

Received April 28, 2021, accepted June 22, 2021, date of publication June 25, 2021, date of current version July 19, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3092645

Using GANCNN and ERNET for Detection of Non Technical Losses to Secure Smart Grids

NADEEM JAVAID¹, (Senior Member, IEEE), HIRA GUL¹,
SOBIA BAIG², (Senior Member, IEEE), FAISAL SHEHZAD¹,
CHENGJUN XIA³, (Senior Member, IEEE),
LIN GUAN³, (Member, IEEE), AND TANZEELA SULTANA¹

¹Department of Computer Science, COMSATS University Islamabad, Islamabad 44000, Pakistan

²Energy Research Center, COMSATS University Islamabad, Lahore 54000, Pakistan

³School of Electric Power, South China University of Technology, Guangzhou 510640, China

Corresponding authors: Nadeem Javaid (nadeemjavaidqau@gmail.com) and Chengjun Xia (cjxia@scut.edu.cn)

This work was supported in part by the Pakistan Science Foundation under Project PSF/NFSC -II/Eng/P-COMSATS-Lhr(08) and in part by the COMSATS University Isb, Pakistan.

ABSTRACT In this paper, two supervised learning models based solutions are proposed for Electricity Theft Detection (ETD). In the first solution, Adaptive Synthetic Edited Nearest Neighbor (ADASYNENN) is used to solve class imbalanced problem. For feature extraction, Locally Linear Embedding (LLE) technique is utilized. Moreover, Self-Attention Generative Adversarial Network (SAGAN) is used in combination with Convolutional Neural Network (CNN) for the classification of electricity consumers. In the second solution, Synthetic Minority Oversampling Technique Edited Nearest Neighbor (SMOTEENN) is proposed. Moreover, a novel classification model, named as ERNET, which is based on EfficientNet, Residual Network (ResNet) and Gated Recurrent Unit (GRU), is used to detect Non-Technical Losses (NTLs). We also used a Sparse Auto Encoder (SAE) for effective feature extraction that makes the classification more robust and easy. Furthermore, a robust Root Mean Square Propagation (RMSProp) optimizer is used to improve the learning rate of the model. To validate the proposed models, simulations are performed using different performance metrics, such as precision, recall, F1-score, Area Under the Curve (AUC), FPR and Root Mean Square Error (RMSE). All simulations are performed using State Grid Corporation of China (SGCC) dataset. The proposed models are compared with benchmark models, such as SAGAN, Wide and Deep Convolutional Neural Network (WDCNN), CNN and Long Short Term Memory (LSTM). The simulation results prove that the proposed models outperform the existing models in terms of the aforementioned performance metrics.

INDEX TERMS Electricity theft detection, gated recurrent unit, non-technical losses, random undersampling, random oversampling, smart grid, SGCC.

NOMENCLATURE

AdaGrad	Adaptive Gradient Algorithm
ADASYNENN	Adaptive Synthetic Edited Nearest Neighbor
ANN	Artificial Neural Network
AUC	Area Under the Curve
a	Input Layer
b	Bias
CDF	Cumulative Distribution Function
CNN	Convolutional Neural Network

CatBoost	Category Boosting
CNN-LSTM	Convolutional Neural Network-Long Short Term Memory
CNN-RF	Convolutional Neural Network-Random Forest
d	Degree of imbalance ratio
d_s	Preset threshold value, which decides the tolerable value of class imbalance ratio
ECNN	Enhanced Convolutional Neural Network
ELR	Enhanced Logistic Regression
ENN	Edited Nearest Neighbor
ERELM	Enhanced Recurrent Extreme Learning Machine

The associate editor coordinating the review of this manuscript and approving it for publication was Yunfeng Wen¹.

ESVR	Enhanced Support Vector Regression
FPR	False Positive Rate
GMM	Gaussian Mixture Model
GPU	Graphical Processing Unit
g_t	Gradient at time t
GRU	Gated Recurrent Unit
G	Number of minority class samples that need to be generated
HABACO	Hybrid Artificial Bee Ant Colony Optimization
h	Hidden layer
IoT's	Internet of Things
KNN	K Nearest Neighbor
LightBoost	Light Gradient Boosting Machine
LLE	Locally Linear Embedding
LR	Logistic Regression
LR	Linear Regression
MARA	Multivariate Auto Regressive Algorithm
MAE	Mean Absolute Error
MSE	Mean Square Error
MLP	Multi Layer Perceptron
MLP-LSTM	Multi Layer Perceptron-Long Short Term Memory
MODWPT	Maximal Overlap Discrete Wavelet Packet Transform
NTLs	Non-Technical Losses
RF	Random Forest
RMSProp	Root Mean Square Propagation
RMSE	Root Mean Square Error
ROC	Receiver Operating Characteristic
ROS	Random Oversampling
RUS	Random Undersampling
SAE	Sparse Auto Encoder
SAGAN	Self-Attention Generative Adversarial Networks
SGD	Stochastic Gradient Descent
S_i	Minority class samples
S_j	Majority class samples
SGCC	State Grid Corporation of China
SMOTE	Synthetic Minority Oversampling Technique
SMOTEENN	Synthetic Minority Oversampling Technique Edited Nearest Neighbor
SVM	Support Vector Machine
P	Probability
$Q1$	First Quartile
$Q2$	Second Quartile
$Q3$	Third Quartile
TLs	Technical Losses
TPR	True Positive Rate
v_t	Exponential average of square gradients
WDCNN-RF	Wide and Deep Convolutional Neural Network-Random Forest
XGBoost	eXtreme Gradient Boosting
xz_i	A randomly chosen sample from minority group i

x	Consumer's observation
x_i	Daily electricity consumption of a consumer
\hat{y}	Predicted output
y	Actual output
z	Input sample
Z	A randomly selected input from data
λ	Random number between 0 and 1
σ	Sigmoid function
η	Learning rate
ρ	Exponential average of last updated value
ϵ	To avoid ending with zero

I. INTRODUCTION

With the increase in the number of residential homes and industries, the demand of energy increases manifolds. Therefore, power generation companies need to generate more electricity [1]. Moreover, there should be a balance between electricity generation and consumption to mitigate the issue of energy shortage [2]. Due to the latest advancements in Advanced Metering Infrastructure (AMI), traditional grids are converted into smart grids where data is collected through smart meters. The balance between demand and supply is also established using bi-directional flow of energy and information [3]. In energy transmission systems, two types of losses occur, which are known as Technical Losses (TLs) and Non-Technical Losses (NTLs). The former losses occur due to poor infrastructure and energy dissipation. Whereas, the latter losses are defined as the difference between total electricity transmitted through distribution lines and the electricity consumed by the users. Due to the NTLs, power utilities face losses worth millions of dollars, which highly affect the country's economy [4]. The manual inspection of these losses is both time consuming and expensive [5], [6].

There are different reasons for the occurrence of NTLs, which are broadly categorized in two categories: human and non-human. The former includes tampering the meter readings, hooking with the main lines, etc. Whereas, the latter includes errors in smart meters, fluctuating energy flow, meter inaccuracies, etc., [6]. With the NTLs, other losses also occur, such as unbearable load on electrical systems, load shedding, economical loss, etc., [7]. With the use of smart meters, flow of both energy and information becomes automated. For the utility companies, the smart meters remotely provide data related to readings of electricity consumption on real time basis. Therefore, it becomes easy to steal the electricity by manipulating the electricity consumption data [8], [9].

To handle NTLs, several solutions have been proposed in the literature [10], [11]. These solutions are broadly categorized into three types: hardware based solutions, game theory based solutions and data-driven based solutions. Hardware based solutions focus on designing smart devices and sensors to detect electricity thieves. These solutions require hardware equipment and devices that are expensive and involve high maintenance cost. Moreover, these devices are less efficient and more time consuming [12]. On the other hand, in game

theory based solutions, there exists a game between players, i.e., utility and electricity users [13]. Both entities try to maximize their benefits. However, these solutions are based on assumptions that can be inappropriate if wrong utility function is formulated. Moreover, these solutions have less accuracy and high False Positive Rate (FPR) [14]. In contrast, many data-driven based solutions are adopted in literature for NTL detection [15]. These solutions include techniques that are based on artificial intelligence and machine learning. These techniques perform pattern analysis on the electricity consumption data without requiring additional hardware cost and human resources. Furthermore, the data-driven based solutions are more robust, efficient and easy to understand as compared to hardware based and game theory based solutions [16]. However, in the data-driven based solutions, there are some problems, which are needed to be addressed. The problems are class imbalanced problem, low accuracy and high FPR. Therefore, this study is carried out keeping in view the data-driven solutions.

The remainder of this paper is organized as follows. Section II discusses the work done in the literature on the detection and prevention of NTLs and also highlights the limitations. Whereas, the proposed solutions are elaborated in Section III. Simulation results along with their discussion are presented in Section IV. Conclusion and future work are given in Section V.

II. STATE-OF-THE-ART METHODOLOGIES

In smart grids, anomaly is defined as the deviation from regular or normal electricity consumption patterns. It occurs due to many factors like arrival of more family members at home, occurrence of a special occasion, illegal use of electricity, etc. In anomaly detection, data-driven models are used that learn the normal patterns and detect the abnormal patterns to identify the electricity thieves.

Maamar *et al.* [17] have proposed a hybrid technique, which is based on k-means clustering and deep neural network for anomaly detection. However, users have to select the value of k (number of clusters) at the start, which is not suitable in a dynamic environment. Also, the imbalance class problem is not resolved. Yip *et al.* [18] have proposed a novel technique, named as loss factor and error term, to detect anomaly in smart meter's data. Loss function is used to calculate NTLs; whereas, error term is used to detect noise in transmission and distribution lines. However, the class imbalanced problem is not addressed. In a recent study [19], Cheng *et al.* have proposed an auto encoder technique for anomaly detection in electricity consumption data. Auto encoder is used to learn the patterns in an unsupervised manner while discarding the noise. However, the class imbalanced problem is not addressed and overfitting is not solved, especially when there is no sufficient level of diversity in the data. Giuseppe *et al.* have proposed a concept of drift aware based approach for the detection of anomaly from electricity consumption patterns. The authors have used Long Short Term Memory (LSTM) to capture the periodicity of normal consumers based on

their previous consumption history [20]. However, the class imbalanced problem is not solved. The LSTM method may require high memory bandwidth to feed its computational units. Ding *et al.* [21] have proposed a hybrid model, which is based on Gaussian Mixture Model (GMM) and LSTM for the detection of real time anomaly. However, it is difficult to determine for certain the number of clusters to be created. Also, the class imbalanced problem is not tackled. Authors in [22] have proposed Jaya-LSTM for the forecasting of electricity load. All of the above mentioned methods perform better in terms of anomaly detection. However, the methods are not feasible enough to accurately detect electricity fraudsters. Zheng [23] have proposed Wide and Deep Convolutional Neural Network (WDCNN) for ETD. They have used State Grid Corporation of China (SGCC) dataset, which consists of verified electricity thieves. However, the class imbalanced problem is not addressed.

Zheng *et al.* [24] have proposed a hybrid technique, which is a combination of maximum information coefficient and clustering technique, to find density peaks for the detection of electricity thieves. However, it is tedious to generate clusters from local densities of data points that are randomly distributed. Thus, it is difficult for the cluster heads to be selected. Moreover, the class imbalanced problem is not solved. Li *et al.* [25] have performed ETD for Internet of Things (IoTs) enabled smart homes. The method is not suitable for solving the future changes in electricity consumption data as it depends on the past data. Moreover, the class imbalanced problem is not addressed. Fleury *et al.* [26] have proposed genetic programming algorithm for theft detection. Data is collected from more than 4000 consumers for experiments. The authors have focused on feature engineering rather than classification. However, it is difficult to tune the parameters of genetic programming algorithm and also selecting wrong number of clusters may affect the accuracy of the algorithm. Micheli *et al.* [27] have proposed multiple linear regression model for the detection of NTL. However, the model is not efficient for real life scenario as the relationship between covariates and response variables may not be linear. In another study, Coma-Puig *et al.* [28] have implemented and compared three machine learning techniques: eXtreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LightBoost) and Category Boosting (CatBoost) for NTL detection. However, the techniques are not suitable for smaller datasets as the methods may create overfitting problem. Viegas *et al.* [29] have proposed fuzzy clustering model for the identification of electricity thieves. They have used Gustafson-Kessel fuzzy algorithm and have attained 63% True Positive Rate (TPR). Buzau *et al.* [30] have applied several machine learning models, such as K Nearest Neighbor (KNN), Linear Regression (LR) and Support Vector Machine (SVM) for NTLs detection. However, there is a problem of selecting k in KNN, and SVM is not suitable for large dataset.

Moreover, electricity theft is a crucial problem for utility companies, as they have to bear huge losses every year. Many

TABLE 1. Analysis of different NTL detection models and their performance measures.

Proposed model	Dataset used	Evaluation measures	Accuracy
Detection of suspicious users by using fuzzy logic [18]	Customer's bill data	Z-shaped and S-shaped member function	N-A
Hybrid of Gaussian mixture and LSTM [21]	Numenta Anomaly Benchmark (NAB) and synthetic dataset	Mean Square Error (MSE), F1-score, recall and precision	N-A
MODWPT [31]	Households' data of Honduras	Matthews correlation coefficient, specificity, accuracy, F1-score, AUC and recall	94%
Anomaly detection system [32]	Consumption data of Tanzania	Accuracy, cross validation and F1-score	87%
Binary black hole algorithm [33]	Electricity consumption dataset of Brazil	Recognition rate and time complexity	N-A
Artificial Neural Network (ANN) based customer filtering and classification and regression tree [34]	Endesa's dataset	Classification, regression and self-organizing map	N-A
Semi-supervised embedded system for NTL [35]	SGCC	Accuracy, FPR, precision, recall and F1-score	95%
Hybrid of MLP and LSTM [36]	Endesa's smart meter data	PR, AUC and precision	N-A
Kullback leibler divergence detector for optimal attack [37]	Electricity consumption data of Ireland	AUC and FPR	N-A
Theft detection system based on entropy [38]	Dataset is taken from Irish Social Science Data Archive	FPR and detection rate	N-A
ECNN and ESVR [39]	ISO-NE	RMSE, Mean Absolute Error (MAE) and MSE	N-A
LSTM and RUSBoost [40]	SGCC	Precision, recall, F1-score and ROC	87%
ELR and ERELM [41]	UCI	RMSE, MAE, MSE	85.51%
MARA [42]	DAYTOWN and EKPC	Mean absolute percentage error and variance	98.76%
HABACO [43]	N-A	Execution time	N-A

data-driven based solutions are proposed in the literature for ETD. However, there exists some limitations in these solutions, which are needed to be addressed. Li *et al.* [44] have proposed a hybrid model, which consists of CNN and Random Forest (RF) for ETD. However, the computational complexity of RF is very high as it takes more time to construct decision trees. In addition, FPR is also not calculated. Hasan *et al.* [45] have proposed a hybrid technique by combining CNN and LSTM for ETD. The proposed technique efficiently performed in terms of accuracy. However, LSTM requires a lot of memory for storing long-term sequences. Moreover, LSTM is not hardware friendly because it needs more resources as compared to CNN and Gated Recurrent Unit (GRU). In [45], authors have used Synthetic Minority Oversampling Technique (SMOTE) to balance the data for training CNN and LSTM models to perform classification. However, SMOTE generates synthetic data samples, which cause overfitting problem. In [31], authors have proposed Random Undersampling (RUS) with adaptive boosting method to solve the problem of imbalanced data. They have proposed Maximal Overlap Discrete Wavelet Packet Transform (MODWPT) for classification. However, RUS leads to loss of relevant samples that are important for the training of machine learning model and prevent it from underfitting problem. Authors in [23] have proposed WDCNN model for the detection of electricity thieves. However, authors do not provide a mechanism to handle the class imbalanced problem, which leads to misclassification. Furthermore, the model has high FPR value, which leads to high inspection cost. Table 1 shows the analysis of some of the existing techniques and their performance measures while Table 2 presents the advantages and disadvantages of existing techniques.

III. PROPOSED MODELS

To overcome the issues identified from the literature, we propose two deep learning models in this work: GANCNN and ERNET. The former is the combination of Self-Attention Generative Adversarial Network (SAGAN) and CNN. Whereas, the latter is a hybrid of EfficientNet, Residual Network (ResNet) and GRU. In the GANCNN model, data sampling and feature extraction are done using Adaptive Synthetic Edited Nearest Neighbor (ADASYNENN) and Locally Linear Embedding (LLE), respectively. In the ERNET model, GRU is applied for the classification of honest and dishonest consumers. GRU does not require separate memory cell and excessive parameters to train the model.

A. PROPOSED GANCNN MODEL

The proposed GANCNN model comprises of five steps: data collection, data pre-processing, data sampling, feature extraction and classification. The proposed model is illustrated in Figure 1. The flowchart of the proposed model is presented in Figure 2. Moreover, the steps and flowchart are discussed in different sections.

1) DATA COLLECTION

The data used in this work is acquired from SGCC [46]. It is a labeled dataset with a known number of electricity thieves. It consists of customers' ID, daily consumption and flagged (i.e., target attribute) either as 0 or 1. Daily electricity consumption data from January 2014 to October 2016 is considered. The consumption records of 42,372 consumers are present in the dataset. Out of these, 3,615 are electricity thieves and remaining 38,733 are normal consumers [23]. Table 3 shows the detail of SGCC dataset.

TABLE 2. The advantages and disadvantages of existing techniques.

Ref.	Advantages	Disadvantages
[17]	A robust hybrid approach to detect anomaly in the behavior of electricity consumers using the k-means clustering mechanism and DNN algorithm	Users have to select the value of k (number of clusters) at the start, which is not suitable in a dynamic environment. The class imbalanced problem is also not resolved
[18]	Anomaly detection framework based on loss factor and error term to detect NTLs	Class imbalanced problem is not addressed
[19]	An autoencoder based ensemble method for anomaly detection in electricity consumption data	Class imbalanced problem is not addressed and overfitting is not solved, especially when there is no sufficient level of diversity in the data
[20]	A LSTM method for anomaly detection in electricity consumption data	Class imbalanced problem is not solved. The LSTM method may require high memory bandwidth to feed the computational units
[21]	A hybrid system based on GMM and LSTM for detecting real time anomaly in electricity consumption data	It is difficult to determine for certain the number of clusters to be created. Class imbalanced problem is not tackled
[22]	A Jaya-LSTM is used for detecting anomaly in electricity consumption data	Class imbalanced problem is not addressed
[23]	A WDCNN method is used for ETD	Class imbalanced problem is not tackled
[24]	A maximum information coefficient, and clustering technique by fast search and find of density peaks are used for ETD	It is tedious to generate clusters from local densities of data points that are randomly distributed. Thus, it is difficult for the cluster heads to be selected. Class imbalanced problem is not solved
[25]	A simple moving average method is used for ETD	Class imbalanced problem is not addressed. The method is not suitable for solving the future changes in electricity consumption data as it depends on the past data
[26]	A gradient boosting machine algorithm, mixed model clustering and genetic programming algorithm, are used for ETD	It is difficult to tune the parameters of genetic programming algorithm and also selecting wrong number of clusters may affect the accuracy of the results
[27]	A multiple linear regression model is used for ETD	The model is not efficient in real life scenario as the relationship between covariates and response variables may not be linear
[28]	Three machines learning techniques, such as XGBoost, LightBoost and CatBoost, are used for ETD	The techniques are not suitable for smaller dataset as the methods may create overfitting problem
[30]	The methods of KNN, LR and SVM are used for ETD	There is a problem of selecting k in KNN and SVM is not suitable for large dataset

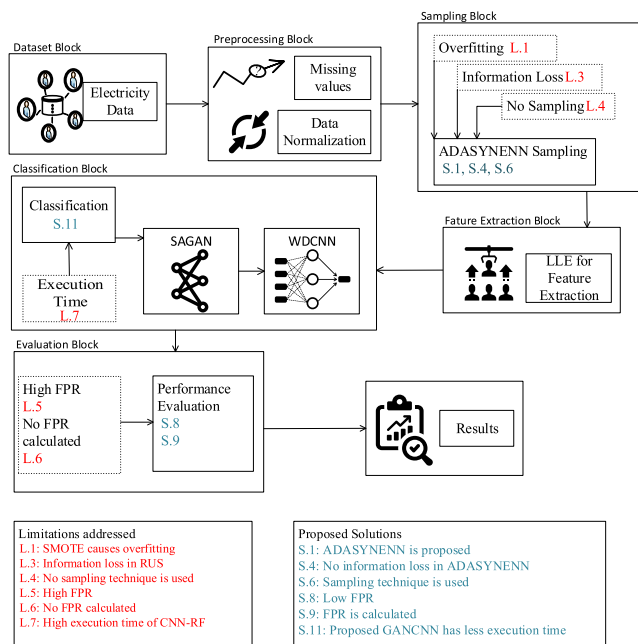


FIGURE 1. Detailed view of the proposed GANCNN model.

2) DATA PRE-PROCESSING

The dataset consists of missing values due to the faulty meters, unscheduled repairing of electricity equipment, data tampering, etc. In this research, we use imputation method for handling the missing values [47]. The working of imputation method is shown in Figure 3. Initially, it checks the

TABLE 3. The detail of SGCC dataset.

Total number of observations	42,372
Normal consumers	38,373
Fraudulent consumers	3,615
Duration	01-01-2014 to 31-10-2016

current value in the dataset. If the value is not null, it remains unchanged; otherwise, the algorithm checks the neighboring values to fill the missing value. If the neighboring values are null, then they are replaced with zeroes. In case, if previous and next values are not null, it takes the mean of both values and replaces the current null value with the mean value. Here, one thing to be noted is that even if the neighboring values are null and replaced with zeroes, the overall pre-processing method is not affected negatively. It is because the data distribution remains the same. To normalize the data, we use *Min-Max* normalization method in this research, as neural networks are sensitive to diverse data. The method is applied using following Equation 1 [23].

$$f(x_i) = \frac{x_i - \min(x)}{\max(x) - \min(x)}, \tag{1}$$

x_i denotes daily electricity consumption of a consumer. Where $\max(x)$ and $\min(x)$ show maximum and minimum values of a consumer’s consumption, respectively.

3) DATA SAMPLING

SGCC dataset is imbalanced in nature, as shown in Figure 4a. If the data in both majority and minority classes is not

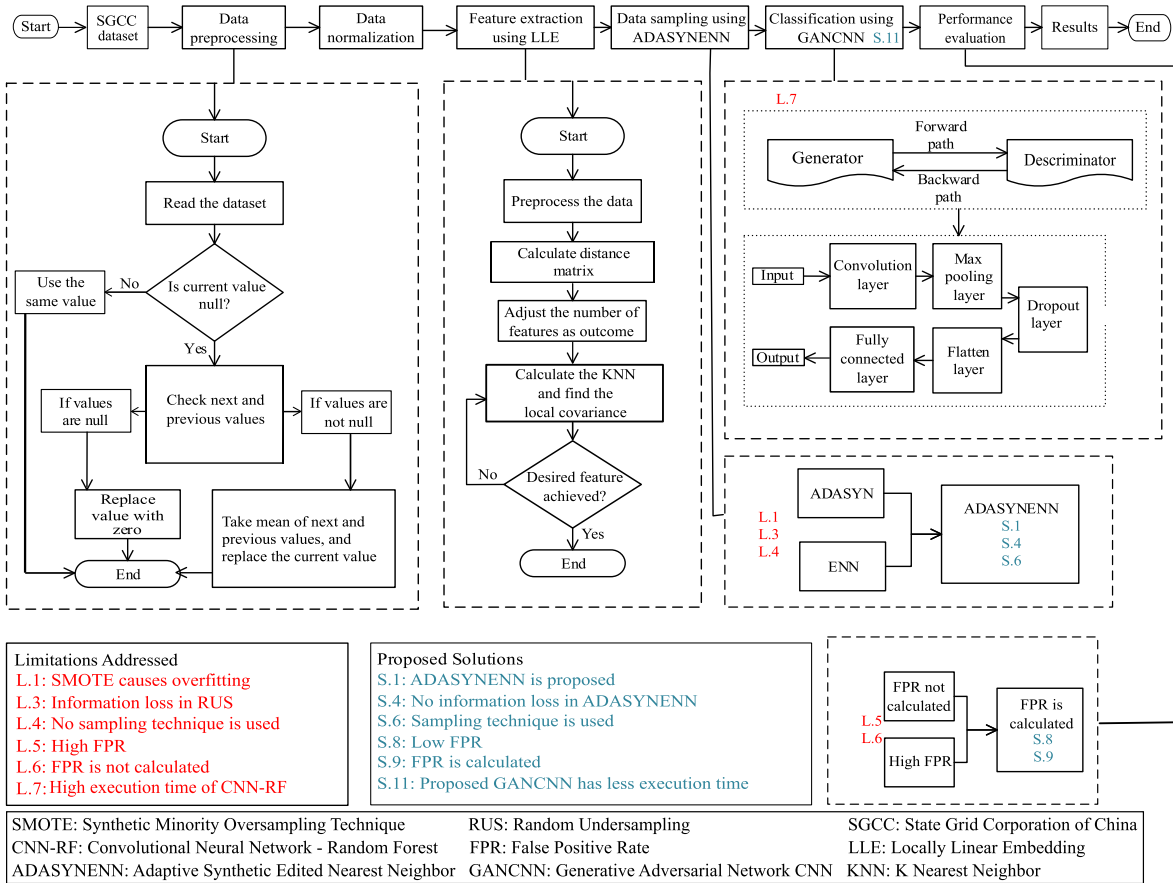


FIGURE 2. Detailed flow diagram of the proposed GANCNN model.

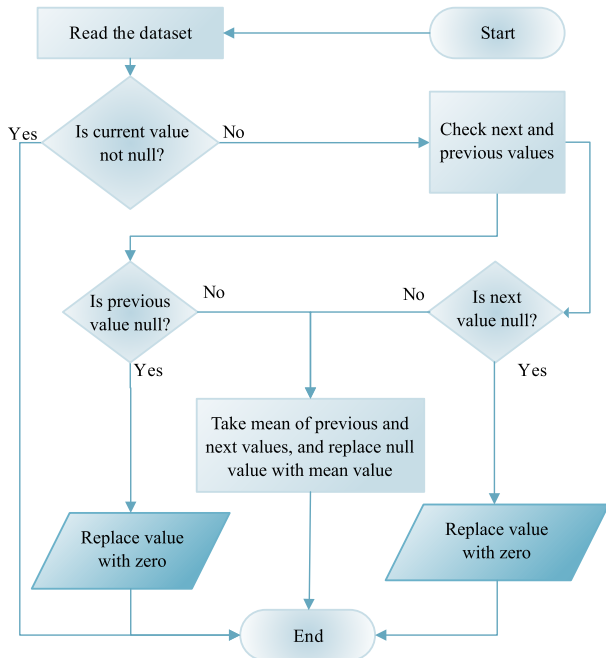


FIGURE 3. Flow chart of data cleaning process.

balanced, the model is biased towards the majority class. The biasness leads to the reduction in classification accuracy,

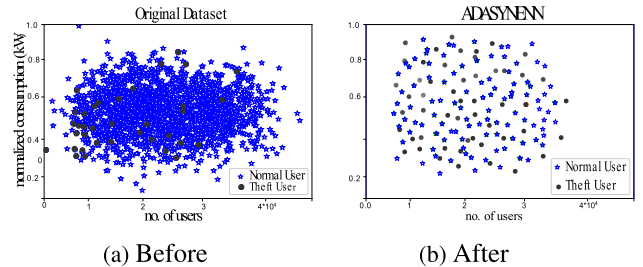


FIGURE 4. Apply sampling technique.

which further leads to performance degradation. In our scenario, normal consumers (i.e., majority class) are more in number as compared to electricity thieves (i.e., minority class). Therefore, it is important to balance the data before passing it to the classification model. If the model is trained on the normal consumers, it becomes ineffective and shows less accuracy in detecting the electricity thieves. In existing studies, many data sampling techniques have been used to handle imbalanced data. The techniques are broadly categorized into two types: oversampling and undersampling. The former is performed when the number of samples of one class are less and they have to be increased. Whereas, the latter is the opposite of it and involves removing the samples, which are more in number. Both oversampling and undersampling

are performed to remove the class imbalanced issue and have their respective advantages and disadvantages. In this work, the hybrid technique, named as ADASYNENN, is used for data sampling. It is a combination of ADASYN (oversampling) and ENN (undersampling) techniques, as shown in Figure 4b. ADASYN is used to increase minority samples and make them equal to majority samples [48]. Whereas, ENN is used to reduce the majority samples and make them equal to the minority samples. The following steps are involved in ADASYN.

- Calculate the number of minority and majority class samples represented as S_j and S_l , respectively.
- Measure the degree of imbalance ratio d using Equation 2.

$$d = S_j - S_l. \quad (2)$$

- Equation 3 is used for calculating the number of samples required to be generated G . The minority samples should be equal to the majority samples for handling class imbalance issue.

$$G = S_l - S_j. \quad (3)$$

- Find KNNs of each minority class sample. After this step, each minority sample is associated with a specific group.
- Give more importance to those minority class samples, which have more majority class samples in their specific groups.
- If $d < d_s$, new minority class samples are generated. (d_s is the preset threshold value, which decides the tolerable value of the class imbalanced ratio).
- A new sample is generated using Equation 4.

$$S_n = x_i + (x_i - x_{z_i}) * \lambda. \quad (4)$$

S_n is a newly generated sample. x_i and x_{z_i} are randomly chosen from minority group i . λ is any random number between 0 and 1.

Moving ahead, ENN is an undersampling technique used to remove majority class samples [49]. In this method, majority class samples are removed, which are near to borderline of minority class samples. Figures 4a and 4b show the dataset before and after applying ADASYNENN sampling technique, respectively.

4) FEATURE EXTRACTION

The efficiency of a classification model depends upon the successful execution of the feature extraction process in which the significant and most important features are extracted. In the literature, different methods have been developed for performing feature extraction. In the first proposed model of the underlying work, LLE is used for extracting useful information from the dataset [50]. The required number of features are extracted using KNN and covariance between features. The closely related features are extracted from the feature space, which are then used for the classification purpose. A detailed stepwise feature extraction process is shown

in Figure 6. It is used for non-linear data and is based on manifold technique, which extracts the important features from the feature space in an iterative manner.

5) CLASSIFICATION

For the classification of electricity thieves and normal consumers, a hybrid GANCNN model is proposed, which is a combination of SAGAN and WDCNN. SAGAN is a deep learning model and is considered as the best training model. It has two modules: generator and discriminator [51]. The former creates synthetic data similar to original data by selecting random input samples from the dataset. The latter discriminates between fake and original data [52]. During GAN's process, both generator and discriminator modules are trained until discriminator is failed half of the time to distinguish between fake and original samples, which means that generator is successful in creating fake samples. The random input samples are selected on the basis of inverse transform technique in which Cumulative Distribution Function (CDF) is used. CDF is given in Equation 5

$$CDF = f_Z(z) = P(Z \leq z) \quad (5)$$

where P is the probability, Z is randomly selected input from data and z is the input sample. The architecture of SAGAN is shown in Figure 5.

The output of SAGAN is passed as an input to deep CNN, which comprises of many layers: convolutional layer, max pooling layer, dropout layer, fully connected layer and flatten layer [44]. Where the convolutional layer overcomes the limitations of traditional neural networks by connecting a neuron to its neighboring neurons, known as receptive field. Convolutional operation is performed on input samples and convoluted feature maps are sent to the max pooling layer, which selects optimal features by reducing high dimensionality. The dropout layer is used to prevent the model from overfitting problem. The optimal features selected in the max pooling layer are used by the fully connected layer to change the dimensions of the vector. The flatten layer connects the input layers to the output layers. The architecture of CNN is shown in Figure 7.

The network structure of the proposed GANCNN model is given in Table 4 while Table 5 provides the hyperparameter of the model. In this study, GAN is used as the front end to process the input data while CNN is used as back end to process the non-linear features of the abstracted data. The GANCNN model has 13 layers, which alternate between convolution and maxpooling layers.

6) ADAM OPTIMIZER

In the literature, many optimization techniques are presented for tuning hyperparameters of the deep learning models. In [53], authors have analyzed different algorithms used for hyperparameters' tuning, which include genetic algorithm, cross validation, simulated annealing, etc. In this work, we use Adam optimizer to tune the hyperparameters of CNN, because it is considered as the best optimizer. It requires

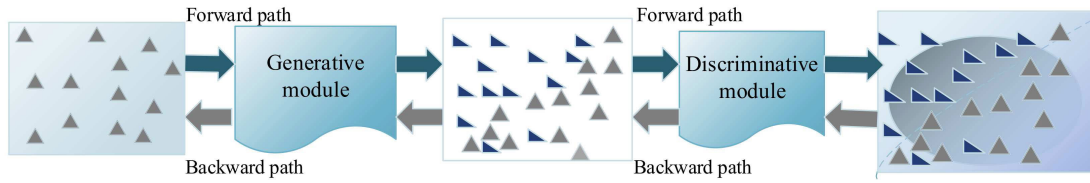


FIGURE 5. Architecture of SAGAN.

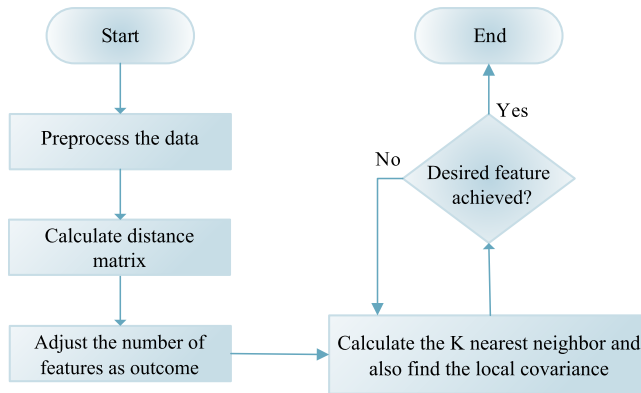


FIGURE 6. Flowchart of LLE technique for feature extraction.

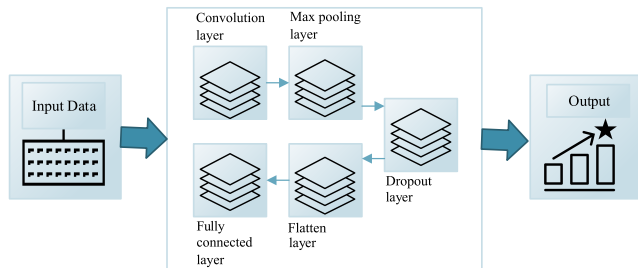


FIGURE 7. Architecture of CNN.

less memory, less execution time and is also efficient for large datasets. It possesses the properties of both Adaptive Gradient Algorithm (AdaGrad) and Root Mean Square Propagation (RMSProp) optimizers, which give the best output [54], [55]. It also calculates the learning rate for each parameter. Moreover, it is also easy to implement as compared to other optimization techniques.

B. PROPOSED ERNET MODEL

The second model proposed in this work for ETD is comprised of five stages, as shown in Figure 8. The stages are same as defined in the GANCNN model. The SGCC dataset is used for this model as well. For dimensionality reduction, Sparse Auto Encoder (SAE) technique is used as a feature extractor. For data sampling, SMOTEENN is proposed. A hybrid of EfficientNet, ResNet and GRU, named as ERNET, is proposed for classification of theft and normal consumers. A detailed flowchart of the ERNET model is shown in Figure 9.

TABLE 4. Network structure of the proposed GANCNN model.

Layer	Type	Number of neurons	Size of kernel	Stride
0	SGCC input	42,372 × 1	-	-
1	Convolution	42,122 × 5	251	1
2	Max pooling	21,065 × 5	2	2
3	Convolution	21,040 × 5	150	1
4	Max pooling	10,520 × 5	2	2
5	Convolution	10,420 × 10	101	1
6	Max pooling	5210 × 10	2	2
7	Convolution	5130 × 20	81	1
8	Max pooling	2565 × 20	2	2
9	Convolution	2500 × 20	66	1
10	Max pooling	1250 × 20	2	2
11	Convolution	1150 × 10	101	1
12	Max pooling	575 × 10	2	2
13	Dense	6	-	-

TABLE 5. Hyperparameter of the proposed GANCNN.

Dropout	0.3
Activation function	Sigmoid
Learning rate	0.01
Number of epochs	300
Learning rate for discriminator	0.0004
Learning rate for generator	0.0001

1) DATA PRE-PROCESSING

SGCC data is collected through smart meters, which are installed across the country. It contains missing or erroneous values. Therefore, we have applied same pre-processing steps, which are performed in the first proposed model. Linear interpolation method [56] is used to fill the missing values, which improves the training and performance of the proposed model.

Robust scaler normalization method [57] is used for normalizing the data. It is similar to *Min-Max* normalization. The only difference is that it uses quartile ranges and is considered more robust for normalization. Formula of robust scaler is given in Equation (6)

$$RobustScaler = \frac{x_i - Q2(x)}{Q3(x) - Q1(x)} \tag{6}$$

x is divided into three parts: Q_1 , Q_2 and Q_3 that represent first, second and third quartile ranges, respectively.

2) DATA SAMPLING

As discussed earlier, a hybrid of two sampling techniques is used in this model for data sampling, known as SMOTEENN.

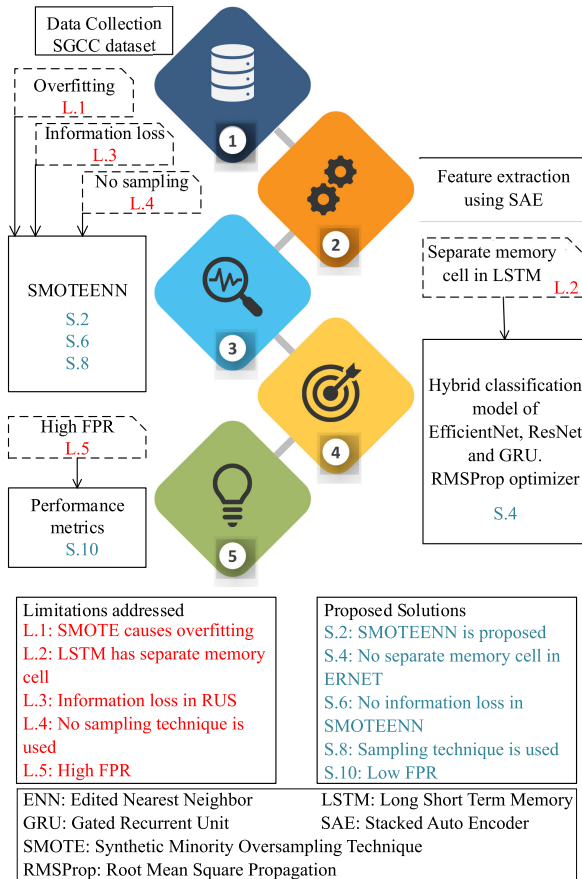


FIGURE 8. Abstract view of the proposed ERNET model.

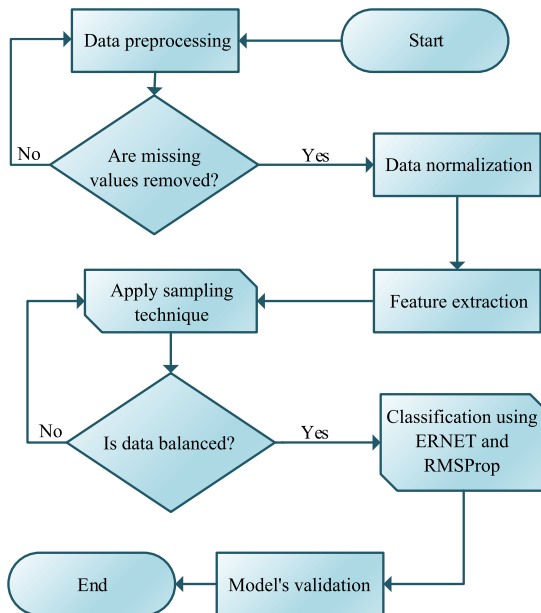


FIGURE 9. Flowchart of ERNET model.

Existing sampling techniques have some limitations. For example, undersampling technique discards important information by randomly removing samples. Whereas, in oversampling technique, overfitting problem arises due to the duplication of samples, which further leads to poor

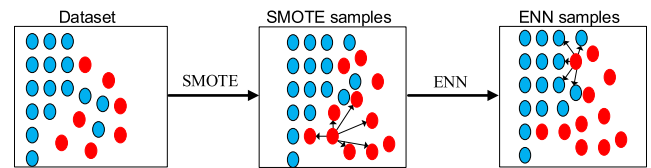


FIGURE 10. SMOTEENN sampling technique.

generalization and misclassification. To resolve the aforementioned issues, SMOTEENN is used in the proposed work. The working of SMOTEENN is presented in Algorithm 1. It can be seen from the algorithm that SMOTEENN comprises of two sampling techniques: SMOTE (lines 1-9) and ENN (lines 10-14). The algorithm is presented to give a better understanding to the readers that how SMOTEENN would work. The technique starts with the oversampling of the minority class using SMOTE, which is an enhanced version of Random Oversampling (ROS). In SMOTE, new samples are generated by calculating the KNNs. If the data is highly imbalanced, it generates new samples of minority class equal in number to the majority class. The newly generated samples create overhead and cause overfitting problem. Afterwards, the majority class samples are removed using ENN, which also solves the aforementioned problems. SMOTEENN is used to efficiently solve the class imbalanced issue, which is not efficiently solved by SMOTE and ENN individually. An abstract view of SMOTEENN is shown in Figure 10.

Algorithm 1 SMOTEENN Algorithm

Input: Dataset D

Output: Balance dataset D

Theft sample is represented as T

LOOP Process

- 1: **for** each a in T **do**
- 2: Search the KNNs in T
- 3: Obtain b stochastically from k samples
- 4: $difference = a - b$
- 5: $gap =$ random number between 0 and 1
- 6: $n = a + difference * gap$
- 7: Addition of $n + D$
- 8: **end for**
- 9: **End SMOTE**
- LOOP Process
- 10: **for** each a_i in D **do**
- 11: Remove a_i from D only if misclassified using KNN with prototype
- 12: $D - a_i$
- 13: **end for**
- 14: **End ENN**
- 15: **return** D

3) FEATURE EXTRACTION

In the feature extraction module, SAE is used to improve the performance of the training model, decrease the loss value

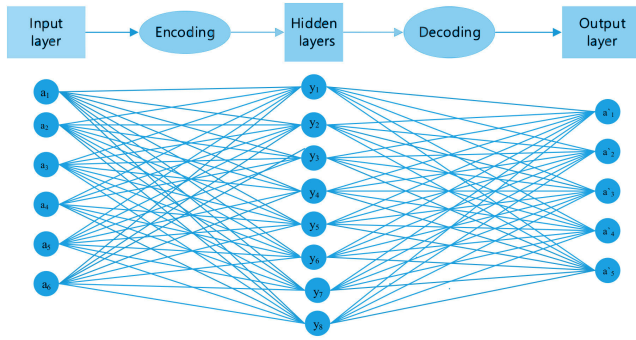


FIGURE 11. SAE structure.

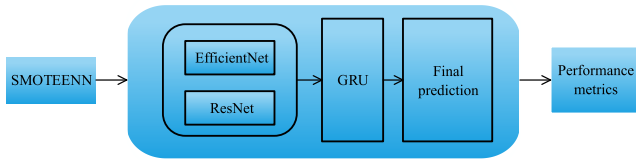


FIGURE 12. Architecture of the proposed ERNET model.

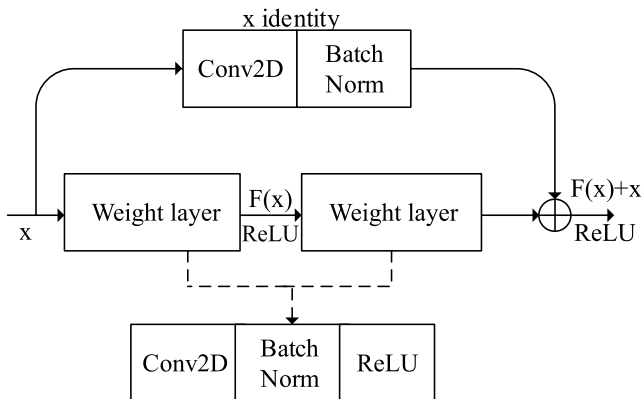


FIGURE 13. Residual block.

and reduce the risk of overfitting. The data dimensionality is reduced by creating new features from the existing dataset. The number of newly created features is less than the existing features. SAE has been applied in many research fields like audio processing, subtitle processing, human body image detection, natural language processing, etc., for extracting optimal features [58]. In this method, input features are encoded before passing them to the hidden layer. After processing in hidden layer, the features are decoded at the output layer. As shown in Figure 11, features passed to the output layer are less in number as compared to those passed to the input layer. Encoding and decoding functions are defined using Equations (7a) and (7b), respectively [58].

$$h = \sigma(Wa + b) \quad (7a)$$

$$a_i = \sigma(Wh + b) \quad (7b)$$

where a is used for input layer, h is hidden layer, W is used to connect parameters among two layers and b is denoted as the

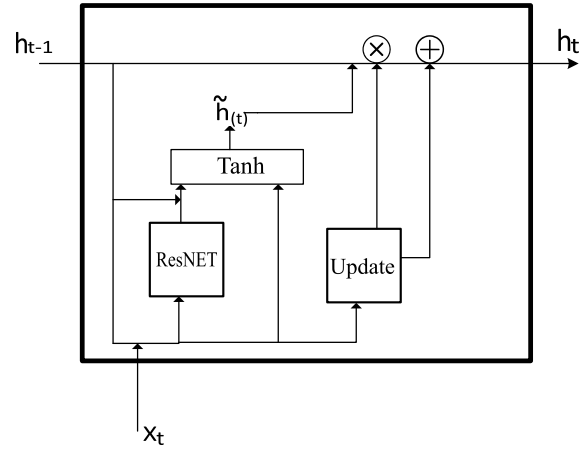


FIGURE 14. GRU block.

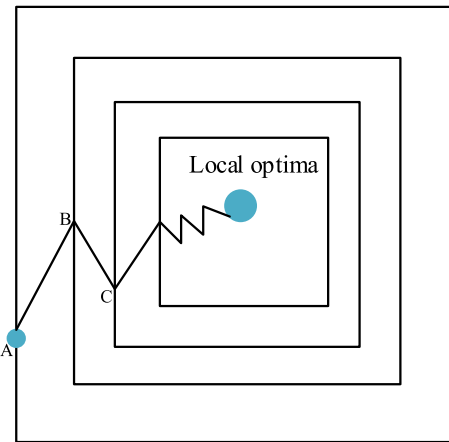


FIGURE 15. RMSprop.

bias value. For non-linear mapping, we use sigmoid function (σ), which is calculated using Equation (8)

$$\sigma(\hat{y}) = \frac{1}{1 + e^{-a_i}} \quad (8)$$

Sparsity penalty is an important step in SAE. It is calculated by adjusting the loss function, given in Equation (9).

$$Obj = L(y, \hat{y}) + regularization + \lambda \sum_i |a_i^{(h)}| \quad (9)$$

y and \hat{y} represent the actual and predicted outputs, respectively. Loss function's value is reduced by adjusting W , b and number of hidden layers. Whereas, overfitting is avoided using regularization.

4) ERNET BASED CLASSIFICATION MODEL

The proposed classification model consists of three sub-models including EfficientNet, ResNet and GRU, as shown in Figure 12. ResNet is an advanced variant of CNN that uses skip connections after each layer, as shown in Figure 13. Moreover, additional weight matrix is used in the model to learn the skip connections. The purpose of skip connections is to prevent gradient vanishing problem

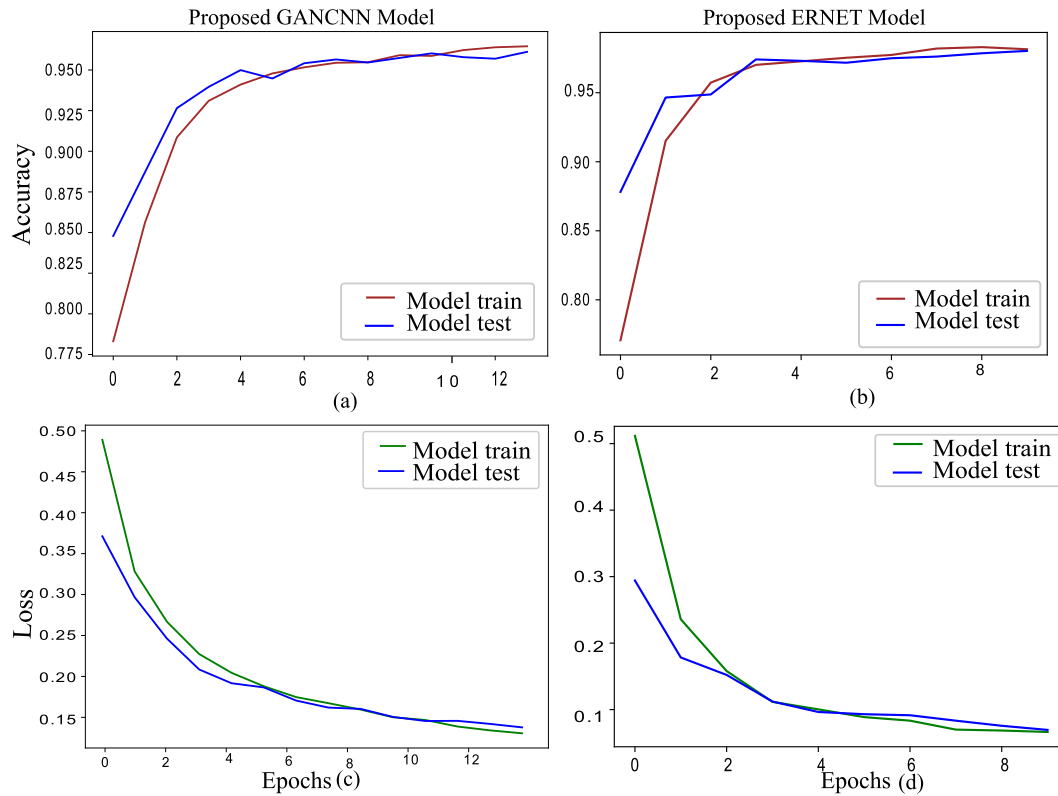


FIGURE 16. Accuracy and loss of the GANCNN and ERNET models.

TABLE 6. Hyperparameter of the proposed ERNET model.

Batch size	32
Number of iterations	100
Adam optimizer	0.001

by reusing activation functions from previous layer until the adjacent layer learns its weights. These skip connections help in simplifying the model as it takes only few layers at the start of the training phase. The model restores the skip connections when it learns all feature space. EfficientNet is combined with ResNet to improve effectiveness of the entire ERNET model. To prevent the loss of information from Rectified Linear Unit (ReLU) activation function, the linear activation function is used in the last layer of each block. The final classification is made by using GRU model, which is an advanced version of RNN, as shown in Figure 14. GRU solves the problem of vanishing gradient by learning long-term dependencies. It has two gates: update gate and reset gate. Update gate is used for long-term dependencies while reset gate is used for short-term dependencies. GRU requires less data for high generalization, needs few tuning parameters and has fast training process. Table 6 shows the hyperparameter of the proposed ERNET model.

5) RMSprop OPTIMIZER

The basic idea of RMSprop is same as that of gradient descent. The only difference is of momentum. RMSprop works in a perpendicular position and its convergence speed

is faster because it moves in a vertical direction [59], [60]. Due to this, the learning rate of model is increased. The main advantage of RMSprop is that it chooses several learning rates for each parameter. The scenario of local optima is illustrated in Figure 15. As shown in figure, the model starts moving from unit A and after one loop of gradient descent, it reaches unit B in the next square. In the next loop, the model reaches at point C that is more closer to the local optima.

Formula of RMSprop optimizer is given in Equations (10)-(12). For each parameter W_i

$$v_t = \rho v_{t-1} + (1 - \rho) * g_t^2 \tag{10}$$

$$\delta W = -\frac{\eta}{\sqrt{v_t + \epsilon}} * g_t \tag{11}$$

$$W_{t+1} = W_t + \delta W_t \tag{12}$$

- ρ : exponential average of last updated value,
- η : initial learning rate,
- g_t : gradient at time t ,
- v_t : exponential average of square gradients,
- ϵ : to avoid ending with zero.

IV. SIMULATION RESULTS

In this section, simulations are performed to validate the performance of the proposed models. Google Colaboratory tool is used to perform simulations and it also provides free access to Graphical Processing Unit (GPU) for data processing and storage [61]. This online tool is very helpful for performing high computational tasks.

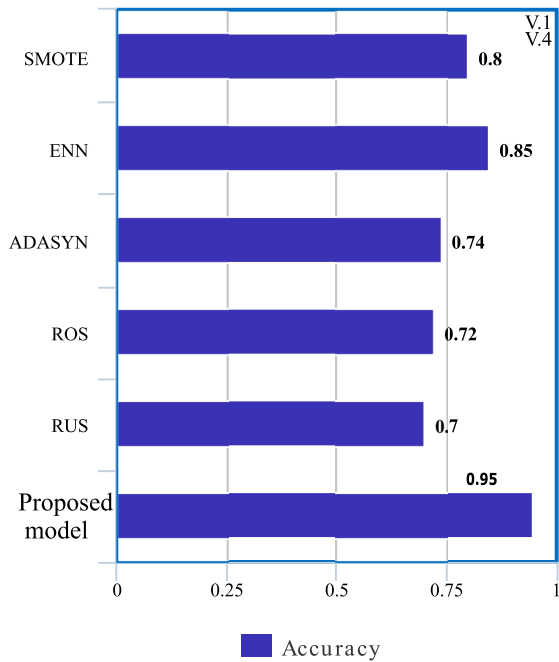


FIGURE 17. Training accuracy of the proposed model and benchmark sampling techniques.

A. ACCURACY AND LOSS VALUES OF THE PROPOSED GANCNN AND ERNET MODELS

In this section, accuracy and loss of both GANCNN and ERNET are calculated. Moreover, the models are compared with each other and existing models. To calculate the accuracy and loss of the proposed GANCNN model, dataset is divided into training and testing sets as 75% and 25%, respectively. The performance metrics and splitting criteria are set as in [62]. Early stopping method is applied to stop the training process when learning rate of the model does not improve. We measure the accuracy of GANCNN and ERNET models to check the closeness of the measured values with the known values. The accuracy of GANCNN is 0.95, as shown in Figure 16. Whereas, the accuracy of ERNET model is 0.98, which means it is accurately trained on the given data. Accuracy and loss are inversely proportional to each other. The training and testing loss of GANCNN and ERNET models are also calculated. In GANCNN, the training loss starts decreasing from 0.48 to 0.15; whereas, the testing loss starts decreasing from 0.37 to 0.16, as shown in Figure 16. On the other hand, the training and testing loss for the ERNET model is also shown in Figure 16. The loss value of training starts decreasing from approximately 0.5 and reaches below 0.1. Whereas, the loss value of testing decreases from 0.3 to 0.1.

B. COMPARISON OF ADASYNENN WITH EXISTING SAMPLING TECHNIQUES

In this section, we compare the proposed ADASYNENN with the existing sampling techniques, such as SMOTE, ENN, ADASYN, ROS and RUS in terms of accuracy and loss. In Figure 17, accuracy of the ADASYNENN is compared

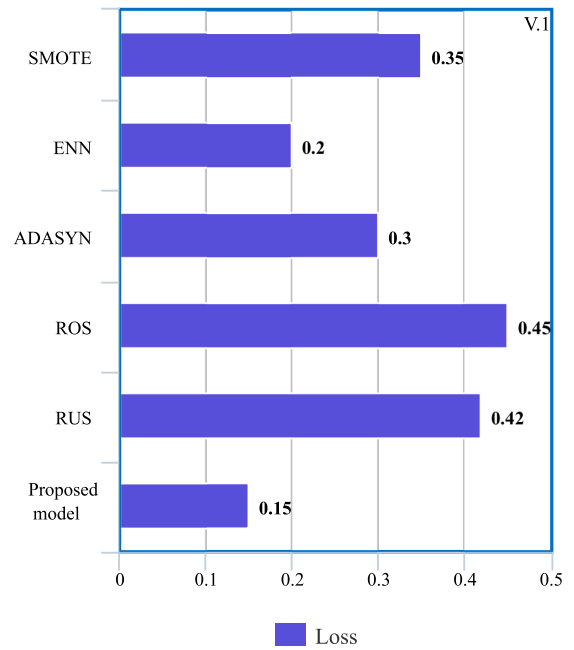


FIGURE 18. Training loss of the proposed model and benchmark sampling techniques.

with the above mentioned benchmark techniques. The results show that ADASYNENN achieves the highest accuracy, i.e., 0.95. Whereas, RUS achieves 0.7 accuracy, which is the lowest among all. The figure is labeled as V.1 and V.4 and it validates the solutions S.1 and S.4. In Figure 18, ADASYNENN is compared with the existing sampling techniques in terms of loss. As shown in the figure, ADASYNENN has the lowest loss of 0.15. Whereas, ROS has the highest loss as compared to other techniques. Moreover, the figure is labeled as V.1 and it validates the solution S.1.

C. ACCURACY OF GANCNN WITH DIFFERENT SAMPLING TECHNIQUES

In this section, we measure the accuracy of GANCNN model by applying different sampling techniques and is shown in Figure 19. The sampling techniques include ROS, RUS, ENN and SMOTE. By applying ROS, the training accuracy of GANCNN is approximately 0.75 and its testing accuracy is 0.76. The training and testing accuracy of GANCNN with RUS is 0.75 and 0.77, respectively. Then, SMOTE is applied with the proposed model that yields 0.78 and 0.76 accuracy on training and testing data, respectively. Moreover, ENN technique is also applied on the proposed model, which achieves 0.78 training and 0.77 testing accuracy. Ultimately, the proposed GANCNN model achieves the highest accuracy by applying the proposed SMOTEENN technique. The accuracy value for both training and testing is 0.95. From the simulation results, it is concluded that SMOTEENN outperforms ROS, RUS, SMOTE and ENN in terms of accuracy. Moreover, Figure 19 is labeled with V.2 and V.6, which validate the solutions S.2 and S.6, respectively.

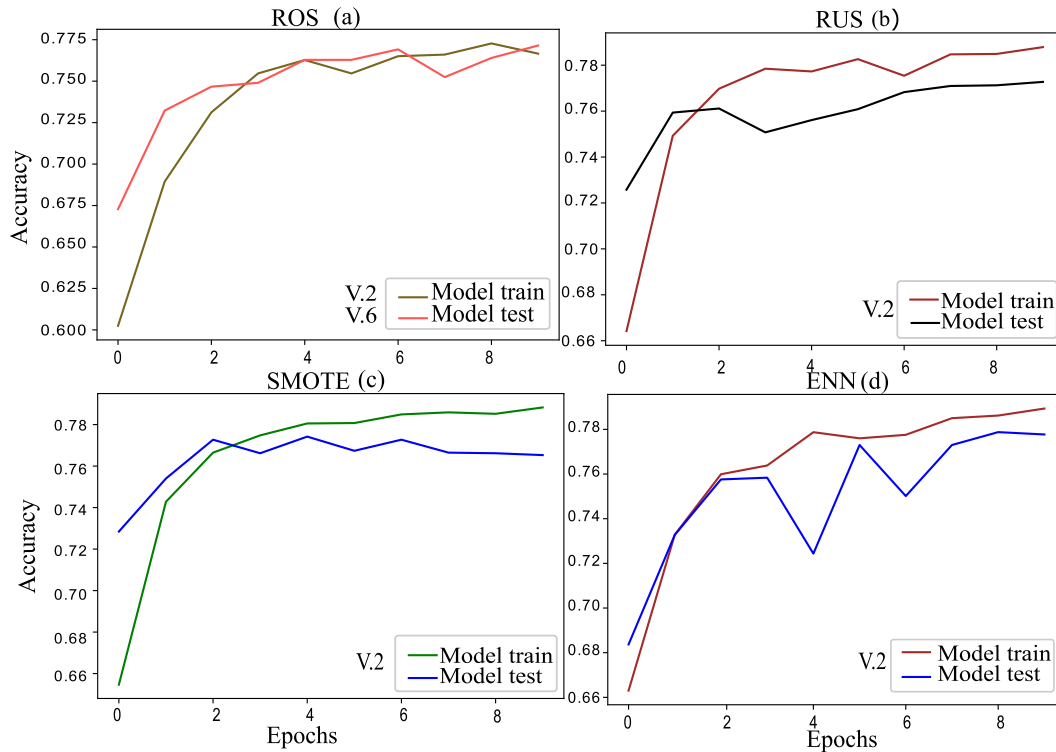


FIGURE 19. Comparison of accuracy of ROS, RUS, SMOTE and ENN.

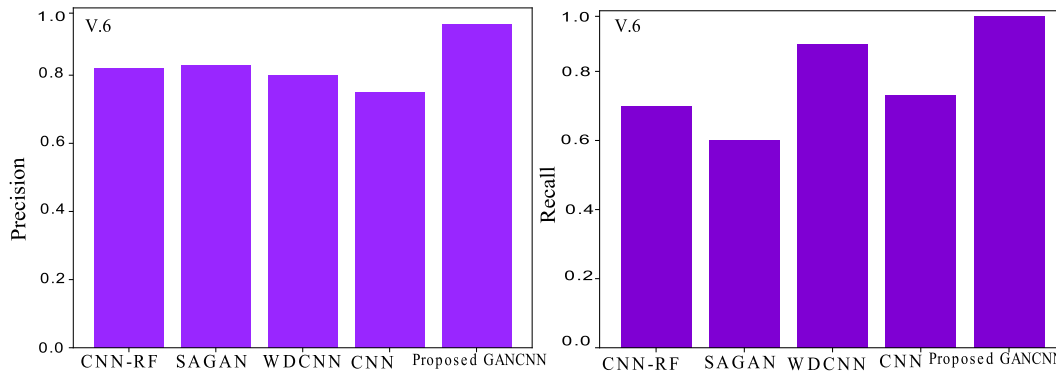


FIGURE 20. Precision and recall of proposed GANCNN model and benchmark models.

D. PERFORMANCE METRICS AND COMPARISON OF THE PROPOSED GANCNN MODEL WITH EXISTING TECHNIQUES

In this section, the proposed GANCNN model is validated in terms of precision, recall, F1-score, Area Under the Curve (AUC), FPR and RMSE. In the existing literature, authors use these metrics to validate the performance of their models. AUC is used in [24], [26], [29], [63], FPR is measured in [26], [64]. Precision, recall and F1-score are used in [44]. Moreover, GANCNN model is compared with SAGAN, WDCNN, CNN and CNN-RF models in terms of aforementioned metrics.

Precision of the proposed and existing models is calculated and compared in this section. It is used as a performance metric in many classification problems. Precision is defined as

the total number of true positive samples divided by the total number of false positive samples and true positive samples. It is also defined as the ratio of correctly predicted positive observations to the total predicted positive observations. It is calculated using Equation (13) [63]

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \tag{13}$$

The precision values of GANCNN and existing models are shown in Figure 20, which is labeled as V.6. The results show that the proposed model has the highest precision value; whereas, CNN has the lowest precision value.

Recall is defined as the total number of true positive divided by the total number of false negative and true positive. Figure 20 shows the recall value of the proposed

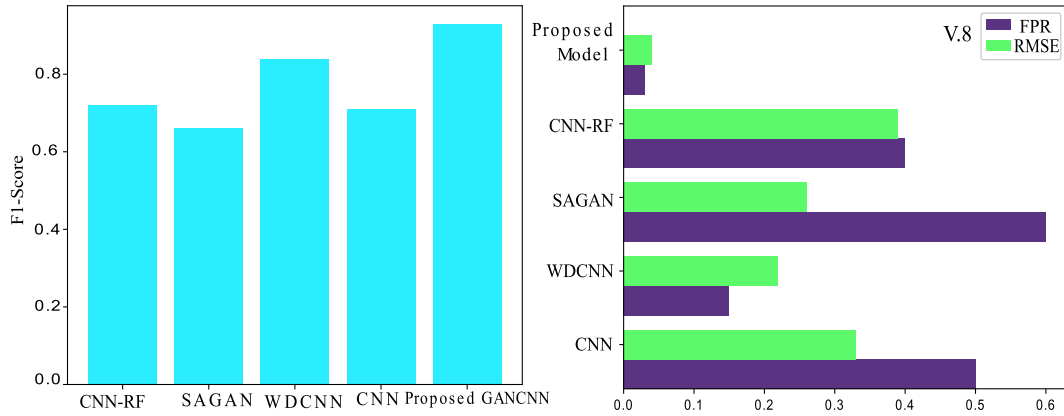


FIGURE 21. F1-Score, FPR and RMSE of proposed GANCNN and existing models.

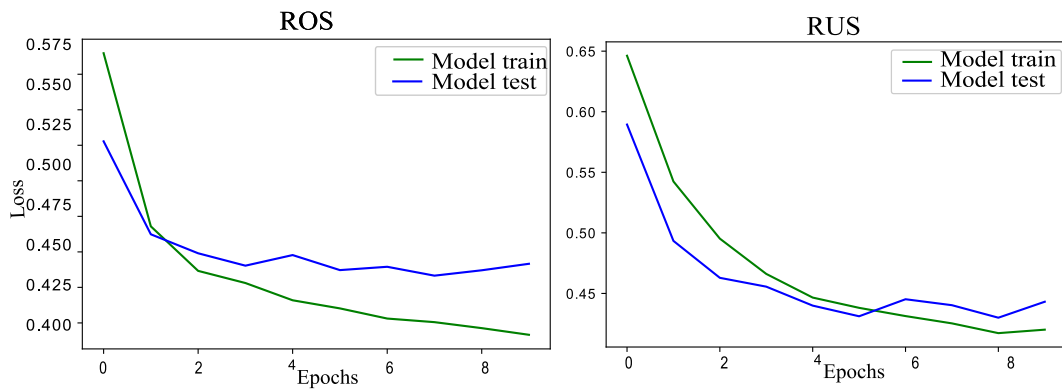


FIGURE 22. Loss of ROS and RUS.

GANCNN and existing models. The results show that recall of GANCNN model is the highest. The recall value of SAGAN is the lowest among all. The figure is labeled as V.6 and it validates the solution S.6. Formula of recall is given in Equation (14) [63]

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative}. \quad (14)$$

F1-score is the mean of precision and recall. It is defined as the ratio of the product of precision and recall to their sum, which is multiplied by 2. In Figure 21, GANCNN is compared with the existing models in terms of F1-score. It is clear from the figure that GANCNN achieves better F1-score as compared to the existing models. F1-score is calculated using Equation (15) [65]

$$F1 - score = 2 * \frac{Precision * Recall}{Precision + Recall}. \quad (15)$$

In Figure 21, FPR and RMSE for GANCNN and existing models are calculated. FPR is defined as the model’s inability to detect actual electricity thieves. High FPR increases the chance of low classification accuracy. It is known as an important performance measure for classification. As, it tells about the users that are misclassified as thieves. If the FPR is

high, it increases the on-site inspection cost. The formula of FPR is given in Equation 16

$$FPR = \frac{False\ Positive}{False\ Positive + True\ Negative}. \quad (16)$$

Figure 21 is labeled as V.8 and it validates the solution S.8.

E. PERFORMANCE METRICS AND COMPARISON OF THE PROPOSED ERNET MODEL WITH EXISTING TECHNIQUES

In this section, the proposed ERNET model is evaluated by applying different sampling techniques. Moreover, ERNET is compared with the existing classification models. Figures 22 and 23 show the performance of ERNET with different sampling techniques: ROS, RUS, SMOTE and ENN. The performance is evaluated in terms of loss. With ROS, training loss of ERNET is 0.40 and its testing loss is 0.45. By applying RUS, ERNET achieves 0.43 training loss and 0.45 testing loss. Moreover, with SOMTE, ERNET achieves 0.42 training loss and 0.45 testing loss. Finally, with the ENN, the training and testing loss are 0.40 and 0.45, respectively. Whereas, the training and testing loss of ERNET with proposed SMO-TEENN technique is 0.5. The results show that the proposed technique outperforms the aforementioned existing techniques.

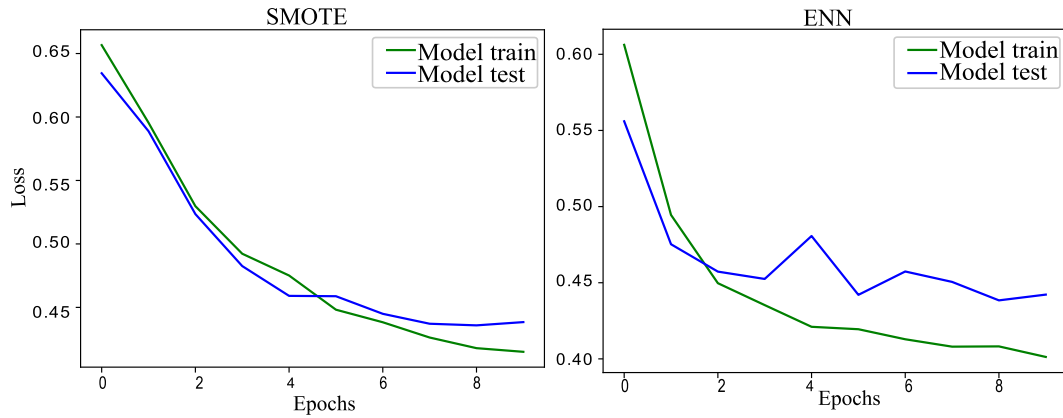


FIGURE 23. Loss of SMOTE and ENN.

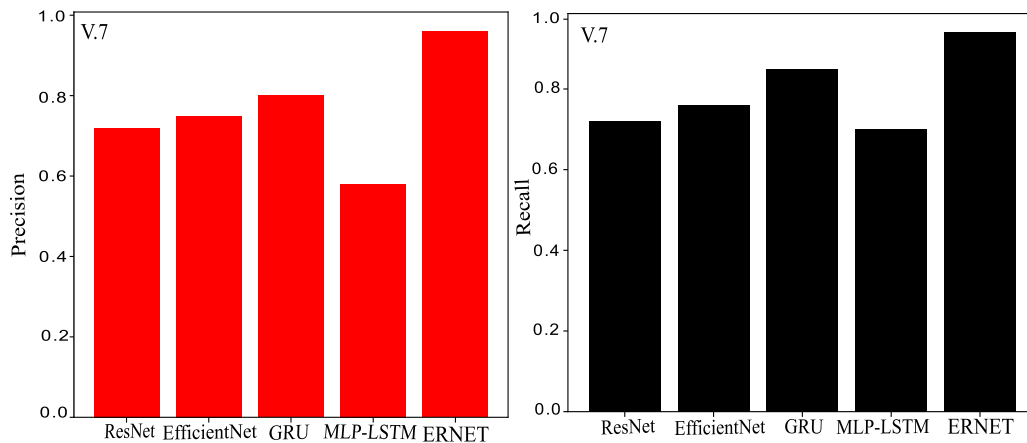


FIGURE 24. Precision and recall of proposed ERNET model and other comparison techniques.

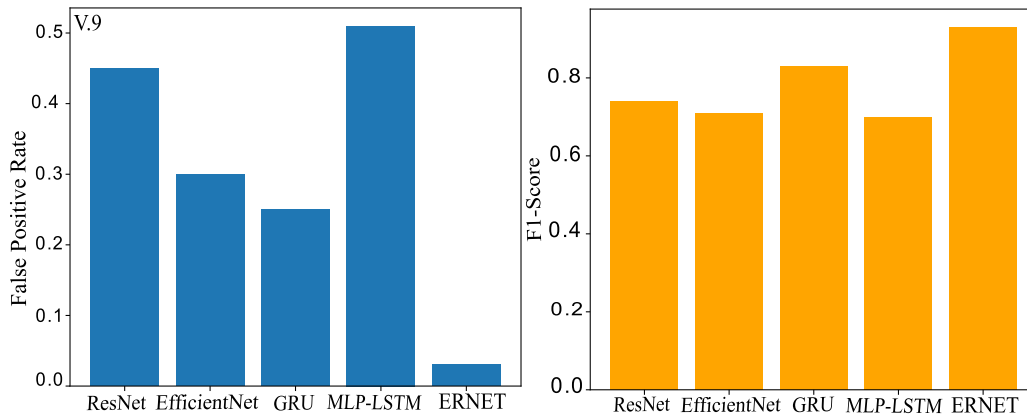


FIGURE 25. FPR and F1-score of proposed ERNET model and other benchmark techniques.

The ERNET model is compared with ResNet, EfficientNet, GRU and MLP-LSTM in terms of precision, recall, FPR and F1-score. Figure 24 shows the values of precision and recall for ERNET and other models. As shown in the figure, precision of ERNET is 0.96, which is the highest than other models, which have less than 0.8 precision. Moreover, the recall score of the ERNET model is 0.94, which is the highest among all existing models. Whereas, recall score of

ResNet, EfficientNet, GRU and MLP-LSTM is 0.72, 0.76, 0.85 and 0.70, respectively. The figure is labeled as V.7 and it validates the solution S.7.

Furthermore, the comparison of ERNET with other models in terms of FPR and F1-score is shown in Figure 25. As it is illustrated in the figure, FPR of ERNET is below 0.1 that is the lowest among all. Whereas, the FPR of MLP-LSTM is 0.51 and is the highest. The FPR of EfficientNet, GRU

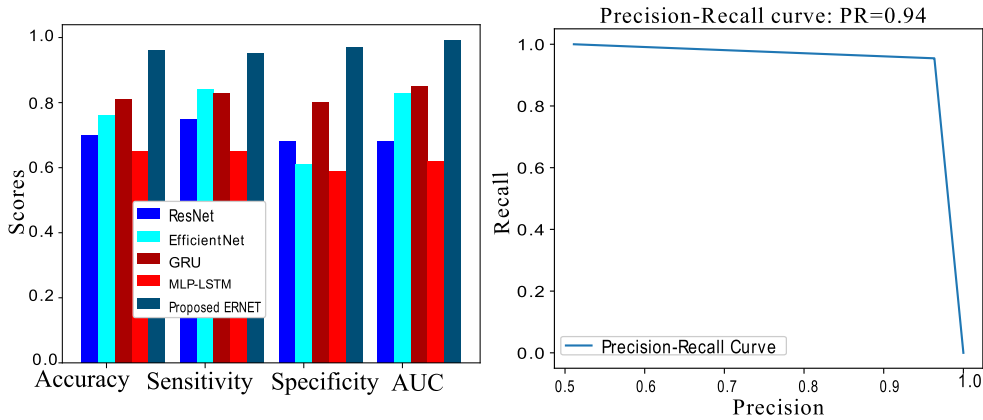


FIGURE 26. Accuracy, sensitivity, specificity and AUC of the proposed ERNET model and other comparison techniques and precision-recall curve of ERNET model.

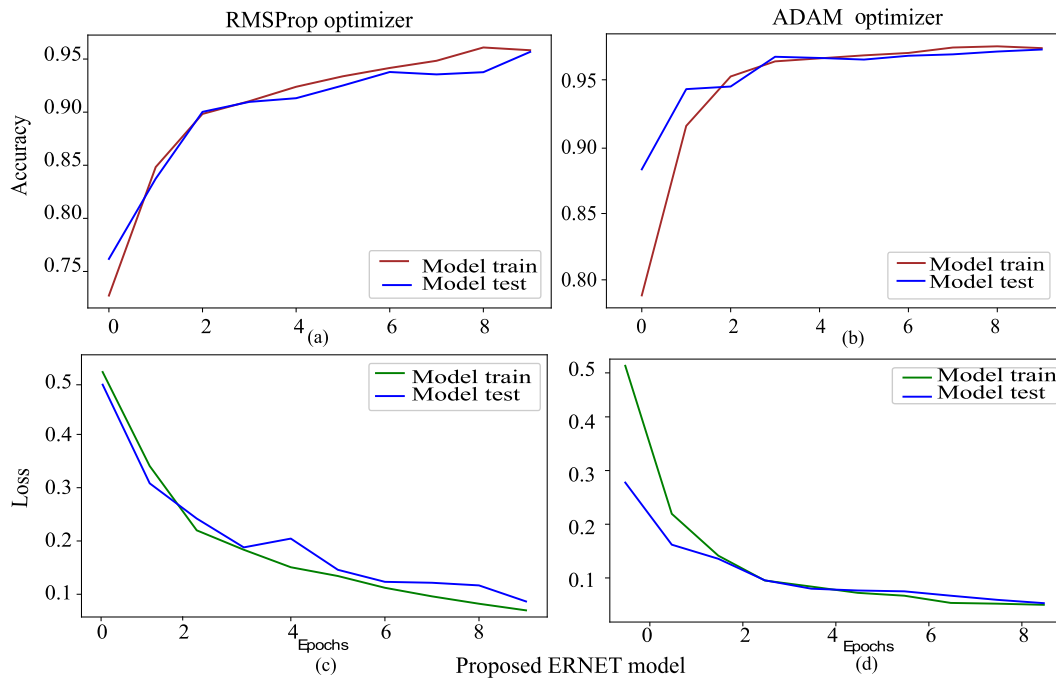


FIGURE 27. Accuracy and loss with Adam and RMSprop optimizer.

and ResNet is 0.30, 0.25 and 0.45, respectively. Moreover, comparison results of the models in terms of F1-score are also shown in Figure 25. The F1-score of ERNET is 0.93, which is the highest than other models. Whereas, F1-score of other models is less than 0.85. The figure is labeled as V.9 and its validates the solution S.9. Moreover, Figure 26 shows the comparison of ERNET with existing models in terms of accuracy, sensitivity, specificity and AUC. The results show that ERNET outperforms all of the existing models in terms of aforementioned metrics. Figure 26 also shows that the value of precision-recall curve of the ERNET model is 0.94.

F. COMPARISON BETWEEN PROPOSED AND EXISTING OPTIMIZERS

Figure 27 shows the performance of the ERNET model with RMSProp and Adam optimizers in terms of accuracy and

loss. The results show that with the RMSProp optimizer, ERNET achieves 0.95 accuracy for both training and testing. Whereas, with the Adam optimizer, accuracy of ERNET for both training and testing is 0.98. Furthermore, the training and testing loss of ERNET with both RMSProp and Adam optimizers is less than 0.1.

Figure 28 shows the comparison of GANCNN with existing models in terms of AUC. As shown in the figure, AUC of GANCNN is 0.9, which is the highest than other models. Whereas, all of the other models have less than 0.83 AUC value. Figure 28 also shows the comparison of Adam optimizer with other optimizers, which include Ada-Grad, Stochastic Gradient Descent (SGD) and RMSProp. As shown in the figure, Adam has the highest accuracy and the lowest loss value as compared to other techniques. Because for each parameter, it uses momentum and adaptive

TABLE 7. Performance analysis of the proposed models.

Model	AUC	Precision	Accuracy	Recall	F1-score	FPR	Execution time (min)
GANCNN	0.98	0.95	0.95	0.99	0.90	0.05	35
ERNET	0.98	0.94	0.98	0.93	0.89	0.02	40

TABLE 8. Mapping of the results with proposed solutions and identified limitations.

Identified limitations	Proposed solutions	Validations
L.1: SMOTE causes overfitting problem	S.1: ADASYNENN is proposed S.2: SMOTEENN is proposed	V.1: As shown in Figure 17 and Figure 18, the ADASYNENN performs well in creating samples while reducing overfitting V.2: As shown in Figure 19, GANCNN has the highest accuracy and resolves overfitting problem
L.2: LSTM has separate memory cell, which consumes a lot of memory	S.3: No separate memory cell is required in ERNET model	V.3: As shown in Figure 30, execution time of proposed ERNET is less as compared to MLP-LSTM
L.3: RUS removes random samples from dataset and information is lost	S.4: No information loss in ADASYNENN S.5: No information loss in SMOTEENN	V.4: As shown in Figure 17, RUS has 0.70 accuracy and proposed model has 0.95 accuracy V.5: As shown in Figure 17, RUS has 0.75 accuracy and proposed model has 0.98 accuracy
L.4: No sampling technique is used, model is biased towards majority class	S.6: ADASYNENN is proposed S.7: SMOTEENN is proposed	V.6: Precision and recall of GANCNN model are high because of sampling technique, as shown in the Figure 20 V.7: As shown in Figure 24, ERNET has high precision and recall values because of the sampling technique
L.5: FPR is high, which leads to high inspection cost	S.8: FPR is low in proposed GANCNN model S.9: FPR is low in proposed ERNET model	V.8: In Figure 21, FPR is low in GANCNN as compared to existing techniques V.9: In Figure 25, FPR is low in ERNET model as compared to existing techniques
L.6: FPR is not calculated	S.8: FPR is low in proposed GANCNN model S.9: FPR is low in proposed ERNET model	V.9: In Figure 21, FPR is low as compared to existing techniques V.10: In Figure 25, FPR is low in the proposed ERNET model
L.7: RF takes more execution time to build the trees	S.11: Execution time of GANCNN model is less as compared to CNN-RF	V.11: As shown in Figure 30, execution time of the proposed GANCNN model is less than CNN-RF model

learning rate mechanism, which lead to faster convergence. The parameters of the Adam optimizer include η , β_1 , β_2 and ϵ . In this scenario, the values of these parameters for GANCNN are set as: $\eta = 0.001$, $\beta_1 = 0.9$, $\beta_2 = 0.999$ and $\epsilon = 1e-08$. In GANCNN, Adam optimizer is applied for better performance of the CNN model. Adam is also more suitable for large, noisy or sparse datasets. Whereas, in the ERNET model, RMSProp optimizer is used in ResNet. η , ρ , $momentum$, ϵ and $centered$ are the parameters of RMSProp optimizer. We set the values of the parameters as: $\eta = 0.001$, $\rho = 0.9$, $momentum = 0.0$, $\epsilon = 1e-07$ and $centered = False$. The comparison between Adam and RMSProp optimizer is as follows. RMSProp performs parameter updation using a momentum on the rescaled gradient. Whereas, in Adam optimizer, the parameters are updated through the running average of first and second gradient moment. Moreover, the bias-correction term is absent in RMSProp. However, Adam involves the correction term, which leads to fast convergence. The accuracy and loss comparison of different optimizers is observed from Figure 28.

G. AUC OF THE PROPOSED MODELS

In this section, AUC of both GANCNN and ERNET models is calculated, as shown in Figure 29. AUC dictates that how well a model performs in classifying the electricity thieves. The range of AUC is between 0 and 1. If AUC is near to 1,

it means that model performs well in terms of classification and if it is near to 0, it means that model does not perform well. The threshold value is set as 0.5. The choice of selecting AUC of 0.5 is to examine the performance of the proposed models in terms of classifying electricity theft. As shown in the figure, AUC of GANCNN is 0.985, which is near to 1 that means classification of the model is good. Whereas, the ERNET model accurately separates both classes (majority and minority) and its value of AUC is 0.988. The results show that ERNET model outperforms GANCNN in terms of AUC. The AUC of 0.985 for the proposed GANCNN means that there is no discrimination in the classification of electricity theft, which means that any AUC greater than 0.5 is acceptable. The same explanation is given for the AUC of 0.988 for the ERNET model. Generally, the AUC of 0.985 means that the ROC curve falls on the diagonal line of the curve. It implies that the ROC curve on the diagonal line has no discriminatory ability. Whereas, the ROC curve above the diagonal line has discriminatory ability to classify electricity theft.

H. COMPARISON OF PROPOSED AND BENCHMARK MODELS' EXECUTION TIME

In this section, the execution time of GANCNN and ERNET models is discussed. We compared GANCNN with CNN-RF, MLP-LSTM, SAGAN and WDCNN models. As shown

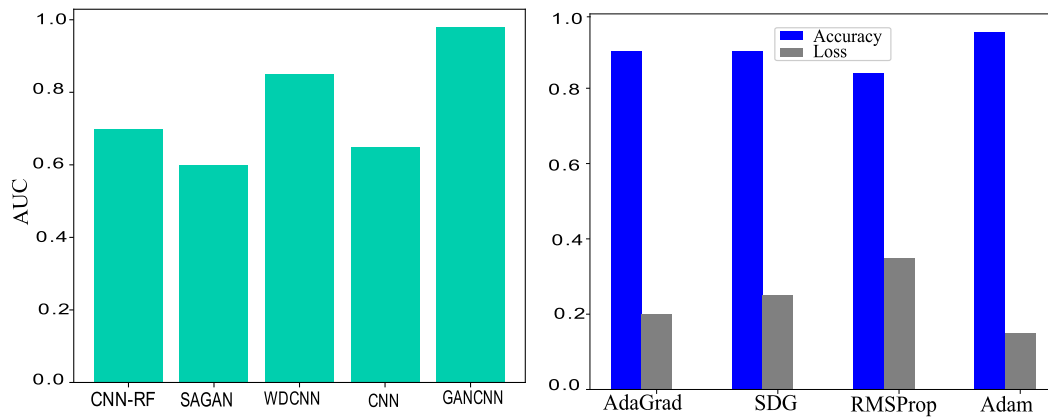


FIGURE 28. AUC of proposed GANCNN model with existing comparison techniques and comparison of Adam with other optimizers.

in Figure 30, MLP-LSTM has the highest execution time because of separate memory cell of LSTM. Whereas, the proposed GANCNN model has the lowest execution time because no additional memory cell is required to store the information for long-term. Moreover, the ERNET model is also compared with the above mentioned benchmark models. The figure shows that ERNET has low execution time as compared to CNN-RF and MLP-LSTM. Figure 30 is labeled as V.3 and V.11, and it validates the solutions S.3 and S.11.

In Table 7, the performance analysis of the proposed models in terms of AUC, precision, accuracy, recall, F1-score, FPR and execution time, is given. From the analysis results, the proposed ERNET model outperforms GANCNN model in terms of accuracy while the proposed GANCNN model performs better than ERNET in terms of precision, recall, F1-score, FPR and execution time. The reasons for better performance of GANCNN model over the ERNET model are given as follows. The GANCNN model generates data that is similar to the actual data. It means that the generated data is not distinguishable from the actual data. Thus, a real synthetic data can be generated in order to address the data imbalance problem. Also, the GANCNN model learns the internal representation of the data such that any difficulty in the data is learned easily. Furthermore, after training the data, the discriminator of the GANCNN model classifies the data efficiently. On the other hand, the ERNET model requires more computational time and memory during training. However, the ERNET model achieves a better accuracy because it can perform well in a larger network with high dept and width. Also, ERNET preserves the computational power through proper scaling of network's depth and width; thereby, increasing its accuracy.

Table 8 shows the mapping of identified limitations with the proposed solutions and their validations. In L.1, the overfitting issue of SMOTE is identified. To solve this issue, two different sampling techniques are proposed: ADASYNENN and SMOTEENN as S1 and S2, respectively. The solutions

are validated as V.1 and V.2, as shown in Figures 17, 18 and 19. Next, in L.2, additional memory requirement issue of LSTM is identified. Therefore, in S.3, GRU is used in ERNET because of its less time complexity. The solution is validated as V.3, as shown in Figure 17. In L.3, the information loss issue in RUS technique is identified. The issue is addressed by presenting ADASYNENN and SMOTEENN with no loss property as S.4 and S.5, respectively. The solutions are validated as V.4 and V.5, as shown in Figure 17. Furthermore, in L.4, no sampling technique is used, which causes biasness of model. Therefore, ADASYNENN and SMOTEENN are proposed as S.6 and S.7. The solutions are validated as V.6 and V.7, as shown in Figures 20 and 24. In L.5 and L.6, FPR is high and is not calculated. To overcome this issue, FPR of the proposed models is calculated and compared with FPR of benchmark models. The solutions are labeled as S.8 and S.9. The validation of these solutions is presented in Figures 21 and 30 as V.8 and V.9, respectively. In L.7, high execution time of RF is identified. This issue is resolved in S.11. The solution is validated as V.11, as shown in Figure 30.

I. DISCUSSION

To efficiently analyze consumers' energy consumption data using classical machine learning methods, the class imbalanced problem must be addressed. The problem occurs when the overall number of one class of data (honest consumers' data) is more than the overall number of another class of data (fraudulent consumers' data). So, this paper proposes two deep neural networks: GANCNN and ERNET. The former is used to generate synthetic data using the actual data. The purpose of generating synthetic data is to solve the class imbalanced problem. Whereas, the latter is used for increasing the network's width and depth. The aim of increasing the width and depth of the network is to achieve higher prediction accuracy from a large dataset.

The effectiveness of the proposed GANCNN and ERNET models is evaluated using the following performance metrics:

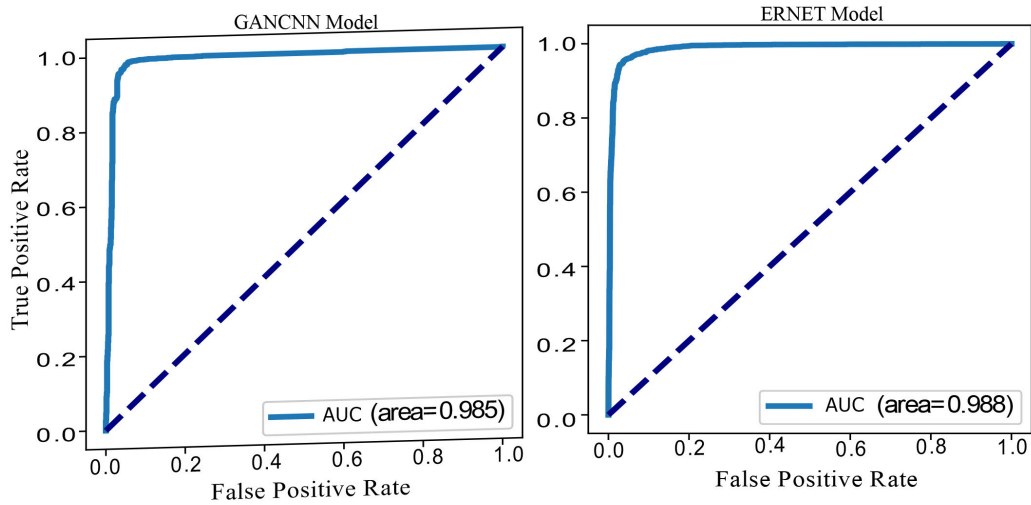


FIGURE 29. AUC of proposed GANCNN and ERNET models.

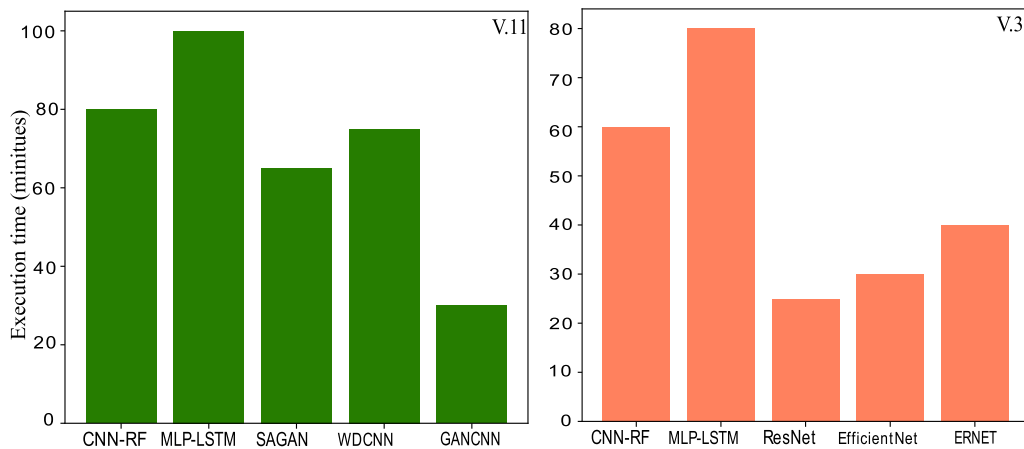


FIGURE 30. Execution time of proposed GANCNN and ERNET models.

precision, AUC, FPR, accuracy, F1-score and recall. In this paper, two case studies are considered for analyzing the proposed models. For the first case, the proposed ERNET model is compared with existing models, such as EfficientNet, MLP-LSTM and GRU. The simulation results show that the proposed ERNET model outperforms the existing models in terms of all of the performance metrics. Figure 17 shows that the training accuracy of the proposed ERNET model is higher than other existing models. Figure 24, Figure 25 and Figure 26 show that the proposed ERNET model outperforms other models in terms of precision, recall, FPR, F1-score, accuracy, sensitivity, specificity and AUC. The reason for the better performance of the proposed ERNET model over other models is as follows. The ERNET model can perform efficiently in a large network with high dept and width. Besides, it conserves its computational power via scaling the network’s width and depth. For the second case, the proposed GANCNN model is compared with other models, such as CNN-RF, SAGAN, WDCNN and CNN. From the simulation results, it is observed that the proposed GANCNN model outperforms other models in terms of all of the performance

metrics. Figure 20 shows that the proposed GANCNN model is better than other models in terms of precision and recall. Figure 28 shows that the proposed GANCNN outperforms other models in terms of AUC and accuracy. Figure 30 shows the performance of the proposed GANCNN model in terms of execution time. The reason for the better performance of the proposed GANCNN model is as follows. The GANCNN model can generate synthetic data that is similar to the actual data, which is extremely important to solve the class imbalanced problem. Generally, the focus of the proposed work is to provide an anomaly detection mechanism that compares the electricity consumption behavior of different consumers regarding the trend of prediction error. If the current pattern of energy consumption of a consumer is different from the previous one, then an anomaly is detected for that consumer. The practical actualization of the proposed work is constrained for different available consumers’ electricity consumption data, computational capacities of consumers and lack of privacy about the consumers’ data. However, we do not consider the privacy preservation of consumers in our proposed scenario, but it will be considered in the future work.

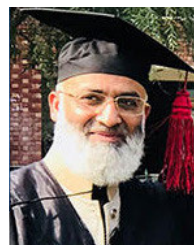
V. CONCLUSION

With increase in the electricity demand over the years, two types of losses are faced by the power utilities: TLs and NTLs. These losses lead to other problems as well, such as huge revenue loss, class imbalanced problem, low accuracy and high FPR for detecting electricity thieves. To solve these problems, two models are proposed in this work. In the first model, ADASYNENN is used to solve the class imbalanced problem; whereas, LLE is used for feature extraction. Moreover, a hybrid technique based on SAGAN and WDCNN, termed as GANCNN, is introduced for ETD. The second proposed model consists of five stages. Firstly, interpolation is used to remove the missing values. Secondly, robust scaler method is used for data normalization. Thirdly, SAE is applied for reduction of data dimensionality. Fourthly, SMOTEENN is applied to solve the class imbalanced problem. Finally, for classifying honest and theft consumers, a hybrid ERNET model is introduced, which is a combination of EfficientNet, ResNet and GRU. Additionally, RMSProp optimizer is used to enhance the performance of the model. To validate the proposed models, extensive simulations are performed using SGCC dataset. Different performance metrics, such as precision, recall, F1-score, FPR, accuracy and AUC are used for evaluation. The results of GANCNN for precision, recall, F1-score, FPR, accuracy and AUC are 0.95, 0.99, 0.9, 0.05, 0.95 and 0.985, respectively. On the other hand, the results of ERNET for precision, recall, F1-score, FPR, accuracy and AUC are 0.94, 0.93, 0.89, 0.02, 0.98 and 0.988, respectively. Moreover, the proposed models are compared with the state-of-the-art models, which include SAGAN, WDCNN, CNN, CNN-RF, MLP-CNN, MLP-LSTM, LSTM and CNN-LSTM. The results show that the proposed models outperformed the aforementioned benchmark models. In future, proposed models will be used for other theft related areas, such as banking. Moreover, different datasets will be used to determine the effectiveness of the proposed models.

REFERENCES

- [1] S. Aslam, A. Khalid, and N. Javaid, "Towards efficient energy management in smart grids considering microgrids with day-ahead energy forecasting," *Electr. Power Syst. Res.*, vol. 182, pp. 106232–106243, May 2020.
- [2] H. M. Hussain, N. Javaid, S. Iqbal, Q. U. Hasan, K. Aurangzeb, and M. Alhussein, "An efficient demand side management system with a new optimized home energy management controller in smart grid," *Energies*, vol. 11, no. 1, pp. 190–218, 2018.
- [3] O. Samuel, F. A. Alzahrani, R. J. U. Hussien Khan, H. Farooq, M. Shafiq, M. K. Afzal, and N. Javaid, "Towards modified entropy mutual information feature selection to forecast medium-term load using a deep learning model in smart homes," *Entropy*, vol. 22, no. 1, pp. 68–99, 2020.
- [4] J. Li and F. Wang, "Non-technical loss detection in power grids with statistical profile images based on semi-supervised learning," *Sensors*, vol. 20, no. 1, pp. 236–255, 2020.
- [5] P. F. M. Simões, R. C. Souza, R. F. Calili, and J. F. M. Pessanha, "Analysis and short-term predictions of non-technical loss of electric power based on mixed effects models," *Socio-Econ. Planning Sci.*, vol. 71, pp. 100804–100811, Sep. 2020.
- [6] G. M. Messinis, A. E. Rigas, and N. D. Hatziaargyriou, "A hybrid method for non-technical loss detection in smart distribution grids," *IEEE Trans. Smart Grid*, vol. 10, no. 6, pp. 6080–6091, Nov. 2019.
- [7] R. R. González, M. J. Jiménez, and M. C. G. Quintero, "Computational intelligent approaches for non-technical losses management of electricity," *Energies*, vol. 13, no. 9, pp. 2393–2418, 2020.
- [8] K. M. Ghorri, M. Imran, A. Nawaz, R. A. Abbasi, A. Ullah, and L. Szathmary, "Performance analysis of machine learning classifiers for non-technical loss detection," *J. Ambient Intell. Hum. Comput.*, pp. 1–16, Jan. 2020.
- [9] M. Ismail, M. F. Shaaban, M. Naidu, and E. Serpedin, "Deep learning detection of electricity theft cyber-attacks in renewable distributed generation," *IEEE Trans. Smart Grid*, vol. 11, no. 4, pp. 3428–3437, Jul. 2020.
- [10] D. Leite, J. Pessanha, P. Simões, R. Calili, and R. Souza, "A stochastic frontier model for definition of non-technical loss targets," *Energies*, vol. 13, no. 12, pp. 3227–3247, 2020.
- [11] M. Nabil, M. Ismail, M. M. E. A. Mahmoud, W. Alasmay, and E. Serpedin, "PPETD: Privacy-preserving electricity theft detection scheme with load monitoring and billing for AMI networks," *IEEE Access*, vol. 7, pp. 96334–96348, 2019.
- [12] V. Krysanov, A. Danilov, V. Burkovsky, P. Gusev, and K. Gusev, "Optimization of electric transmission lines (ETL) operation modes based on hardware solutions of process platform FACTS," in *Proceedings of 14th International Conference on Electromechanics and Robotics 'Zavalishin's Readings'*. Singapore: Springer, 2020, pp. 623–630.
- [13] Z. Qu, H. Li, Y. Wang, J. Zhang, A. Abu-Siada, and Y. Yao, "Detection of electricity theft behavior based on improved synthetic minority over-sampling technique and random forest classifier," *Energies*, vol. 13, no. 8, pp. 2039–2059, 2020.
- [14] Y. Wang, Q. Chen, and C. Kang, "Electricity theft detection," in *Smart Meter Data Analytics*. Singapore: Springer, 2020, pp. 79–98.
- [15] M. S. Saeed, M. W. Mustafa, N. N. Hamadneh, N. A. Alshammari, U. U. Sheikh, T. A. Jumani, and I. Khan, "Detection of non-technical losses in power utilities—A comprehensive systematic review," *Energies*, vol. 13, no. 18, pp. 4727–4752, 2020.
- [16] Y. Liu, T. Liu, H. Sun, K. Zhang, and P. Liu, "Hidden electricity theft by exploiting multiple-pricing scheme in smart grids," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 2453–2468, 2020.
- [17] 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.
- [18] 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.
- [19] C. Fan, F. Xiao, Y. Zhao, and J. Wang, "Analytical investigation of autoencoder-based methods for unsupervised anomaly detection in building energy data," *Appl. Energy*, vol. 211, pp. 1123–1135, Feb. 2018.
- [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] N. Ding, H. Ma, H. Gao, Y. Ma, and G. Tan, "Real-time anomaly detection based on long short-term memory and Gaussian mixture model," *Comput. Electr. Eng.*, vol. 79, pp. 106458–106469, Oct. 2019.
- [22] R. Khalid, N. Javaid, F. A. Al-Zahrani, K. Aurangzeb, E. U. H. Qazi, and T. Ashfaq, "Electricity load and price forecasting using Jaya-long short term memory (JLSTM) in smart grids," *Entropy*, vol. 22, no. 1, pp. 10–31, 2020.
- [23] 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.
- [24] 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.
- [25] 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.
- [26] 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.
- [27] G. Micheli, E. Soda, M. T. Vespucci, M. Gobbi, and A. Bertani, "Big data analytics: An aid to detection of non-technical losses in power utilities," *Comput. Manage. Sci.*, vol. 16, nos. 1–2, pp. 329–343, Feb. 2019.
- [28] B. Coma-Puig and J. Carmona, "Bridging the gap between energy consumption and distribution through non-technical loss detection," *Energies*, vol. 12, no. 9, pp. 1748–1765, 2019.

- [29] 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.
- [30] M. M. Buzau, J. Tejedor-Aguilera, P. Cruz-Romero, and A. Gómez-Expósito, "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.
- [31] 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.
- [32] X. Wang, I. Yang, and S.-H. Ahn, "Sample efficient home power anomaly detection in real time using semi-supervised learning," *IEEE Access*, vol. 7, pp. 139712–139725, 2019.
- [33] C. C. O. Ramos, D. Rodrigues, A. N. D. Souza, and J. P. Papa, "On the study of commercial losses in Brazil: A binary black hole algorithm for their characterization," *IEEE Trans. Smart Grid*, vol. 9, no. 2, pp. 676–683, Mar. 2018.
- [34] J. I. Guerrero, I. Monedero, F. Biscarri, J. Biscarri, R. Millan, and C. León, "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.
- [35] X. Lu, Y. Zhou, Z. Wang, Y. Yi, L. Feng, and F. Wang, "Knowledge embedded semi-supervised deep learning for detecting non-technical losses in the smart grid," *Energies*, vol. 12, no. 18, pp. 3452–3470, 2019.
- [36] M.-M. Buzau, J. Tejedor-Aguilera, P. Cruz-Romero, and A. Gómez-Expósito, "Hybrid deep neural networks for detection of non-technical losses in electricity smart meters," *IEEE Trans. Power Syst.*, vol. 35, no. 2, pp. 1254–1263, Mar. 2020.
- [37] 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.
- [38] S. K. Singh, R. Bose, and A. Joshi, "Entropy-based electricity theft detection in AMI network," *IET Cyber-Phys. Syst., Theory Appl.*, vol. 3, no. 2, pp. 99–105, Jun. 2018.
- [39] M. Zahid, F. Ahmed, N. Javaid, R. A. Abbasi, H. S. Z. Kazmi, A. Javaid, 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, pp. 122–154, 2019.
- [40] M. Adil, N. Javaid, U. Qasim, I. S. M. Ullah, and J. G. Choi, "LSTM and Bat-based RUSBoost approach for electricity theft detection," *Appl. Sci.*, vol. 10, no. 12, pp. 4378–4399, 2020.
- [41] A. Naz, M. U. Javed, N. Javed, T. Saba, M. Alhussein, and K. Aurangzeb, "Short-term electric load and price forecasting using enhanced extreme learning machine optimization in smart grids," *Energies*, vol. 12, no. 5, pp. 866–896, 2019.
- [42] A. Ahmad, N. Javaid, A. Mateen, M. Awais, and Z. A. Khan, "Short-term load forecasting in smart grids: An intelligent modular approach," *Energies*, vol. 12, no. 1, pp. 164–185, 2019.
- [43] S. Zahoor, S. Javaid, N. Javaid, M. Ashraf, F. Ishmanov, and M. K. Afzal, "Cloud-fog-based smart grid model for efficient resource management," *Sustainability*, vol. 10, no. 6, pp. 2079–2100, 2018.
- [44] 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.
- [45] M. Hasan, R. N. Toma, A. A. Nahid, M. M. Islam, and J. M. Kim, "Electricity theft detection in smart grid systems: A CNN-LSTM based approach," *Energies*, vol. 12, no. 17, pp. 3310–3322, 2019.
- [46] B. Chai, Q. Zhang, Q. Chen, T. Zhao, and K. Gao, "Research on applications of artificial intelligence in business management of power grid enterprises," in *Proc. IEEE 4th Adv. Inf. Technol., Electron. Autom. Control Conf. (IAEAC)*, vol. 1, Dec. 2019, pp. 683–688.
- [47] S. J. Choudhury and N. R. Pal, "Imputation of missing data with neural networks for classification," *Knowl.-Based Syst.*, vol. 182, pp. 104838–104844, Oct. 2019.
- [48] T. Xu, G. Coco, and M. Neale, "A predictive model of recreational water quality based on adaptive synthetic sampling algorithms and machine learning," *Water Res.*, vol. 177, pp. 115788–115799, Jun. 2020.
- [49] L. Gong, S. Jiang, and L. Jiang, "Tackling class imbalance problem in software defect prediction through cluster-based over-sampling with filtering," *IEEE Access*, vol. 7, pp. 145725–145737, 2019.
- [50] J. Wang, R. K. W. Wong, and T. C. M. Lee, "Locally linear embedding with additive noise," *Pattern Recognit. Lett.*, vol. 123, pp. 47–52, May 2019.
- [51] A. Hatua, T. T. Nguyen, and A. H. Sung, "Dialogue generation using self-attention generative adversarial network," in *Proc. IEEE Int. Conf. Conversational Data Knowl. Eng. (CDKE)*, Dec. 2019, pp. 33–38.
- [52] T. Hu, Q. Guo, H. Sun, T. E. Huang, and J. Lan, "Nontechnical losses detection through coordinated BiWGAN and SVDD," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 5, pp. 1866–1880, May 2021.
- [53] R. Khalid and N. Javaid, "A survey on hyperparameters optimization algorithms of forecasting models in smart grid," *Sustain. Cities Soc.*, vol. 61, pp. 102275–102300, Oct. 2020.
- [54] I. K. M. Jais, A. R. Ismail, and S. Q. Nisa, "Adam optimization algorithm for wide and deep neural network," *Knowl. Eng. Data Sci.*, vol. 2, no. 1, pp. 41–46, 2019.
- [55] F. Hoseini, A. Shahbahrami, and P. Bayat, "AdaptAhead optimization algorithm for learning deep CNN applied to MRI segmentation," *J. Digit. Imag.*, vol. 32, no. 1, pp. 105–115, 2019.
- [56] X. Ma, S. Luan, C. Ding, H. Liu, and Y. Wang, "Spatial interpolation of missing annual average daily traffic data using copula-based model," *IEEE Intell. Transp. Syst. Mag.*, vol. 11, no. 3, pp. 158–169, Fall 2019.
- [57] A. Reurink and J. Garcia-Bernardo, "Competing for capitals: The great fragmentation of the firm and varieties of FDI attraction profiles in the European Union," *Rev. Int. Political Econ.*, pp. 1–34, Mar. 2020.
- [58] L. Hou, V. Nguyen, A. B. Kanevsky, D. Samaras, T. M. Kurc, T. Zhao, and J. H. Saltz, "Sparse autoencoder for unsupervised nucleus detection and representation in histopathology images," *Pattern Recognit.*, vol. 86, pp. 188–200, Feb. 2019.
- [59] M. Huk, "Stochastic optimization of contextual neural networks with RMSprop," in *Proc. Asian Conf. Intell. Inf. Database Syst. Cham, Switzerland: Springer*, 2020, pp. 343–352.
- [60] K. Adem and S. Kiliçarslan, "Performance analysis of optimization algorithms on stacked autoencoder," in *Proc. 3rd Int. Symp. Multidisciplinary Stud. Innov. Technol. (ISMSIT)*, Oct. 2019, pp. 1–4.
- [61] E. Bisong, "Google colab," in *Building Machine Learning and Deep Learning Models on Google Cloud Platform*. Berkeley, CA, USA: Apress, 2019, pp. 59–64.
- [62] X. Gong, B. Tang, R. Zhu, W. Liao, and L. Song, "Data augmentation for electricity theft detection using conditional variational auto-encoder," *Energies*, vol. 13, no. 17, pp. 4291–4305, 2020.
- [63] 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, pp. 3151–3172, 2020.
- [64] 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.
- [65] C. H. Park and T. Kim, "Energy theft detection in advanced metering infrastructure based on anomaly pattern detection," *Energies*, vol. 13, no. 15, pp. 3832–3842, 2020.



NADEEM JAVAID (Senior Member, IEEE) received the bachelor's degree in computer science from Gomal University, Dera Ismail Khan, Khyber Pakhtunkhwa, 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 127 master's and 20 Ph.D. theses. He has authored over 900 articles 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 an Associate Editor of IEEE Access journal and an Editor of the *Sustainable Cities and Society*.



HIRA GUL received the master's degree in computer science 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 interest includes data science in smart cities.



SOBIA BAIG (Senior Member, IEEE) received the B.S. degree from the University of Engineering and Technology, Taxila, Pakistan, in 1995, and the master's and Ph.D. degrees in electronics engineering from the Ghulam Ishaq Khan Institute, Topi, Pakistan, in 2001 and 2008, respectively. She is currently the Head of the Energy Research Center and an Associate Professor at the Department of Electrical and Computer Engineering, COMSATS University Islamabad, Lahore, Pakistan. Her research interests include work on efficient and stable energy systems, voltage and frequency control for micro-grids, smart grid communication techniques non-orthogonal multiple access (NOMA) techniques in 5G, and power line communication (PLC) networks.



FAISAL SHEHZAD is currently pursuing the master's degree in computer science with the Communications over Sensors (ComSens) Research Laboratory, Department of Computer Science, COMSATS University Islamabad, Islamabad Campus, under the supervision of Dr. Nadeem Javaid. His research interest includes data science in smart cities.



CHENGJUN XIA (Senior Member, IEEE) was born in Huanggang, Hubei, China, in January 1974. He received the B.S. degree from Xi'an Jiao Tong University, Xi'an, China, in 1995, and the Ph.D. degree from the Huazhong University of Science and Technology, Wuhan, China, in 2003, both in electrical engineering. From 2003 to 2005, he was a Postdoctoral Fellow at the Jiangsu Provincial Electric Power Research Institute, Southeast University. He was a Visiting Scholar at the University of Manitoba, Canada, from 2016 to 2017. He is currently an Associate Professor at the School of Electric Power, South China University of Technology, Guangzhou, China. His research interests include power system stability analysis and control, HVDC and FACT, VSC-HVDC, and power system simulation.



LIN GUAN (Member, IEEE) received the B.S. and Ph.D. degrees in electric power engineering from the Huazhong University of Science and Technology, Wuhan, China, in 1990 and 1995, respectively. She is currently a Professor with the Electric Power College, South China University of Technology, Guangzhou, China. From 2014 to 2015, she was a Visiting Scholar with Stanford University. She is the author of more than 120 articles and the principal investigator of more than 50 projects. Her research interests include application of artificial intelligence technology in electrical engineering, power system security and control, and power system planning and reliability.



TANZEELA SULTANA received the bachelor's degree in computer science from The University of Azad Jammu and Kashmir (UAJK), Kotli campus, Kotli, Azad Kashmir, in 2016. She is currently pursuing the master's degree in computer science from COMSATS University Islamabad, Islamabad, Pakistan. She is currently working as a Research Associate with the Communications over Sensors (ComSens) Research Laboratory, Department of Computer Science, COMSATS University Islamabad. She has authored over three articles in technical journals and 11 proceedings in international conferences. Her research interests include data analytics in smart grids, cloud computing, and blockchain in IoT and vehicular networks.

...