synthetic theft attacks. Secondly, the encoder module efficiently encodes and extracts the important features. Thirdly, the BiGRU module is used to efficiently identify the drift due to its long term memory that keeps long-distance dependency. Moreover, the encoder module, synthetic theft attacks, and dropout regularization efficiently handle the overfitting problem. Finally, handling the model's local optima trapping issue using Adam optimizer further enhances the proposed model's performance.

The FPR value is computed and displayed in Figure 11 for all the models: proposed and existing. The proposed model has the minimum FPR value among all other schemes, i.e., 0.102, whereas, LR has the maximum FPR value among all the schemes, i.e., 0.348. The proposed model outperforms other existing benchmark schemes in terms of FPR due to the effectively dealing with the drift using the long term memory in BiGRU. Secondly, the encoder module perfectly extracts the necessary features.

The area under the receiver operating characteristic curve (AUC-ROC) or in short, AUC, is considered another important performance parameter. It measures and quantifies a model's overall performance. It is obtained by plotting the

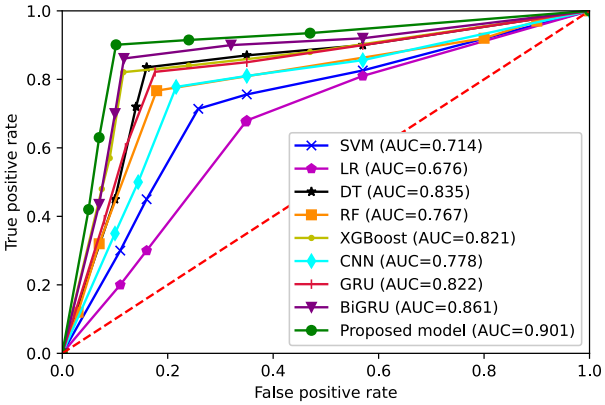


FIGURE 12. Performance comparison using AUC-ROC.

FPR on x-axis and TPR on y-axis. The AUC values lie in the range of 0 and 1. The AUC curve for the proposed and other benchmark models is given in Figure 12. The AUC score for the proposed model is 0.901, which is enhanced considerably through synthesized theft attacks. On the other hand, the AUC score for the LR classifier is 0.676, which is the lowest of all the other classifiers, which makes it the worst performing classifier.

VI. CONCLUSION AND FUTURE WORK

Aiming to deal with the problem of ETD, this article proposes an AE-BiGRU model. The synthetic theft attacks, for balancing the data, are being implemented in the proposed study. As AE has a strong ability of extracting the important features, it is employed for feature extraction. Moreover, BiGRU is used for theft classification and achieves minimum FPR value because it has long term memory, due to which it can easily identify the drift. Furthermore, eight benchmark ML and DL schemes, RF, SVM, XGBoost, LR, DT, CNN, GRU, and BiGRU, are implemented for comparison purpose. The simulation results depict that our proposed model gives excellent performance in ETD with 91.3% precision, 90.1% accuracy, 90.1% AUC score, 89.9% F1 score, 88.6% recall, and 10.2% FPR value, which are better than all the benchmarks. Hence, it is concluded that AE-BiGRU is an efficient model for ETD with maximum accuracy and minimum FPR value. Furthermore, in the future, we will consider additional non-sequential information of consumers to find the accurate location of the energy theft. In addition, the high sampling frequency of EC values and wrapper feature selection methods will be taken into account to further enhance the performance of the proposed model.

ACKNOWLEDGMENT

The authors would like to acknowledge Taif University Researchers Supporting Project number (TURSP-2020/292) Taif University, Taif, Saudi Arabia. The authors would like also to acknowledge Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2022R193), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

REFERENCES

[1] A. Khalid, N. Javaid, A. Mateen, B. Khalid, Z. A. Khan, and U. Qasim, "Demand side management using hybrid bacterial foraging and genetic algorithm optimization techniques," in *Proc. 10th Int. Conf. Complex, Intell., Softw. Intensive Syst. (CISIS)*, Fukuoka, Japan, 2016, pp. 494–502.

[2] The Scientific World—Let's Have a Moment of Science. (2022). *What are the Uses of Electricity in Modern Life?* Accessed: Dec. 18, 2021. [Online]. Available: https://www.scientificworldinfo.com/2020/05/uses-of-electricity-in-our-daily-life.html

[3] A. Aldegheishem, M. Anwar, N. Javaid, N. Alrajeh, M. Shafiq, and H. Ahmed, "Towards sustainable energy efficiency with intelligent electricity theft detection in smart grids emphasising enhanced neural networks," *IEEE Access*, vol. 9, pp. 25036–25061, 2021.

[4] M. Nabil, M. Ismail, M. M. E. A. Mahmoud, W. Alasmary, and E. Serpedin, "PPETD: Privacy-preserving electricity theft detection scheme with load monitoring and billing for AMI networks," *IEEE Access*, vol. 7, pp. 96334–96348, 2019.

[5] J. Nagi, K. S. Yap, S. K. Tiong, S. K. Ahmed, and M. Mohamad, "Nontechnical loss detection for metered customers in power utility using support vector machines," *IEEE Trans. Power Del.*, vol. 25, no. 2, pp. 1162–1171, Apr. 2010.

[6] Y. Huang and Q. Xu, "Electricity theft detection based on stacked sparse denoising autoencoder," *Int. J. Elect. Power Energy Syst.* vol. 125, Feb. 2021, Art. no. 106448.

[7] G. M. Messinis and N. D. Hatziargyriou, "Review of non-technical loss detection methods," *Electr. Power Syst. Res.*, vol. 158, pp. 250–266, May 2018.

[8] B. Yildiz, J. I. Bilbao, J. Dore, and A. B. Sproul, "Recent advances in the analysis of residential electricity consumption and applications of smart meter data," *Appl. Energy*, vol. 208, pp. 402–427, Dec. 2017.

[9] A. Ghasempour and J. Lou, "Advanced metering infrastructure in smart grid: Requirements, challenges, architectures, technologies, and optimizations," in *Smart Grids: Emerging Technologies, Challenges and Future Directions*. Hauppauge, NY, USA: Nova Science Publishers, 2017, pp. 1–8.

[10] Y. Peng, Y. Yang, Y. Xu, Y. Xue, R. Song, J. Kang, and H. Zhao, "Electricity theft detection in AMI based on clustering and local outlier factor," *IEEE Access*, vol. 9, pp. 107250–107259, 2021.

[11] Z. Zheng, Y. Yang, X. Niu, H.-N. Dai, and Y. Zhou, "Wide and deep convolutional neural networks for electricity-theft detection to secure smart grids," *IEEE Trans. Ind. Informat.*, vol. 14, no. 4, pp. 1606–1615, Apr. 2018.

[12] M.-M. Buzau, J. Tejedor-Aguilera, P. Cruz-Romero, and A. Gómez-Expósito, "Hybrid deep neural networks for detection of non-technical losses in electricity smart meters," *IEEE Trans. Power Syst.*, vol. 35, no. 2, pp. 1254–1263, Mar. 2020.

[13] P. Jokar, N. Arianpoo, and V. C. M. Leung, "Electricity theft detection in AMI using customers' consumption patterns," *IEEE Trans. Smart Grid*, vol. 7, no. 1, pp. 216–226, Jan. 2016.

[14] S. McLaughlin, B. Holbert, A. Fawaz, R. Berthier, and S. Zonouz, "A multi-sensor energy theft detection framework for advanced metering infrastructures," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 7, pp. 1319–1330, Jul. 2013.

[15] H. O. Henriques, A. P. L. Barbero, R. M. Ribeiro, M. Z. Fortes, W. Zanco, O. S. Xavier, and R. M. Amorim, "Development of adapted ammeter for fraud detection in low-voltage installations," *Measurement*, vol. 56, pp. 1–7, Oct. 2014.

[16] H. O. Henriques, R. L. S. Corrêa, M. Z. Fortes, B. S. M. C. Borba, and V. H. Ferreira, "Monitoring technical losses to improve non-technical losses estimation and detection in LV distribution systems," *Measurement*, vol. 161, Sep. 2020, Art. no. 107840.

[17] Z. Zhou, J. Bai, M. Dong, K. Ota, and S. Zhou, "Game-theoretical energy management design for smart cyber-physical power systems," *Cyber-Phys. Syst.*, vol. 1, no. 1, pp. 24–45, 2015.

[18] A. A. Cardenas, S. Amin, G. Schwartz, R. Dong, and S. Sastry, "A game theory model for electricity theft detection and privacy-aware control in AMI systems," in *Proc. 50th Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, Allerton, Liverpool, Oct. 2012, pp. 1830–1837.

[19] S. Amin, G. A. Schwartz, and H. Tembine, "Incentives and security in electricity distribution networks," in *Proc. Int. Conf. Decis. Game Theory Secur.* Berlin, Germany: Springer, 2012, pp. 264–280.

[20] Z. Qu, H. Liu, Z. Wang, J. Xu, P. Zhang, and H. Zeng, "A combined genetic optimization with AdaBoost ensemble model for anomaly detection in buildings electricity consumption," *Energy Buildings*, vol. 248, Oct. 2021, Art. no. 111193.

[21] S. K. Gunturi and D. Sarkar, "Ensemble machine learning models for the detection of energy theft," *Electr. Power Syst. Res.*, vol. 192, Mar. 2021, Art. no. 106904.

[22] M. Asif, O. Nazeer, N. Javaid, E. H. Alkhammash, and M. Hadjouni, "Data augmentation using BiWGAN, feature extraction and classification by hybrid 2DCNN and BiLSTM to detect non-technical losses in smart grids," *IEEE Access*, vol. 10, pp. 27467–27483, 2022, doi: 10.1109/ACCESS.2022.3150047.

[23] M. N. Hasan, R. N. Toma, A.-A. Nahid, M. M. M. Islam, and J.-M. Kim, "Electricity theft detection in smart grid systems: A CNN-LSTM based approach," *Energies*, vol. 12, no. 17, p. 3310, Aug. 2019.

[24] H. Liu, Z. Li, and Y. Li, "Noise reduction power stealing detection model based on self-balanced data set," *Energies*, vol. 13, no. 7, p. 1763, Apr. 2020.

[25] X. Gong, B. Tang, R. Zhu, W. Liao, and L. Song, "Data augmentation for electricity theft detection using conditional variational auto-encoder," *Energies*, vol. 13, no. 17, p. 4291, 2020.

[26] N. Javaid, N. Jan, and M. U. Javed, "An adaptive synthesis to handle imbalanced big data with deep Siamese network for electricity theft detection in smart grids," *J. Parallel Distrib. Comput.*, vol. 153, pp. 44–52, Jul. 2021.

[27] S. Hussain, M. W. Mustafa, T. A. Jumani, S. K. Baloch, H. Alotaibi, I. Khan, and A. Khan, "A novel feature engineered-CatBoost-based supervised machine learning framework for electricity theft detection," *Energy Rep.*, vol. 7, pp. 4425–4436, Nov. 2021.

[28] X. Feng, H. Hui, Z. Liang, W. Guo, H. Que, H. Feng, Y. Yao, C. Ye, and Y. Ding, "A novel electricity theft detection scheme based on text convolutional neural networks," *Energies*, vol. 13, no. 21, p. 5758, 2020.

[29] X. Kong, X. Zhao, C. Liu, Q. Li, D. Dong, and Y. Li, "Electricity theft detection in low-voltage stations based on similarity measure and DT-KSVM," *Int. J. Electr. Power Energy Syst.*, vol. 125, Feb. 2021, Art. no. 106544.

[30] R. Punmiya and S. Choe, "ToU pricing-based dynamic electricity theft detection in smart grid using gradient boosting classifier," *Appl. Sci.*, vol. 11, no. 1, p. 401, Jan. 2021.

[31] A. Arif, N. Javaid, A. Aldegheishem, and N. Alrajeh, "Big data analytics for identifying electricity theft using machine learning approaches in microgrids for smart communities," *Concurrency Comput., Pract. Exper.*, vol. 10, p. e6316, Sep. 2021.

[32] Z. Yan and H. Wen, "Electricity theft detection base on extreme gradient boosting in AMI," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–9, 2021.

[33] S. Dong, Z. Zeng, and Y. Liu, "FPETD: Fault-tolerant and privacy-preserving electricity theft detection," *Wireless Commun. Mobile Comput.*, vol. 2021, Jun. 2021, Art. no. 6650784.

[34] A. U. Pamir, S. Munawar, M. Asif, B. Kabir, and N. Javaid, "Synthetic theft attacks implementation for data balancing and a gated recurrent unit based electricity theft detection in smart grids," in *Proc. Conf. Complex, Intell., Softw. Intensive Syst.* Cham, Switzerland: Springer, 2021, pp. 395–405.

[35] *State Grid Corporation of China*. Accessed: Jun. 6, 2021. [Online]. Available: https://www.sgcc.com.cn

[36] C. Genes, I. Esnaola, S. M. Perlaza, L. F. Ochoa, and D. Coca, "Recovering missing data via matrix completion in electricity distribution systems," in *Proc. IEEE 17th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Edinburgh, U.K., Jul. 2016, pp. 1–6.

[37] N. Vandewalle, M. Ausloos, and P. Boveroux, "The moving averages demystified," *Phys. A, Stat. Mech. Appl.*, vol. 269, no. 1, pp. 170–176, 1999.

[38] S. Luan, Z. Gu, L. B. Freidovich, L. Jiang, and Q. Zhao, "Out-of-distribution detection for deep neural networks with isolation forest and local outlier factor," *IEEE Access*, vol. 9, pp. 132980–132989, 2021.

[39] I. M. Pires, F. Hussain, N. M. M. Garcia, P. Lameski, and E. Zdravevski, "Homogeneous data normalization and deep learning: A case study in human activity classification," *Future Internet*, vol. 12, no. 11, p. 194, Nov. 2020.

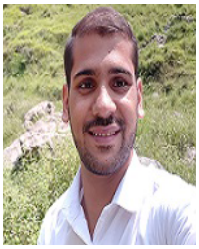
[40] R. Punmiya and S. Choe, "Energy theft detection using gradient boosting theft detector with feature engineering-based preprocessing," *IEEE Trans. Smart Grid*, vol. 10, no. 2, pp. 2326–2329, Mar. 2019.

[41] Maksoud, E. A. Abdel, S. Barakat, and M. Elmogy, "Medical images analysis based on multilabel classification," in *Machine Learning in Bio-Signal Analysis and Diagnostic Imaging*. New York, NY, USA: Academic, 2019, pp. 209–245.

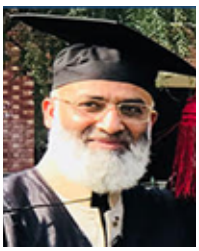
[42] (2022). *Feature Extraction*. Deepai. Accessed: Aug. 12, 2021. [Online]. Available: https://deepai.org/machine-learning-glossary-and-terms/feature-extraction

[43] H. Gul, N. Javaid, I. Ullah, A. M. Qamar, M. K. Afzal, and G. P. Joshi, "Detection of non-technical losses using SOSTLink and bidirectional gated recurrent unit to secure smart meters," *Appl. Sci.*, vol. 10, no. 9, p. 3151, Apr. 2020.

[44] (2022). *Auto-Encoder: What is it? And What is it Used for? (Part 1)*. Accessed: Aug. 18, 2021. [Online]. Available: https://towardsdatascience.com/auto-encoder-what-is-it-and-what-is-it-used-for-part-1-3e5c6f017726
[45] X. Liu, Y. Wang, X. Wang, H. Xu, C. Li, and X. Xin, "Bi-directional gated recurrent unit neural network based nonlinear equalizer for coherent optical communication system," *Opt. Exp.*, vol. 29, no. 4, pp. 5923–5933, 2021.
[46] C. Xu, J. Shen, X. Du, and F. Zhang, "An intrusion detection system using a deep neural network with gated recurrent units," *IEEE Access*, vol. 6, pp. 48697–48707, 2018.
[47] J. Yang, F. Yang, Y. Zhou, D. Wang, R. Li, G. Wang, and W. Chen, "A data-driven structural damage detection framework based on parallel convolutional neural network and bidirectional gated recurrent unit," *Inf. Sci.*, vol. 566, pp. 103–117, Aug. 2021.
[48] H.-G. Lee, G. Park, and H. Kim, "Effective integration of morphological analysis and named entity recognition based on a recurrent neural network," *Pattern Recognit. Lett.*, vol. 112, pp. 361–365, Sep. 2018.
[49] Avila, N. Fabian, G. Figueroa, and C.-C. Chu, "NTL detection in electric distribution systems using the maximal overlap discrete wavelet-packet transform and random undersampling boosting," *IEEE Trans. Power Syst.*, vol. 33, no. 6, pp. 7171–7180, Nov. 2018.
[50] Z. Aslam, N. Javaid, A. Ahmad, A. Ahmed, and S. M. Gulfam, "A combined deep learning and ensemble learning methodology to avoid electricity theft in smart grids," *Energies*, vol. 13, no. 21, p. 5599, 2020.
[51] M. S. Saeed, M. W. Mustafa, U. U. Sheikh, T. A. Jumani, I. Khan, S. Atawneh, and N. N. Hamadneh, "An efficient boosted C5.0 decision-tree-based classification approach for detecting non-technical losses in power utilities," *Energies*, vol. 13, no. 12, p. 3242, Jun. 2020.
[52] J. Pereira and F. Saraiva, "Convolutional neural network applied to detect electricity theft: A comparative study on unbalanced data handling techniques," *Int. J. Elect. Power Energy Syst.*, vol. 131, Oct. 2021, Art. no. 107085.



PAMIR received the B.S. degree in software engineering from the National University of Modern Languages (NUML), Islamabad, Pakistan, in 2016, and the M.S. degree in software engineering from the Communication Over Sensors (ComSens) Research Laboratory, Department of Computer Science, COMSATS University Islamabad (CUI), Islamabad, Pakistan, under the supervision of Dr. Nadeem Javaid, in 2018, where he is currently pursuing the Ph.D. degree in computer science. His research interests include data science, smart grids, and optimal power flow.



NADEEM JAVAID (Senior Member, IEEE) received the bachelor's degree in computer science from Gomal University, Dera Ismail Khan, Pakistan, in 1995, the master's degree in electronics from Quaid-i-Azam University, Islamabad, Pakistan, in 1999, and the Ph.D. degree from the University of Paris-Est, France, in 2010. He is currently a Professor and the Founding Director of the Communications Over Sensors (ComSens) Research Laboratory, Department of Computer Science, COMSATS University Islamabad, Islamabad Campus. He is also working as a Visiting Professor with the School of Computer Science, University of Technology Sydney, Australia. He has supervised 146 master's and 27 Ph.D. theses. He has authored over 900 articles in technical journals and international conferences. His research interests include energy optimization in smart/microgrids and in wireless sensor networks using data analytics and blockchain. He was a recipient of the Best University Teacher Award (BUTA'16) from the Higher Education Commission (HEC) of Pakistan, in 2016, and the Research Productivity Award (RPA'17) from the Pakistan Council for Science and Technology (PCST), in 2017. He is an Associate Editor of IEEE Access and an Editor of Sustainable Cities and Society.

UMAR QASIM received the B.S. degree in computer science and the M.B.A. degree from Hamdard University, and the M.S. and Ph.D. degrees in information systems from the New Jersey Institute of Technology, USA. He has been in the IT field for over 20 years and has taught at various post-secondary institutions in North America and Pakistan. He has extensive experience in the field of information technology. He worked at Dalhousie University, McMaster University, and various other software development companies in USA, for more than ten years. He headed the Digital Preservation Program with the University of Alberta, for eight years, and was responsible for internal operations and external preservation partnerships. He is currently working as an Associate Professor with the University of Engineering and Technology (UET) at Lahore, where he is involved in teaching and research and serving on various committees both at the department and university level. He maintained and shared expertise on digital preservation with the university community, and the professional community of practice at large. His research interests include the various areas of information systems, including but not limited to wireless sensor networks, databases, digital preservation, data science, and information management.

ADAMU SANI YAHAYA received the B.S. degree from Bayero University, Kano, Nigeria, in 2011, and the M.S. degree from Meliksah University (Erciyes University), Turkey, in 2014. He is currently pursuing the Ph.D. degree with the Communications Over Sensors (ComSens) Research Laboratory, Department of Computer Science, COMSATS University Islamabad, Islamabad, under the supervision of Prof. Nadeem Javaid. He is also a Lecturer with the Department of Information Technology, Bayero University. He has authored research publications in international journals and conferences. His research interests include data science, optimization, security and privacy, energy trading, blockchain, and smart grid.

EMAN H. ALKHAMMASH received the M.Sc. and Ph.D. degrees in computer science from the University of Southampton, U.K. She is currently working as an Associate Professor in computer science at Taif University, Saudi Arabia. She was awarded a Senior Fellow of the Higher Education Academy (FHEA), in March 2020. Her research interests include formal methods, AI, and data science.

MYRIAM HADJOUNI received the M.Sc. degree (Hons.) from the Higher Institute of Management of Tunis, University of Tunis, Tunisia, in 2005, and the joint Ph.D. degree (Hons.) in computer science from Paris XI (actual new name Paris Saclay) University, France, and Manouba University, Tunisia, in 2012. She is currently working as an Assistant Professor with the Computer Sciences Department, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia. Her research interests include but not restricted to information retrieval, artificial intelligence, data science, data analytic, big data, and image retrieval.