Data Augmentation Using BiWGAN, Feature Extraction and Classification by Hybrid 2DCNN and BiLSTM to Detect Non-Technical Losses in Smart Grids

MUHAMMAD ASIF[ID][1], (Graduate Student Member, IEEE), **OROOJ NAZEER**[1,2],
NADEEM JAVAID[ID][1,3], (Senior Member, IEEE), **EMAN H. ALKHAMMASH**[4],
AND MYRIAM HADJOUNI[5]

[1]Department of Computer Science, COMSATS University Islamabad, Islamabad 44000, Pakistan
[2]Department of Computing and Technology, Abasyn University, Islamabad 44000, Pakistan
[3]School of Computer Science, University of Technology Sydney, Ultimo, NSW 2007, Australia
[4]Department of Computer Science, College of Computers and Information Technology, Taif University, Taif 21944, Saudi Arabia
[5]Department of Computer Sciences, College of Computer and Information Science, Princess Nourah Bint Abdulrahman University, Riyadh 11671, Saudi Arabia

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

ABSTRACT In this paper, we present a hybrid deep learning model that is based on a two-dimensional convolutional neural network (2D-CNN) and a bidirectional long short-term memory network (Bi-LSTM)to detect non-technical losses (NTLs) in smart meters. NTLs occur due to the fraudulent use of electricity. The global integration of smart meters has proven to be beneficial for the storage of historical electricity consumption (EC) data. The proposed methodology learns the deep insights from the historical EC data and informs power utilities about the presence of NTLs. However, the effective detection of NTLs faces the problem of class imbalance that occurs due to the rare availability of fraudulent electricity consumers. To solve this issue, an evolutionary bidirectional Wasserstein generative adversarial network (Bi-WGAN) is employed. Bi-WGAN synthesizes the most plausible fraudulent EC samples by integrating an auxiliary encoder module. Besides, the inevitable curse of high dimensional data reduces the generalization ability of classifiers. The proposed hybrid model efficiently handles the highly dynamic data by utilizing its potent feature extracting capabilities. The one-dimensional daily EC data is passed to Bi-LSTM model for capturing the non-malicious changes from consumers' profiles. Meanwhile, 2D-CNN takes 2D weekly EC data as input to extract the potential features by applying different convolutions and pooling operations. Extensive experiments are conducted on a realistic smart meters dataset to prove the effectiveness of the proposed model. The results show that the proposed model outperforms the state-of-the-art models by achieving area under the curve receiver operating characteristics of 0.97 and precision-recall area under the curve of 0.98, which make it suitable for real-world scenarios.

INDEX TERMS Bidirectional generative adversarial network, convolutional neural network, data augmentation, deep learning, electricity theft detection, feature extraction, long short-term memory network, non-technical losses, smart grids.

I. INTRODUCTION

Nowadays, the major activities of human lives are dependent on the electricity. It has become an important part of human life. In the modern era, varieties of ways are introduced to generate electricity, such as production through hydro power, wind power, fuel power and thermal power. However, different losses occur during the generation of electricity [1]. The most common losses are classified into technical losses (TLs) and non-technical losses (NTLs). TLs happen because of the

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

heat production in electrical distribution lines, short circuits in transformers or other grids components, etc. Whereas, NTLs occur due to energy theft, meter bypassing, meter malfunctioning, billing errors, etc. The major source of NTL is electricity theft (ET). The power utilities around the globe accounted for billions of dollars per annum due to NTLs. The electric utilities in the United States of America bear almost $6 billion every year because of NTLs [2]. Similarly, the Chinese power companies lost almost $15 million till 2018 as a result of energy theft [3]. The underdeveloped countries are also affected by NTLs, such as Brazil and India and they lose approximately 16% and 25% of their total energy supply, respectively [4]. Besides the huge financial loss, NTLs also disturb the normal flow of electricity by overloading the transformers and grid's internal components.

The recent enhancement in advanced metering infrastructure (AMI) integrates communication flow with energy flow to enable the cooperation between consumers and electric utilities. The integration of AMI brings potential benefits, such as efficient recording of electricity usage, remote controlling of electricity consumption (EC), real-time pricing and providing grids' status information for power utilities to detect NTLs. However, it introduces numerous ways for electricity thieves to remotely compromise the smart metering systems and manipulate meters' reading [5]. Keeping the above concerns in view, electricity theft detection (ETD) has become essential for the modern era. In addition, the availability of massive EC data enables researchers to exploit state-of-the-art data driven methods for better ETD.

According to literature, different researchers performed ETD using varieties of statistical and machine learning (ML) methods [6], [7]. In general, three methods are commonly used for ETD. These methods are enlisted as follows: i) state based methods, ii) game theory based methods and iii) data driven based methods. In state or hardware based methods, special devices and sensors are integrated with the smart meters to detect the abnormal consumers [8]. However, these methods are costly in terms of both time and money. Moreover, extra maintenance cost is required for installation and management of these devices. Whereas, in game theory based methods, a virtual environment is initially created. Then, a game is played between electric utilities and consumers to perform ETD [9]. A special utility function is formulated where the rules and regulations are defined. The game is stopped when the equilibrium state is achieved. However, these methods are not proven to be effective because designing a suitable utility function for complex scenarios is a challenging task for researchers. In contrast, the data driven based methods demand only data for model's training so they become cost effective solutions to perform ETD. The massive availability of EC data enables the application of numerous data driven based solutions. The researchers put their efforts by adopting different supervised and unsupervised ML solutions to detect electricity thieves and support the power industries to reduce revenue loss.

TABLE 1. List of Acronyms.

Abbreviation	Description
AMI	Advanced metering infrastructure
ADASYN	Adaptive synthetic
AUC	Area under the curve
AUC-ROC	AUC receiver operating characteristics
Bi-GRU	Bidirectional gated recurrent unit
Bi-LSTM	Bidirectional long short-term memory
Bi-WGAN	Bidirectional wasserstein GAN
CNN	Convolutional neural network
DR	Detection rate
EC	Electricity consumption
ETD	Electricity theft detection
FP	False positive
FPR	FP rate
FN	False negative
GAN	Generative adversarial network
ICS	Internal covariate shift
LR	Logistic regression
ML	Machine learning
MCC	Matthews correlation coefficient
MLP	Multilayer perceptron
NTLs	Non-technical losses
1D	One-dimensional
PR-AUC	Precision-recall area under the curve
RF	Random forest
RUS	Random under sampling
SMOTE	Synthetic minority oversampling technique
SGCC	State grid corporation of China
SVM	Support vector machine
TLs	Technical losses
TN	True negative
TP	True positive
TPR	TP rate
2D	Two-dimensional
WD-CNN	Wide and deep CNN
b	Bias factor
$\bigtriangledown$	Gradient
α	Learning rate
z	Latent noise
$\mathbb{P}_r$	Probability distribution of real dataset
$\mathbb{P}_z$	Probability distribution of latent noise
S_{prep}	Preprocessed dataset
S_{real}	Real electricity consumption dataset
σ	Sigmoid function

In recent literature, varieties of supervised and unsupervised methods are adopted to detect energy thieves in smart grids. In this regard, several machine and deep learning based solutions are proposed by researchers to perform ETD [1], [10]–[12], [13]. However, these solutions do not provide satisfactory results because of inefficient feature engineering. Poor feature engineering also degrades the generalization ability of models. Moreover, limited amount of labeled EC data is another underlying cause that decreases the detection accuracy. Furthermore, in deep learning models, the problem of internal covariate shift (ICS) adversely affects the stable learning of hidden layers [1], [14]. ICS occurs when the input distribution of a hidden neural layer is transferred to other layers. The severe lack of fraudulent electricity consumers in real-world scenarios creates a class imbalance problem, which is an important concern for efficient ETD [1], [5], [14], [15]. In addition, the noisy and high dimensional data leads

to the curse of dimensionality issue, which is confronted by the researchers during ETD [14].

Keeping the above concerns in view, we propose a novel deep learning solution to improve the detection accuracy of ETD in power grids. The proposed model consists of a two-dimensional convolutional neural network (2D-CNN) and a bidirectional long short-term memory (Bi-LSTM). A bidirectional Wasserstein generative adversarial network (Bi-WGAN) is exploited for synthesizing the minority class theft samples. The one-dimensional (1D) daily EC data is converted into a 2D manner according to weeks. 2D-CNN is developed to capture the weekly insights and periodicity from 2D weekly data. Meanwhile, Bi-LSTM takes 1D data as input and extracts the long-term temporal correlation from EC profiles. It also overcomes the effects of non-malicious factors and consequently, reduces the high false positive rate (FPR). Finally, a single feature vector is devised by merging the outcomes of both models. Then, a sigmoid function is employed for final ETD. It is worth mentioning that this work is the extension of [16].

The major contributions of this study are enlisted as follows.

- A novel state-of-the-art methodology is introduced, which combines 2D-CNN and Bi-LSTM models. The proposed model efficiently performs feature extraction and resolves the curse of dimensionality issue.
- The Bi-WGAN model is employed to resolve the inevitable class imbalance problem. The samples generated by the model are closely related to real-world theft patterns. To the best of our knowledge, we apply Bi-WGAN first time in the ETD domain for augmenting the theft class samples.
- The Bi-LSTM model is leveraged to handle the problem of high FPR, which occurs due to several non-malicious factors. The model intelligently captures long-term tendency and temporal correlations from the EC data to minimize the effects of non-malicious changes.
- For comprehensive analysis of the proposed model, area under the curve (AUC), precision, recall, AUC receiver operating characteristics (AUC-ROC), precision-recall AUC (PR-AUC), F1-score and Matthews correlation coefficient (MCC) metrics are considered.

The organization of the manuscript is as follows. The related work is presented in Section II. The formulation and analysis of the problem statement are given in Section III. The proposed scheme is explained in Section IV. Section V describes the experimental results of the proposed and benchmark schemes. In last, the manuscript is concluded in Section VI.

II. RELATED WORK

The literature is saturated with numerous statistical and ML models where ETD is performed. In fact, these models require handcraft feature engineering and pertinent domain expertise, which is a difficult and time-consuming task. The existing ML models under-performed while capturing temporal correlations and complex non linearities from EC profiles. In general, most of the ML models performed ETD by

utilizing only 1D EC data. However, catching latent features and periodicity from 1D data is a difficult process [1]. In [14], it is referred that all conventional schemes are centered around manual feature engineering in order to identify NTL patterns. Moreover, in the existing work, no mathematical based solutions are established to distinguish shunt and double tapping attacks. The authors of [17] examine that the existing ML algorithms are not taken into account for the proper feature engineering step, which consequently leads to the poor generalization issue.

The authors of [5] identify that many conventional ML techniques are exploited to detect NTLs in power grids. However, they neglect an efficient feature engineering process that results in poor generalization and low detection accuracy. Many classification and clustering techniques make an early decision about the abrupt changes in consumers' consumption that results in a high FPR because it may happen due to several non-malicious factors, e.g., weekends, change of residents, change of appliances, change of seasonality, etc. Moreover, the existing techniques perform poorly in the detection of zero-day attacks. Similarly, the authors of [12] and [18] highlight the issue of inappropriate feature engineering. The process of handcraft feature engineering demands the involvement of domain expert, which is a time intensive and difficult task. In [18], the most prominent features are extracted through autoencoder from highly dynamic EC data to perform efficient ETD. However, further improvement is needed to recognize some intelligent attacks, such as shunt attack, zero data attack, double tapping attack and so forth.

In [19], numerous clustering based techniques are exploited for anomaly detection in smart meters' data. However, the fluctuations and variations in the normal and theft load profiles are not properly detected, which yield poor detection results. Similarly, the authors in [20] analyze some traditional techniques that are applied to detect data poisoning attacks. However, these techniques add up an additional stage of data filtering, which first removes any available false label and then performs the detection step. In [21]–[24], the authors discuss that many pattern recognition and conventional ML techniques are employed for NTL detection. These techniques demand extensive handcraft feature engineering, which is a laborious, time-consuming and financially expensive task. Moreover, the re-involvement of the domain experts is needed when new features are to be required. In addition, these techniques poorly perform to extract vital features from the available high dimensional EC data.

According to [25], many conventional anomaly detection algorithms mistakenly detect the normal user as abnormal because of several non-malicious factors: changes of home residents, weekends, changes in the number of appliances, etc. These non-malicious factors also become the reason of high FPR. Moreover, in [21], it is mentioned that many researchers exploit deep learning models for theft identification and self feature learning from the highly dynamic EC data. However, these models are tested and evaluated on the artificially generated data, which is not effective for a

reliable assessment. According to [26] and [27], the manual creation of features is not sufficient to properly detect the NTL behavior because of stochastic changes in EC profiles. In [28], the problem of maintaining temporal correlation in the existing ML models is highlighted. Moreover, the learning algorithms are unable to learn the potential features from 1D raw EC data.

The study of [29] demonstrates that many researchers propose different electricity theft detectors. However, these detectors have low detection accuracy because the EC data is a highly dynamic and rapidly growing time-series data. In [30], the authors discuss that many conventional data mining and ML techniques are exploited to filter customers' consumption patterns for the detection of irregular electricity profiles. However, these techniques under-perform because of improper feature engineering. Moreover, different non-malicious factors mislead the classification model in a wrong direction, which is a quite serious issue in the existing research. From [31] and [32], numerous non-malicious factors degrade the detection accuracy of traditional ML models. In [33], bidirectional gated recurrent unit (Bi-GRU) is used for extracting the high level features from the electricity load profile in order to detect NTLs. However, synthetic minority oversampling technique (SMOTE) and SMOTE over sampling tomik link are used for data balancing, which raise overfitting issue because of generating duplicate records and vanishing the temporal correlation between consumption patterns. In addition, the authors of [34] discover that the existing deep learning techniques are not suitable for anomaly detection in electricity power data because of interpretability and practicality concerns. On the other hand, the authors of [2], [12], [17], [21], [29] and [35] highlight a critical class imbalance issue that occurs in ETD because of less availability of fraudulent consumers. Consequently, the majority class dominates the minority class, which leads to high FPR. Moreover, the learning algorithms are skewed towards the majority class. As a result, the misclassification rate is increased to a greater extent. According to [4], [11] and [19], the problem of limited amount of labeled EC data becomes challenging for ML algorithms to perform efficient ETD. Similarly, the authors of [22], [26] and [28] examine that the severe imbalance proportion of classes adversely affect the generalization power of classifiers. Due to this, the classification algorithms have higher chance to suffer from the overfitting issue.

From [32] and [36], the existing literature is teemed with various oversampling techniques that are employed to handle the problem of class imbalance. In oversampling techniques, the minority class samples are augmented and the proportion of classes is equalized. SMOTE, K-mean SMOTE, adaptive synthetic (ADASYN) and so forth are well known oversampling techniques that are used to synthesize the minority class instances. The GAN model is also exploited to augment the minority class samples. It becomes popular due to its tremendous success in generating artificial data. However, the above mentioned techniques lack in capturing the arbitrary fluctuation and probabilistic curve from EC patterns while generating fraudulent samples. Consequently, the final classification results do not provide real-world assessment.

III. PROBLEM ANALYSIS

With the advent of AMI, the energy flow is integrated with the communication flow in order to establish two way real time coordination between consumers and power industries. However, with the involvement of the Internet, the communication flow can be prone to different contamination attacks, which are harmful for power utilities and become one of the reasons for NTLs. So, there is an important need for a robust ETD model. In [1], wide and deep convolutional neural network (WD-CNN) is proposed to reduce the curse of dimensionality. However, a single layer of neural network is integrated inside the wide component that does not learn the temporal correlation and hidden features from 1D EC data and also gets stuck in local optima. Moreover, the models presented in [2], [4] and [14] do not use any feature extraction module to reduce the data dimensionality. The rapid growth in the dimensions of time series data degrades the model's accuracy and increases the computational overhead. Therefore, if data dimensionality is not handled correctly, the deep or ML models memorize the noise and redundant features that lead toward poor generalization problem. Furthermore, the ICS is another common issue that occurs in deep neural networks. It happens due to the shifting of input distribution between different layers of neural networks and the changing of network parameters on each hidden layer. However, in [1] and [14], no mechanism is presented to handle the ICS problem, which adversely affects the stable learning of neural networks. It also degrades the hidden layers' feature learning capabilities, increases the training time and slows down the convergence rate. Another major issue faced by the researchers is the high FPR that occurs due to several non-malicious factors and false injection of noise in data by the intelligent attackers. For instance, the deep learning models used in [1] and [21] are unable to capture the non-malicious changes and long-term temporal correlation from the EC data, which increases the FPR and onsite inspection cost as well.

The imbalanced nature of data is another major concern that occurs when detecting energy thieves. It raises the overfitting and poor generalization issues. In [1], [14] and [15], the problem of imbalance data is not handled. As a result, the classification model is skewed towards the larger class. Furthermore, in [11] and [29], the dataset is balanced through random under sampling (RUS), which overlooks the important information. Moreover, in [4] and [22], the authors exploit SMOTE approach for data balancing. It generates the synthetic samples without considering the overlapping of neighboring samples. Therefore, it introduces an additional noise and increases the ratio of duplicate records, which lead the models towards overfitting. Furthermore, in ETD, the selection of appropriate performance metric is a necessary task for better evaluation of a model.

However, in [2] and [19], the appropriate metrics are not considered for performing a comprehensive analysis.

IV. PROPOSED ELECTRICITY THEFT DETECTION MODEL

This section describes the architecture of the proposed electricity theft detection model, which is divided into four stages.

1) In the first stage, data preprocessing is performed in which missing values are filled through linear interpolation method, outliers are handled by three sigma rule (TSR) and feature scaling is done using Min-Max normalization.
2) In the second stage, class imbalance issue is resolved by augmenting the minority class theft samples using Bi-WGAN.
3) In the third stage, a hybrid deep learning model is designed in which two modules, termed as 2D-CNN and Bi-LSTM, are integrated in a parallel manner to perform efficient feature extraction and memorization of temporal EC patterns.
4) In the fourth stage, a hybrid module is developed to perform the classification of theft and benign consumers.

Further explanation about the above mentioned steps is given in the upcoming subsections. Moreover, the complete representation of the proposed scheme is shown in Fig. 1. For easy understanding, a unique step number is assigned to each stage. In the first step, data preprocessing is carried out. In the second step, the preprocessed data is separated into minority theft class and majority benign class. In the third step, the data augmentation is performed by simulating theft samples. The balanced dataset is produced at step four by concatenating the augmented theft samples with benign ones. In the fifth and sixth steps, feature extraction and memorization of temporal EC patterns are preformed by 2D-CNN and Bi-LSTM, respectively. Finally, the classification is performed in the seventh step by leveraging a fully connected neural network.

A. DATA PREPROCESSING MODULE

The EC data recorded through AMI may contain noisy, erroneous and missing values. This is because of the metering faults, problem in storage devices, meter tampering, etc. The erroneous values in the dataset should be removed for achieving accurate results. Therefore, the data preprocessing techniques are adopted to handle the above issues. Missing values are tackled through a linear interpolation method [1]. The equation used for filling the missing values is given below.

$$f(x_i) = \begin{cases} \dfrac{x_{i\text{-}1} + x_{i+1}}{2}, & x_i == NaN, x_{i\pm 1} \neq NaN, \\ 0, & x_i == NaN, x_{i\text{-}1} \text{ or } x_{i+1} == NaN, \\ x_i, & x_i \neq NaN. \end{cases}$$

(1)

where x_i represents the electricity usage of a consumer over a period i (e.g., a day). The equation has three parts. The first part ensures that the EC value of a user at period $i \pm 1$ should

not be equal to NAN. If the condition is satisfied, the missing EC value of the consumer x_i is filled by taking the average of $i \pm 1$ EC values. Otherwise, the missing value is filled by zero, which is the second part of equation. The third part of the equation states that if x_i is not NAN then do not change it. Similarly, some unusual values are also found in the EC dataset. These values are referred to as outliers. The outliers badly degrade the system performance. In this case, we handle the outlier using a well known method, termed as TSR [37]. The mathematical equation of TSR is given below.

$$f(x_i) = \begin{cases} \bar{x} + 2 \times \sigma(x), & if \ x_i > \bar{x} + 2 \times \sigma(x), \\ x_i, & otherwise. \end{cases}$$

(2)

where x shows the real EC vector of a consumer and $\bar{x}$ represents the average value of real usage. σ denotes the standard deviation. In equation 2, the expression $x_i > \bar{x} + 2 \times \sigma(x)$ states that if x_i does not follow the Gaussian distribution, it will be declared as an outlier and will be handled by filling with $\bar{x} + 2 \times \sigma(x)$. After incorporating outliers and missing values, there is a need to scale the EC data. If we pass EC data to neural networks without proper feature scaling, it may raise the gradient exploding issue and increase the computational overhead. The convergence rate of the neural network is also suffered. Therefore, we adopt Min-Max normalization technique to scale the EC data in the range of 0 to 1. The equation of Min-Max normalization is given below.

$$x_{\text{new}} = \frac{x_i - min(x)}{max(x) - min(x)}.$$

(3)

In equation 3, $max(x)$ and $min(x)$ represent the maximum and minimum EC of a user, respectively. Algorithm 1 describes the complete workflow of data preprocessing steps. The input, output, variables and functions of the algorithm are described in lines 1 to 7. The lines 8 to 15 define the linear interpolation method used for handling the missing values present in the electricity load profiles. Similarly, the lines from 17 to 21 and 23 deal with outliers and features scaling, respectively.

B. DATA AUGMENTATION MODULE

The problem of data imbalance adversely affects the performance of classification algorithms. This issue is raised when the data samples of one class is higher than the other class. In ETD, this problem commonly occurs because the data samples of theft consumers are rarely available. As a result, the classification algorithms get biased towards the majority class and ignore the minority class. Keeping this in view, Bi-WGAN model is opted in this work to resolve the class imbalance problem by simulating the EC patterns of fraudulent consumers. In [28], it is used for extracting the rich task-targeting features from the EC data and shows satisfactory performance. Moreover, in [38], it performs efficiently while synthesizing the fake image samples. Hence, we are inspired and motivated from [28] and [38] and exploited Bi-WGAN for generating the theft class samples. The synthesized theft patterns of Bi-WGAN closely mimic

---

Algorithm 1 Data Preprocessing

---

1 **Input**: Real dataset
   $S_{real} = \{(x^1, y^1), (x^2, y^2), \ldots, (x^n, y^n)\}, x, y \in \mathbb{R}$
2 **Output**: Preprocessed dataset S_{prep}
3 **Variables and Functions**: EC of user $x \subseteq S_{real}$
4 $min(x)$: minimum consumption value of user x
5 $max(x)$: maximum consumption value of user x
6 S_{prep}: store preprocessed data
7 σ: standard deviation, $avg(x)$: average value of x
   Handling missing values:
8 **for** $n=1$ to $S_{real}.length$ **do**
9   **for** $i=1$ to $x.length$ **do**
10   **if** $x_i^n == NaN$ && $x_{i-1}^n || x_{i+1}^n \neq NaN$ **then**
11        $x_i^n = (x_{i-1}^n + x_{i+1}^n)/2$
12   **end**
13   **if** $x_i^n == NaN$ && $x_{i-1}^n || x_{i+1}^n == NaN$ **then**
14        $x_i^n = 0$
15   **end**
16   **Outlier detection:**
17   **if** $x_i^n > avg(x^n) + 2 * \sigma(x^n)$ **then**
18        $x_i^n = avg(x^n) + 2 * \sigma(x^n)$
19   **else**
20        $x_i^n = x_i^n$
21   **end**
22   **Normalization:**
23      $x_i^n = (x_i^n - min(x^n))/(max(x^n) - min(x^n))$
24   **end**
25    $S_{prep}^n = x^n$
26 **end**

---

the patterns of real-world electricity thieves. Moreover, the auxiliary encoder model strengthens the augmentation ability of Bi-WGAN model through inverse mapping of original input to the latent dimension.

Bi-WGAN is the advanced version of Bi-GAN and WGAN [39], [40]. It is introduced to mitigate the drawbacks of traditional GAN [41]. The traditional GAN suffers from mode collapse, vanishing gradient and nash equilibrium problems. The mode collapse issue occurs when the generator model generates almost the same data. In GAN, the Jensen divergence loss function is used, which raises the vanishing gradient issue during the adversarial training. Furthermore, both generator and discriminator try to update their loss functions, simultaneously, which affect the convergence speed of the GAN model. Moreover, in traditional GAN, only the mapping from latent space to the samples exists, while the inverse mapping is not present. In Bi-WGAN, an external encoder module is attached with the generator network for performing the inverse mapping of the real input to the latent space. Moreover, an updated loss function, known as Wasserstein distance (WD) [35], is used instead of Jensen divergence. This function assists the model to obtain an optimal solution within minimum time. In this manner, the convergence speed of the model towards the global optimum solution

is enhanced. The overall working of Bi-WGAN by augmenting electricity theft samples is explained below.

The available electricity theft data is selected as an input for the training of Bi-WGAN model. It utilizes the objective function and loss function of Bi-GAN and WGAN, respectively. Equation 4 presents the objective function of Bi-WGAN [32].

$$\min_{GE} \max_{D} V(G, E, D) = \mathbb{E}_{x \sim Px(x)}[logD(x, E(x))]$$
$$+ \mathbb{E}_{z \sim Pz(z)}[log(1 - D(G(z), z))]. \quad (4)$$

where G, E, D represent generator, encoder and discriminator models, respectively. The original distribution of electricity theft samples is denoted by $Px(x)$. $Pz(z)$ indicates the distribution of latent noise z. $\mathbb{E}_x$ and $\mathbb{E}_z$ depict the overall expected values of discriminator and generator models, respectively. $E(x)$ represents the encoded representation of the real electricity theft data x. A zero-sum game is conducted among G, E and D to achieve an optimal output, which is the high resemblance electricity theft patterns. G is responsible for generating those samples, which mimic the patterns of real-world thieves. Whereas, the goal of D is to check either the generated theft data is real or fake. We pass real theft samples along with the generated samples of G to D for differentiating between real and fake samples. The role of E is to improve the capabilities of G by adding the encoded representation $E(x)$ back to the latent dimension z. The training process continues until $Pz(z)$ becomes similar to $Px(x)$. To measure the differences between the real and the fake probability distributions of theft samples, WD is utilized. It shifts the small amount of $Px(x)$ to $Pz(z)$ for generating those theft samples, which are closely related to the real-world thieves. In this way, WD improves the convergence speed and the stable learning of Bi-WGAN model. The mathematical formulation of WD [35] is given below.

$$W(Px(x), Pz(z)) = \inf_{\gamma \in \Pi(Px(x), Pz(z))} \mathbb{E}_{(x,z) \sim \gamma}[\|x - z\|]. \quad (5)$$

where $\Pi(Px(x), Pz(z))$ demonstrates the set of joint distributions $\gamma(x, z)$. Whereas, $|x, z|$ denotes the mass transported from the value of x to z. The overall aim of $W(Px(x), Pz(z))$ is to reduce the difference between $Px(x)$ and $Pz(z)$ to a minimal level, so that the generated EC samples of G have a high resemblance with the real-world electricity thieves.

In Algorithm 2, the process of handling class imbalance problem is presented. The lines from 1 to 7 describe the input, output, variables and functions for the algorithm. The preprocessed data is split into honest and theft consumers at line 8. In lines 9 and 10, the probability distribution for Bi-WGAN is formulated using the real EC data of energy thieves and random noise, respectively. The lines 11 to 25 present the training process of both generator and discriminator models. The training process is not stopped until the model finds the optimal weight parameters and minimum loss value. Afterwards, the lines 27 and 28 describe the sample generation of theft class through Bi-WGAN after after successfully training
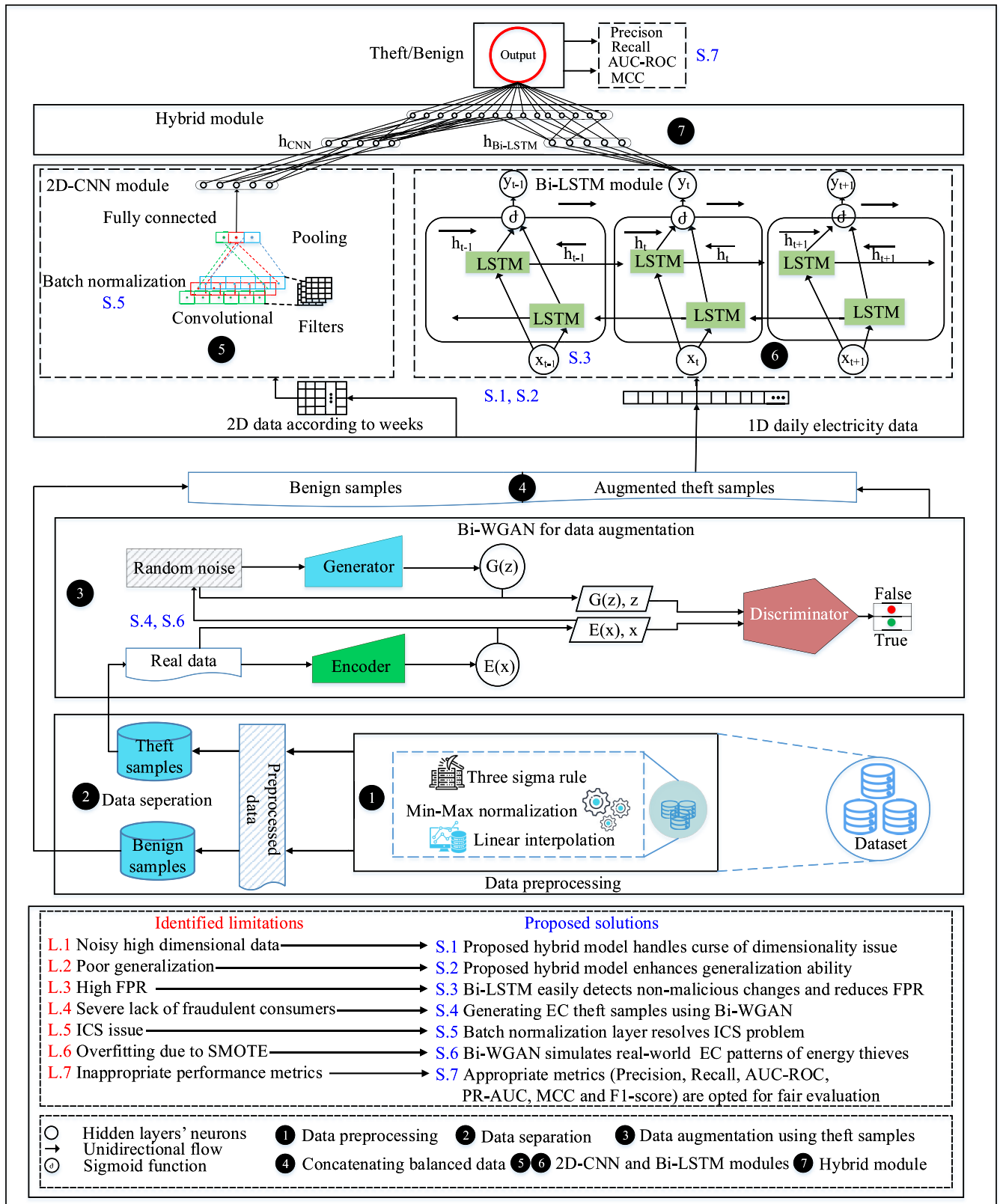
FIGURE 1. The proposed electricity theft detection model.

Algorithm 2 Bi-WGAN for Data Augmentation

1   **Input**: Preprocessed dataset
    $S_{prep} = \{(x^1, y^1), (x^2, y^2), \dots, (x^n, y^n)\}, \ x, y \in \mathbb{R},$
2   **Output**: Parameters after training θ_G, θ_D, trained
    Bi-WGAN model G_{train}, balanced dataset S_{bal}
3   **Variables and Functions**: $S_{bal}, X_{theft}, X_{honest}$,
    $\alpha = 0.00005, c = 0.01, \theta_G$ initial generator parameter,
    θ_D initial discriminator parameter, size of batch m,
    discriminator's counter $n_{critics}$, encoder ε, encoded input
    e_{in}
4   RMSprop(α): optimizer
5   split(): splitting theft and honest users' data
6   clip(): for clipping weights
7   Bi-WGAN process:
8   $X_{theft}, X_{honest} = split(S_{prep})$
9   $\mathbb{P}_r = P_{distribution}(X_{theft})$
10   $\mathbb{P}_z = P_{distribution}(z)$
11   **while** θ_G *has not converged* **do**
12   **for** *j=0 to* $n_{critics}$ **do**
13   Sample from real data distribution $x^m_{i=1} \sim \mathbb{P}_r$
14   Sample from latent data distribution $z^m_{i=1} \sim \mathbb{P}_z$
15      $e_{in} = \varepsilon(x)$
16      $\hat{x} = G(z)$
17      $l_d =$
       $\nabla_d \left[\frac{1}{m} \sum_{i=1}^m D_w(x, e_{in}) - \frac{1}{m} \sum_{i=1}^m D_w(\hat{x}, z) \right]$
18      $\theta_d = \theta_d + \alpha.RMSProp(\theta_d, l_g)$
19      $\theta_d = clip(\theta_d, -c, c)$
20   **end**
21   Sample a batch from latent variable $z^m_{i=1} \sim \mathbb{P}_z$
22    $l_g = - \nabla \frac{1}{m} \sum_{i=1}^m D_w(z)$
23    $\theta_g = \theta_g + \alpha.RMSProp(\theta_g, l_g)$
24   update $G_{train}(\theta_g)$
25   **end**
26   After training of generator, theft samples are generated
27   $X_{gen} = G_{train}.predict(N_{sample})$
28   $S_{bal} = concatenate(X_{gen}, X_{theft})$

---

the model. In addition, notations and symbols used in the algorithm is taken from [42].

C. ARCHITECTURE OF THE PROPOSED HYBRID MODEL

In this study, a hybrid deep learning model is developed, which is the combination of 2D-CNN and Bi-LSTM. The hybrid model performs better than standalone model that is proved in [43]. Both 2D-CNN and Bi-LSTM models are integrated in a parallel manner. 2D-CNN takes 2D weekly EC data for extracting the potential feature and periodicity from consumers' profiles. Meanwhile, 1D daily electricity data is passed to Bi-LSTM for memorizing the global and temporal correlated features. At the end, both models' outcomes are combined in the hybrid module for final classification. The detailed working of these modules is provided in the following subsections.

1) 2D CONVOLUTIONAL NEURAL NETWORK

CNN is introduced to automatically capture the complex feature representation and non-linearity from highly dynamic data. It is mostly used in the domain of image processing and computer vision. However, the authors of [44] employed it for a speech recognition task. The results showed the superior performance of CNN by capturing the latent correlations from the speech data. In [1], a 2D-CNN is constructed with the help of 2D convolution and pooling layers to explore the electricity load profiles. It extracts the promising EC patterns for efficient ETD. Therefore, motivated from [1] and [44], we design a 2D-CNN model to investigate the electricity load profiles. The major task of 2D-CNN is to learn the hidden representations and potential features from the highly dynamic feature space. Most of the EC datasets are provided in 1D raw form. They contain the daily EC records of different consumers. Since the 1D EC data has limited periodicity and associations in EC patterns, so there is a need to transform 1D daily EC profiles of consumers into 2D weekly profiles. Therefore, 1D data is converted into 2D weekly data. 2D-CNN takes this data as input and passes it through various filtrations, convolutions and pooling operations to capture the latent trends and hidden fluctuations for better generalization. In convolutional operations, different filters are incorporated. They learn hidden feature representations and generate feature maps accordingly. Afterwards, pooling operations are performed to diminish the spatial dimensions of generated feature maps. In particular, we opt a max pooling strategy in this work. The max pooling strategy picks up the highest values from the given receptive field of the specific feature map and drops the remaining values. The dropout layers are added in 2D-CNN to avoid overfitting issue. Moreover, we add batch normalization layers in 2D-CNN to prevent it from the ICS problem. Furthermore, the deep learning models are very sensitive to diverse data, so the data should be in a normalized form before passing it to the next layer. Otherwise, they will become vulnerable to the gradient exploding or overfitting problems. The mathematical formulation of the convolutional layer [1] of 2D-CNN is as follows.

$$y^i = \sigma_i(w_i * x_i + b_i). \qquad (6)$$

where σ_i depicts the sigmoid activation function and y^i represents the output of ith convolutional layer. x_i refers to the input, which is basically 2D weekly EC data. Similarly, w_i denotes the weight of ith convolutional layer and b_i depicts the bias factor. The output y^i stores feature maps after the convolving operations are performed. Afterwards, the pooling operations are performed through a max pooling strategy. The equation of the max pooling layers is shown below.

$$y^m = max_{i,j \in R}(y^{i,j}). \qquad (7)$$

where y^m denotes the outcomes of max pooling layers, which contain the reduced feature maps. Similarly, j depicts the jth neurons of a specific convolutional layer. The dropout and batch normalization layers are added to prevent the model from overfitting and ICS issues. Moreover, the flatten layer is

utilized to convert the feature map into 1D vector for establishing connectivity between the following pooling layers and the upcoming fully connected layer. The mathematical derivation of the fully connected layer is as follows.

$$y^f = g_i(w_i^f * y^m + b_i^f). \tag{8}$$

where g_i represents the activation function. w_i^f and b_i^f denote the weight and bias factors of the fully connected layer, respectively. y^f shows the output of the fully connected layer, which contains the most important feature set that is extracted from the 2D EC data. This feature set is further passed to the hybrid module where it is concatenated with the feature set of Bi-LSTM for the final classification as a honest or a malicious consumer.

2) BIDIRECTIONAL LONG SHORT-TERM MEMORY NETWORK

The EC data contains lots of fluctuations in the EC profiles of consumers. We observed that the electricity patterns of consumers have a strong association with each other. In this regard, we opt a Bi-LSTM model to capture the long-term trend from the EC data for better NTL detection. The selection of Bi-LSTM is made because the authors of [45] prove that its performance is outstanding in predicting the traffic routes. The traffic routes dataset belongs to the time series data. In the case of ETD, the EC data is also associated with the time series data [46]. Moreover, the other reason of using Bi-LSTM is that it stores the EC patterns for a long time in its memory states to identify the effects of non-malicious changes. As a result, it reduces the false detection of electricity consumers to a minimal level.

Bi-LSTM is the extension of the traditional LSTM model in which two sub-models are trained simultaneously. The first sub-model works in the forward direction and the other one works in the backward direction. Both sub-models are aimed to learn long-term periodicity and temporal correlation in EC load profiles. In Bi-LSTM, the provision of context about EC patterns in both directions further improves its feature learning capabilities. It also memorizes the long-term historical EC patterns of consumers' profiles, which are beneficial to deal with the non-malicious changes. Consequently, the high FPR is reduced to a greater extent. The reduction in FPR helps the power utilities to save the maximum monetary cost that is incurred in unnecessary onsite inspections.

Moreover, Bi-LSTM maintains the long-term sequence in EC patterns through the collaboration of both short and long-term memory states. The long-term memory state stores the historical information for a long time. This state is updated at each time step with the updated information. Whereas, the short-term memory state consists of different memory gates that keep the output at current time step. There are three memory gates that work in the short-term memory state. The input gate decides how much input data should be kept and how much will be thrown away. It employs sigmoid function for making the decision. Moreover, it utilizes both current and previous state input data during decision process.

Similarly, the unnecessary information is discarded by the forget gate. It passes only important information to the cell state. In last, the final decision about how much information is passed to the next hidden state is taken by the output gate. In addition, the long-term historical information is stored in cell state for future decisions. The process of storing the information in both directions increases the detection accuracy and reduces the high FPR. The mathematical representations of different memory gates [14] are given as follows.

$$f_t = \sigma_z(W_f x_t + U_f h_{t-1} + b_f), \tag{9}$$
$$i_t = \sigma_z(W_i x_t + U_i h_{t-1} + b_i), \tag{10}$$
$$o_t = \sigma_z(W_o x_t + U_o h_{t-1} + b_o), \tag{11}$$
$$\hat{c}_t = \sigma_z(W_c x_t + U_c h_{t-1} + b_c), \tag{12}$$
$$c_t = f_t * c_{t-1} + i_t * \hat{c}_t, \tag{13}$$
$$h_t = o_t * \sigma_z(c_t). \tag{14}$$

where i_t, f_t and o_t denote the values of input, forget and output gates at current time step, respectively. Similarly, σ_z denotes the sigmoid activation function of the corresponding gate, which decides about the activation of the gates. W and U indicate the weights matrices, which are integrated with the input of current and previous time steps, respectively. Moreover, $\hat{c}_t$ and c_t signify the values in cell state at current and overall timestamps, respectively. h_t represents hidden state at time t. The factor b shows the bias term.

3) HYBRID MODULE

The hybrid module refers to a combined module where the outcomes of both Bi-LSTM and 2D-CNN modules are integrated into a unique feature vector. A joint weight matrix is constructed for the hybrid training of both models. Finally, a sigmoid function is applied on the combined feature vector for the detection of NTL patterns.

$$NTL_{det} = \sigma_h(W[h_{2D-CNN}, h_{Bi-LSTM}] + b), \tag{15}$$

where σ_h denotes the sigmoid activation function. h_{2D-CNN} and $h_{Bi-LSTM}$ represent the final output of 2D-CNN and Bi-LSTM models, respectively. Similarly, W denotes the joint weight for a hybrid model and b is the bias factor. Algorithm 3 describes the process of feature learning and NTL detection through the hybrid 2D-CNN and Bi-LSTM model. The lines 1 to 3 describe the input, output, variables and functions of the algorithm. In lines 4 and 5, the transformation of data from 1D to 2D is given. The lines from 6 to 12 present the overall working mechanism of 2D-CNN model. Similarly, the lines 14 to 30 indicate the learning process of Bi-LSTM model. The lines from 19 to 24 describe the updating process of memory gates and cell states. These gates keep or throw current state information according to the previous cell state information. The lines from 31 to 38 show the backpropagation and the weight updating processes for different memory gates. Finally, the line 40 describes detection of energy thieves.

Algorithm 3 The Proposed Hybrid Model

---

1 **Input**: Balanced dataset
   $S_{bal} = \{(x^1, y^1), (x^2, y^2), \ldots, (x^n, y^n)\}, x, y \in \mathbb{R}$
2 **Output**: NTL_{det}
3 **Variables and Functions**: Weights
   W^l, U^l, b^l and $\forall\, l,\ h_{t-1}^L$
4 $X_{1D} = S_{bal}$
5 $X_{2D} = transform(X_{1D})$
6 **2D-CNN working**:
7 Input layer $x_i = Input(X_{2D})$
8 Convolutional layers: x_i denotes input of convolutional
   layers
9 $y_i = \sigma_i(w_i * x_i + b_i)$
10 Max pooling layers: $y^m = max_{i,j \in R}(y^{i,j})$
11 Fully connected layer: $y^f = g_i(w_i^f * y^m + b_i^f)$
12 $h_{2D-CNN} = y^f$
13 **Bi-LSTM working mechanism**:
14 **while** W^l, U^l and b^l not converge **do**
15   **for** $x \in X_{1D}$ **do**
16   **Same process for forward and backward pass**
17   **for** *each hidden layer l=1 to l/2* **do**
18   **for** *each time step t* **do**
19         $i_t = \sigma(W_i^l x_t^l + U_i^l h_{t-1} + b_i^l)$
20         $f_t = \sigma(W_f^l x_t^l + U_f^l h_{t-1} + b_i^l)$
21         $o_t = \sigma(W_o^l x_t^l + U_o^l h_{t-1} + b_i^l)$
22         $\hat{c}_t = \sigma(W_c^l x_t^l + U_c^l h_{t-1} + b_c^l)$
23         $c_t^l = f_t^l * c_{t-1}^l + i_t^l * \sigma\ (\hat{c}_t^l)$
24         $h_t^l = o_t^l * \sigma(c_t^l)$
25   **end**
26       $h'^l = h_t^l$
27   **end**
28   **Fully connected**:
29   Compute: $z^l \leftarrow W^l \sigma(h'^l) + b^l$
30     $h_{Bi-LSTM} = tanh(z^l)$
31   **Back propagation**:
32     $\nabla U^l\ T(x),\ \nabla W^l\ T(x)$ and $\nabla b^l\ T(x)$
33   **end**
34 **end**
35 **Hybrid layer**:
36 $NTL_{det} = \sigma(W[h_{2D-CNN}, h_{Bi-LSTM}] + b)$

V. EXPERIMENTS AND RESULTS

In this section, the experimental results of the proposed and the existing schemes are presented. The experiments are conducted on a realistic smart meters dataset, which is released by the State grid corporation of China (SGCC). The detailed description of the dataset is provided in Section V-A. Moreover, Python 3.0 and Google Colab are used for the training of deep learning models. All the deep learning models are developed through TensorFlow and Keras, which are open source libraries that build deep neural networks. The baseline models are fitted using scikit-learn library.

A. DATASET DESCRIPTION

The EC dataset is a publicly available realistic smart meters' dataset, which is released by SGCC. It comprises of daily EC of 42,372 consumers from 1 Jan 2014 to 31 Oct 2016. In the dataset, each row represents the complete electricity profile of a consumer and every column depicts daily EC at a specific date. The normal and abnormal users in the dataset are labeled as 0 and 1, respectively. The meta information about the dataset is given in Table 2.

TABLE 2. Information of SGCC dataset.

Total No. of consumers	42,372
Fraudulent consumers	3,615
Honest consumers	38,757
Time duration	2014-01-01 to 2016-10-31

B. PERFORMANCE EVALUATION

In the ETD scenario, the available EC data is imbalanced. Therefore, the selection of appropriate performance metrics is a necessary task for fair and better evaluations of the model. In the case of class imbalance, the accuracy metric is not suitable because it only focuses on the correct predictions. Moreover, both false positive (FP) and false negative (FN) are important in the case of ETD. Therefore, in this study, the selection of AUC metric is made to properly distinguish between honest and dishonest consumers. Moreover, FN is also important for power utilities because it increases the financial loss. Hence, the selection of MCC metric is made because it takes into account all the positive and negative classes. It tells about how well true positive (TP), FP, true negative (TN) and FN are separated. In particular, the range of MCC score is between 0 and 1. The model performs well if the value of MCC score is closer to 1. The interaction towards 1 shows that the classification model efficiently detects the positive and negative class samples. In addition, we consider precision, recall, PR-AUC and F1-score metrics for comprehensive analysis of the proposed scheme. Precision tells about the correct predictions of the model, which assist the electric utilities to save the extra onsite inspection cost. Similarly, recall provides the overall suspicious list of energy thieves, which also reduces the financial loss. Whereas, PR-AUC focuses on both precision and recall, and measures the ratio among them.

The mathematical formulation of the aforementioned performance metrics is given as follows [22].

$$Precision = \frac{TP}{TP + FP}, \tag{16}$$

$$Recall = \frac{TP}{TP + FN}, \tag{17}$$

$$F1 - score = 2 * \frac{Precision * Recall}{Precision + Recall}, \tag{18}$$

$$T = (TP + FP)(TP + FN),$$

$$P = (TN + FP)(TN + FN),$$

$$MCC = \frac{TP * TN - FP * FN}{\sqrt{T * P}}, \quad (19)$$

$$AUC = \frac{\sum i \in positiveclassRANK_i - \frac{P(1+P)}{2}}{P * N}. \quad (20)$$

where P and N represent positive and negative class samples, respectively. TP refers to the correctly identified positive class users, which are actually normal electricity users. Similarly, TN depicts the accurately identified abnormal class users. Whereas, FN and FP represent the misclassified normal and abnormal class users, respectively.

C. MEASURING EFFECTS OF IMBALANCE DISTRIBUTION ON PERFORMANCE RESULTS

Table 3 presents the analysis of the proposed methodology using different sampling techniques to analyze the significance of the balanced and the imbalanced data distributions. The performance results depict that the hybrid 2D-CNN and Bi-LSTM model obtains the highest performance on the Bi-WGAN's generated data distribution. The near miss and SMOTE based balanced data does not provide satisfactory performance results because these schemes randomly remove and synthesize duplicate data records, respectively, which raise information loss and overfitting issues. Moreover, Bi-WGAN utilizes an auxiliary encoder module to improve the stable learning and the convergence speed. That is why the Bi-WGAN generated samples have close resemblance with the real-world theft patterns, which enable the classification model to perform efficient ETD.

TABLE 3. Proposed model performance on imbalance distribution.

Performance metrics	Bi-WGAN	Imbalance dataset	SMOTE (Over-sampling)	Near Miss (Under-sampling)
Accuracy	0.95	0.92	0.89	0.58
Precision	0.97	0.83	0.58	0.40
Recall	0.92	0.53	0.64	0.93
F1-score	0.94	0.60	0.58	0.52
MCC score	0.93	0.62	0.62	0.49
AUC score	0.97	0.88	0.84	0.58
PR-AUC score	0.98	0.72	0.72	0.61

D. COMPARATIVE ANALYSIS WITH BENCHMARK MODELS

In this section, the proposed model is compared with the state-of-the-art benchmark models for efficient ETD. For fair comparison, the same data preprocessing techniques are opted for them. The description of the benchmark models is given below.

1) SUPPORT VECTOR MACHINE
The support vector machine (SVM) is the most popular ML classifier. Both classification and regression tasks are performed through SVM. In general, it is exploited for binary classification. However, it also performs multi classification using a kernel trick. In [2], SVM is exploited for final NTL detection. Therefore, we select SVM as a baseline classifier in this work.

2) RANDOM FOREST
The random forest (RF) classifier is an ensemble learning approach. It integrates several decision trees together that make a forest. It follows a bagging method. In the bagging method, the final outcome is decided by taking the average or majority voting of different weak learners. In [21], it is used to perform ETD.

3) LOGISTIC REGRESSION
Logistic regression (LR) is a simple and well known ML classifier. It is used for binary classification and follows the principle of neural networks. It contains a single layer of neural network and a sigmoid activation function on the output layer for binary classification. If the value on the output layer is closer to 1, then the electricity user is classified as an honest user and vice versa [21].

4) WIDE AND DEEP CNN
WD-CNN [1] is a hybrid deep learning approach. It is proposed to detect electricity thieves in power grids. It consists of two deep learning models, known as wide and deep components. The wide component contains a single fully connected layer of the neural network. It is used for extracting the abstract features from the 1D daily EC data. Meanwhile, the deep component captures the local features and periodicity from the 2D weekly consumption data.

5) LSTM AND MLP
For efficient ETD, a hybrid of LSTM and multi layer perception (MLP) is proposed in [14]. In the proposed model, the sequential time series data is passed to LSTM for capturing the temporal correlation from the EC profiles of consumers. Similarly, the non-sequential additional data is fed to the MLP model for better detection of energy thieves. Afterwards, the outputs of both models are combined into a unique feature vector. Then, final NTL detection is performed by applying the sigmoid activation function.

E. PERFORMANCE ANALYSIS AND DISCUSSION

This section presents the analysis of the experimental results. First of all, we discuss the analysis of data augmentation using Bi-WGAN. In Fig. 2(a), the loss curves of discriminator on both real and fake samples along with the loss of generator model are shown. The blue and the orange curves exhibit the discriminator loss on real and fake samples. The gradual decay in discriminator loss indicates that the discriminator model efficiently discriminates the real samples and the samples that are synthesized by the generator model. The reason is that the discriminator model is trained more than the generator model in Bi-WGAN. In particular, the weights of discriminator model are updated by utilizing the half batch of real samples and the half batch of fake samples at each round of the training process. On the other hand, the
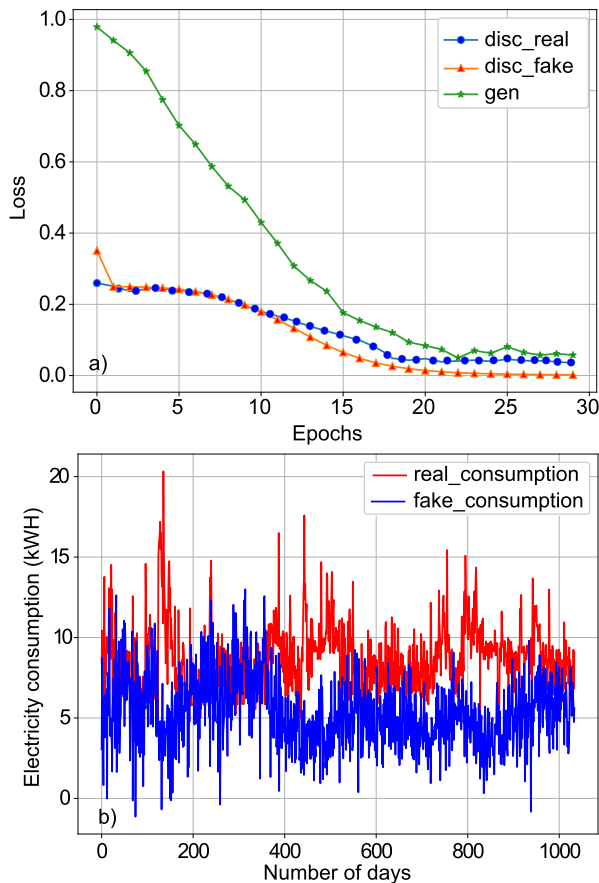
FIGURE 2. (a) Training loss of Bi-WGAN generator and discriminator. (b) Real and Bi-WGAN generated EC patterns.

loss of generator model during the training phase is shown by the green curve. The addition of an external encoder module in Bi-WGAN strengthens its power towards generating the most plausible EC samples. Due to this addition, it efficiently captures the complex probability distribution curve from EC profiles. That is why the loss of generator model is gradually reduced after few iterations of training. Consequently, the generated patterns have close resemblance with the real-world theft patterns. More specifically, in Bi-WGAN, the Wasserstein loss function is used instead of Jensen divergence loss function.

The Wasserstein loss function measures the score of realness or fakeness of given samples while the regular GAN loss function predicts the probability of generated samples as real or fake. Hence, the addition of Wasserstein loss function, integrating auxiliary encoder module in generator network and the process of training discriminator model boost the performance of Bi-WGAN towards generating prominent electricity theft samples. Fig. 2(b) illustrates the performance of Bi-WGAN during the generation of fake electricity theft patterns. The red curve shows the real theft pattern of an electricity user. Similarly, the blue curve demonstrates Bi-WGAN generated theft patterns. From the figure, it is seen that Bi-WGAN efficiently learns the objective laws from

the real electricity theft profiles and generates the real-world synthetic theft patterns with high precision. Moreover, it is proved that the integration of an external encoder module in Bi-WGAN helps in simulating realistic real-world theft patterns.

Table 4 describes the performance results of the proposed model and the benchmark models on 70% training data and 30% testing data. From the results, it is seen that the proposed model shows superior performance on all the existing models. In the proposed hybrid model, the concurrent usage of 2D-CNN and Bi-LSTM boosts its performance towards achieving the best performance results. It obtains 0.97 AUC-ROC score, which is the best achievement for efficient ETD. It also beats the existing schemes, such as SVM, LR, RF, WD-CNN and LSTM-MLP in terms of AUC-ROC. Higher AUC-ROC means that a classification model efficiently distinguishes the two classes. Moreover, the proposed model achieves PR-AUC of 0.98. This score states that how well the model correctly identifies the electricity thieves. Our model obtains the highest PR-AUC because of the powerful capabilities of Bi-LSTM and 2D-CNN. Whereas, SVM obtains the lowest AUC-ROC score of 0.77 because it does not perform well on high dimensional data. It draws $n - 1$ hyperplanes, where n denotes the number of features. Therefore, the selection of an optimal hyperplane in the case of highly dynamic data is very difficult for it. That is why SVM obtains the lowest AUC-ROC score as compared to other baseline models. In contrast, RF achieves a suitable AUC-ROC of 0.94 because it follows the ensemble learning procedure. In RF, the outcomes of several weak learners are combined for the final prediction using the majority voting phenomenon. Moreover, it uses a random subset of data samples and features for training each weak learner. This process improves its performance results. Therefore, it performs better than the conventional ML techniques. It obtains AUC-ROC and PR-AUC of 0.94 and 0.96, respectively, which is higher than SVM and LR predictions. LR does not achieve satisfactory results because it has one single hidden layer. WD-CNN and LSTM-MLP models achieve 0.92 and 0.95 AUC-ROC scores, respectively. LSTM-MLP obtains better results than WD-CNN because it uses the strong memorization and feature extraction abilities of LSTM and MLP, respectively.

Fig. 3(a) shows the loss of the proposed hybrid model during the training phase. The orange curve depicts the loss on validation data and the blue curve demonstrates the loss on training data. It is clearly seen that the hybrid model performs well on both training and validation data. We analyze that the loss value decreases when the epoch value increases. However, after running 10 iterations of the training phase, the loss value on training data starts decreasing gradually; meanwhile, the loss value on validation data becomes smooth. This implies that the model has good generalization ability before the 10th iteration. Moreover, a threshold must exist for epoch value to optimize the training process. For instance, in our case, the best performance of training is achieved when the epoch value reaches 10.

TABLE 4. Comparison analysis of the proposed model with benchmark schemes.

Model	Accuracy	Precision	Recall	F1-score	AUC-ROC	MCC	PR-AUC
SVM	0.77	0.80	0.77	0.77	0.77	0.89	0.83
LR	0.83	0.85	0.83	0.83	0.83	0.68	0.87
RF	0.94	0.95	0.95	0.94	0.94	0.89	0.96
WD-CNN	0.84	0.91	0.87	0.84	0.92	0.93	0.91
LSTM-MLP	0.95	0.95	0.92	0.93	0.95	0.90	0.96
Proposed hybrid model	**0.95**	**0.97**	**0.92**	**0.94**	**0.97**	**0.93**	**0.98**
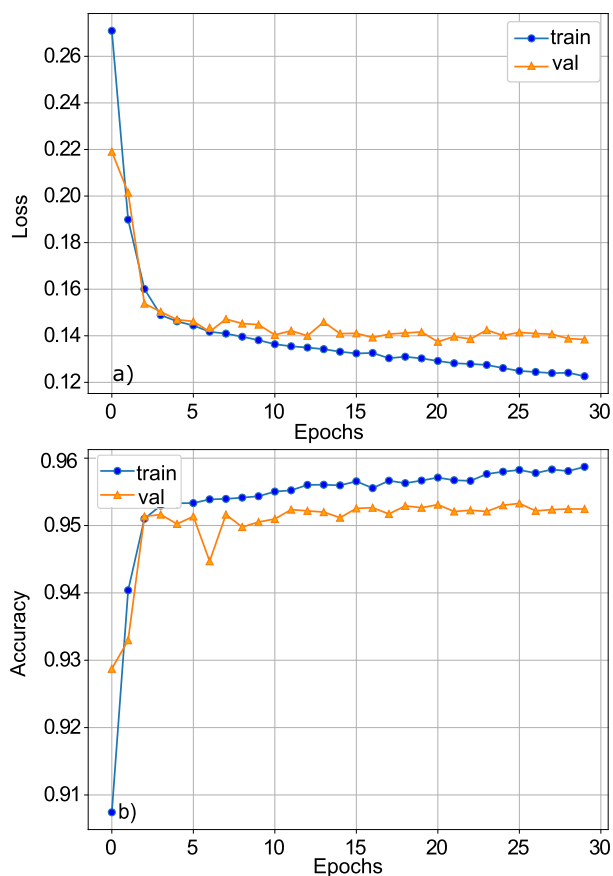
FIGURE 3. (a) Training and validation losses of hybrid 2D-CNN and Bi-LSTM. (b) Training and validation accuracy of hybrid 2D-CNN and Bi-LSTM.

FIGURE 4. (a) AUC-ROC score of the proposed hybrid model. (b) MCC score of the proposed hybrid model.

Fig. 3(b) illustrates the accuracy of the hybrid model during the training phase. It is seen that the hybrid model performs well on both training and validation datasets because of the effective gated configuration and the integration of both forward and backward passes in Bi-LSTM model. In particular, the powerful feature extraction capabilities of 2D-CNN model also improve the classification results. The performance of the hybrid model on validation data is more stable than training data. This implies that the proposed hybrid model efficiently detects electricity thieves and honest consumers from the EC data due to the hybrid functionalities of 2D-CNN and Bi-LSTM. Its training accuracy gradually increases when the epoch value increases. The optimal
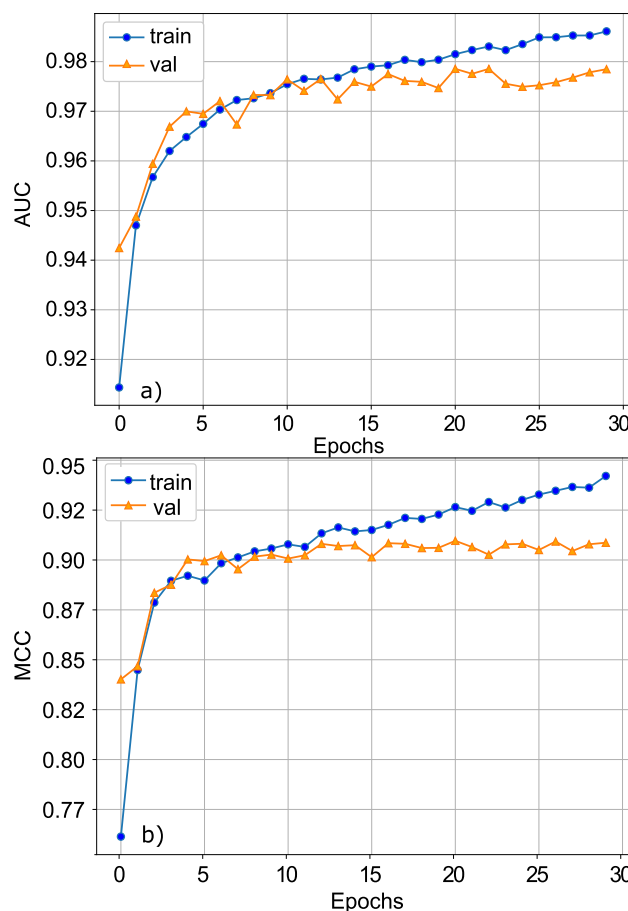
performance is obtained when the number of epoch hits 10. Furthermore, a large fluctuation is seen in the accuracy value at epoch 6. It is because of a noisy batch of samples during the model's training. However, the model stabilizes its learning after the 6th epoch. Similarly, Fig. 4(a) depicts the AUC-ROC score of the hybrid model during the training and validation phases. It is seen that the model obtains an AUC-ROC score of 0.97, which is an excellent achievement. This achievement implies that the hybrid model effectively discriminates normal and theft classes due to its best learning mechanism. Fig. 4(b) exhibits the MCC score. MCC metric is opted because it equally incorporates all findings of confusion matrix. It finds the correlation between TP, FP,
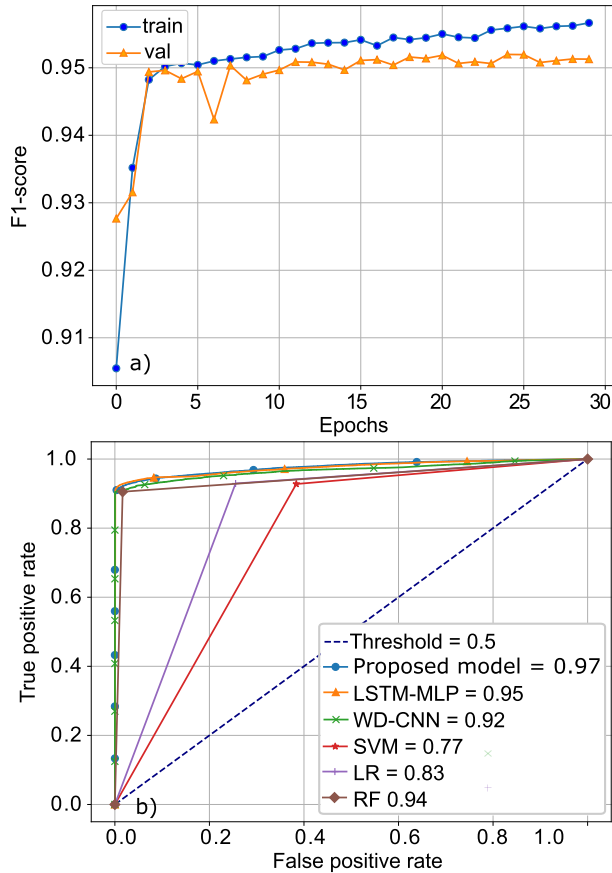
FIGURE 6. PR-AUC based benchmark comparison.
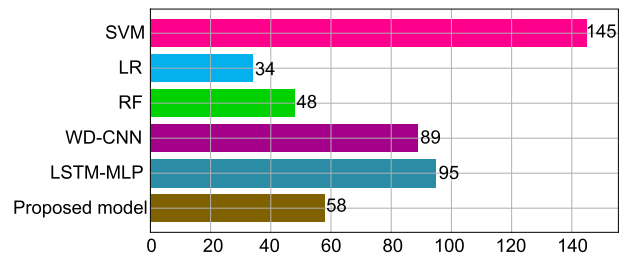
FIGURE 7. Training time (sec) of the proposed hybrid model and baseline models.

FIGURE 5. (a) F1-score of hybrid 2D-CNN and Bi-LSTM. (b) AUC-ROC based benchmark comparison.

TN and FN. FN and TN are also important for electric utilities because they help utilities to restore maximum monetary cost. From the figure, it is observed that MCC score is increasing at each iteration, which shows that the proposed model perfectly deals with FN and TN. It obtains MCC score of 0.93, which is satisfactory in case of detecting electricity thieves. Consequently, it will be beneficial for power utilities to recover maximum revenue by identifying the energy thieves. The F1-score is depicted in Fig. 5(a) on both validation and training datasets. It is determined by computing the harmonic means of precision and recall values. During training, an abrupt change is seen in the 6th epoch. This is because of noise in the training batch. HBesides, the proposed model obtains F1-score of 0.94, which depicts its superior performance on validation dataset. The higher F1-score helps the electric utilities to accurately identify and locate the energy thieves. It also becomes beneficial to increase the detection rate (DR) and reduce the high FPR.

The AUC-ROC scores of the proposed scheme and the baseline models are illustrated in Fig. 5(b). The proposed scheme obtains an AUC-ROC score of 0.97, which is satisfactory as compared to the existing classifiers, such as SVM, LR, RF, WD-CNN and LSTM-MLP. This achievement implies that the proposed scheme efficiently distinguishes the
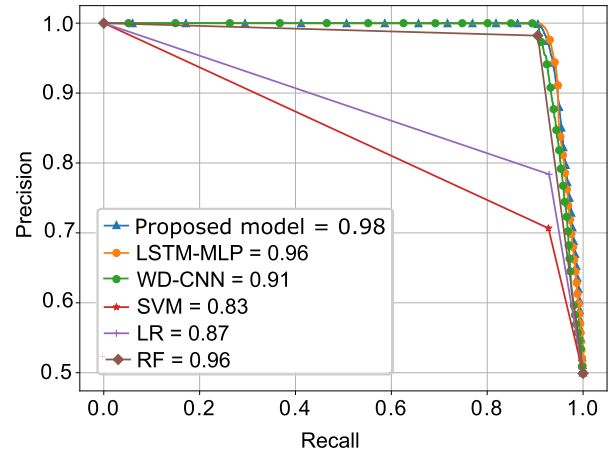
two classes due to its hybrid feature learning mechanism. Moreover, the powerful gated configuration along with the integration of both forward and reverse feature learning paths in Bi-LSTM increases its performance towards capturing the non-malicious changes. Consequently, the high FPR is reduced to a minimum extent. The PR-AUC scores of the proposed and baseline models are shown in Fig. 6. It equally focuses on both precision and recall. In the case of detecting electricity frauds, these both factors are dominant for electric utilities. A high PR-AUC score proves the efficacy of models. The proposed scheme achieves PR-AUC of 0.98, which is higher than all baseline models. This implies that the proposed scheme is proven to be beneficial for power industries to accurately identify the energy frauds and help them to recover maximum income. Moreover, Fig. 7 illustrates the training time of the proposed and baseline models. It is seen that the proposed model takes less time for training as compared to other deep models. The reason is that the proposed model efficiently discards the redundant and noisy features from the high dimensional EC data and reduces the computational overhead to a greater extent. The model obtains the highest performance results as compared to the baseline models. Moreover, LR takes least time for training because it contains one layer of neural networks. However, it does not obtain satisfactory results. The SVM model takes the highest training time because it first draws multiple hyperplanes and then selects an optimal hyperplane from them to perform the

TABLE 5. Mapping between identified limitations, proposed solutions and validation results.

Identified limitations	Proposed solutions	Validations
L.1 Noisy high dimensional EC data with inappropriate feature engineering adversely affect the detection accuracy of classification models and increase the computational time as well [2], [4]	S.1 The extraction of the most potential features from the highly dynamic EC data is performed by proposing a hybrid approach, which is the combination of 2D-CNN and Bi-LSTM. Moreover, the proposed model reduces the computational overheard to a greater extent	V.1 The suitable key performance indicators, such as MCC, AUC-ROC, PR-AUC and execution time are opted to validate the proposed solution and the results are exhibited in Figs. 4, 5, 6 and 7, respectively
L.2 Without using proper feature extraction, the model memorizes noise and redundant features, which raise poor generalization issue [11]	S.2 2D-CNN extracts the hidden latent representations from the EC data. Meanwhile, Bi-LSTM learns the most significant temporal correlation from EC patterns to enhance the generalization ability	V.2 The solution is validated through final classification results, as given in Table 4
L.3 High FPR occurs because of several non-malicious factors and abrupt changes in EC patterns [1], [21]	S.3 Bi-LSTM intelligently captures the temporal correlation from EC data and easily detects the non-malicious changes using the memory states. Consequently, it diminishes the FPR to a minimal level	V.3 The solution is validated through plotting FPR and TPR using AUC-ROC, as shown in Fig. 5(b)
L.4 The severe lack of fraudulent electricity samples leads to a class imbalance problem, which negatively affects the classifier's performance [14], [15]	S.4 Fraudulent electricity samples are generated by utilizing the Bi-WGAN model. Its simulated samples are closely associated with the real-world theft patterns	V.4 Bi-WGAN efficiently generates fake EC data, as shown in Fig. 2(b)
L.5 In [4], [22], SMOTE generates synthetic samples of minority theft class to balance the data, however, it raises overfitting issue due to the duplication of records	S.5 Bi-WGAN prevents the overfitting issue by intelligently generating the synthetic theft samples	V.5 The performance of Bi-WGAN during the generation of theft samples is shown in Fig. 2(b)
L.6 ICS issue affects the convergence speed of the deep neural networks because it updates input distribution on each hidden layer [1], [14]	S.6 Adding batch normalization layers and regularization penalties in both 2D-CNN and Bi-LSTM modules helps to resolve ICS issue	V.6 The performance and convergence speed of the proposed solution is enhanced to a greater extent, as shown in Figs. 3, 4 and 5
L.7 Improper selection of performance metrics for imbalance classification does not provide real-world assessment [19]	S.7 Appropriate performance metrics are adopted for better evaluation	V.7 Figs. 3-6 depict the results using appropriate performance metrics

classification task. This process increases the computational complexity to a greater extent.

F. MAPPING BETWEEN LIMITATIONS, SOLUTIONS AND THEIR VALIDATIONS

The mapping of identified limitations with their proposed solutions and validations is given in Table 5. L1 is about the noisy high dimensionality issue, which is solved by proposing a hybrid of 2D-CNN and Bi-LSTM model and their results are validated through suitable key performance indicators, as shown in Figs. 4, 5 and 6. The poor generalization issue is highlighted in L2. It occurs because of noisy and duplicate features in the EC data. The issue is solved through the proposed hybrid model. The proposed model captures only potential features and discards the irrelevant features. Moreover, it efficiently extracts the temporal correlated features from the EC data. Table 5 validates this solution. In L3, the problem of high FPR is discussed. This problem occurs due to several non-malicious factors and abrupt changes in EC load profiles. It may happen because of false data injection by the intelligent attacker. Hence, the problem of high FPR is resolved by utilizing the Bi-LSTM model. It maintains the context of the long-term temporal correlation in memory states. In this manner, the effects of various non-malicious factors are easily identified by the model. The solution is validated through AUC-ROC that is shown in Fig. 5(b). The class imbalance issue is highlighted in L4. Bi-WGAN is employed to synthesize the fraudulent electricity samples. The solution

is validated through the generated sample of Bi-WGAN, as shown in Fig. 2(b). L5 is about the overfitting issue, which occurs when using SMOTE due to the duplication of EC records. Bi-WGAN simulates plausible theft samples because of their powerful feature learning capabilities. The solution is validated in Fig 2(b) where the learning process of Bi-WGAN is presented. In L6, the ICS issue is discussed that occurs in neural network while transferring the input distribution from one hidden layer to the others. To solve ICS, we add batch normalization layers and regularization penalties in the neural network. The solution is validated by analyzing the convergence speed of the proposed model, which is shown in Figs. 3, 4 and 5. In L7, it is mentioned that the improper selection of performance metrics in ETD does not provide fair assessment. Therefore, the selection of appropriate metrics is made for the fair evaluation of the proposed model. The solution is validated by suitable performance indicators, which are shown in Figs. 3-6.

VI. CONCLUSION AND FUTURE WORK

In this article, we have proposed a hybrid deep learning model for the detection of ET in power grids. The proposed model combines 2D-CNN and Bi-LSTM models. The noisy high dimensionality issue is tackled through the hybrid capabilities of both Bi-LSTM and 2D-CNN modules. Furthermore, the challenge of the severe lack of fraudulent samples is solved by generating realistic theft samples using Bi-WGAN. All the experiments are conducted on the realistic smart meters

dataset, which is released by the SGCC. The comparison with other baseline models proves that the proposed scheme surpasses the performance of the state-of-the-art models, such as LR, SVM, RF, WD-CNN and LSTM-MLP. Moreover, the simulation results illustrated that the proposed model achieves higher AUC-ROC, PR-AUC, F1-score and MCC score as compared to the baseline models. Our model obtains AUC-ROC and PR-AUC of 0.97 and 0.98, respectively that make it more suitable for real-world scenarios. Furthermore, the proposed model can be used in different industrial applications to detect anomalies and frauds. In the future, we will consider the high sampling EC data to enhance the performance of the proposed hybrid model.

DATASET AVAILABILITY
Dataset used in this study is publically available at "https://github.com/henryRDlab/Electricity TheftDetection/".

REFERENCES
[1] 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.

[2] 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.

[3] Q. Chen, K. Zheng, C. Kang, and F. Huangfu, "Detection methods of abnormal electricity consumption behaviors: Review and prospect," *Autom. Electr. Power Syst.*, vol. 42, no. 17, pp. 189–199, 2018.

[4] 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.

[5] R. Razavi, A. Gharipour, M. Fleury, and I. J. Akpan, "A practical feature-engineering framework for electricity theft detection in smart grids," *Appl. Energy*, vol. 238, pp. 481–494, Mar. 2019.

[6] A. S. Iwashita, D. Rodrigues, D. S. Gastaldello, A. N. de Souza, and J. P. Papa, "An incremental optimum-path forest classifier and its application to non-technical losses identification," *Comput. Electr. Eng.*, vol. 95, Oct. 2021, Art. no. 107389.

[7] S.-V. Oprea and A. Bâra, "Machine learning classification algorithms and anomaly detection in conventional meters and Tunisian electricity consumption large datasets," *Comput. Electr. Eng.*, vol. 94, Sep. 2021, Art. no. 107329.

[8] C.-H. Lo and N. Ansari, "CONSUMER: A novel hybrid intrusion detection system for distribution networks in smart grid," *IEEE Trans. Emerg. Topics Comput.*, vol. 1, no. 1, pp. 33–44, Jun. 2013.

[9] 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.

[10] N. Javaid, H. Gul, S. Baig, F. Shehzad, C. Xia, L. Guan, and T. Sultana, "Using GANCNN and ERNET for detection of non technical losses to secure smart grids," *IEEE Access*, vol. 9, pp. 98679–98700, 2021.

[11] M. M. Buzau, J. Tejedor-Aguilera, P. Cruz-Romero, and A. Gomez-Exposito, "Detection of non-technical losses using smart meter data and supervised learning," *IEEE Trans. Smart Grid*, vol. 10, no. 3, pp. 2661–2670, May 2019.

[12] 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.

[13] S. I. Popoola, B. Adebisi, M. Hammoudeh, H. Gacanin, and G. Gui, "Stacked recurrent neural network for BotNet detection in smart Homes," *Comput. Electr. Eng.*, vol. 92, Jun. 2021, Art. no. 107039.

[14] M.-M. Buzau, J. Tejedor-Aguilera, P. Cruz-Romero, and A. Gomez-Exposito, "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.

[15] D. Yao, M. Wen, X. Liang, Z. Fu, K. Zhang, and B. Yang, "Energy theft detection with energy privacy preservation in the smart grid," *IEEE Internet Things J.*, vol. 6, no. 5, pp. 7659–7669, Oct. 2019.

[16] M. Asif, B. Kabir, A. Ullah, S. Munawar, and N. Javaid, "Towards energy efficient smart grids: Data augmentation through BiWGAN, feature extraction and classification using hybrid 2DCNN and BiLSTM," in *Proc. Int. Conf. Innov. Mobile Internet Services Ubiquitous Comput.*, Cham, Switzerland: Springer, 2021, pp. 108–119.

[17] 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.

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

[19] K. Zheng, Q. Chen, Y. Wang, C. Kang, and Q. Xia, "A novel combined data-driven approach for electricity theft detection," *IEEE Trans. Ind. Informat.*, vol. 15, no. 3, pp. 1809–1819, Mar. 2019.

[20] A. Takiddin, M. Ismail, U. Zafar, and E. Serpedin, "Robust electricity theft detection against data poisoning attacks in smart grids," *IEEE Trans. Smart Grid*, vol. 12, no. 3, pp. 2675–2684, May 2021.

[21] S. Li, Y. Han, X. Yao, S. Yingchen, J. Wang, and Q. Zhao, "Electricity theft detection in power grids with deep learning and random forests," *J. Electr. Comput. Eng.*, vol. 2019, pp. 1–12, Oct. 2019.

[22] 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.

[23] R. R. Bhat, R. D. Trevizan, R. Sengupta, X. Li, and A. Bretas, "Identifying nontechnical power loss via spatial and temporal deep learning," in *Proc. 15th IEEE Int. Conf. Mach. Learn. Appl. (ICMLA)*, Dec. 2016, pp. 272–279.

[24] B. Kocaman and V. Tümen, "Detection of electricity theft using data processing and LSTM method in distribution systems," *Sādhanā*, vol. 45, no. 1, pp. 1–10, Dec. 2020.

[25] G. Fenza, M. Gallo, and V. Loia, "Drift-aware methodology for anomaly detection in smart grid," *IEEE Access*, vol. 7, pp. 9645–9657, 2019.

[26] X. Lu, Y. Zhou, Z. Wang, Y. Yi, L. Feng, and F. Wang, "Knowledge embedded semi-supervised deep learning for detecting non-technical losses in the smart grid," *Energies*, vol. 12, no. 18, p. 3452, Sep. 2019.

[27] C. C. O. Ramos, D. Rodrigues, A. N. de Souza, and J. P. Papa, "On the study of commercial losses in Brazil: A binary black hole algorithm for theft characterization," *IEEE Trans. Smart Grid*, vol. 9, no. 2, pp. 676–683, Mar. 2018.

[28] T. Hu, Q. Guo, H. Sun, T.-E. Huang, and J. Lan, "Nontechnical losses detection through coordinated BiWGAN and SVDD," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 5, pp. 1866–1880, May 2021.

[29] N. F. Avila, 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.

[30] J. I. Guerrero, I. Monedero, F. Biscarri, J. Biscarri, R. Millan, and C. Leon, "Non-technical losses reduction by improving the inspections accuracy in a power utility," *IEEE Trans. Power Syst.*, vol. 33, no. 2, pp. 1209–1218, Mar. 2018.

[31] M. S. Saeed, M. W. Mustafa, U. U. Sheikh, T. A. Jumani, and N. H. Mirjat, "Ensemble bagged tree based classification for reducing non-technical losses in multan electric power company of Pakistan," *Electronics*, vol. 8, no. 8, p. 860, Aug. 2019.

[32] 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, Aug. 2020.

[33] 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.

[34] X. Wang, I. Yang, and S.-H. Ahn, ''Sample efficient home power anomaly detection in real time using semi-supervised learning,'' *IEEE Access*, vol. 7, pp. 139712–139725, 2019.

[35] 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.

[36] 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.

[37] V. Chandola, A. Banerjee, and V. Kumar, ''Anomaly detection: A survey,'' *ACM Comput. Surv.*, vol. 41, no. 3, pp. 1–58, 2009.

[38] U. Mutlu and E. Alpaydın, ''Training bidirectional generative adversarial networks with hints,'' *Pattern Recognit.*, vol. 103, Jul. 2020, Art. no. 107320.

[39] M. Arjovsky, S. Chintala, and L. Bottou, ''Wasserstein generative adversarial networks,'' in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 214–223.

[40] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville, ''Improved training of Wasserstein GANs,'' 2017, *arXiv:1704.00028*.

[41] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, ''Generative adversarial networks,'' 2014, *arXiv:1406.2661*.

[42] M. Arjovsky, S. Chintala, and L. Bottou, ''Wasserstein generative adversarial networks,'' in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 214–223.

[43] J. Yu, X. Zhang, L. Xu, J. Dong, and L. Zhangzhong, ''A hybrid CNN-GRU model for predicting soil moisture in maize root zone,'' *Agricult. Water Manage.*, vol. 245, Feb. 2021, Art. no. 106649.

[44] J. Zhao, X. Mao, and L. Chen, ''Speech emotion recognition using deep 1D & 2D CNN LSTM networks,'' *Biomed. Signal Process. Control*, vol. 47, pp. 312–323, Jan. 2019.

[45] Z. Cui, R. Ke, Z. Pu, and Y. Wang, ''Stacked bidirectional and unidirectional LSTM recurrent neural network for forecasting network-wide traffic state with missing values,'' *Transp. Res. C, Emerg. Technol.*, vol. 118, Sep. 2020, Art. no. 102674.

[46] N. Javaid, A. Naz, R. Khalid, A. Almogren, M. Shafiq, and A. Khalid, ''ELS-Net: A new approach to forecast decomposed intrinsic mode functions of electricity load,'' *IEEE Access*, vol. 8, pp. 198935–198949, 2020.

MUHAMMAD ASIF (Graduate Student Member, IEEE) received the B.S. degree in information technology from the University of Gujrat, Gujrat, Pakistan, in 2017. He is currently pursuing the M.S. degree in computer science with the Communications Over Sensors (ComSens) Research Laboratory, COMSATS University Islamabad, Islamabad Campus, under the supervision of Prof. Nadeem Javaid. His research interests include electricity load forecasting, financial market forecasting, and smart grids.

OROOJ NAZEER received the M.S. degree in computer science from Abasyn University, Islamabad, under the supervision of Prof. Nadeem Javaid.
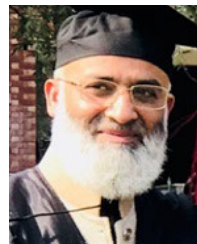
NADEEM JAVAID (Senior Member, IEEE) received the bachelor's degree in computer science from Gomal University, Dera Ismail Khan, Pakistan, in 1995, the master's degree in electronics from Quaid-i-Azam University, Islamabad, Pakistan, in 1999, and the Ph.D. degree from the University of Paris-Est, France, 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 at 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 the Editor of *Sustainable Cities and Society*.

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 of computer science with Taif University, Saudi Arabia. Her research area includes formal methods, AI, data science, and so on. She was awarded as a Senior Fellow of the Higher Education Academy (FHEA) in March 2020.

MYRIAM HADJOUNI received the Ph.D. degree (Hons.) in computer science from Paris XI (actual new name Paris Saclay) University, France, and Manouba University, Tunisia, in 2012, and the M.Sc. degree (Hons.) from the Higher Institute of Management of Tunis, University of Tunis, Tunisia, in 2005. 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, Kingdom of Saudi Arabia. Her research includes but not restricted to information retrieval, artificial intelligence, data science, data analytic, big data, and image retrieval.

• • •