IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received January 19, 2021, accepted January 31, 2021, date of publication February 2, 2021, date of current version February 12, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3056566

Towards Sustainable Energy Efficiency With Intelligent Electricity Theft Detection in Smart Grids Emphasising Enhanced Neural Networks

ABDULAZIZ ALDEGHEISHEM[1], MUBBASHRA ANWAR[2], NADEEM JAVAID[2], (Senior Member, IEEE), NABIL ALRAJEH[3], MUHAMMAD SHAFIQ[4], AND HASAN AHMED[5]

[1]Urban Planning Department, College of Architecture and Planning, King Saud University (KSU), Riyadh 11574, Saudi Arabia
[2]Department of Computer Science, COMSATS University Islamabad, Islamabad 44000, Pakistan
[3]Biomedical Technology Department, College of Applied Medical Sciences, King Saud University (KSU), Riyadh 11633, Saudi Arabia
[4]Department of Information and Communication Engineering, Yeungnam University, Gyeongsan 38541, South Korea
[5]School of Computing and Communication, Lancaster University, Lancaster LA1 4WA, U.K.

Corresponding authors: Nadeem Javaid (nadeemjavaidqau@gmail.com) and Muhammad Shafiq (shafiq.pu@gmail.com)

This work was supported by the King Saud University, Riyadh, Saudi Arabia, under Project RSP-2020/295.

ABSTRACT In smart grids, electricity theft is the most significant challenge. It cannot be identified easily since existing methods are dependent on specific devices. Also, the methods lack in extracting meaningful information from high-dimensional electricity consumption data and increase the false positive rate that limit their performance. Moreover, imbalanced data is a hurdle in accurate electricity theft detection (ETD) using data driven methods. To address this problem, sampling techniques are used in the literature. However, the traditional sampling techniques generate insufficient and unrealistic data that degrade the ETD rate. In this work, two novel ETD models are developed. A hybrid sampling approach, i.e., synthetic minority oversampling technique with edited nearest neighbor, is introduced in the first model. Furthermore, AlexNet is used for dimensionality reduction and extracting useful information from electricity consumption data. Finally, a light gradient boosting model is used for classification purpose. In the second model, conditional wasserstein generative adversarial network with gradient penalty is used to capture the real distribution of the electricity consumption data. It is constructed by adding auxiliary provisional information to generate more realistic data for the minority class. Moreover, GoogLeNet architecture is employed to reduce the dataset's dimensionality. Finally, adaptive boosting is used for classification of honest and suspicious consumers. Both models are trained and tested using real power consumption data provided by state grid corporation of China. The proposed models' performance is evaluated using different performance metrics like precision, recall, accuracy, F1-score, etc. The simulation results prove that the proposed models outperform the existing techniques, such as support vector machine, extreme gradient boosting, convolution neural network, etc., in terms of efficient ETD.

INDEX TERMS Electricity theft detection, generative adversarial network, GoogLeNet, imbalanced data, Urban planning, SMOTEENN.

NOMENCLATURE

AdaBoost	Adaptive boosting
ADASYN	Adaptive synthetic
AMI	Advanced metering infrastructure
AUC	Area under the curve
ANN	Artificial neural network
AI	Artificial intelligence
ARIMA	Auto regressive integrated moving average
BBHA	Binary black hole algorithm
BGRU	Bidirectional gated recurrent unit
CatBoost	Categorical boosting

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

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

CDCGAN	Conditional deep convolutional generative adversarial network
CNN	Convolutional neural network
CGAN	Conditional generative adversarial network
CM	Confusion matrix
CWGAN-GP	Conditional wasserstein generative adversarial network gradient penalty
DT	Decision tree
DL	Deep learning
DERs	Distributed energy resources
DR	Detection rate
DWPT	Discrete wavelet packet transform
ENN	Edited nearest neighbor
ETD	Electricity theft detection
FMM	Finite mixture model
FPR	False positive rate
FP	False positive
FN	False negative
GAN	Generative adversarial network
GAN-NETBoost	Generative adversarial network GoogLeNet adaptive boosting
GBTD	Gradient boosting theft detector
GP	Gradient penalty
ILSVRC	ImageNet large-scale visual recognition challenge
IoT	Internet of things
KLD	Kullback-Leibler divergence
LGB	Light gradient boosting
LC	Lipschitz constraint
LSTM	Long short-term memory
ML	Machine learning
MODWPT	Maximal overlap discrete wavelet packet transform
MCC	Matthews correlation coefficient
MLP	Multilayer perceptron
NM	Near miss
NNs	Neural networks
NTL	Non-technical loss
PCA	Principal component analysis
PR	Precision recall
ROC	Receiver operating characteristic
ROS	Random oversampling
RF	Random forest
ReLU	Rectified linear unit
RUS	Random undersampling
RUSBoost	Random undersampling boosting
SG	Smart grid
SMs	Smart meters
SMOTE	Synthetic minority oversampling technique
SMOTEENN	Synthetic minority oversampling technique and edited nearest neighbor
SMOTE-Link	Synthetic minority oversampling technique and tomek-link
SGCC	State grid corporation of China
SVM	Support vector machine
SEAI	Sustainable energy authority of Ireland
SETS	Smart energy theft system
SALM	SMOTEENN-AlexNet-LGB model
TL	Technical loss
TP	True positive
TN	True negative
TPR	True positive rate
WD-CNN	Wide and deep convolutional neural network
WGAN	Wasserstein generative adversarial network
WGAN-GP	Wasserstein generative adversarial network gradient penalty
XGBoost	Extreme gradient boosting
y_m	Missing value
x	Real electricity consumption data
Z	A randomly selected input from
σ	Standard deviation
μ	Mean
s	Sample
k	Value of K used in KNN
r	Radius of circle in KNN
w^m	Identical weight
c_m	Weak Classifier
ϵ	Error rate
b	Bias
w	Weight

I. INTRODUCTION

Electricity has become inevitable for human life as almost all daily activities are dependent on it, such as communication, transportation, domestic appliances, heating and cooling systems, etc. It is considered vital for providing convenience, comfort, and consolation to life [1]. With the increase in world population and the usage of smart appliances, the electricity consumption is increasing rapidly. So, there is a need to optimize electric power usage, improve energy reliability, increase efficiency and reduce power losses for both users and utilities [2]. To achieve the aforementioned objectives, existing grids are transformed into the smart grids (SGs). One of the fundamental advancements in the power grid is the inclusion of Advanced Metering Infrastructure (AMI) that enables bi-directional communication between customers and utilities. In reality, smart meters (SMs) are considered a vital part of the AMI that encourage productive and effective data trading in power utility systems [3]–[5]. Almost all electricity companies face electrical power losses, which lead to huge economic losses [6]–[8]. The difference between the total electric energy generated at the generation side and the energy distributed to consumers is referred as electrical power losses. Technical losses (TLs) and commercial loss are two types of electrical power losses [9]–[11]. Commercial loss is also known as non-technical losses (NTLs) [9]. In Figure 1, the losses in electrical power systems are illustrated.

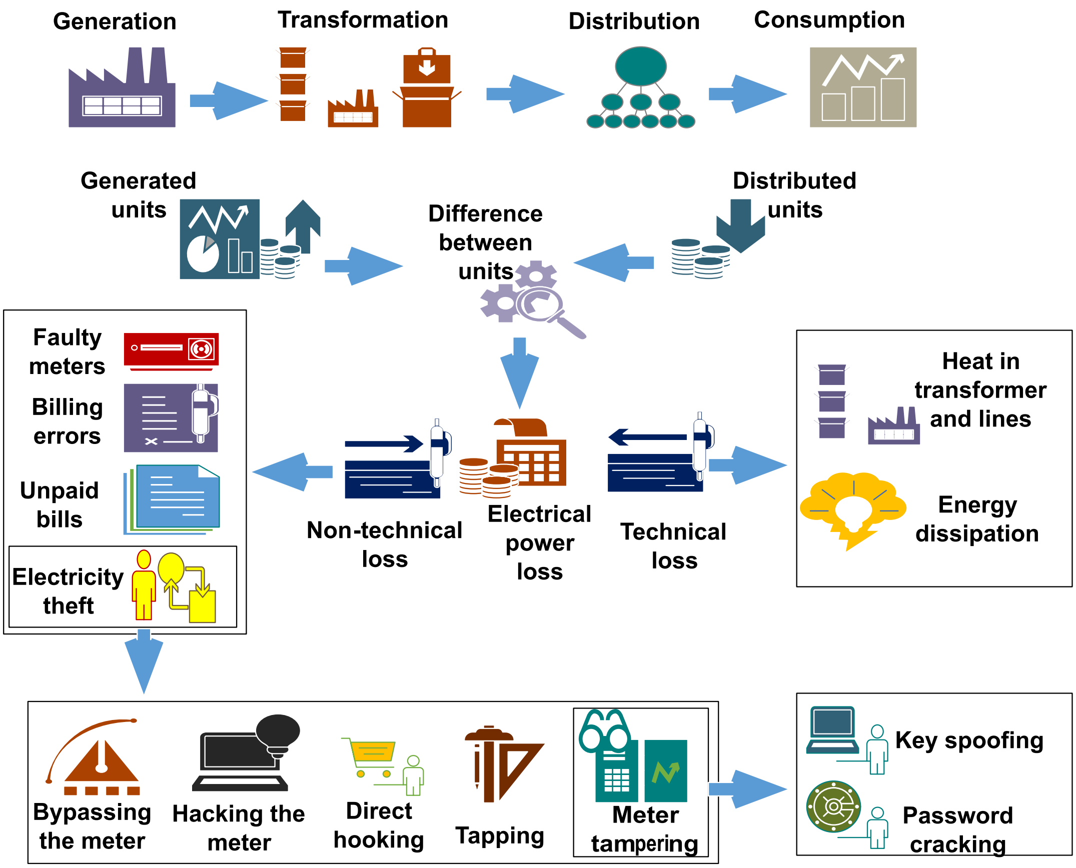


FIGURE 1. Illustration of power system and electrical power losses.

TLs occur due to energy dissipation in distribution and transmission lines, transformers, and other electrical equipments [12], [13]. The losses depend on network characteristics, planning, and operations [6]. On the other hand, NTLs are calculated as the consumed electrical power that is not billed. The main causes of these losses are faulty meters, unpaid and erroneous bills, and electricity theft [12], [14]–[17]. For electricity theft, key spoofing and password cracking are the conventional techniques, which are used to tamper the meters [12], [13]. The magnets are also used to manipulate the meter readings. Electricity theft leads to the increase in electricity demand, significant and massive revenue loss, heavy load, and lack of safety for both users and utilities [10], [18], [19]. Electricity theft also affects the honest consumers. The unpaid electricity bills are splitted and their burden is put on the honest consumers. So, they pay extra bills [9]. According to a recent report of the Northeast Group, utility companies lose $96 billion annually around the world due to NTLs [20], [21]. Moreover, electrical power theft has become a critical issue for energy utilities globally. A considerable amount of research has been done on electricity theft detection (ETD). Traditional techniques for identifying electricity theft are based on human efforts. Inspections are performed by technical staff who compare the anomalous meter's electricity consumption pattern with the normal ones [10]. However, it is difficult and expensive to conduct multiple on-site inspections for ETD [22]. Such approaches are labor-intensive, time-consuming, costly, and ineffective [10], [14], [22]. Therefore, there is a need for a smart and expert system that can accurately predict electricity fraudsters.

Recently, energy specialists adopted different approaches to identify electricity theft in the power systems. Some studies used the data obtained from the SGs to identify electricity theft [10], [16]. In SGs, bi-directional flow of information and electricity establishes a connection between utilities and customers [23], [24]. Artificial intelligence (AI) [25], machine learning (ML) [26], deep learning (DL) [4], hybrid models [10] and game theory based approaches [27] play major roles in ETD. However, most of these approaches have some limitations. The ETD in AMI is a special type of anomaly detection. It involves highly imbalanced data that influences the theft detection rate (DR) of the models. Here, the imbalanced data means that the number of samples are not distributed equally and honest customers are more than dishonest customers. When ML algorithms are trained to detect the electricity theft in the power systems, most of them ignore the data distribution between both classes and become biased towards the majority class. As a result, these algorithms provide a high false DR [22].

In some studies, sampling techniques, such as undersampling, and oversampling are used to overcome the data imbalance problem. Both techniques work differently. Samples from the larger class are omitted in undersampling to balance the dataset [22]. However, oversampling techniques promote data balancing by generating data for the smaller class [28]. Synthetic minority oversampling technique (SMOTE) is a common technique to solve the balancing problems using the nearest neighbor approach [29]. Another common technique is adaptive synthetic (ADASYN) [30]. Both the aforementioned techniques are oversampling algorithms, which are used to deal with imbalanced dataset problem. The random oversampling (ROS) approach leads to overfitting. Whereas, the random undersampling (RUS) method loses information that affects the learning process and causes underfitting problem. Therefore, the mentioned sampling techniques are insufficient in terms of balancing the data. Another problem with data driven methods is the low theft DR. Besides, a high false-positive rate (FPR) makes the ETD approaches expensive and inefficient. Therefore, it is also a challenging part of ETD. Furthermore, accuracy is not considered a reasonable performance measure in the case of data imbalance problems. Hence, ETD models must be evaluated using appropriate performance metrics like precision, recall, F1 Score, etc. [22].

In this study, two different models are presented to overcome the problems identified in existing approaches. In the first model, we present a combination of two different sampling approaches, such as SMOTE and edited nearest neighbor (ENN), known as SMOTEENN, to balance the dataset. Moreover, a variant of convolutional neural network (CNN) architecture, i.e., AlexNet, is also used in the proposed work for feature extraction, which gained much popularity in the imageNet large scale visual recognition competition (ILSVRC) over the past few years [31]. Thereafter, a light gradient boosting (LGB) model is used for classification. The proposed model is named as SMOTEENN-AlexNet-LGB (SALM) for ETD. In the second model, a novel approach is proposed for ETD that consists of a Generative Adversarial Network (GAN), GoogLeNet, and Adaptive Boosting (AdaBoost), named as GAN-NETBoost. In the model, conditional Wasserstein GAN (CWGAN) with gradient penalty (CWGAN-GP) is used to balance the data by synthesizing the fake electricity consumption profiles of the minority class. GAN has gained much attention for anomaly detection [32]. It is proficient at learning the distribution of provided data to generate synthesized data close to the real data [33], [34]. CWGAN-GP is trained using the labeled electricity consumption data. Inception module based GoogLeNet is used for feature extraction while AdaBoost is used for final classification.

In the proposed work, we consider the publicly available electricity consumption data taken from the state grid corporation of China (SGCC). The dataset contains a daily electricity consumption report of both malicious and honest customers. However, the data is highly imbalanced (the ratio of dishonest consumers and honest consumers is 1:10), which is the biggest hurdle in ETD. The performance of the models is evaluated using Area under the curve (AUC), recall, precision, etc., and the comparison is performed with different traditional sampling and boosting algorithms.

This work is an extension of [35]. The main contributions of this work are listed below.

- A hybrid sampling technique SMOTEENN is proposed that prevents the loss of important information and overfitting problems caused by undersampling and oversampling techniques, respectively.
- AlexNet module of the proposed SALM model is used to extract important features from high dimensional data.
- To improve ETD and to reduce High FPR, LGB module of the proposed SALM model is used.
- The conventional GAN model stucks in training instability, vanishing and exploding gradient problems; therefore, CWGAN-GP is used in this study to deal with the issues of conventional GAN and efficiently balance the data by synthesizing the minority class samples.
- To extract the characteristics of customers' electricity consumption behavior, a version of GoogLeNet based on three inception layers is used in the GAN-NETBoost model.
- Using the suitable performance measures, the results of the proposed models are evaluated and compared with traditional models. The Simulation results shows that the proposed models outperform the existed models in term of efficient ETD.

The list of abbreviations used in this article is presented in the Nomenclature section given at the start of this manuscript. The remaining paper is organized as follows. Section II presents the existing studies related to ETD. The problems highlighted and addressed in this work are discussed in Section III. Section IV presents the proposed approach in detail. Furthermore, Section V describes the performance measures used for the models' evaluation in this work. Section VI compares the simulation results of the proposed methods against existing methods. Finally, Section VII concludes the proposed work.

II. RELATED WORK

In the last few decades, electrical power has become a backbone for the development of any country [36]. It has the potential to either raise or reduce the country's economy. Over the past few years, NTL detection has become a major problem in energy systems across the world. NTL is defined as the difference between TLs and total losses [37]. Electricity theft is the major reason for NTLs. Different approaches have been developed for detection and estimation of electrical power theft. The state-of-the-art techniques to detect electricity theft in AMI system are categorized into three groups: state based, classification based, and game theory based [38].

In [4], ML algorithms and statistical models are used to develop smart energy theft system (SETS) for securing Internet of things (IoT) based smart homes [39], [40]. Whereas, in [6], an artificial neural network (ANN) based scheme is presented to detect illegal electricity consumption patterns using probabilistic ANN and Levenberg-Marquardt algorithm. In NTLs detection, support vector machine (SVM) is a famous and extensively used model. In [41], authors tried to balance transmission and distribution level in SGs by merging decision tree (DT) and SVM in a top down manner. DT is used to estimate the power consumption rate and the output is then fed into SVM that classifies malicious and normal customers. In [27], authors introduced a security feature in AMI system by using SVM and named the model as consumption pattern based energy theft detector. It was evaluated using the dataset released by the sustainable energy authority of Ireland (SEAI). The proposed work is efficient in terms of low sampling rate for balancing the data. However, the dataset contains patterns that are collected manually from users to train the model, which is tedious and time consuming task.

Big data is used for the detection of electrical power loss. The selection of appropriate features is beneficial as it can enhance the model's performance. A variety of algorithms for the selection of an optimal set of features are available in literature. To improve the DR, authors in [9] used a binary black hole algorithm (BBHA) for selecting the appropriate features. Moreover, results are compared with other existing algorithms used for the selection of useful features from electricity consumption profiles of customers. A BBHA is a parameter free algorithm that makes the respective model less prone to errors than other algorithms. Authors in [10] proposed a deep network, named as wide and deep CNN (WD-CNN), for the detection of fraudulent customers in an electric power system. The model is trained and tested using electricity theft consumption data provided by the utility of China. The wide component of CNN is designed to derive the concept of memorization and generalization. Moreover, data is converted into two dimensions to achieve better accuracy.

A hybrid method is proposed, which combines two deep networks to analyze the electricity theft data [29]. Authors proposed a model by combining CNN and long short-term memory (LSTM) named as CNN-LSTM. In the model, CNN is used to extract the useful features from electricity consumption data and LSTM is used for final classification. Also, SMOTE is used to balance the dataset for satisfactory results. In [42], a hybrid approach of CNN and random forest (RF) is adopted to analyze the data of 5,000 consumers released by SEAI. Moreover, inspired by [10], authors reshaped the dimensions of data, which provide generalized feature extraction by CNN. A data-driven model is proposed for ETD in [13]. Clustering is performed for the extraction of similar groups for different types of consumers' consumption behavior. Moreover, a fuzzy based distance is measured to check whether the customers belong to a similar group or not. So, if the customers do not belong to the similar group, than it is considered as electricity theft. In [14], a data driven approach is proposed to detect cyber attacks in power systems by combining both state based and classification based schemes. Correlation between NTLs and a consumer's profile is captured with the help of maximum information coefficient. In [15], both the auxiliary database and the data of SMs are exploited to detect electrical power losses. Extreme gradient boosting (XGBoost) is used, which shows better performance as compared to other ML algorithms. Authors in [43] designed a gradient boost theft detector (GBTD) model that focused on feature engineering. Simulation results show that LGB performed better than XGBoost and categorical boosting (CatBoost). Furthermore, GBTD achieved higher accuracy in terms of ETD and improved time complexity. However, the model's performance is evaluated using the synthetic data.

In [12], authors introduced genetic programming algorithm to detect electricity theft. Finite mixture model (FMM) based clustering and Pearson's coefficient are used to detect abnormalities in AMI. The main aim is to achieve sustainability and efficiency that are considered the key factors in every energy system [44], [45]. In [20], LSTM and statistics modeling are used to achieve efficiency. The real electricity consumption patterns and billing data are fed to the model for training and testing purposes. However, time series analysis is not considered. Moreover, the feed forward network is used to identify cyber attacks in the AMI system [31]. Imbalanced nature of data is a challenging problem while handling electricity theft. Different sampling algorithms are used to mitigate this problem. In [22], random undersampling boosting (RUSBoost) is used to balance the dataset. Moreover, maximal overlap discrete wavelet packet transform (MODWPT) is used for feature extraction. However, for the massive imbalanced data, RUS is not considered an efficient approach due to loss of important information during elimination of instances from majority class. It also results in high FPR that renders it expensive and inefficient. The deployment of distributed energy resources (DERs) is increasing day by day. The study in [46] presented meter fraud detection in DERs and provided a set of electricity fraud attacks. An assumption is made in this article that the electricity thief is aware of the ETD method. The proposed method consisted of auto regressive integrated moving average (ARIMA) model and Kullback-Leibler divergence (KLD). Moreover, feature extraction is performed using principal component analysis (PCA). Data from five different resources is collected for the detection of meter fraud in DERs.

In [47], authors combined supervised and unsupervised learning by applying ANN and text mining to reduce NTLs in power utilities. ANN is used in data preprocessing step; whereas, an unsupervised technique is introduced in the second module. Authors in [48] proposed a combination of SMOTE and tomek-link (SMOTE-Link) to balance the dataset. Bidirectional gated recurrent unit (BGRU) is used for classification of consumers, which integrated the feature extraction property of PCA. SGCC dataset is explored for experimentation. Table 1 provides an overview of the existing electrical power loss detection approaches, focusing on the dataset used for testing and training of the model.

TABLE 1. Related work summary.

Task (References)	Techniques	Dataset	Performance Measures
SETS is developed to secure IoT-based smart home [4]	MLP, Recurrent Neural Network (RNN), GRU, LSTM, Simple Moving Average	Evaluated in Singapore home environment	Acc = 99.96%
Detection of illegal consumers in SG environment [6]	Probabilistic ANN, Levenberg-Marquardt	PJM	Acc = 96.11%
Determination of the most relevant features to characterize illegal customers [9]	BBHA, Optimal Power Flow	Two datasets provided by Brazilian utility	Recognition rate = 86%, 91.6%
Identification of periodicity and non-periodicity [10]	Wide and deep CNN	SGCC	AUC = 0.78%, (Mean Absolute Percentage Error) MAPE = 0.90%
Feature-engineering based ETD model [12]	FMM, GP, LGB	Irish CBT	AUC = 0.87%, F1-score = 0.92%, Precision = 0.85%, Acc = 0.86%
Detection of loss in SGs [13]	Fuzzy Gustafson Kessel clustering	4,000 Irish households	AUC = 0.74%, TPR = 63.6%, FPR = 24.3%
ETD against AMI in the energy Internet [14]	CFSFDP, MIC	Irish smart meter dataset	AUC = 81.6%, MAPE = 83.1%
Formulation of various characteristics of the customer's consumption behavior [15]	K-Means, K Nearest Neighbor, LR, SVM, NN, XGBoost	Electricity utility in Spain (Endesa)	AUC = 0.91%
Anomaly detection and distinguishing drift [20]	K-Means, LSTM	Electricity load diagrams	Precision = 78%, Recall = 88%
Detection of loss in electric distribution systems [22]	MODWPT, RUSBoost	2,271 selected consumers.	MCC = 0.73%, AUC = 0.81%, F1-score = 0.82%, Precision = 0.88%, Recall = 0.65%, Specificity = 0.98%, Acc = 0.94%
Identification of the suspicious customers [27]	Consumption pattern based energy theft detector, SVM	Electric Ireland and SEAI	DR= 94%, FPR = 11%
Detection of theft using historical power consumption patterns [29]	SMOTE, CNN, LSTM	SGCC	Acc = 89%, Precision = 0.90%, Recall = 0.87%
Detection of power theft attacks [31]	DNN	Irish Smart Energy Trial	DR= 93%, FPR = 2.3%
Detection of the malicious consumers [41]	DT, SVM	Power consumption of various homes in USA	Acc = 92.5%, FPR = 5.12%
Discovering the anomalous consumption behavior of suspicious customers [42]	SMOTE, CNN, RF	Electric Ireland and SEAI	AUC = 0.99%
ETD focusing on feature engineering [43]	GBTD, XGBoost, CatBoost, LGB	Electric Ireland SEAI	DR= 90%, FPR = 5%
Meter fraud detection in DER [46]	ARIMA, KLD	CER, Ausgrid solar, NREL solar, Engie wind	AUC = 99%
NTL detection [48]	SMOTE-LINK, Kernel PCA, BGRU	SGCC	AUC = 0.86, Precision = 0.80, Recall = 0.89

Moreover, different performance measures used for the models' evaluation are also given along with their values.

III. PROBLEM ANALYSIS

One of the main issues that affect the economic stability of a country is the electricity theft [10]. It is reported by Northeast Group LLC that the world loses $89.3 billion annually due to electricity theft [49]. Moreover, it also negatively affects the electricity supply and tariffs. Furthermore, as a result of electricity theft, revenue loss is divided among all customers, including the honest (legal) customers at the time of electricity tariff calculations [9]. Therefore, ETD and prevention are compulsory for a stable energy system. In order to detect electricity theft, the significant challenge is the imbalanced nature of the data. In [10], [15], supervised learning techniques are used for ETD. However, highly imbalanced data is used, which causes misclassification and low DR. In [43], ROS technique is used that causes overfitting and leads to computational complexity. Whereas, in [22], authors used RUS technique to balance the data. However, false DR is high and RUS also discards useful information from data while removing the electricity consumption instances.

Moreover, the performance of ML algorithms can be enhanced by extracting the most relevant features from raw data. Authors in [9], [22] used MODWPT and BBHA for feature extraction from electricity consumption data. However, MODWPT leads to computational complexity that requires more memory and time to extract features from a large amount of data. On the other hand, BBHA gets stuck in local optima and fails to provide efficient results for ETD. In [27] and [42], the ETD models are not evaluated using reasonable performance metrics for the identification of electricity fraudsters.

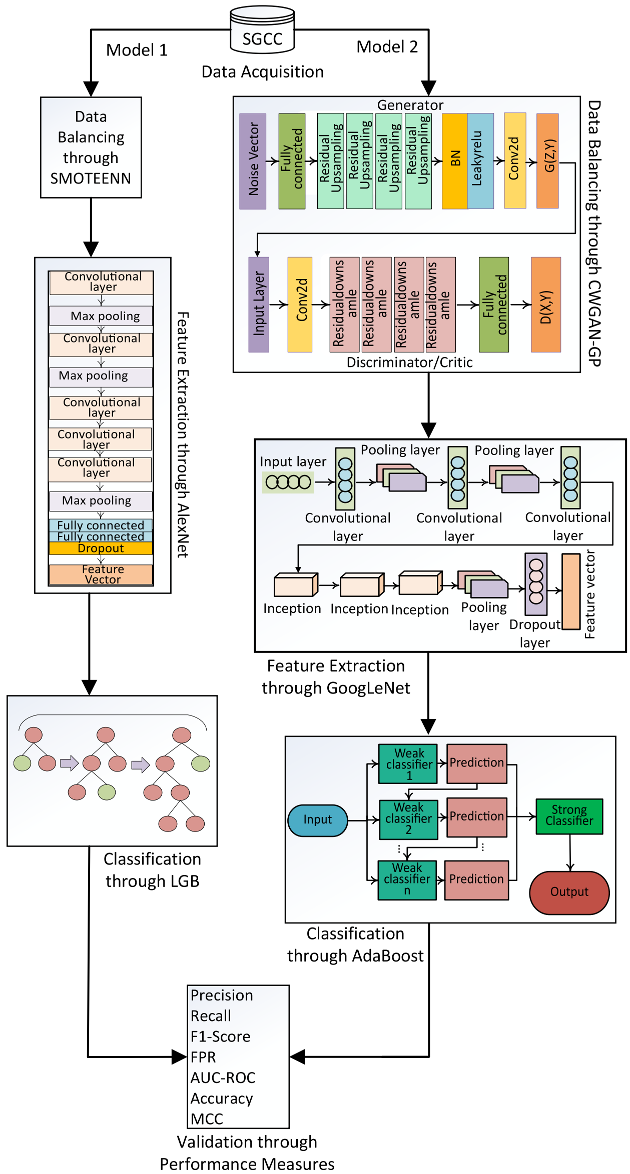


FIGURE 2. Proposed system model of SALM and GAN-NETBoost.

IV. PROPOSED SYSTEM MODEL

This section presents the proposed solutions for the problems highlighted in Section III. Two ETD models are proposed, as shown in Figure 2. The proposed SALM is presented in Figure 3. It is designed to address the identified limitations in Section III. We have mentioned all the limitations using red color in the model. Solution for each limitation is explicitly mentioned with blue color. The flowchart of the proposed SALM is presented in Figure 4. In contrast, the second ETD model is represented in Figure 5. The limitation identified in the existing literature are shown with red color and their proposed solutions are represented by blue color. The flowchart of GAN-NETBoost model is presented in Figure 6. In both models, data preprocessing, balancing the dataset, and feature extraction are the main steps, which are performed before final classification of the consumers. The following are identified limitations from the existing literature, which are referred as L1-L12:

- **L1:** presence of imbalanced data,
- **L2:** overfitting caused by duplicating the minority class instances,
- **L3:** loss of information due to the removal of instances from majority class,
- **L4:** high FPR due to the inability of model to learn high dimensional power consumption data,
- **L5:** generalization (poor performance of the model on unseen data),
- **L6:** difficult to extract meaningful information from large size time-series data,
- **L7:** inaccuracy in ETD,
- **L8:** optimization problem or being stuck in local optima,
- **L9:** improper validation of the proposed model or lack of performance measures regarding binary classification,
- **L10:** overfitting in deep learning models,
- **L11:** gradient vanishing problem, and
- **L12:** authenticity of data.

In the following subsections, the models are discussed in detail.

A. DATA AVAILABILITY

Real and verified electricity consumption patterns acquired from SGCC are used for training and testing of the proposed models. The dataset contains daily electricity consumption reports of both dishonest and honest consumers. The time duration of data is from Jan 2014 to Oct 2016. In this study, electricity consumption data of approximately 42,372 customers is presented of which 3,615 consumers are dishonest and 38,757 are honest. So, the ratio is 1:10, which proves that the data set is highly imbalanced. The dataset cannot be used without preprocessing as imbalanced dataset negatively affects the performance of a classifier [35].

B. DATA PREPROCESSING

The dataset consists of labeled real electricity consumption patterns. The presence of noise and missing values in the data along with data diversity degrade the model's performance. Therefore, data preprocessing is performed before training both models on the data. Figure 7 shows the techniques that are exploited to reduce the complexity of data. These techniques are discussed as follows.

- **Missing Values:** interpolation is used to fill data values that are lost or missing. In this work, linear interpolation is used to recover the data using the formula given in equation 1, modified from [10].

$$y_m = y_{1m} + \frac{y_{2m} - y_{1m}}{x_{2m} - x_{1m}} * (x_m - x_{1m}). \tag{1}$$

  where, y_m is the missing value in electricity consumption data over period x_m. y_{2m} is prior value to y_m and y_{1m} is the value that proceeds y_m. Moreover, x_m, x_{1m}, and x_{2m} show the time of data for y_m, y_{1m}, and y_{2m}, respectively.

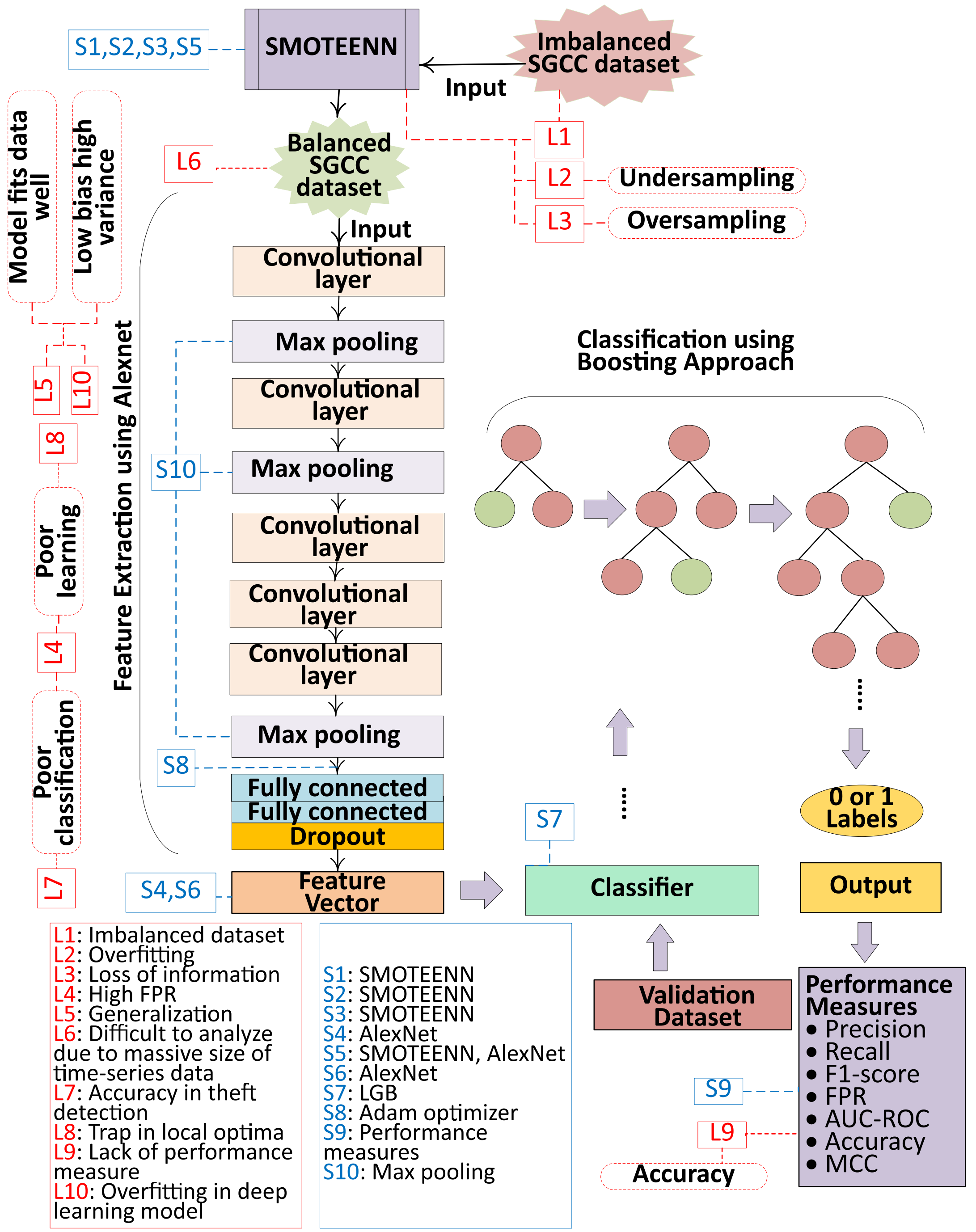


FIGURE 3. The proposed SALM model.

- **Noisy Data:** erroneous values that are also called noise and sometimes outliers may lead the classifier towards poor accuracy. Z-score is a simple and powerful method that deals with this issue and it is calculated using equation 2 [50].

$$z = \frac{x_m - \mu}{\sigma}. \tag{2}$$

where, μ is mean and σ is standard deviation.

- **Data Scaling:** min-max scaler method is used for scaling the data between 0 and 1. Equation 3 is used for scaling, taken from [10].

$$c = \frac{x_i - x_{min}}{x_{max} - x_{min}}. \tag{3}$$

where, x_{min} and x_{max} represent minimum and maximum values of x, respectively. In x_i, x belongs to the electricity

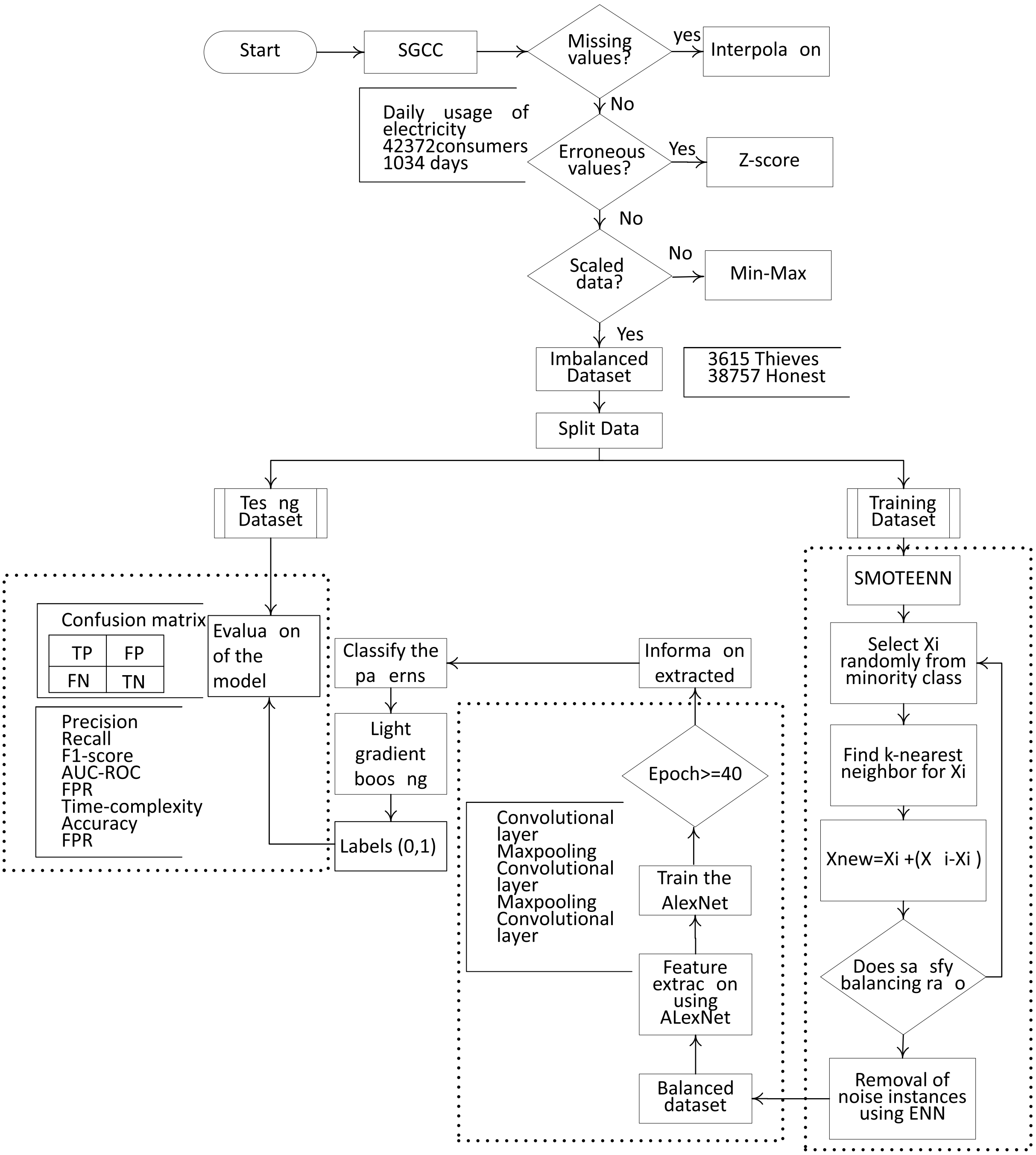


FIGURE 4. **Flowchart of the proposed SALM model.**

consumption of single user and *i* represents the specific interval.

C. DATA SAMPLING

In this section, the sampling techniques used in both models are discussed in detail. In ML, balancing the dataset is considered as an important step for better classification results. Different sampling techniques are used in the literature to balance the data. These sampling techniques create a balanced version of the dataset. ML algorithms are designed to maximize the accuracy and to reduce the error rate. However, imbalance nature of the data makes a classifier biased towards majority class. The classifier always has the tendency to predict the dominant class while ignoring the minority class. Algorithms perform well when there are equal number of samples in each class. The following are the commonly used techniques to balance the dataset.

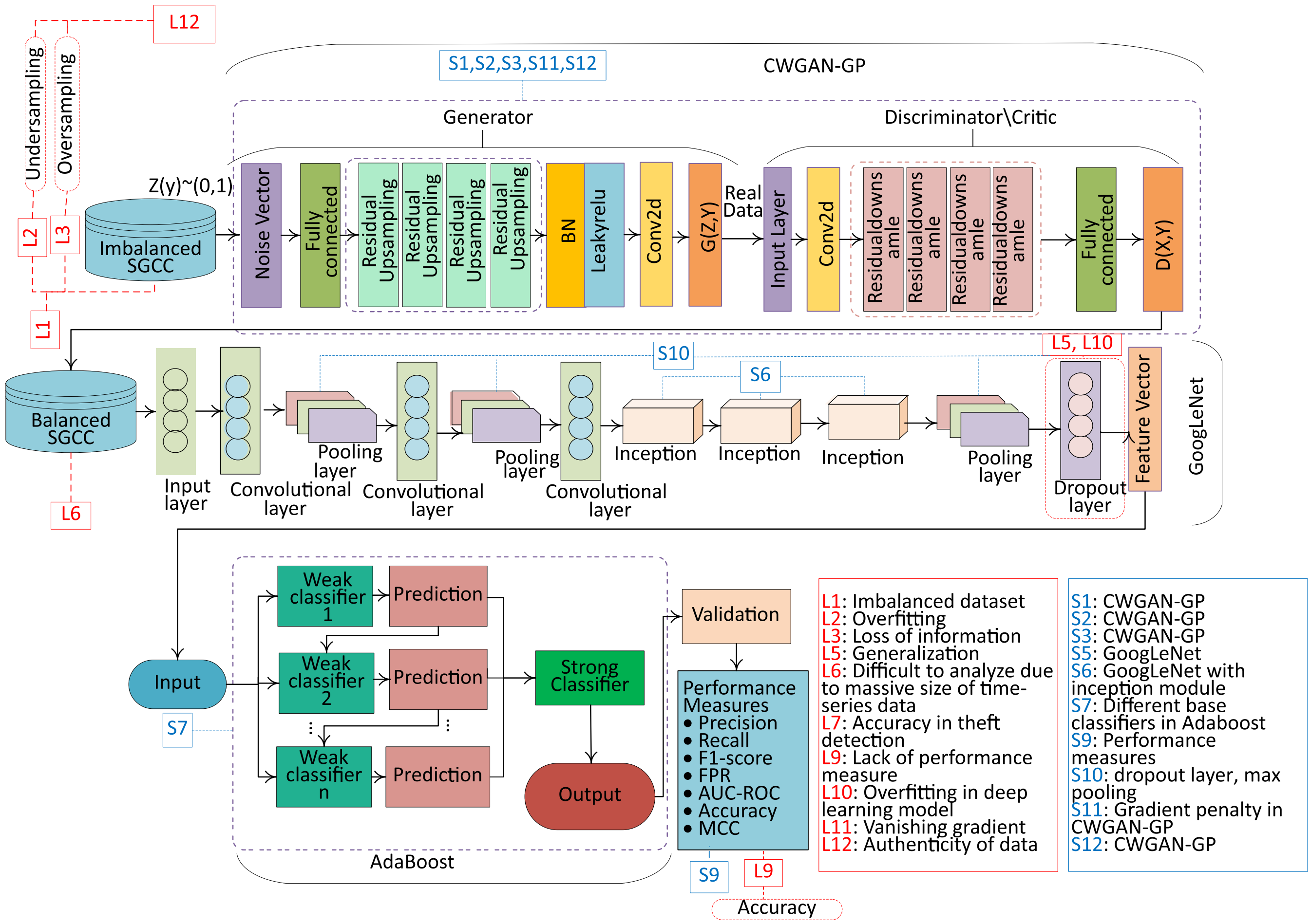


FIGURE 5. The proposed GAN-NETBoost model.

- **Oversample Minority Class:** in this approach, minority class is duplicated randomly to balance the majority class. The duplicated data are generated using the existing data. ROS is used for the duplication of the samples of minority class. Therefore, duplication of data causes overfitting and also increases computational complexity.
- **Undersample Majority Class:** RUS refers to the random selection of the samples of majority class without considering the usefulness of the samples. The selected samples are discarded to balance the number of samples in both classes. The major drawback of undersampling is the loss of important information. Thus, it makes the dataset small and leads to underfitting.
- **Hybrid Sampling:** it combines oversampling and undersampling to generate a balanced dataset. In this technique, partial sampling is performed in which minority class is oversampled and majority class is undersampled. In this way, negative impacts of both techniques are minimized. However, it may lead to poor generalization due to ignorance of distribution of samples in both classes.

1) SYNTHETIC MINORITY OVERSAMPLING TECHNIQUE EDITED NEAREST NEIGHBOR

In SALM, a combination of both oversampling and undersampling techniques is utilized to balance the class distribution uniformly as well as to overcome the drawbacks of both sampling algorithms. It is known as SMOTEENN. It is hybrid of the following algorithms.

Synthetic minority oversampling technique: it is used to balance the data by using the data of minority class. The working of SMOTE is described as follows.

- Step 1: minority class is selected as vector V.
- Step 2: for each sample s, $s \in V$, find the nearest neighbor.
- Step 3: sampling rate R is a set that constructs N by selecting the nearest neighbors (i.e., $s1, s2, \ldots, sn$) randomly.
- Step 4: equation 4 is used to generate the samples

$$x' = x + rand(0, 1) * |x - x_k|. \tag{4}$$

where, x', represents the generated samples and x_k represents the nearest value.
- Step 5: steps are repeated until data is balanced.

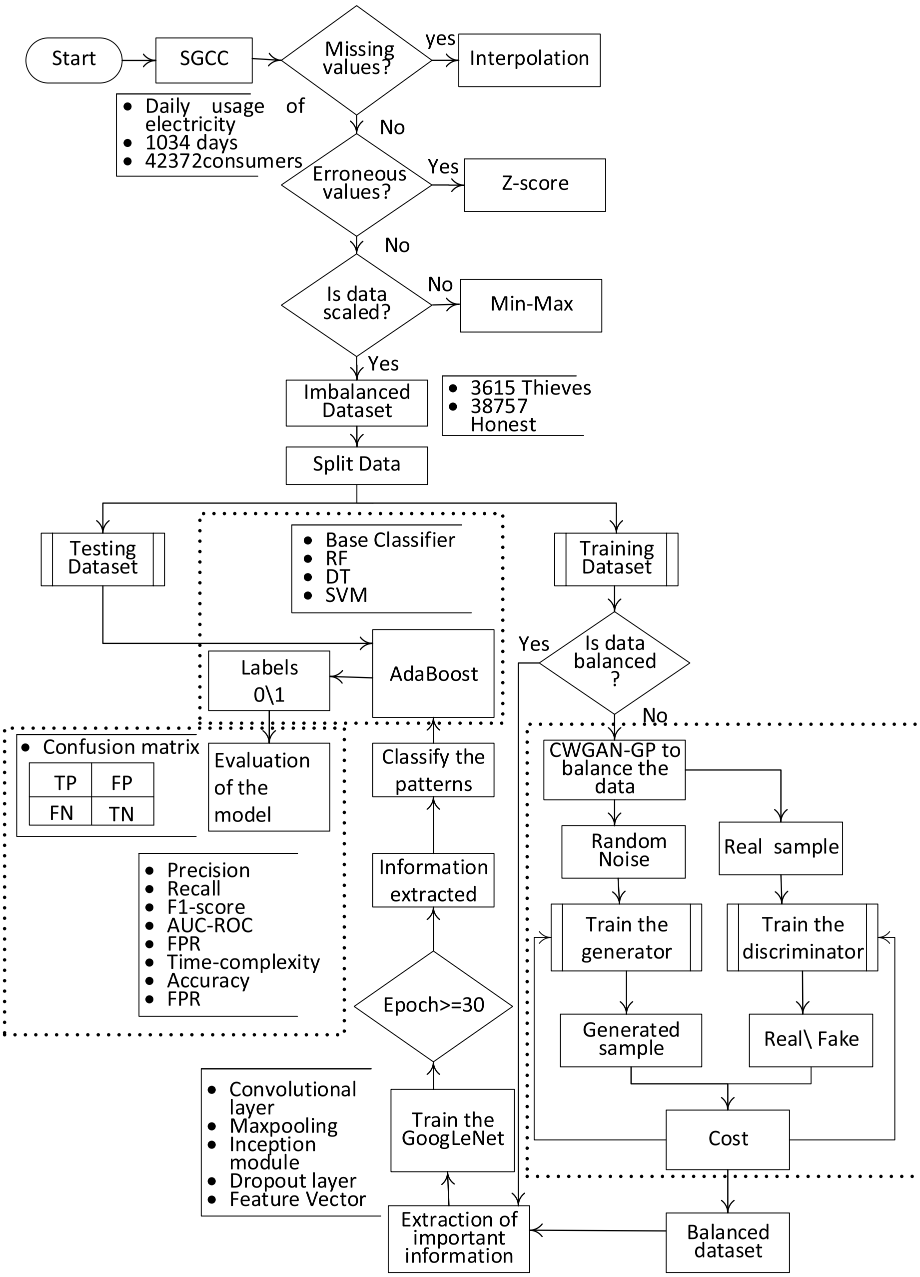


FIGURE 6. Flowchart of the GAN-NETBoost model.

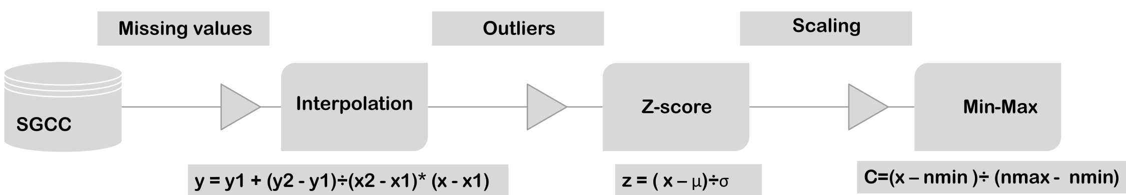


FIGURE 7. Overview of data preprocessing.

Edited nearest neighbors: it removes samples from the majority class to balance the dataset. The borderline and noisy samples are removed from the majority class in order to get a balanced class. ENN is based on KNN that removes those samples that act differently. If a sample has more nearest neighbors from different class, then it will be discarded. ENN works according to the steps given below.

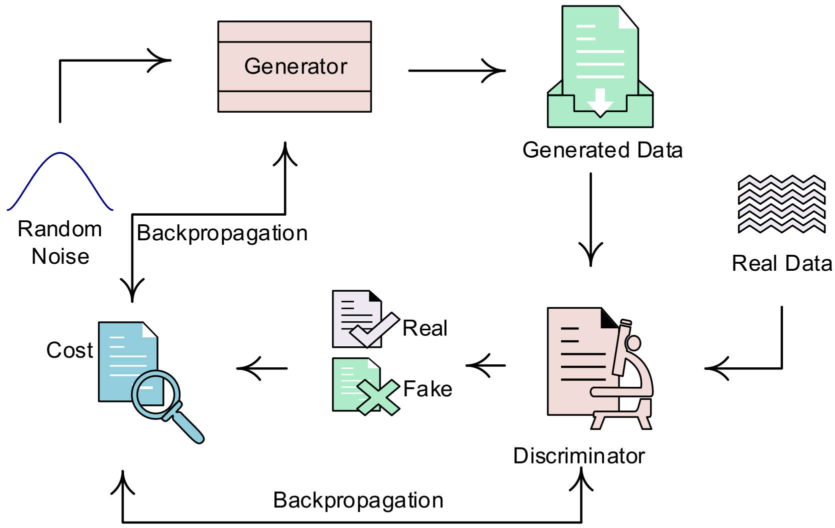


FIGURE 8. General view of GAN.

- Step 1: select the majority class as N.
- Step 2: for all given samples v from N, find the nearest neighbors.
- Step 3: if the more number of nearest neighbors of a sample belong to the majority class, then discard it.
- Step 4: repeat the steps for each sample from the majority class.

2) CONDITIONAL WASSERSTEIN GENERATIVE ADVERSARIAL NETWORK-GRADIENT PENALTY

To compensate for an imbalanced dataset in GAN-NETBoost, CWGAN-GP is used. It is a modified version of GAN, which is capable of learning the distribution of the provided data that helps in data generation. GAN belongs to the class of neural networks (NNs) and has the ability to generate new data from the existing data [51]. Goodfellow proposed GAN in 2014, which comprises of two NNs: generator and discriminator. Figure 8 illustrates a simple structure of GAN. CWGAN-GP is the combination of conditional generative adversarial network (CGAN) and wasserstein generative adversarial network gradient penalty (WGAN-GP). The results prove that CWGAN-GP achieves enhanced performance as compared to conventional data sampling techniques.

- **Generator:** the generator network learns the data distribution and generates the data, which is close to the real data. It tries to confuse the discriminator, which is used to distinguish fake samples from real data samples. Random noise z is fed to the generator as an input that is either Gaussian noise or an arbitrary point in latent space. The cost function for the generator model is measured using the following equation 5. It was used by Goodfellow in [51].

$$\min_G V(G) = \mathbb{E}_{x \sim px(x)}[log(1 - D(G(z)))], \tag{5}$$

where $G(z)$ is the value generated using noise z, $D(G(z))$ shows the discriminator's prediction whether $G(z)$ is real or fake and $\mathbb{E}_x$ is the expected value over all real data instances. x belongs to real data and $px(x)$ represents the distribution of real data.
- **Discriminator:** it is used to discriminate between real and the samples generated by GAN. $G(z)$ represents the instances produced by a generator and x represents the instances that belong to actual dataset. The instances are fed into discriminator as input of the network. The loss function used for training the discriminator is given in equation 6, taken from [51].

$$\max_D V(D) = \mathbb{E}_{x \sim Pdata(x)}[logD(x)] + \mathbb{E}_{z \sim Pz(z)}[log(1 - D(G(z)))], \tag{6}$$

where $D(x)$ depicts the discriminator's estimated probability, x represents real data instances and $\mathbb{E}_z$ shows the expected value for overall generated fake instances $G(z)$. $pz(z)$ represents the distribution of input noise z.

Both generator and discriminator fight to achieve the desired equilibrium, which is achieved in a game theory context using a minimax game. Moreover, simultaneous training of both generator and discriminator is done in an adversarial manner. The objective function of GAN is given in equation 7 [51].

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim Pdata(x)}[logD(x)] + \mathbb{E}_{z \sim Pz(z)}[log(1 - D(G(z)))]. \tag{7}$$

a: CONDITIONAL GENERATIVE ADVERSARIAL NETWORK

It compiles generator and discriminator under some conditions discussed later in this section. It is an advanced design of GAN. Data is generated in a supervised manner by using a condition that is in the form of some additional information [52]. In our case, the condition is referred to as labels

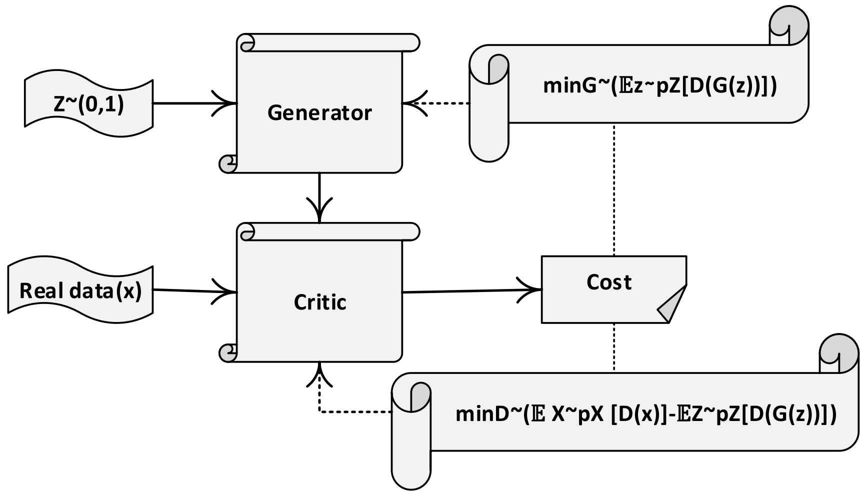


FIGURE 9. General overview of WGAN.

and is represented by y. These labels control the mode of the data to be generated. The additional information is fed to G as input along with noise z and can be presented as $G(z, y)$. Equation 8 gives the cost function for CGAN [52].

$$\underset{G}{min}\underset{D}{max} V(D, G) = \mathbb{E}_{x\sim Pdata(x)}[logD(x|y)] + \mathbb{E}_{z\sim Pz(z)}[log(1 - D(G(z|y)))]. \quad (8)$$

where, y shows the auxiliary condition (information about the labels), $G(z, y)$ is the generated value and label, and $D(x, y)$ is the discriminator's estimate along with label. Training instability and failure to converge are two big challenges that occur while training a GAN model. The following are some limitations of GAN that are tackled by CGAN.

- **Mode Collapse:** during data generation, a stage comes where almost the same data is generated. The generated data can fool the discriminator. However, it is unable to fulfill the requirements of the real world data distribution. Therefore, it is considered as a failure of the GAN model.
- **Vanishing Gradient:** if a discriminator is not trained well and during the training process the loss becomes zero. It means that no gradient is left for further iterations, which leads to the vanishing gradient problem. On the other hand, if the discriminator is not trained properly, then the generator would not generate a real sample due to the lack of appropriate feedback from the discriminator.
- **Nash Equilibrium:** GAN is based on minimax, which means gradient descent is used for training both discriminator and generator networks. Both networks are trained together to achieve a Nash equilibrium. These networks update their loss functions, simultaneously. Thus, creating difficulty in the model's convergence.

b: WASSERSTEIN GENERATIVE ADVERSARIAL NETWORK

It is proposed by Martin Arjovsky, as shown in Figure 9. In WGAN, discriminator is replaced by critics; whereas, minimax loss is replaced by Wasserstein's loss, also known as earth-mover. During the training of WGAN, Wasserstein's loss helps to learn the data distribution and improve the stability of both generator and discriminator by calculating the probability distribution. Wasserstein's loss function is taken from [53], and is given in equation 9.

$$W(p_r, p_g) = \underset{\gamma \epsilon \Pi(p_r, p_g)}{inf} \mathbb{E}_{(x,y)\sim\gamma}[\|x - y\|]. \quad (9)$$

where, $\Pi(p_r, p_g)$ is the set of joint distributions $\gamma(x, y)$. p_r and p_g denote the starting and ending marginals, respectively. Π contains all the possible transport plans, γ and (x, y) indicates the transported mass from x to y. In the output layer of the critic model, a linear function is used to activate the model. The objective of replacing the sigmoid function with a linear function is to use the Lipschitz function. Furthermore, (1,0) labels are replaced with (-1,1) for real and generated data, respectively. Equation 10 gives the Lipschitz function, which is taken and modified from [53].

$$|f(x_1) - f(x_2)| \leq |x_1 - x_2|, \quad (10)$$

where x_1 and x_2 are the real values of x. $f()$ is the function that bounds the real values. The main challenge in WGAN is the implementation of the Lipschitz restriction. Clipping enforces the Lipschitz constraint (LC) on the critic's model meanwhile introducing some additional problems. Poor quality samples and weak convergence of the model make its performance questionable.

c: WASSERSTEIN GENERATIVE ADVERSARIAL NETWORK-GRADIENT PENALTY

It is used in this work. The LC given in equation 11 was introduced in [54]. The below equation is used in this work after modification.

$$|f(x_1) - f(x_2)| \leq k|x_1 - x_2|, \tag{11}$$

where k is an independent constant that shows the maximum gradient norm. In [54], gradient penalty (GP) is also used in WGAN instead of the weight clipping that helps to implement the above mentioned LC. The objective of WGAN-GP is mentioned in equation 12 [54].

$$L = \underbrace{\mathop{\mathbb{E}}_{\widetilde{x}\sim P_g} [D(\widetilde{x})] - \mathop{\mathbb{E}}_{x\sim P_r} [D(x)]} + \underbrace{\lambda \mathop{\mathbb{E}}_{\hat{x}\sim P_{\hat{x}}} [(\| \nabla_{\hat{x}} D(\hat{x})\|_2 - 1)^2]}. \tag{12}$$

where, P_g is generator distribution and P_r is data distribution. $\hat{x}$ is the random sample, x is real sample, $\widetilde{x}$ defines model distribution, and λ denotes GP coefficient, which is set to 10. First part of equation 12 shows original critical loss while the second part shows GP. The following are the modifications made in the construction of the CWGAN-GP model.

- **Leaky ReLU:** to deal with the vanishing gradient problem of ReLU, leaky ReLU is used in the proposed model as an activation function.
- **Batch Normalization:** it is used to enhance the stability of the model and also to normalize the input within the network. The intention behind using batch normalization is to avoid vanishing and exploding gradients as well as mode collapsing to optimize the model's training. It is applied after the convolutional layer.
- **Adam Optimizer:** it is the most widely used optimizer due to its speed as it achieves optimal results quickly and minimizes the loss function as well. It offers the advantages of an adaptive gradient algorithm. Therefore, Adam optimizer is used in this model.
- **Gaussian Weight Initialization:** in NNs, weight initialization is considered an essential task because large weight values can lead to convergence problem. Also, the use of larger weight values in the network forces the activation layer to produce vanishing and exploding outputs, which result in very small or very large gradient update. To mitigate this issue, a Gaussian distribution function is used having values 0 and 1 for mean and standard deviation, respectively. It is considered a reasonable approach for weight initialization. In our model, we are using the Xavier approach to ensure the distribution of the inputs to each activation function with zero mean.

D. FEATURE EXTRACTION

In this section the feature extraction techniques used in both models are discussed in detail. In the models, the feature selection is not considered because the dataset contains only electricity consumption data.

1) AlexNet

The launch of ILSVRC has brought revolutionary changes in the field of DL. The ILSVRC is an annual software contest where different models are presented as detectors and classifiers. In 2012, AlexNet won the ILSVRC and was named after the developer, Alex Krizhevsky [30]. In SALM, AlexNet with five convolutional layers, three maxpooling layers, two fully connected layers and a dropout layer is exploited. AlexNet learns the features from massive daily electricity consumption data. Moreover, Adam optimizer is used to update the weights of the network.

2) INCEPTION MODULE BASED GoogLeNet

Inception module is a deep CNN architecture that was introduced in ILSVRC, held in 2014 (ILSVRC14) [55]. In GAN-NETBoost, it is used for feature extraction. The breakthrough performance of this model in ILSVRC14 made it popular in the creation and innovation of deep architectures to achieve high accuracy. In the development of the inception module, we focus on the kernel size. Basically, the inception module uses three types of kernels at once, as depicted in the Figure 10. It also represents the basic concept of the inception module with dimensionality reduction [55].

In the inception layer, the purpose of 1×1 convolution is the reduction of data instances. Besides, the use of rectified linear activation function improves the performance of the model. Higher computational requirements are decreased by using low dimensional embedding. It also extracts information at earlier level. Usually, it becomes difficult to extract information from dense layer where information gets compressed. The presented GoogLeNet module consists of 27 layers. The addition of inception module in GoogLeNet enhance its performance, as shown in Figure 10. In the proposed work, a streamlined version of GoogLeNet with inception module is used for the sake of feature extraction. The proposed architecture consists of 3 inception layers. Figure 5 illustrates the proposed architecture of GoogLeNet used for dimensionality reduction.

E. CLASSIFIERS

The classifiers of SALM and GAN-NETBoost models are discussed in this section.

1) LIGHT GRADIENT BOOSTING MODEL

In SALM, LGB model is used for classification. It uses histogram based algorithms that pace up its learning. Extracted features are fed to the model as input for training purpose. Boosting algorithms split the tree depth wise, whereas LGB model splits the tree leaf wise that makes it different from others and improves its accuracy. LGB performs well on large data, requires low memory for execution and has good accuracy.

2) ADAPTIVE BOOSTING MODEL

In GAN-NETBoost, AdaBoost is used for classification, which is a famous ML based boosting approach.

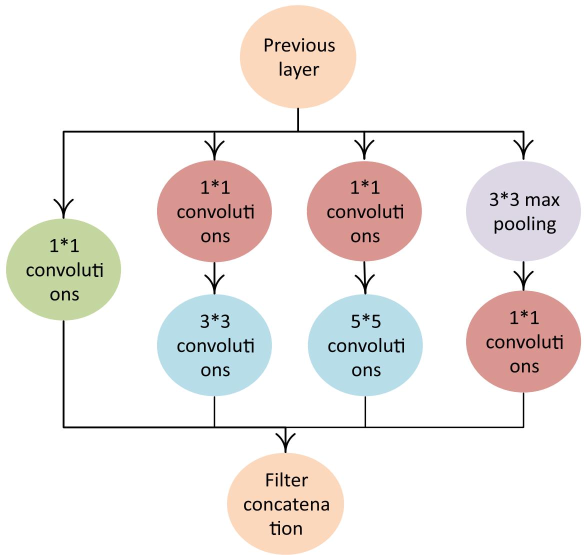


FIGURE 10. Inception module with dimensionality reduction.

AdaBoost combines multiple weak classifiers to make an efficient and robust classifier. It works by assigning higher weight to those instances that are miss classified and are lower to those who are already handled well. The basic idea behind AdaBoost's working is the training of various weak classifiers using the same training set and then combining them to build a stronger classifier. Weights are assigned to each sample on the basis of correct classification and the updated dataset having new weighted instances is fed to the next classifier for training and classification. Moreover, the continuous training process in AdaBoost improves data classification ability by reducing both bias and variance. It is used for both classification and regression problems [56]. The working of AdaBoost is given below.

- Step 1: AdaBoost selects the training samples (x_n) with labels.
- Step 2: identical weights ($w^m I$) are assigned to all samples.
- Step 3: a weak classifier (c_m) is build to classify the data samples.
- Step 4: the error rate (ϵ_m) is evaluated using equation 13 and t_n in equation 13 depicts the true class label.

$$\epsilon_m = \sum_{n=1}^{N} w_n^m I(c_m(x_n) \neq t_n). \tag{13}$$

- Step 5: the weight (α_m) is updated using equation 14 according to (ϵ_m) for next iteration.

$$\alpha_m = \frac{1}{2} \ln \frac{1-\epsilon_m}{\epsilon_m}. \tag{14}$$

- Step 6: steps 1-5 are repeated to assemble a strong classifier (C_M).

$$C_M(x) = sign(\sum_{m}^{M} = (\alpha_m c_m(x)). \tag{15}$$

Extracted features are used as inputs of AdaBoost to distinguish innocent consumers from fraudulent consumers. As AdaBoost is sensitive to erroneous data and missing values, therefore, data is cleaned and normalized in the proposed work. In this article, AdaBoost is trained using different base learners using the same dataset to achieve the optimal base classifier. RF, DT and SVM are the weak learners that are tested as base classifiers in AdaBoost and referred as ABRF, ABDT, and ABSVM, respectively. The performance of these base learners is evaluated using accuracy and AUC. The purpose behind using different base classifiers is to experimentally verify the impact of these models on the accuracy of binary classification using electricity theft data.

- **Adaptive Boosting with Random Forest:** RF is used as a base classifier in AdaBoost to build a prediction model. RF is one of the most widely used ensemble learning techniques and is considered as a successful technique for high dimensional classification.
- **Adaptive Boosting with Decision Tree:** DT is a simple algorithm used for classification problems and pattern recognition. ABDT is a combination of AdaBoost and DT as a base classifier. A tree is structured using simple and understandable rules. The learning process of DT is based on the practical method of inductive inference. Classification is performed from roots to leaves and the nodes are tested by variables. Branches are extended to assign appropriate nodes to the variables.
- **Adaptive Boosting with Support Vector Machine:** it is a combination of AdaBoost and SVM as a base classifier, named as ABSVM. SVM has gained tremendous popularity due to its good performance. It is flexible and used for both classification and regression problems. Structural risk minimization was the motivation behind the development of SVM. The decision function is given in equation 16.

$$f(s) = (w, \Phi(s)) + b, \tag{16}$$

where w is the weight factor, b represents bias and $\Phi(s)$ shows the mapping of samples.

V. MODEL EVALUATION

In this section different performance measures used for the evaluation of the proposed model are discussed. ETD is the binary classification task to differentiate between honest and dishonest consumers. The results are presented in the form of a confusion matrix (CM), which is a combination of predicted and actual values that involves: true positive (TP), true negative (TN), false positive (FP), and false negative (FN). These terms are defined below.

- **TP:** when a classifier predicts fraud and the actual value is also fraud.
- **TN:** when a classifier predicts no-fraud and the actual value is also no-fraud.
- **FP:** when a classifier predicts fraud, but the actual value is no-fraud.

- **FN:** when a classifier predicts no-fraud, but the actual value is fraud.

TP and TN are the correct values while FP and FN indicate error or misclassification of the classifier. The basic aim of ETD is to detect maximum fraud cases and to avoid costly on-site inspections [26]. The performance metrics used for model's evaluation are discussed below.

A. PRECISION

Precision describes the ratio of consumers that are correctly predicted as fraudulent and are fraudulent. It refers to the measure of a classifier's exactness. Precision gives the information about how many theft cases are identified. It is calculated using equation 17 [22]. Low precision means a high ratio of FP, which is to be avoided.

$$Precision = \frac{TP}{TP + FP}. \tag{17}$$

B. RECALL

Recall is also known as sensitivity. It shows the proportion of thieve consumers predicted as thieve by the classifier [33]. It is defined as the measure of how much a classifier detects TP correctly. TP and FN collectively show the total number of thieves while TP is the predicted number of thieves. Recall indicates a classifier's performance concerning FN. It gives information about how many theft cases are missed. Equation 18 is used for calculating recall [22]. Low recall means a high number of FN.

$$Recall = \frac{TP}{TP + FN}. \tag{18}$$

C. FALSE POSITIVE RATE

FPR is an important performance measure. A model with low FPR performs better in terms of ETD. Because once a user is detected as thief, on-site inspection is performed to catch him. False detection is harmful in terms of both cost and time. Equation 19 is used to calculate FPR.

$$FPR = \frac{FP}{FP + TN}. \tag{19}$$

D. ACCURACY

Accuracy is not considered a good performance metric when the dataset is imbalanced. Suppose the dataset contains records of 100 customers where 95 are honest and 5 are fraudulent. The algorithm predicts every case as honest, including fraudulent as well. The algorithm shows brassiness towards majority class in case of imbalanced dataset. Although the algorithm's performance in terms of anomaly detection is poor, still the accuracy will be 95%. Equation 20 is used to calculate accuracy of a model [22].

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}. \tag{20}$$

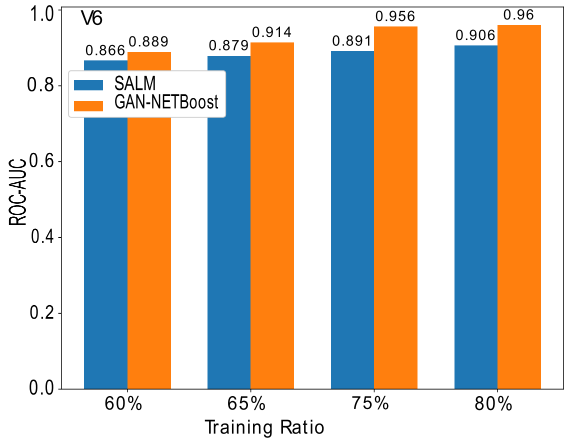


FIGURE 11. AU-ROC of SALM and GAN-NETBoost models using different data sizes.

E. AREA UNDER THE RECEIVER OPERATING CHARACTERISTICS

It is considered as the most important measure for any classification model mostly in the case of data imbalance problem. Receiver operating characteristic (ROC) is known as the probability curve. Moreover, AUC is also a famous measure of separability and used as a summary of the ROC curve. It plots the FPR and true positive rate (TPR) on x and y axis, respectively. Higher AUC refers to high TPR and leads to better prediction of the model.

F. F1-SCORE

It is the combination of precision and recall. If F1-score is high, it means that the model is perfectly predicting the consumers as honest or thieve. Equation 21 is used to calculate F1-score.

$$F1 - score = 2 * \frac{Precision * Recall}{Precision + Recall}. \tag{21}$$

G. MATTHEWS CORRELATION COEFFICIENT

B.W. Matthews introduced Matthews correlation coefficient (MCC) [57], which is used as a quality measure in ML. It is considered as a more reliable statistical rate. It only produces a high score if TP, TN, FP, and FN obtain good results. Equation 22 is taken from [57] and is modified for calculating MCC in this work.

$$MCC = \frac{TP * TN - FP * FN}{Precision + Recall}. \tag{22}$$

VI. SIMULATION RESULTS

Simulations are performed using a machine having 2.3 GHz Intel Core i5 CPU and Python programming language in Google Colaboratory. Experiments are conducted using four different ratios of training and testing datasets, such as 60:40, 65:35, 75:25, and 80:20. The selection of suitable hyperparameters is important for efficient performance of ETD model. The models are trained using Adam optimizer. Figure 11 shows the performance of the proposed models, i.e., SALM and GAN-NETBoost, for different training datasets. The obtained results imply that both models offer higher accuracy when the training ratio is 80%. Moreover, the performance of GAN-NETBoost is better than SALM for all data ratios. The figure depicts that SALM has achieved AUC of 90.6% and the GAN-NETBoost model has achieved 96% AUC at 80:20 ratio of training and testing dataset. The tag of V6 in Figure 11 indicates the validation of the solution proposed against limitation L6. A large amount of power consumption patterns are analyzed correctly by AlexNet and GoogLeNet. AUC increases with the increase in training dataset's size, which means that models can correctly detect abnormality in power consumption data.

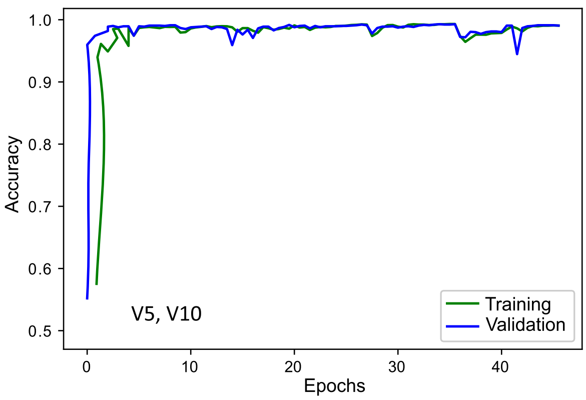


FIGURE 12. Accuracy of SALM.

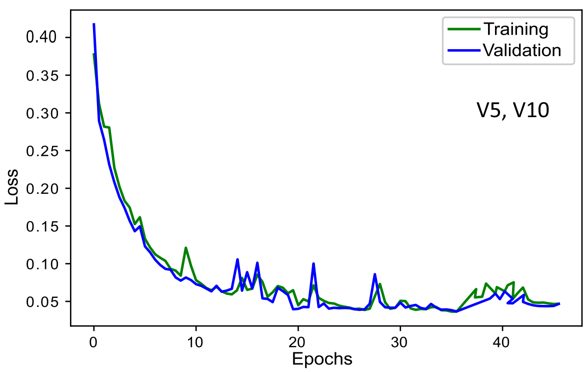


FIGURE 13. Loss of SALM.

Overfitting problem is defined as the inability of a model to perform well on unseen data; whereas, the model performs exceptional on training dataset. This limitation makes the model inefficient. Figure 12 shows the accuracy of SALM; whereas, loss is presented in Figure 13. These figures depict better convergence of the training process. The closeness of training and validation scores show that the model learns properly, which means SALM is generalized and there is no overfitting problem. The tags V5 and V10 in Figures 12 and 13 demonstrate the validation of solutions proposed against the limitations L5 and L10, respectively.

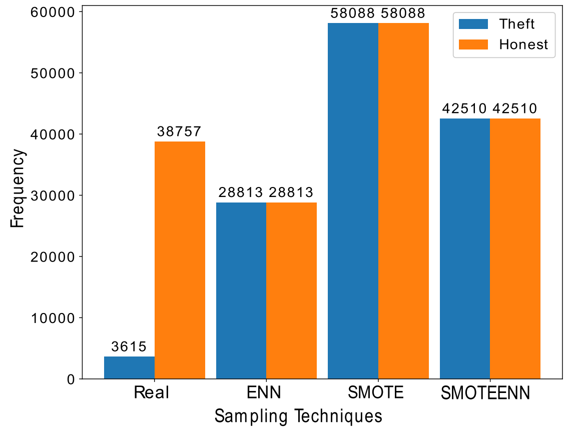


FIGURE 14. Results of different sampling techniques.

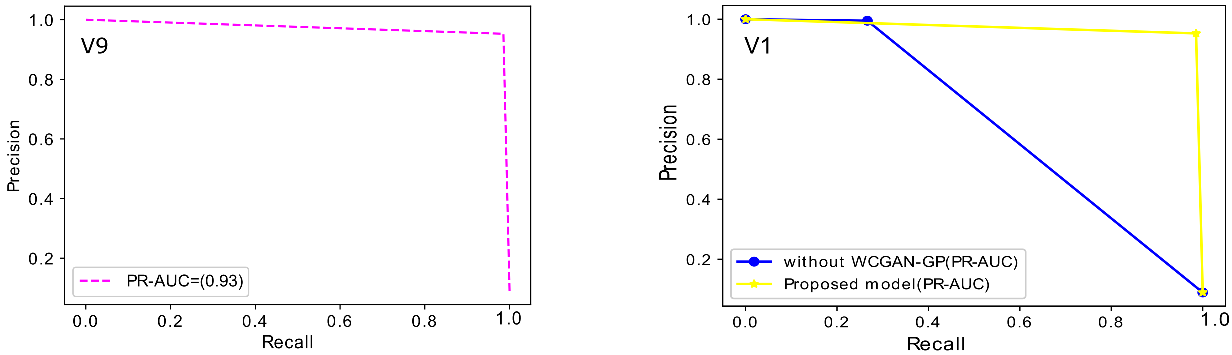
TABLE 2. V1: Comparison of GAN-NETBoost model with and without balancing technique CWGAN-GP.

Metrics	Balanced Dataset	Imbalanced Dataset
Accuracy	0.95	0.74
Precision	0.96	1.00
Recall	0.94	0.60
MCC	0.91	0.75
F1-score	0.95	0.75

TABLE 3. V1: Comparison of SALM With and without balancing technique SMOTEENN.

Metrics	Balanced Dataset	Imbalanced Dataset
Accuracy	0.90	0.70
Precision	0.94	1.00
Recall	0.91	0.58
MCC	0.87	0.73
F1-score	0.93	0.72

Figure 14 represents the original distribution of complete dataset. Moreover, it also shows the obtained results after applying SMOTEENN, SMOTE, and ENN on training dataset. The number of nearest neighbors, i.e., k is selected as 5 for both SMOTE and ENN. L1 tag is assigned to an imbalanced dataset problem, which is considered the most important limitation related to the dataset for binary classification. In binary classification problems, the imbalanced nature of data affects the performance of the model and it becomes biased towards the majority class by ignoring the minority class. Tables 2 and 3 show the importance of balanced dataset for ETD. Table 2 shows the values against different performance measures for the GAN-NETBoost model with CWGAN-GP (balanced dataset) and without CWGAN-GP (imbalanced dataset). It can be observed that CWGAN-GP enhanced the performance of the model. Table 3 shows the improved performance of SALM using SMOTEENN. Tables 2 and 3 validate the results of solutions proposed for L1. Hence, these tables are tagged as V1.



(a) PR Curve of SALM Model

(b) PR Curve of the GAN-NETBoost Model

FIGURE 15. PR-AUC of both models using SGCC dataset.

Figure 15 shows the trade-off between TPR and the positive prediction rate in the form of a precision-recall (PR) curve. More area under the PR curve means that the classifier efficiently distinguishes regular and irregular samples. In Figure 15a, it is clearly shown that the SALM is capable of explicitly detecting electricity theft. In Figure 15b, PR curves for balanced and imbalanced datasets are presented. A balanced dataset improves the performance of GAN-NETBoost and helps to classify the instances efficiently and accurately.

To demonstrate how the proposed methods efficiently perform in terms of ETD, Figure 16 presents the AUC-ROC curve. The classifier with AUC value closer to 1 achieves superior performance than other classifiers. The 0.90 value of AUC proves better performance of the proposed SALM model, as shown in Figure 16a. Figures 16b and 16c show the simulation results for GAN-NETBoost. In Figure 16b, the performance of AdaBoost is tested by using different base classifiers, such as ABDT, ABRF, and ABSVM. These models are trained and tested using the SGCC dataset. Performance of all models is evaluated using AUC and accuracy. All results are generated using 10-fold cross-validation mechanism. The comparative analysis shows that ABSVM achieved better performance than ABRF and ABDT. The AUC value of ABSVM is 0.96, which is closer to 1. It validates better performance of AdaBoost with SVM as a base classifier. ABRF and ABDT achieved AUC values of 0.92 and 0.89, respectively. Moreover, Figure 16c shows the AUC score achieved by the proposed model using CWGAN-GP against imbalanced data. In contrast, the PR curve for the proposed model with balanced and imbalanced dataset is shown in Figure 15b. AUC shows better performance of ABRF than ABDT using electricity consumption data. ABSVM outperforms both ABDT and ABRF by achieving high accuracy. So, ABSVM is considered as a final base classifier. The remaining results are obtained using ABSVM. Figure 16d shows the AUC obtained after applying different sampling techniques and compares their performance with SMOTEENN. Figure 17 represents the accuracy score for base models of AdaBoost. ABSVM achieved accuracy of 0.95; whereas, accuracy of ABRF and ABDT is 0.93 and 0.90, respectively. It proves that SVM performs better than RF and DT as a base classifier in the present scenario.

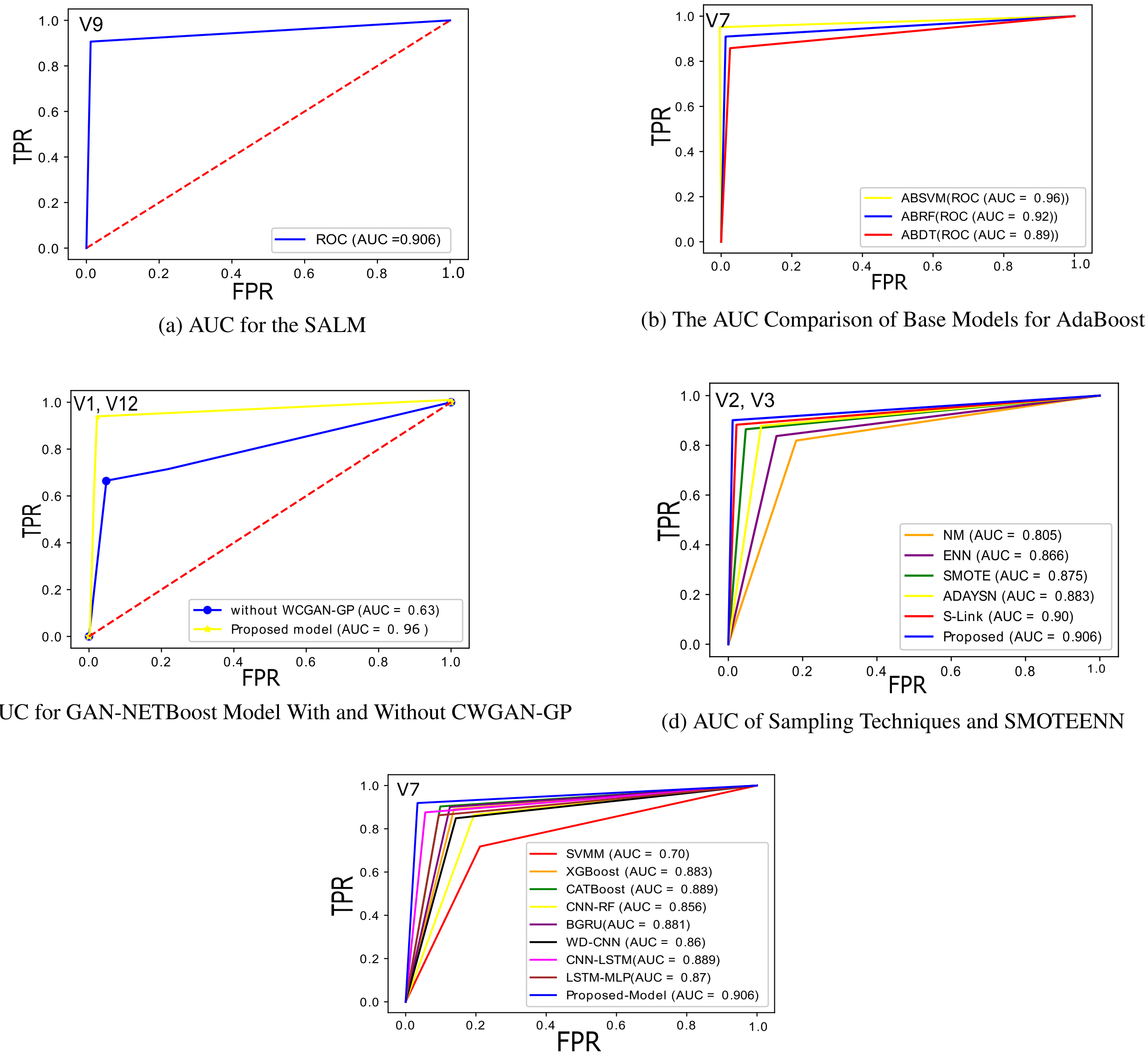
A. PERFORMANCE COMPARISON WITH SAMPLING TECHNIQUES

In this section, different sampling techniques are explored to balance the data distribution in both majority and minority classes. A comparison of undersampling, oversampling and hybrid techniques is performed for fair validation. Moreover, performance of the proposed techniques is evaluated without using any sampling technique. Undersampling and oversampling techniques help to balance the dataset. The undersampling techniques decrease the size of the majority class. Conversely, the oversampling approaches work to expand minority class to balance the dataset. Different techniques and algorithms are used to generate data, including repetition, bootstrapping or SMOTE, etc. The proposed methods works well when the quantity of data is adequate. The outstanding performance of CWGAN-GP and SMOTEENN is shown by comparing them with existing sampling algorithms discussed below.

1) UNDERSAMPLING TECHNIQUES

The following undersampling techniques are used for comparison.

- **Edited Nearest Neighbor:** it belongs to the class of undersampling techniques. The working of ENN technique is mentioned in Section IV-C1.
- **Near-Miss:** it performs undersampling by calculating distance between samples. NM-1, NM-2 and NM-3 are the variants of NM. We used final variant NM-3 with $k = 3$ in this study. NM is the most commonly used undersampling technique that shrinks the larger class to balance the dataset. In NM, near-neighbor method is used to prevent information loss. Initially, the distance between all instances of both classes, i.e., majority and minority, is measured. Then instances from majority class with small distance to the minority class are selected.



(e) ROC Curve of the Benchmarks and SALM Based on the SGCC Dataset

FIGURE 16. **ROC-AUC of both models using SGCC dataset.**

2) OVERSAMPLING TECHNIQUES

The following are the oversampling techniques that are explored for balancing the data.

- **Synthetic Minority Over Sampling Technique:** it is the most common and widely used oversampling approach. The working of this technique is mentioned in section IV-C1.
- **Adaptive Synthetic:** it is used to balance the dataset by generating synthetic samples for the minority class. In oversampling, data from the minority class is copied to balance it with data of the majority class that causes overfitting. ADASYN was proposed to overcome this problem. It is considered as an extension of SMOTE. It creates synthetic examples where the instances of minority class are less. The great advantage of ADASYN is that it does not copy the same minority data [58].

3) HYBRID TECHNIQUE

The combination of different sampling techniques is also useful for balancing the dataset. In this study, a hybrid approach is used to balance the data. The performance of the proposed technique is compared with existing hybrid techniques.

- SMOTE+Tomek-link: it is a combination of SMOTE and tomek-link.

TABLE 4. V2, V3: Performance comparison of SALM with different sampling algorithms.

Metrics	UnderSampling		OverSampling		Hybrid	
	NM	ENN	SMOTE	ADASYN	SMOTE-LINk	SMOTEENN
Precision	0.95	0.84	0.87	0.95	0.87	0.94
MCC	0.75	0.74	0.80	0.81	0.84	0.87
F1-score	0.85	0.86	0.89	0.86	0.90	0.93
Specificity	0.99	0.85	0.84	0.86	0.86	0.88
Recall	0.61	0.88	0.91	0.82	0.90	0.91
Accuracy	0.94	0.86	0.88	0.88	0.90	0.90

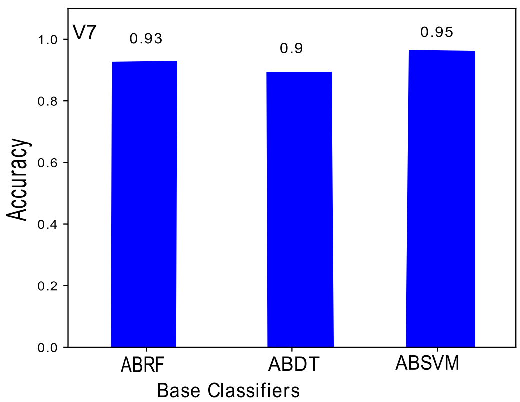


FIGURE 17. Accuracy of base models.

4) CONDITIONAL DEEP CONVOLUTIONAL GENERATIVE ADVERSARIAL NETWORK

Figure 18 represents an overview of the conditional deep convolutional GAN (CDCGAN) that is designed for data generation and results' comparison.

5) CONDITIONAL WASSERSTEIN GENERATIVE ADVERSARIAL NETWORK

CWGAN is designed for data generation to balance the data. The working of CWGAN is discussed in section IV-C2 and its architecture is presented in Figure 9. Additional information is fed into CWGAN for generating the data for the minority class.

Table 4 shows the results obtained after applying different sampling techniques. Moreover, the score achieved by the SALM model shows superior performance as compared to existing schemes. FPR is considered as an important performance measure in terms of ETD because it has a potential to increase and decrease the cost of ETD process. Thereafter, considering the importance of FPR, impact of different sampling techniques on FPR is presented in Figure 19. The figure depicts that the value of FPR for SALM is lowest, which proves that it is better than other sampling techniques. In GAN-NETBoost, CWGAN-GP is used to cope with the imbalanced nature of the dataset. The evaluation of the CWGAN-GP model is presented using reasonable performance metrics. Its comparison with existing data sampling techniques is performed that are extensively used in the literature. Figure 20 shows the AU-ROC and PR scores of the sampling algorithms. It is clear that CWGAN-GP has the highest score as compared to other sampling algorithms. Table 5 shows the values of F1-score, precision, recall, and accuracy for different sampling techniques. These values show that CWGAN-GP generates more realistic and accurate samples for the minority class, which help to boost the performance of the classifier. For a fair comparison, CWGAN-GP is used to balance the instances in both classes, and then classification is performed.

B. COMPARISON WITH EXISTING CLASSIFIERS

In this section, the comparison of SALM and GAN-NETBoost models is performed with existing classifiers using SGCC dataset with the same training ratio of 80:20 for training and testing samples. For a fair comparison, all experiments are performed on balanced dataset using SMOTEENN in SLAM and CWGAN-GP in GAN-NETBoost. Following models are used for comparison using aforementioned performance measures, as discussed in Section V.

- **Support Vector Machine:** it is famous for both regression and classification problems as it is a flexible and a powerful supervised algorithm [35], [59]. Table 6 shows the values of parameters used for simulations of SVM.
- **Extreme Gradient Boosting:** it is a boosting algorithm that supports both regression and classification. It has already proved its superior performance in the context of ETD [37]. The model is tested on publicly available SGCC dataset. Table 7 shows the parameters used for experiment of XGBoost.
- **Categorical Boosting:** it is based on gradient boosting over DT. The vectorized form of DT is the core idea behind CatBoost [37]. Table 8 shows the different parameters for CatBoost.
- **Bidirectional Gated Recurrent Unit:** it is an advanced version of GRU that was proposed by enhancing RNN to solve gradient problem. It is also used for ETD and we used same parameters that were used in [48].
- **Hybrid of Convolutional Neural Network and Random Forest:** a hybrid of CNN and RF is designed using the same architecture and parameters used in [42].

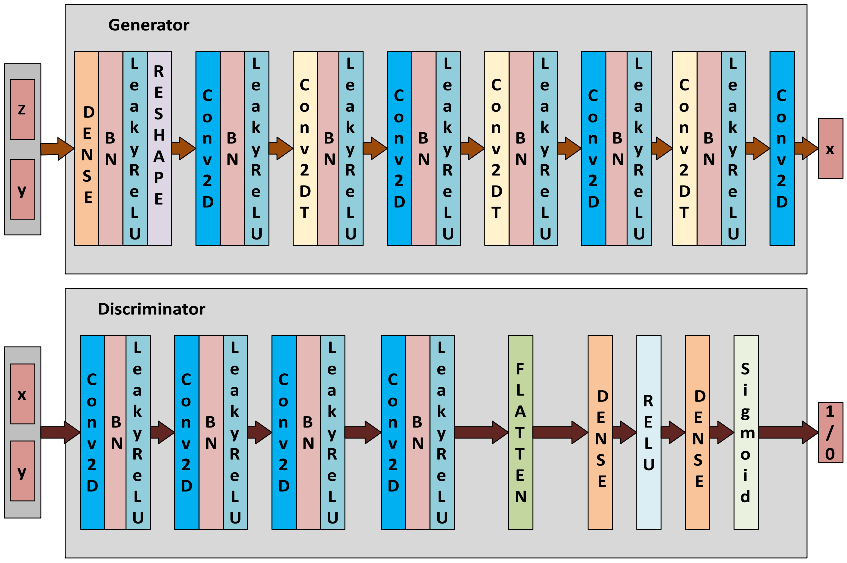


FIGURE 18. Overview of CDCGAN.

TABLE 5. V2,V3: Performance comparison of CWGAN-GP with different data sampling techniques.

Sampling Techniques	F1-score	Precision	Recall	Accuracy	MCC
NM	0.88	0.82	0.82	0.84	0.68
SMOTE	0.88	0.94	0.82	0.88	0.78
ADASYN	0.90	0.87	0.90	0.90	0.81
CDCGAN	0.92	0.93	0.92	0.91	0.90
CWGAN	0.95	0.96	0.92	0.95	0.90
CWGAN-GP	0.95	0.96	0.94	0.95	0.91

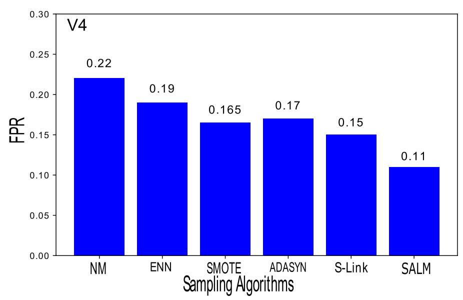


FIGURE 19. FPR of sampling techniques.

TABLE 6. Parameters for SVM.

Parameters	Values
Kernel	RBF
C	0.01

CNN captures high level features while RF performs final classification.

- **Hybrid of Multilayer Perceptron and Long Short-Term Memory:** both MLP and LSTM are extensively used for ETD [16]. A hybrid approach of these algorithms is used for ETD.

TABLE 7. Parameters for XGBoost.

Parameters	Values
No. of trees	1000
LR	0.01
Max. depth	15

TABLE 8. Parameters for CatBoost.

Parameters	Values
LR	0.1
Max. iter	2

- **Wide and Deep Convolutional Neural Network:** MLP is used as a wide component for capturing periodicity of the data. The global features are also extracted from one-dimensional data. Furthermore, a deep component captures periodicity and non-periodicity from two-dimensional data. The WD-CNN model [10] is also used for the comparison.

TABLE 9. V6, V8: Performance comparison of SALM with existing methods.

Metrics	SVM	CatBoost	XGBoost	BGRU	CNN-RF	WD-CNN	CNN-LSTM	LSTM-MLP	SALM
Precision	0.75	0.96	0.95	0.82	0.80	0.84	0.94	0.90	0.95
PR-AUC	0.78	0.88	0.87	0.78	0.87	0.81	0.87	0.90	0.93
MCC	0.67	0.81	0.81	0.68	0.84	0.73	0.78	0.80	0.87
F1-score	0.72	0.86	0.86	0.84	0.85	0.86	0.88	0.85	0.93
Recall	0.71	0.80	0.82	0.82	0.89	0.88	0.82	0.87	0.91
Accuracy	0.60	0.90	0.88	0.84	0.90	0.86	0.88	0.90	0.91

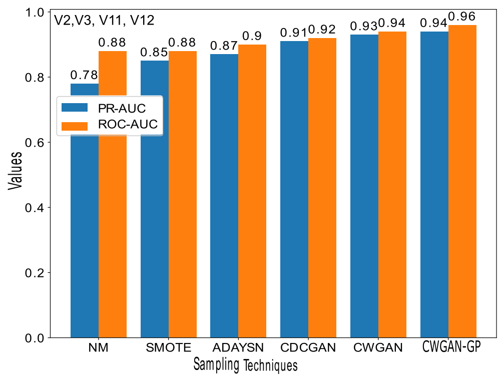


FIGURE 20. Comparison of the ROC-AUC and PR-AUC score against other sampling techniques.

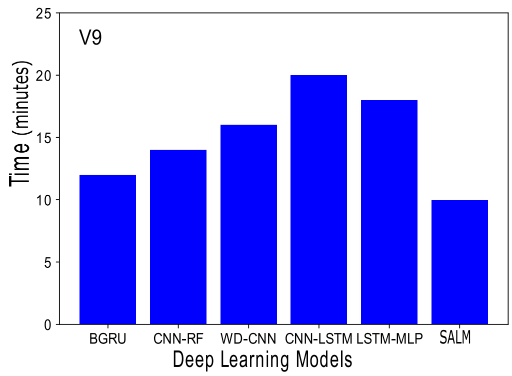


FIGURE 21. Execution time of DL models.

- **Hybrid of Convolutional Neural Network and Long Short-Term Memory:** the former is used for feature extraction; whereas, the latter is used for learning data distribution between fraudulent and honest customers. LSTM is an extensively used technique in ETD. CNN automates feature extraction while LSTM performs final classification. The model is built using the parameters in [29].

Table 9 shows the results of the comparison performed between above mentioned algorithms. In the table, the values for precision, recall, F1-score, precision recall area under the curve (PR-AUC), and MCC of existing algorithms and the proposed SALM are given. Figure 16e shows the ROC of all algorithms. SALM has obtained ROC of 0.90 which is highest as compared to the existing methods. Moreover, Figure 21 shows that its learning time is less than other DL models.

To validate the performance of GAN-NetBoost model, its comparison is performed with some existing classification techniques to identify electricity thieves using real electricity consumption dataset. The techniques include some simple and efficient ML algorithms, DL networks, and hybrid methods that are extensively used in literature. The imbalanced nature of the dataset makes ETD a difficult and a challenging task. It is considered unfavorable for the classifier's performance.

The comparison is given in Table 10. The table shows that the values for Recall, F1-score and accuracy of the GAN-NETBoost are highest, whereas the precision values of both CNN-RF and GAN-NETBoost are the same. The results prove that the GAN-NETBoost outperforms existing models for ETD. Labels V6 and V8 show the validation of the proposed solutions against L6 and L8. Different limitations are addressed in this article. Table 11 shows the mapping of limitations with their proposed solutions along with their validations. Some validations are presented in the form of tables, and some are shown with the help of figures. Labels for the respective validations are mentioned both in figures and tables. A brief description of this mapping is given in Table 11. The proposed approach is scalable as it takes only a few minutes for ETD on a large dataset (i.e., electricity consumption record of 42,372 consumers). Besides, ETD is not a time critical problem, so even if it is applied on a larger dataset, its executional time will be acceptable.

TABLE 10. V6, V8: Comparison of GAN-NETBoost against existing ETD techniques.

Methods	Precision	Recall	F1-score	Acc
SVM	0.54	0.87	0.69	0.57
CNN-RF	0.96	0.80	0.86	0.90
WD-CNN	0.92	0.83	0.87	0.84
CNN-LSTM	0.94	0.82	0.88	0.88
LSTM-MLP	0.84	0.88	0.86	0.86
GAN-NETBoost	0.96	0.94	0.95	0.95

TABLE 11. Mapping of limitations with solutions and results.

Limitations	Solutions	Validations
L1: Imbalanced dataset (number of fraudulent customers is less than number of non-fraudulent customers that leads to misclassification)	S1: Combination of oversampling and undersampling (SMOTEENN) and CWGAN-GP	V1: Tables 2 and 3 verify the performance of both models using balanced dataset. Moreover, Figures 15b and 16c show the comparison results of using balanced and imbalanced dataset.
L2: Overfitting due to duplication of data in oversampling	S2: SMOTEENN and CWGAN-GP are used to generate samples in order to balanced the dataset	V2: Figures 16d and 20, Tables 4 and 5 show the comparison between oversampling techniques and proposed solutions.
L3: Loss of information due to RUS techniques	S3: SMOTEENN and CWGAN-GP are used to balanced the dataset efficiently without losing important information	V3: Figures 16d and 20, Tables 4 and 5 show comparison between undersampling techniques and proposed solutions. The obtained results of proposed techniques are better than existing techniques.
L4: High FPR	S4: AlexNet captures relationship between daily consumption patterns and passes it to LGB, which classifies those patterns	V4: Figure 19 shows the FPR score of existing techniques and the proposed model.
L5: Generalization (bad performance on new (unseen) data)	S5: Dropout layer helps to avoid generalization error	V5: Figures 12 and 13 are the proof of generalization of the proposed SALM.
L6: Difficult to analyze large size time-series data (extraction of meaningful information)	S6: AlexNet and GoogLeNet for learning the meaningful features and patterns from massive size of data	V6: Tables 9 and 10 show the results that how correctly AlexNet and GoogLeNet learn the patterns.
L7: Accuracy in ETD	S7: LGB and AdaBoost (Boosting techniques that focus on the accuracy)	V7: Figures 16b, 16e and 17 show the achieved accuracy in terms of ETD.
L8: Trap in local optima (relative best solutions within a neighbor solution set)	S8: Adam optimizer is used to overcome the gradient vanishing problem because it dynamically change the learning rate during training of a model and achieved the global optima earlier as compared to other existing optimization algorithm	V8: Figures 15, 15b, 16, 19 and Tables 4, 9, 5 and 10 show the values against each performance measure.
L9: Lack of meaningful performance measures to evaluate the model	S9: Precision, Recall, F1-score, MCC, AU-ROC, PR-AUC, time-complexity and accuracy are the performance measures that are used in this study to evaluate the performance of the model	V9: Figures 15a, 16a, and 21 show some performance measures that are used to evaluate the model. Moreover, Tables 4 and 5 present some other measures that are utilized for evaluation of the model.
L10: Overfitting in deep learning models	S10: Dropout layer and maxpooling layer are used to overcome the issue of overfitting	V10: Figures 12 and 13 show the results on the training dataset and testing dataset. The closeness between training and testing dataset validates good performance of the model.
L11: Vanishing gradient in GAN (poor learning)	S11: Leaky ReLU and GP are used to overcome the vainsing gradient problem	V11: Figure 20 shows the performance of the CWGAN-GP and validates the better performance with the use of GP.
L12: Authenticity of data (insufficient realistic data)	S12: CWGAN-GP (GAN learns the true distribution between dataset and generates data accordingly)	V12: Table 5 shows the performance of the proposed CWGN-GP and other sampling techniques. Moreover, Figures 16c and 20 validate the authenticity of the generated data.

VII. CONCLUSION AND FUTURE WORK

ETD in an electric power system is a challenging task. It is highly crucial to avoid different issues like increased electricity cost, imbalanced supply and demand, etc. In this work, two novel models, SALM and GAN-NETBoost, are proposed for ETD using real electricity consumption data. In the first model, SMOTEENN is used to balance the dataset that prevents a classifier from misclassification. Moreover, AlexNet is used for feature extraction, and LGB is used to differentiate legal consumers from illegal consumers. The proposed model is efficient in classifying fraudulent and non-fraudulent customers. In the second model, CWGAN-GP is used to balance the data. It significantly improves the model's performance. Besides, GoogLeNet based architecture is designed to use inception modules for dimensionality reduction. Important features are extracted from the dataset and AdaBoost is used to classify the customers. The performance of AdaBoost is evaluated using different base classifiers (DT, RF, SVM). The simulation results depict that the SVM outperforms both RF and DT as a base classifier. The performance of both SALM and GAN-NETBoost is evaluated using the dataset provided by SGCC and performance metrics like ROC curve, precision, recall, MCC, and F1-score. SALM achieved precision of 0.955, recall of 0.918, MCC of 0.876 and F1-score of 0.939. Whereas, GAN-NETBoost achieved 0.968, 0.94, 0.91 and 0.95 for precision, recall, MCC and F1-score, respectively. The simulation results depict that both models outperform state-of-the-art techniques, such as SVM, BGRU, CNN-RF, SMOTE, ADASYN, CDCGAN, etc. Moreover, the performance of GAN-NETBoost is better than SALM in terms of AUC on all training ratios, i.e., 60%, 65%, 75% and 80%.

In our model, only electricity consumption data is used for ETD; however, in practical applications, the electricity consumption patterns of honest users are affected by several features, such as electricity price, weather conditions, vacations, etc. So, consideration of these features is very important. In future, the affect of these features on electricity consumption patterns will be studied.

REFERENCES

[1] M. Rasheed, N. Javaid, M. Awais, Z. Khan, U. Qasim, N. Alrajeh, Z. Iqbal, and Q. Javaid, "Real time information based energy management using customer preferences and dynamic pricing in smart homes," *Energies*, vol. 9, no. 7, p. 542, Jul. 2016, doi: 10.3390/en9070542.

[2] A. Maamar and K. Benahmed, "A hybrid model for anomalies detection in AMI system combining K-means clustering and deep neural network," *Comput., Mater. Continua*, vol. 60, no. 1, pp. 15–39, 2019.

[3] S.-C. Yip, W.-N. Tan, C. Tan, M.-T. Gan, and K. Wong, "An anomaly detection framework for identifying energy theft and defective meters in smart grids," *Int. J. Electr. Power Energy Syst.*, vol. 101, pp. 189–203, Oct. 2018.

[4] W. Li, T. Logenthiran, V.-T. Phan, and W. L. Woo, "A novel smart energy theft system (SETS) for IoT-based smart home," *IEEE Internet Things J.*, vol. 6, no. 3, pp. 5531–5539, Jun. 2019.

[5] G. Hafeez, N. Javaid, S. Iqbal, and F. Khan, "Optimal residential load scheduling under utility and rooftop photovoltaic units," *Energies*, vol. 11, no. 3, p. 611, Mar. 2018, doi: 10.3390/en11030611.

[6] A. A. Ghasemi and M. Gitizadeh, "Detection of illegal consumers using pattern classification approach combined with Levenberg–Marquardt method in smart grid," *Int. J. Electr. Power Energy Syst.*, vol. 99, pp. 363–375, Jul. 2018.

[7] S. Kazmi, N. Javaid, M. J. Mughal, M. Akbar, S. H. Ahmed, and N. Alrajeh, "Towards optimization of metaheuristic algorithms for IoT enabled smart homes targeting balanced demand and supply of energy," *IEEE Access*, vol. 7, pp. 24267–24281, 2019.

[8] P. Massaferro, J. M. D. Martino, and A. Fernandez, "Fraud detection in electric power distribution: An approach that maximizes the economic return," *IEEE Trans. Power Syst.*, vol. 35, no. 1, pp. 703–710, Jan. 2020.

[9] 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.

[10] 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.

[11] A. L. Shah, W. Mesbah, and A. T. Al-Awami, "An algorithm for accurate detection and correction of technical and nontechnical losses using smart metering," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 11, pp. 8809–8820, Nov. 2020, doi: 10.1109/TIM.2020.2999175.

[12] 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.

[13] J. L. Viegas, P. R. Esteves, and S. M. Vieira, "Clustering-based novelty detection for identification of non-technical losses," *Int. J. Electr. Power Energy Syst.*, vol. 101, pp. 301–310, Oct. 2018.

[14] 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.

[15] 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.

[16] 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.

[17] K. M. Ghori, R. A. Abbasi, M. Awais, M. Imran, A. Ullah, and L. Szathmary, "Performance analysis of different types of machine learning classifiers for non-technical loss detection," *IEEE Access*, vol. 8, pp. 16033–16048, 2020.

[18] S. Aslam, N. Javaid, F. Khan, A. Alamri, A. Almogren, and W. Abdul, "Towards efficient energy management and power trading in a residential area via integrating a grid-connected microgrid," *Sustainability*, vol. 10, no. 4, p. 1245, Apr. 2018, doi: 10.3390/su10041245.

[19] A. Mahmood, N. Javaid, M. A. Khan, and S. Razzaq, "An overview of load management techniques in smart grid," *Int. J. Energy Res.*, vol. 39, no. 11, pp. 1437–1450, Sep. 2015.

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

[21] J. Y. Kim, Y. M. Hwang, Y. G. Sun, I. Sim, D. I. Kim, and X. Wang, "Detection for non-technical loss by smart energy theft with intermediate monitor meter in smart grid," *IEEE Access*, vol. 7, pp. 129043–129053, 2019.

[22] 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.

[23] S. Zahoor, S. Javaid, N. Javaid, M. Ashraf, F. Ishmanov, and M. Afzal, "Cloud–fog–based smart grid model for efficient resource management," *Sustainability*, vol. 10, no. 6, p. 2079, Jun. 2018, doi: 10.3390/su10062079.

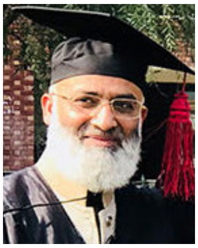
[24] A. Khan, N. Javaid, and M. I. Khan, "Time and device based priority induced comfort management in smart home within the consumer budget limitation," *Sustain. Cities Soc.*, vol. 41, pp. 538–555, Aug. 2018.

[25] P. Glauner, J. A. Meira, P. Valtchev, R. State, and F. Bettinger, "The challenge of non-technical loss detection using artificial intelligence: A survey," 2016, *arXiv:1606.00626*. [Online]. Available: http://arxiv.org/abs/1606.00626

[26] 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.
[27] 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.
[28] Z. Qu, H. Li, Y. Wang, J. Zhang, A. Abu-Siada, and Y. Yao, "Detection of electricity theft behavior based on improved synthetic minority oversampling technique and random forest classifier," *Energies*, vol. 13, no. 8, p. 2039, Apr. 2020, doi: 10.3390/en13082039.
[29] 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, doi: 10.3390/en12173310.
[30] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
[31] M. Ismail, M. Shahin, M. F. Shaaban, E. Serpedin, and K. Qaraqe, "Efficient detection of electricity theft cyber attacks in AMI networks," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Apr. 2018, pp. 1–6.
[32] H. Zenati, C. S. Foo, B. Lecouat, G. Manek, and V. R. Chandrasekhar, "Efficient GAN-based anomaly detection," 2018, *arXiv:1802.06222*. [Online]. Available: http://arxiv.org/abs/1802.06222
[33] Z. Zheng, L. Zheng, and Y. Yang, "Unlabeled samples generated by GAN improve the person re-identification baseline *in vitro*," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 3754–3762.
[34] 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.*, early access, Jun. 4, 2020, doi: 10.1109/TNNLS.2020.2994116.
[35] M. Anwar, N. Javaid, A. Khalid, M. Imran, and M. Shoaib, "Electricity theft detection using pipeline in machine learning," in *Proc. Int. Wireless Commun. Mobile Comput. (IWCMC)*, Jun. 2020, pp. 2138–2142.
[36] N. Javaid, G. Hafeez, S. Iqbal, N. Alrajeh, M. S. Alabed, and M. Guizani, "Energy efficient integration of renewable energy sources in the smart grid for demand side management," *IEEE Access*, vol. 6, pp. 77077–77096, 2018.
[37] J. R. Aguero, "Improving the efficiency of power distribution systems through technical and non-technical losses reduction," in *Proc. PES T&D*, May 2012, pp. 1–8.
[38] R. Jiang, R. Lu, Y. Wang, J. Luo, C. Shen, and X. Shen, "Energy-theft detection issues for advanced metering infrastructure in smart grid," *Tsinghua Sci. Technol.*, vol. 19, no. 2, pp. 105–120, Apr. 2014.
[39] N. Javaid, S. Hussain, I. Ullah, M. Noor, W. Abdul, A. Almogren, and A. Alamri, "Demand side management in nearly zero energy buildings using heuristic optimizations," *Energies*, vol. 10, no. 8, p. 1131, Aug. 2017, doi: 10.3390/en10081131.
[40] R. Khalid, N. Javaid, M. H. Rahim, S. Aslam, and A. Sher, "Fuzzy energy management controller and scheduler for smart homes," *Sustain. Comput., Informat. Syst.*, vol. 21, pp. 103–118, Mar. 2019.
[41] A. Jindal, A. Dua, K. Kaur, M. Singh, N. Kumar, and S. Mishra, "Decision tree and SVM-based data analytics for theft detection in smart grid," *IEEE Trans. Ind. Informat.*, vol. 12, no. 3, pp. 1005–1016, Jun. 2016.
[42] 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, doi: 10.1155/2019/4136874.
[43] 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.
[44] A. Ahmad, N. Javaid, N. Alrajeh, Z. Khan, U. Qasim, and A. Khan, "A modified feature selection and artificial neural network-based day-ahead load forecasting model for a smart grid," *Appl. Sci.*, vol. 5, no. 4, pp. 1756–1772, Dec. 2015.
[45] M. B. Rasheed, N. Javaid, A. Ahmad, M. Awais, Z. A. Khan, U. Qasim, and N. Alrajeh, "Priority and delay constrained demand side management in real-time price environment with renewable energy source," *Int. J. Energy Res.*, vol. 40, no. 14, pp. 2002–2021, Nov. 2016.
[46] V. B. Krishna, C. A. Gunter, and W. H. Sanders, "Evaluating detectors on optimal attack vectors that enable electricity theft and DER fraud," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 4, pp. 790–805, Aug. 2018.
[47] 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.
[48] 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, doi: 10.3390/app10093151.
[49] PR Newswire. (2014). *World Loses $89.3 Billion to Electricity Theft Annually, $58.7 Billion in Emerging Markets*. Accessed: Feb. 10, 2020. [Online]. Available: http://www.prnewswire.com/news-releases/world-loses-893-billion-to-electricity-theft-annually-587-billion-in-emerging-markets- 300006515.html
[50] X. Yang, W. Zhou, N. Shu, and H. Zhang, "A fast and efficient local outlier detection in data streams," in *Proc. Int. Conf. Image, Video Signal Process.*, Feb. 2019, pp. 111–116.
[51] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.
[52] M. Mirza and S. Osindero, "Conditional generative adversarial nets," 2014, *arXiv:1411.1784*. [Online]. Available: http://arxiv.org/abs/1411.1784
[53] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein GAN," 2017, *arXiv:1701.07875*. [Online]. Available: http://arxiv.org/abs/1701.07875
[54] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved training of Wasserstein GANs," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5767–5777.
[55] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.
[56] R. E. Schapire, "Explaining AdaBoost," in *Empirical Inference*. Berlin, Germany: Springer, 2013, pp. 37–52.
[57] S. Boughorbel, F. Jarray, and M. El-Anbari, "Optimal classifier for imbalanced data using matthews correlation coefficient metric," *PLoS ONE*, vol. 12, no. 6, Jun. 2017, Art. no. e0177678, doi: 10.1371/journal.pone.0177678.
[58] H. He, Y. Bai, E. A. Garcia, and S. Li, "ADASYN: Adaptive synthetic sampling approach for imbalanced learning," in *Proc. IEEE Int. Joint Conf. Neural Netw., IEEE World Congr. Comput. Intell.*, Jun. 2008, pp. 1322–1328.
[59] M. Zahid, F. Ahmed, N. Javaid, R. Abbasi, H. Z. Kazmi, A. Javaid, M. Bilal, M. Akbar, and M. Ilahi, "Electricity price and load forecasting using enhanced convolutional neural network and enhanced support vector regression in smart grids," *Electronics*, vol. 8, no. 2, p. 122, Jan. 2019, doi: 10.3390/electronics8020122.

ABDULAZIZ ALDEGHEISHEM received the bachelor's degree in planning and urban design from the College of Architecture and Planning, King Saud University, the master's degree in city planning from the University of Pennsylvania, Philadelphia, USA, in 2001, and the Ph.D. degree in urban planning and spatial information from the University of Illinois in Urbana–Champaign, USA, in 2006. He is also the Dean of the College of Architecture and Planning, King Saud University, and an Associate Professor with the Department of Urban Planning, College of Architecture and Planning, KSU. He has worked as the Head of the Urban Planning Department, in 2012. He also works as an Adviser with Vision Realization Office (VRO) at the university, and he is also the Supervisor of the Traffic Safety Technologies Chair. He has worked as an Adviser in a number of government agencies and supervised many projects and specialized studies. He is interested in the role of spatial information in urban planning and management. His research interests include urban planning, spatial management, and smart city technologies. He also obtained the Certificate in urban and regional planning studies from MIT, Cambridge, MA, USA.

MUBBASHRA ANWAR received the master's degree from the Communications Over Sensors (ComSens) Research Laboratory, Department of Computer Science, COMSATS University Islamabad, Islamabad Campus, under the supervision of Dr. Nadeem Javaid. Her research interests include data science, optimization, security and privacy, energy trading, and smart grids.



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 in computer science from the University of Paris-Est, France, in 2010. He is currently an Associate Professor and the Founding Director of the Communications Over Sensors (ComSens) Research Laboratory, Department of Computer Science, COMSATS University Islamabad, Islamabad Campus. He has supervised 126 master's and 20 Ph.D. theses. He has authored more than 900 papers in technical journals and international conferences. His research interests include energy optimization in smart grids and in wireless sensor networks using data analytics and blockchain. He was a recipient of the Best University Teacher Award from the Higher Education Commission of Pakistan in 2016 and the Research Productivity Award from the Pakistan Council for Science and Technology in 2017. He is also an Associate Editor of IEEE ACCESS and an Editor of the *Sustainable Cities and Society* journal.

NABIL ALRAJEH received the Ph.D. degree in biomedical informatics engineering from Vanderbilt University, USA. He has worked as a Senior Advisor for the Ministry of Higher Education. He is currently a Professor of health informatics with King Saud University and the Rector of Prince Mugrin Bin Abdulaziz University. His role was implementing development programs, including educational affairs, strategic planning, and research and innovation. He is also a Board Member of several private universities in Saudi Arabia.

MUHAMMAD SHAFIQ received the master's degree in information technology (IT) from the University of the Punjab, Gujranwala, Pakistan, in 2006, the M.S. degree in computer science from the University Institute of Information Technology, Arid Agriculture University, Rawalpindi, Pakistan, in 2010, and the Ph.D. degree in information and communication engineering from Yeungnam University, South Korea, in February 2018. He was a Faculty Member with the Faculty of Computing and IT, University of Gujrat, Gujrat, Pakistan, from 2010 to 2014, and formerly held the same position at the Department of Computer Science and IT, Federal Urdu University, Islamabad, Pakistan. His research interests include the Internet of Things (IoT); cognitive radio-based IoT networks-architecture and design; mobile ad hoc networks; wireless sensor networks, performance, management, and security; 5G cellular networks, admission control, and mobility management; device-to-device communications; medium access control protocols; the Internet routing protocols; spectrum trading and auctions; information systems, design, and access control; and human–computer interaction.

HASAN AHMED received the Ph.D. degree in computer vision from the Electrical Engineering Department, Manchester University, U.K. He is currently an Associate Professor with the School of Computing and Communication Systems, Lancaster University, U.K. He has taught courses in computing and communication systems at graduate and postgraduate levels. He held postdoctoral fellowship in automatic target recognition and real-time video enhancement research at Manchester University, in 1998. He has published more than 50 research articles and holds a U.S. patent. He has supervised six Ph.D. and more than ten M.Sc. students. His research interests include computer vision, artificial intelligence and decision-making processes, digital communication systems, machine learning, computer vision, and deep learning.