

Received January 26, 2022, accepted February 4, 2022, date of publication February 8, 2022, date of current version February 22, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3150016

# AlexNet, AdaBoost and Artificial Bee Colony Based Hybrid Model for Electricity Theft Detection in Smart Grids

ASHRAF ULLAH<sup>1</sup>, NADEEM JAVAID<sup>1,2</sup>, (Senior Member, IEEE),

MUHAMMAD ASIF<sup>1</sup>, (Graduate Student Member, IEEE),

MUHAMMAD UMAR JAVED<sup>1</sup>, AND ADAMU SANI YAHAYA<sup>1</sup>

<sup>1</sup>Department of Computer Science, COMSATS University Islamabad, Islamabad 44000, Pakistan

<sup>2</sup>School of Computer Science, University of Technology Sydney, Ultimo, NSW 2007, Australia

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

**ABSTRACT** Electricity theft (ET) is an utmost problem for power utilities because it threatens public safety, disturbs the normal working of grid infrastructure and increases revenue losses. In the literature, many machine learning (ML), deep learning (DL) and statistical based models are introduced to detect ET. However, these models do not give optimal results due to the following reasons: curse of dimensionality, class imbalance problem, inappropriate hyper-parameter tuning of ML and DL models, etc. Keeping the aforementioned concerns in view, we introduce a hybrid DL model for the efficient detection of electricity thieves in smart grids. AlexNet is utilized to handle the curse of dimensionality issue while the final classification of energy thieves and normal consumers is performed through adaptive boosting (AdaBoost). Moreover, class imbalance problem is resolved using an undersampling technique, named as near miss. Furthermore, hyper-parameters of AdaBoost and AlexNet are tuned using artificial bee colony optimization algorithm. The real smart meters' dataset is used to assess the efficacy of the hybrid model. The substantial amount of simulations proves that the hybrid model obtains the highest classification results as compared to its counterparts. Our proposed model obtains 88%, 86%, 84%, 85%, 78% and 91% accuracy, precision, recall, F1-score, Matthew correlation coefficient and area under the curve receiver operating characteristics, respectively.

**INDEX TERMS** AlexNet, adaptive boosting, artificial bee colony, deep driven models, electricity theft detection, feature extraction, smart grids.

## I. INTRODUCTION

Electricity has a greater influence on our daily lives. Several electronic devices, communication devices and modern electric vehicles are dependent on electricity. The government and private companies distribute electricity to various enterprises, agencies and end users. However, electricity losses often occur in transmission and distribution lines. These losses are classified as technical losses (TLs) and non-technical losses (NTLs) [1], [2]. TLs happen because of short circuits in transformers, electric shocks, lethal fires, etc. Whereas, NTLs emerges because of energy theft (ET), direct hooking, unpaid bills, meter tampering, meter

The associate editor coordinating the review of this manuscript and approving it for publication was Yang Li<sup>1</sup>.

hacking, etc. Among them, ET is a major source of NTLs. The study conducted by the northeast group in 2015 stated that the world faces loss of \$89.3 billion yearly due to NTLs [3]. In Pakistan, the power utility companies lose major portion of their total energy supply per year due to ET. The line losses of Peshawar electric supply company (PESCO) are decreasing every year by 2%, but in the year 2018-19, these losses increased up to 36.2% [4]. Moreover, these losses are not limited to underdeveloped countries but they also affect the economy of developed countries as well. According to a report, the United States and the United Kingdom accounted for loss of \$10.5 and £175 billion per annum because of ET, respectively [5].

Advanced meter infrastructure (AMI) is a new concept, which is introduced in traditional power grids. It consists of

TABLE 1. List of abbreviations.

Abbreviation	Description
AMI	Advanced metering infrastructure
AUC	Area under the curve
ABC	Artificial bee colony
BHA	Binary black hole algorithm
CNN	Convolution neural network
ETD	Electricity theft detection
FPR	False positive rate
GRU	Gated recurrent unit
KPCA	Kernel principal component analysis
LSTM	Long short-term memory
LR	Logistic regression
ML	Machine learning
NM	Near miss
MCC	Matthew correlation coefficient
MODWPT	Maximal overlap discrete wavelet-packet transform
NTLs	Non-technical losses
NaN	Not a number
PSO	Particle swarm optimization
PESCO	Peshawar electric supply company
ROS	Random oversampling
RUS	Random under sampling
RF	Random forest
SGCC	State grid corporation of China
SMOTE	Synthetic minority oversampling technique
SVM	Support vector machine
TLs	Technical losses
TPR	True positive rate

smart meters, sensors, computing devices and modern communication technologies, which are used to perform two-way communication between consumers and utilities. AMI is also responsible to collect data about electricity consumption (EC), prices at current time and the status of power grids. The involvement of the Internet in AMI opens numerous ways for attackers to remotely hack the smart metering system and steal electricity. Due to the above-mentioned issue, the detection of NTLs is an important need for the current era. In this regard, the researchers introduced various techniques to deal with NTLs [6]. These techniques are listed as follows.

1. State based: In state based techniques, various hardware components, such as sensors and transformers are integrated with smart meters to detect NTLs. These techniques perform well, however, extra monetary cost is required to maintain the embedded devices [7].

2. Game theory: In game theory based techniques, two players are involved: power utility and electricity thief. Both players play a game with each other for achieving the equilibrium state. These techniques are not feasible because formulating a suitable utility function is a tedious task [8].

3. Data driven: These techniques get the attention of the research community because they only require data for their training. The integration of smart meters produces a substantial amount of data. Therefore, the researchers propose various data driven techniques to perform electricity theft detection (ETD) in power grids. These techniques belong to machine learning (ML), meta learning, ensemble learning and deep learning (DL) [9]. In this era, these techniques

become popular because they only demand data for training. The solutions proposed in the existing work of ETD have the following limitations: (i) In ETD, the imbalance data is one of the main problems in which the records of one class is more than the second class. It raises the problem of poor generalization and overfitting because the classification model skews towards the larger class and ignores the smaller class. (ii) Due to the massive volume of EC data, the diversity and dimensionality of data are increased to a greater extent. The high dimensional data mostly misleads classifiers during the accurate ETD. (iii) The presence of several non-malicious factors increases the misclassification rate of supervised learning models. These factors also become the reason of high false positive rate (FPR). The classification model gets confused and mistakenly classifies the normal consumers as abnormal. The major contributions of this study are listed below.

- An AlexNet model is employed to perform feature extraction, which selects the most suitable patterns from the high dimension EC data.
- The class imbalance problem is resolved using an under-sampling technique, named as near miss (NM).
- An ensemble learning adaptive boosting (AdaBoost) model is exploited to perform the classification of malicious and non-malicious energy consumers.
- The hyper-parameter tuning of the proposed ETD model is done using artificial bee colony (ABC), which belongs to the family of the meta heuristic algorithms.
- The effectiveness of the proposed model is checked using suitable indicators, such as precision, accuracy, recall, F1-score, Matthew correlation coefficient (MCC) and area under the curve receiver operating characteristics (AUC-ROC).

The remaining manuscript is organized as follows: Section II describes the related work while Section III explains the identified problems. The proposed system model is discussed in Section IV. Section V presents the simulation results and the conclusion is presented in Section VI.

## II. RELATED WORK

This section presents the explanation of the existing tools and techniques, which are already proposed in the literature to handle NTLs. These techniques are mostly based on data driven solutions. In [10], the authors employ convolutional neural network (CNN) for performing ETD. The effectiveness of the model is checked using suitable undersampling and oversampling techniques. The authors work on ensemble approaches to differentiate between normal and malicious samples. The main purpose of their proposed model is to reduce bias and variance. In [11], the authors introduce a CNN, gated recurrent unit and particle swarm optimization (CNN-GRU-PSO) based hybrid model to detect NTLs. CNN-GRU extracts abstract and temporal features from the EC data while PSO is leveraged for hyper-parameters' optimization. The underline focus of the proposed solution is to reduce the high FPR because electric utilities have a low budget for

onsite inspection. The authors introduce a binary black hole algorithm (BHA) for handling the curse of dimensionality issue [12]. BHA performs well as compared to other meta-heuristic techniques. The BHA algorithm is a novel and evolutionary algorithm in which multiple stars participate and the star having the highest fitness value is selected for black hole. In the proposed study, the stars represent features. The features that have a large influence on the detection accuracy are selected and the remaining ones are discarded. In [13], the authors develop a hybrid DL model to detect NTLs in power grids. The hybrid model is the combination of GRU and GoogleNet. Both 1D daily and 2D weekly EC data are used to train the model. Moreover, the simulations confirm that the hybrid model performs better than the state-of-the-art models. The authors of [14] employ long short-term memory (LSTM) for NTL detection in the EC data. LSTM captures the long-term tendency from the EC data and avoids the effects of non-malicious factors. Moreover, random undersampling boosting (RUSBoost) is used for data balancing and classification. The RUSBoost model first performs under sampling of majority class and then applies an ensemble boosting strategy to perform final classification of energy thieves. Similarly, the authors of [15] utilize a combined DL-boosting model for efficient detection of NTLs. In the proposed work, feature extraction is carried out through the DL model and final classification is performed using a boosting model. In [16], the authors develop a combined ETD model, which integrates both CNN and random forest (RF) for accurate results. The CNN model performs efficient extraction of potential features while RF is utilized for the classification purpose.

In [17], a hybrid DL model is designed, which uses the advantages of both LSTM and CNN models. The proposed hybrid model efficiently extracts temporal correlations and hidden patterns from the consumers' profiles and performs ETD accordingly. In [18], a framework is developed by combining maximum overlap discrete wavelet packet transform (MODWPT) and RusBoost algorithms. The former is employed to capture time-related patterns from the EC data. While latter is utilized to differentiate between normal and malicious EC patterns. In [19], bidirectional GRU (BiGRU) and kernel principal component analysis (KPCA) are exploited to identify fraudulent electricity consumers. In the proposed study, BiGRU is utilized for the classification purpose and KPCA is leveraged to perform feature extraction. The authors of [20] propose a deep neural network, which combines an ensemble technique and a DL model. The DL model initially resolves the high dimensionality problem and then selects the relevant prominent features. Afterwards, the selected features are passed to the ensemble model for the final detection of fraudulent electricity consumers. The ETD results prove the efficacy of the proposed model over the compared models. In [21], the authors introduce a clustering based deep model for detecting energy thieves in power grids. The proposed technique utilizes density based clustering approach and builds different

clusters, which contain similar EC records. The EC samples which do not lie in any cluster are declared as energy thieves. Similarly, in [22], an ETD model is designed with the help of gradient boosting models. The boosting model increases the ETD accuracy to a greater extent. Moreover, the class imbalance problem is resolved by introducing synthetic theft attacks. The authors claim that the motive of electricity theft is to consume less energy as compared to the normal consumers. Furthermore, supervised learning methods require labeled data to learn patterns from the EC data. However, it is a labor intensive and challenging task to collect labeled data. Due to this reason, unsupervised learning models like auto encoders attract the attention of research community. In [23], the authors introduce a stacked sparse denoising autoencoder (SDAE) for the accurate detection of fraudulent electricity consumers. The results confirm that the 'unsupervised learning models perform well as compared to the supervised learning models in case of less availability of labeled data. Moreover, they introduce some noise and sparsity in the regular autoencoder to enhance its robustness against intelligent attacks. Furthermore, a meta-heuristic technique, termed as PSO, is used to find optimal hyper-parameters of SDAE. By doing this, the performance of SDAE is improved to a greater extent. In [24], the authors propose a novel technique to detect malicious patterns in the EC data. The proposed model is developed for smart homes where statistical and machine learning techniques are integrated to detect the anomalous EC patterns. Similarly, the authors of [25] design a robust ETD model to detect the following types of attacks: zero days, shunt and double tapping. In the proposed model, a pattern recognition technique is used for capturing the fraudulent EC patterns. In [26], the authors develop a novel technique for ETD by combining a clustering method with a statistical technique to explore the EC profiles of consumers. In the proposed work, the correlation among EC patterns is identified by maximum information coefficient method. Meanwhile, anomalous electricity consumers are detected by the density based clustering method. In [27], a new ensemble method is introduced to identify the anomalous patterns from the EC data. The proposed method efficiently detects real time anomaly from the EC data and overcomes the high FPR accordingly. In [28], a finite mixture model is leveraged for the soft clustering while GA is employed for the parameter tuning. The results confirm that the proposed work is an efficient solution for ETD.

### III. PROBLEM ANALYSIS

Electricity thefts are harmful to the power grids. They not only increase the revenue losses, but also generate fatal electricity shocks. In the past, electricity thefts were captured through onsite field inspection, which is a labor intensive and financially expensive task. To overcome these problems, data driven based ML and DL techniques are employed for ETD. However, these techniques face the problems of overfitting, curse of dimensionality and imbalanced proportion of classes [13], [17], [22]. Moreover, a high

FPR problem occurs when the dataset is imbalanced and the classifier has limited abilities to learn long-term temporal patterns [29]–[31]. Keeping the above issues in view, a hybrid model is developed to perform ETD in power grids. Moreover, it is important to mention that this article is the extension of our conference paper [32].

**IV. PROPOSED METHODOLOGY**

This section provides the in depth discussion of the proposed ETD model. Fig. 1 shows the working mechanism of the model. Furthermore, different components of the model are explained below.

**A. DATA PREPROCESSING**

The underlying purpose of preprocessing is to give a standardized structure of the EC data before training and testing of the model. In this stage, various data preprocessing operations are performed, such as filling missing values, handling outliers and normalizing data. The description of each operation is described below.

**1) HANDLING MISSING VALUES**

In this work, the real-time smart meter dataset is used that is collected through onsite inspection. Therefore, it contains missing and noisy values. In most cases, missing values in data are denoted by not a number (NaN). If data with missing values is fed into ML or DL models, it decreases their performance. In the existing studies, different approaches are introduced to cope with the missing values. In this work, a linear interpolation (LI) method is utilized to fill the missing values [33]. The LI method takes the average of the next day and the previous day EC values and fills the missing values accordingly.

**2) REMOVING OUTLIERS**

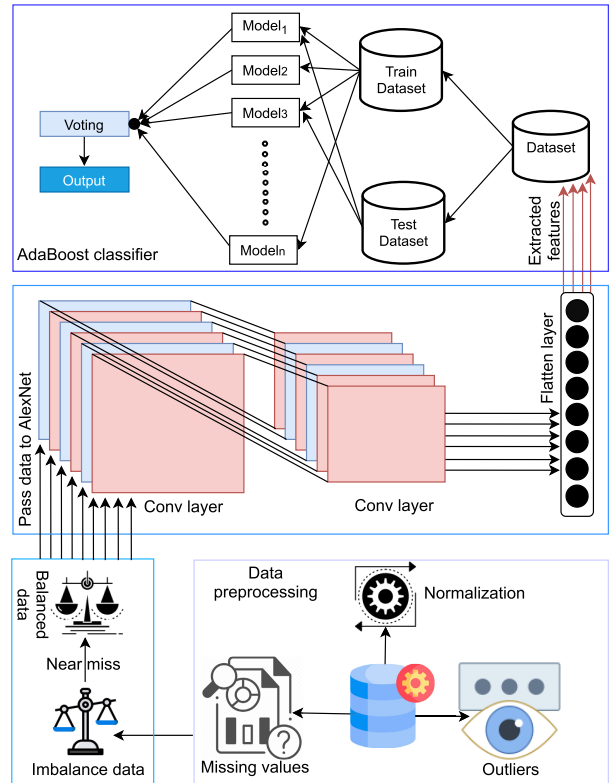
The outlier is a value that has significantly different behavior from other observations or values in a dataset. The presence of outliers disturbs the performance of classification models. In the proposed work, three-sigma rule of thumb is exploited to handle the outliers present in the EC profiles of consumers [33].

**3) NORMALIZING DATA VALUES**

DL models poorly performed on the diverse range of values. Moreover, these values negatively affect the learning process of DL models and push them towards the gradient exploding problem. In this regard, we use min-max normalization to scale the values of features between 0 and 1 [33].

**B. SOLVING UNEVEN DISTRIBUTION OF CLASSES**

The uneven distribution of classes is a critical problem being faced when performing ETD. In real world scenarios, normal electricity consumers are more than the abnormal consumers. The less availability of abnormal consumers creates the problem of imbalance data distribution, which adversely affects the performance of supervised learning models. Furthermore, the classification models are skewed towards the larger class



**FIGURE 1. Proposed ETD model.**

and overlook the smaller class, which generates false results. Therefore, in this study, an undersampling technique is utilized to resolve the problem of imbalance data. Synthetic Minority Oversampling Technique (SMOTE) is an oversampling technique, which is mostly used to handle uneven distribution of class samples [29] and [30]. It takes samples of both classes as input and generates synthetic samples of the minority class until both classes have equal distribution. However, this sampling technique raises an overfitting problem because of creating duplicate records. So, we use an under-sampling technique, termed as NM, to solve the class imbalance problem. This technique intelligently eliminates instances from the larger class to equalize the distribution of both classes. It eliminates those points, which are near the decision boundary. The working of NM is illustrated in Fig. 2. Moreover, the steps of NM are defined below.

- 1) First, it calculates the distance between all points of minority and majority classes.
- 2) Then, it selects those instances, which have least distance to the smaller class. The  $n$  number of instances will be eliminated.
- 3) If there are  $m$  instances in the minority class, then the algorithm will return  $m * n$  instances of the majority class.

**C. PROPOSED DEEP LEARNING BASED ENSEMBLE MODEL**

In this work, a hybrid model is developed for ETD that integrates the benefits of both AlexNet and AdaBoost.



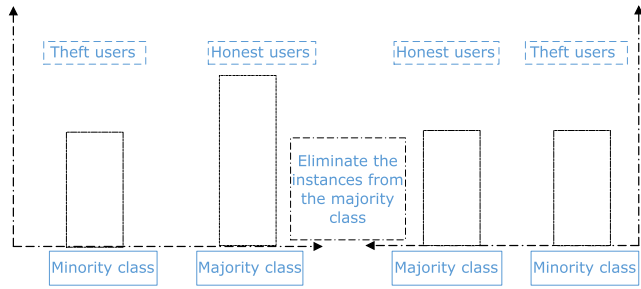


FIGURE 2. Block diagram of NM.

The AlexNet model extracts latent and dense characteristics from the EC profiles of consumers. Whereas, the final ETD is carried out through AdaBoost. In [16], the authors introduce a combined model by integrating CNN and RF in a sequential manner. They prove that the combined models give more satisfactory results than the standalone ML and DL models. Therefore, after getting motivation from [16], we develop a hybrid model to perform ETD in power grids. The proposed model combines two modules. The first module contains AlexNet model, which is used to learn features. The AlexNet model is a variant of CNN, which overcomes its limitations and gives better performance results. While the second module consists of an AdaBoost classifier, which takes extracted features of the AlexNet model as input and differentiates between normal and malicious patterns. The performance of AdaBoost is increased by enhancing the number of decision trees. Moreover, the ABC algorithm is practiced to select suitable hyper-parameters for both AlexNet and AdaBoost models. The description of each module is given below.

1) AlexNet MODEL AS FEATURE EXTRACTOR

The EC data is collected through smart meters. It often contains missing and noisy values. The noise in the EC data may be in terms of anomalies, outliers, overlapping and redundant records, missing records, inconsistent EC values, etc. These noises should be handled, otherwise, the proposed ETD model produces false results and increases the FPR to a greater extent. Therefore, in this work, we adopt essential preprocessing techniques to handle the noises. The anomalies and outliers are handled through three sigma rule, missing values are filled through LI and the inconsistent values are tackled through normalization. Moreover, the irrelevant and noisy features are discarded by the AlexNet model. The AlexNet model auto selects the relevant features and reduces the effects of noises to a minimal level. In addition, the selection of suitable EC features is a necessary task towards performing efficient ETD. Therefore, we utilize AlexNet to extract hidden and dense features from the consumers’ profiles. It was designed to reduce the shortcomings of the traditional models of that time like LeNet [34]. The architecture of AlexNet is similar to the LeNet model. However, it has more number of filters as compared to the LeNet model. It comprises of convolution, pooling and fully

connected layers. Convolution layers are used to attain abstract and latent features while pooling layers are helpful to get high level features, which reduce the curse of dimensionality issues. Moreover, dropout layers are used in place of regularization techniques to control the overfitting problem. However, dropout layers increase the training time of the AlexNet model. The basic architecture of the AlexNet model is illustrated in Fig. 3. Moreover, the explanation of each component is given below.

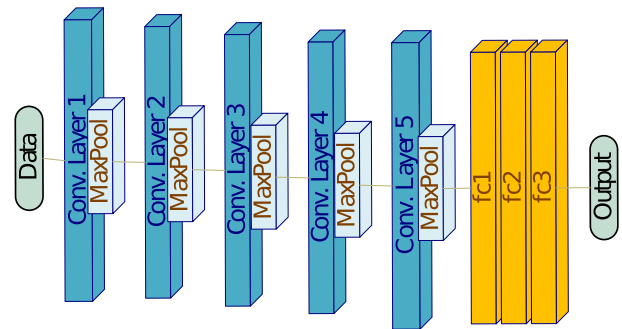


FIGURE 3. Block diagram of AlexNet .

a: CONVOLUTION LAYER

Convolution layers are an important part of CNN. These layers apply two dimensional (2D) filters on input images to extract the high level features that have high relevancy with predictive labels. Moreover, these layers are also used to obtain features from one dimensional (1D) and three dimensional (3D) data. In this study, the EC data is converted into 2D weekly data. In the literature, some studies show that it is possible to extract high level features from EC data if we align it according to weeks. 2D filters are moved on the input feature vector and different feature maps are created accordingly. These feature maps are the high level representation of 2D input data. Moreover, the optimal values of these filters are attained using stochastic gradient descent method, which gives better results as compared to handcrafted filters. Furthermore, Fig. 4 shows the architecture of convolution layer. The mathematical representations of the steps used in the convolution layers are as follows. Dim (input shape) =  $(n_h, n_w)$  where  $n_h$  denotes features and  $n_w$  depicts observations. Filter =  $(f_w, f_h)$  where  $f_w$  and  $f_h$  represent width and height of the filter, respectively. The convolution operation between input shape and filters is described by the equation given below [35].

$$Conv(inputshape, filter) = \sum_{i=1}^{n_h} \sum_{j=1}^{n_w} f_{ij}. \tag{1}$$

The learnable process in the convolution is described as [11].

$$y_{conv}(X_t^{ft}) = \sigma \left( \sum_{ft=1}^{ft} W_t^{ft} * X_t^{ft} + b_t^{ft} \right), \tag{2}$$

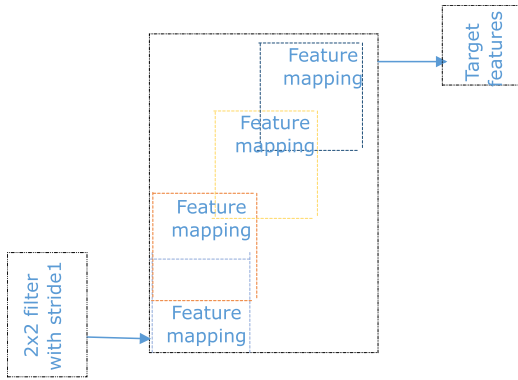


FIGURE 4. Architecture of convolution layer.

where  $X_t^{ft}$  is the output feature map after the application of filter  $f^t$ .  $W_t^{ft}$  and  $b_t^{ft}$  refer to the learnable parameters, which describe the weight and bias factors.  $\sigma$  represents the rectified linear unit (ReLU) activation while  $*$  is a dot product between filters and features.

*b: POOLING LAYER*

Convolution layers are very helpful because they extract the low-level features from the EC data. These layers extract the precise location of features. As a result, a minor change in the input data creates a different feature map. These changes happen due to cropping, rotation and flipping of input images. Researchers use signal processing techniques to decrease the resolution size of feature map. However, these techniques do not give good results and decrease the performance of DL models. Pooling layers are a new concept in the DL models. They are applied after performing the convolution operations. These layers diminish the spatial dimension of feature map without affecting the original resolution of input data. Moreover, they overcome the computational overhead issues. In the literature, several pooling operations are introduced by researchers. Among them, max-pooling performs better than the rest. Therefore, we adopt max-pooling strategy reduce the spatial dimensions of the EC data [36]. The max-pooling strategy only keeps the largest value from the specific area of a feature map and discards the rest of the values. Fig. 5 represents the working mechanism of the max-pooling operation. The mathematical representation of the max-pooling layer is as follows [35].

$$y^m = \max_{i,j} \in R(y_{i,j}), \tag{3}$$

where  $y^m$  shows the outcomes of max-pooling operations and  $R$  represents the set of real numbers.  $i$  and  $j$  denote the  $i_{th}$  convolution layer and the  $j_{th}$  neuron, respectively.

*c: ACTIVATION FUNCTION*

After convolution operations, activation functions are applied on the extracted feature maps. In the literature, different activation functions are available like tanh, linear, sigmoid, ReLU, Leaky ReLU (LReLU), etc. Each function has its own advantages and disadvantages like sigmoid and tanh

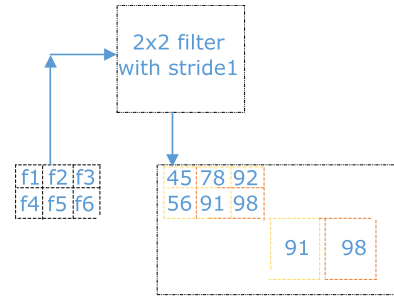


FIGURE 5. Working of max-pooling layer.

functions create exploding and vanishing gradient problems, which disturb the performance of DL models. In this study, we use ReLU as an activation function after each convolutional layer. This function converts negative values into zeros and passes the remaining values to the upcoming convolutional layers without any changes. The research work shows that ReLU function overcomes exploding and vanishing gradients problems of DL models and enhances their performance. The mathematical equation used for ReLU function is given below [33].

$$Relu = \max(0, value). \tag{4}$$

*d: DROPOUT LAYER*

DL models are sensitive to overfitting with a few training examples of datasets. Ensemble DL models are introduced to overcome this issue. However, these models have their own disadvantages and are difficult to maintain the training of multiple models. A single neural network can be simulated with different architectures by introducing the concept of dropout layers. These layers deactivate some neurons of hidden layers to overcome the overfitting and poor generalization issues. The average activation rate of the dropout layer is from 0 to 1. However, the literature shows that DL models give optimal results when the dropout rate is set at 0.5 where half number of neurons are deactivated during the training process. Moreover, Fig. 6 shows the activation and deactivation of neurons before and after the dropout operations.

*e: FLATTEN LAYER*

After performing above mentioned operations, feature maps are converted into 1D data and passed into a simple neural network to differentiate between normal and malicious EC patterns. However, in this study, output of flatten layer is considered as an extracted feature set, which is obtained after removing the overlapping and noisy values. This feature set is a more suitable representation of the EC data. Fig. 7 represents the working of flatten layer.

*f: FULLY CONNECTED LAYER*

This layer is the last layer of AlexNet [37]. It connects the neurons of preceding layers to the neurons of upcoming

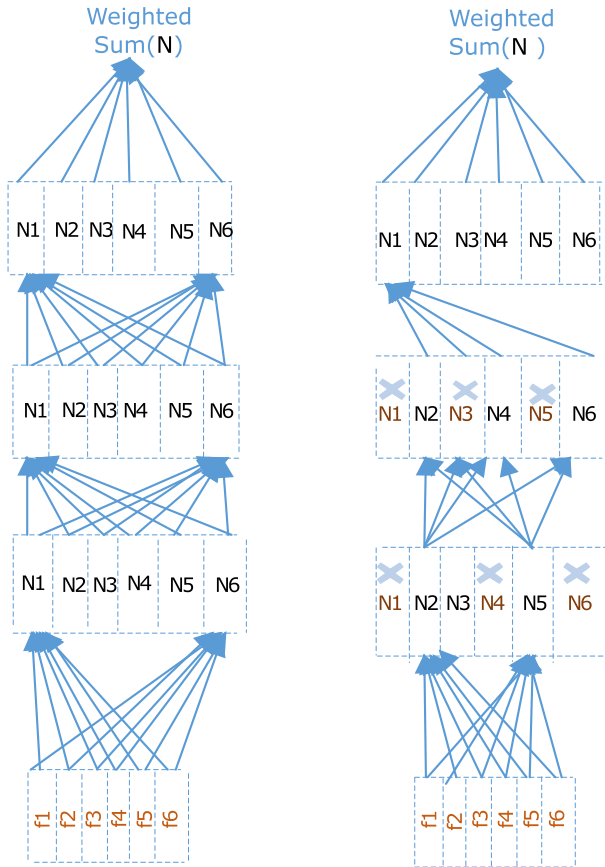


FIGURE 6. Standard layer versus dropout layer.

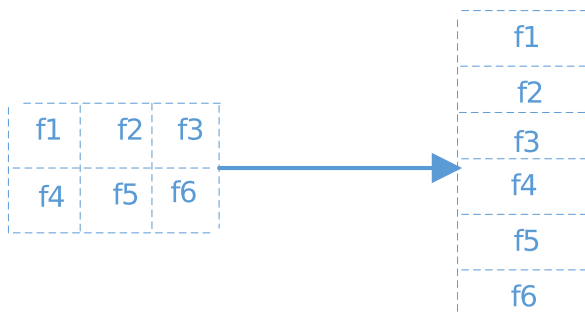


FIGURE 7. Working of flatten layer.

layers. Moreover, it extracts global features from the provided feature maps. It also compiles the result of the previous layer to perform final classification. The architecture of fully connected layer is shown in Fig. 8. Furthermore, the mathematical formula used for the fully connected layer is presented below.

$$f_o = \sigma(W_t \cdot x_t + b_t), \tag{5}$$

where  $x_t$  represents input feature vector.  $W_t$  and  $b_t$  denote weight and bias factors, respectively.  $\sigma$  shows the ReLU function while  $f_o$  represents the final output. ReLU is used after each convolution to avoid the vanishing gradient problem.

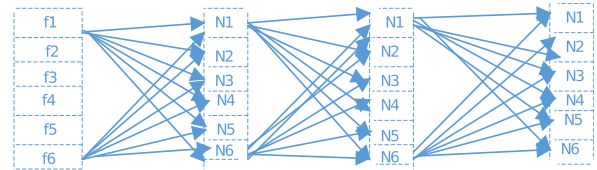


FIGURE 8. Architecture of fully connected layer.

The basic purpose of dropout layer is to solve overfitting and vanishing gradient problems.

### 2) AdaBoost MODEL AS CLASSIFIER

Ensemble methods attract the attention of the research community because they win many Kaggle competitions. Kaggle is a data science community, which launches different competitions to solve real-world problems [38]. Ensemble methods are of two types: bagging and boosting. In former methods, multiple weak learners are initially trained on different ratios of data and then a voting mechanism is used to decide about the normal and malicious samples. RF is a bagging method, which trains multiple decision trees on data. The bagging methods give good results if all weak learners get diverse knowledge from the EC data. However, their performance is decreased when the results of weak classifiers are overlapped. Contrarily, in boosting methods, different weak learners are integrated in a sequential manner to make a strong learner.

The AdaBoost classifier is based on the boosting approach where high weights are assigned to the wrong predictions of a weak learner and passed to the next weak learner for better classification. So, the extracted features of AlexNet are passed to AdaBoost classifier, which uses ensemble approach to differentiate between normal and malicious EC patterns. Through this approach, the performance of the standalone ML classifier is increased by reducing loss function's value iteratively. The working mechanism of AdaBoost classifier is shown in Fig. 9.

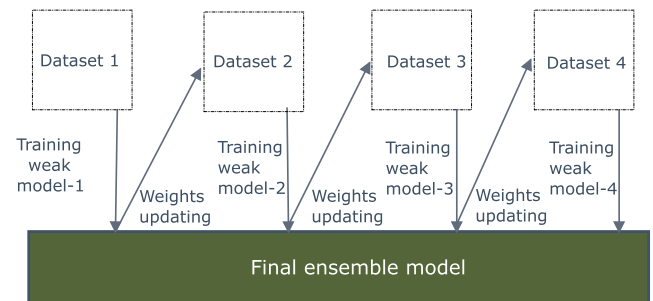


FIGURE 9. Workflow of AdaBoost classifier.

### 3) TUNING HYPER-PARAMETER OF THE PROPOSED MODEL THROUGH ENHANCED ABC ALGORITHM

In [39], the authors proposed a hybrid model by combining a Jaya optimization algorithm with an ensemble learning model. The model is introduced to identify the transient

stability status of power systems. The Jaya algorithm has significantly increased the efficiency of the proposed model by finding the optimal hyper-parameters. Therefore, being motivated from this work, we employ ABC algorithm for hyper-parameters' tuning of the proposed ETD model. The ABC algorithm comes under the umbrella of meta heuristics techniques [40]. It was founded by Karaboga in 2005. Mostly, it is used for optimization tasks. It mimics the working procedure of the honey bees when they search for food. The collection of honey bees is called a swarm. In ABC algorithm, there are three types of bees: employee, onlooker and scout. The employee bee initially searches the food sources using its own intelligence and memory power and then informs the onlooker bees about these sources. The onlooker bee finds the rich source of food through the information provided by the employee bee. The onlooker bee chooses those food sources, which have good quality of fitness. The scout bee is picked from the onlooker bees. This bee tries to find new sources and abandon the old ones for a rich food source [35]. The total number of solutions, onlooker bees and employee bees are equal in the swarm [36]. Fig.10 exhibits the working behavior of ABC.

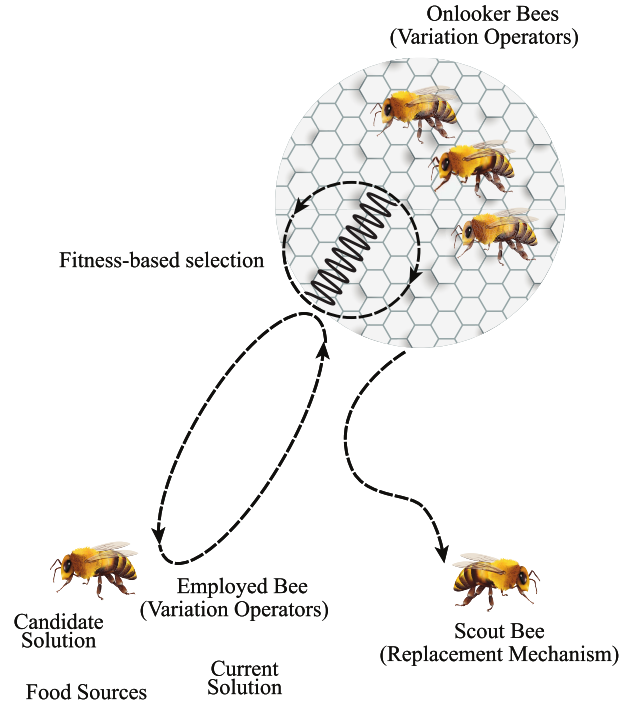


FIGURE 10. Schematic diagram of ABC algorithm.

**Algorithm 1** Proposed Model’s Overall Working

- 1: Take  $Imb_{dst} \in SGCC$ .
- 2: **Data preprocessing phase:**
- 3: Apply  $NM(Imb_{dst})$  to make it  $Bal_{dst}$ .
- 4: Apply min-max, interpolation ( $Bal_{dst}$ )
- 5: **Data splitting phase**
- 6: Split  $Bal_{D_{st}}$  into  $Bal_{d_{train}}$ ,  $Bal_{d_{test}}$
- 7: **Feature extraction phase:** AlexNet is used to select the  $features_{relevant}$  form the  $Bal_{D_{st}}$
- 8: **Hyper-parameter tuning based binary classification:**
  - i.  $features_{relevant}$  are passed to AdaBoost technique for classification
  - ii. The best well tuned ensemble model  $Ens_{AdaBoost-ABC}$  is trained on the  $Bal_{D_{strain}} \in Adaboost_{parameters}$  which are  $max_{depth}$ ,  $n_{estimator}$  for the  $features_{relevant}$  to meet the best convergence
- 9: **Loss calculation:**
- 10: Loss is computed on  $Bal_{dst}$  by the binary cross entropy from the  $features_{relevant}$  on the basis of  $Ens_{AdaBoost-ABC}$
- 11: If loss is minimum then goto step 15 otherwise goto step 8 (ii)
- 12: **Benchmark techniques**
- 13: Performance comparison is performed with benchmark techniques
- 14: **Performance evaluation metrics**
- 15: Performance of the model is measured through suitable metrics

The ABC algorithm is a well-known and powerful optimization algorithm. The selection of ABC is made because of its powerful exploration abilities and some unique properties

compared with other optimization algorithms. These properties are as follows. The ABC algorithm has fewer control parameters as compared to other optimization algorithms like PSO and GA [40]. It efficiently handles stochastic cost objective function. Furthermore, the convergence rate of ABC is fast due to the separate phases for exploitation and exploration as compared to other optimization techniques [41]. Moreover, the authors of [42], [43] exploit ABC algorithm for hyper-parameters' tuning of the proposed deep learning models. The simulation results prove that the selected hyper-parameters of ABC greatly improve the performance results. Therefore, based on these unique properties and being motivated from [42], [43], we utilize ABC algorithm for hyper-parameters' tuning of the proposed ETD model. The founded hyper-parameters of ABC improves the ETD performance to a greater extent.

Besides, it is important to discuss that the traditional ABC algorithm has the problem of premature convergence in which the algorithm gets stuck in exploitation phase due to less diversity in the population. Therefore, in this work, an advance variant of the ABC algorithm is used that is proposed by [44]. In the modified version, a new control parameter, named as cognitive learning factor (CLF), is introduced to regulate the current position of candidate solution in employee and onlooker bee phases. This enhancement significantly reduces the problem of premature convergence in ABC. Through simulations, it is seen that the proposed model efficiently performs on the selected hyper-parameters of the enhanced ABC algorithm. Moreover, the working of the proposed ETD model is given in Algorithm 1.



**TABLE 2. Mapping of identified problems and suggested solutions with validations.**

Identified problems	Suggested solutions	Validations
L1: Uneven distribution of data	S1: NM	V1: Tables IV-VIII
L2: Curse of dimensionality	S2: AlexNet	V2: Figs. 11 and 12
L3: Overfitting	S3: Reducing overlapping samples	V3: Fig. 13
L4: High FPR	S4: AlexNet-AdaBoost-ABC	V4: Figs. 14 and 15

**V. SIMULATION RESULTS**

In this section, the simulation results and discussion of the proposed and baseline models are provided. To perform simulations, we use a real dataset of smart meters. The detail of dataset is provided in Section V-A. Moreover, Python 3.0 and Google Colab are used for the implementation of the proposed and baseline models.

**A. DATASET ACQUISITION**

The dataset used in this work is released by the state grid of China (SGCC), a well-known electric utility in China [33]. The dataset is publicly available on the Internet. It comprises the EC records of 42,372 consumers from Jan 1, 2014 to Oct 31, 2016. In the dataset, rows exhibit the overall records of electricity consumers and columns denote the daily electricity usage. Table 3 presents the explanation of the dataset.

**TABLE 3. Description of dataset.**

Description	Values
Electricity consumption time window	2014/01/01– 2016/10/31
Total customer	42,372
normal users	38,757
fraudulent users	3615
Select dataset size	7230
Samples before applying NM	38,757
Samples after applying NM	3615

**B. PERFORMANCE METRICS**

The suitable selection of performance metrics is a necessary task in the case of imbalance classification. The AUC-ROC metric is a reliable indicator for imbalanced classification problems. It computes the difference between positive and negative class samples. The value of AUC ranges between 0 and 1. Moreover, it plots TPR and FPR at different classification thresholds. The mathematical formula used for AUC is given below [4].

$$AUC = \frac{\sum Rank_{i \in positiveclass} - \frac{S(1+T)}{2}}{S * T}, \tag{6}$$

where *S* shows positive class and *T* represents negative class. The classification model performs well when the value of AUC is 1. Similarly, if the value of AUC is 0.5, the model performs by random guessing.

$$REC = \frac{TruePositive}{TruePositive + FalseNegative}, \tag{7}$$

$$PRE = \frac{TruePositive}{TruePositive + FalsePositive}, \tag{8}$$

$$F1 - Score = 2 * \frac{PRE * REC}{PRE + REC}, \tag{9}$$

where *PRE* and *REC* denote precision and recall, respectively. Moreover, precision is another useful metric to measure the relevancy of total positive predicted results. Similarly, recall calculates the number of truly predicted positive samples out of the actual class labels. In addition, F1-score measures the balance ratio between precision and recall for better evaluation of the model [36]. On the other hand, MCC is another suitable performance metric, which provides the correlation between positive and negative predictions [45]. The formula used to calculate the MCC score is described below [45].

$$MCC = \frac{TP * TN - FP * FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}, \tag{10}$$

where *TP*, *FP*, *FN* and *TN* represent true positive, false positive, false negative and true negative, respectively. All of these outcomes are computed through a confusion matrix, which provides the summary of the overall correct and incorrect predictions of the classification model.

**C. MAPPING OF IDENTIFIED PROBLEMS WITH SUGGESTED SOLUTIONS**

Table 2 presents the mapping of identified problems with suitable suggested solutions. Moreover, the problems, solutions and validations are labeled as L, S and V, respectively. L1 is about class imbalance problem, which is solved through NM. In L2, curse of dimensionality issue is discussed, which is handled through the AlexNet model. The overfitting problem is highlighted in L3, which is solved through generating synthetic theft data. In L4, the problem of high FPR is discussed, which is minimized through the proposed hybrid model.

**D. BENCHMARK MODELS AND THEIR SIMULATION RESULTS**

The description and the performance analysis of the baseline models are provided in this section.

1) SUPPORT VECTOR MACHINE

It is a popular classification technique, which is mostly used in the ETD. It is mostly used for a binary classification task. The simulation results of support vector machine (SVM) [46] using SMOTE, NM and imbalance data distribution are given

TABLE 4. Classification results of SVM.

Evaluation metrics	Imbalanced data distribution	SMOTE	NM
Precision	0.39	0.64	0.65
Recall	0.48	0.60	0.61
Accuracy	0.43	0.58	0.60
F1-score	0.25	0.38	0.40
MCC	0.34	0.52	0.39

in Table 4. The results exhibit that SVM performs well on NM instead of SMOTE and imbalanced data distribution. However, it underperforms on all baseline models. The reason is that it draws  $n - 1$  hyperplanes for making the classification boundary where  $n$  represents number of features. The EC data is high dimensional, which increases the computational overhead of SVM to a great extent. Moreover, the convergence speed of the SVM model is greatly affected. Fig. 13 shows AUC-ROC of SVM during training and validation process.

2) LOGISTIC REGRESSION

Logistic regression (LR) is a well-known ML model, which is mostly used for binary classification tasks [47]. It is a probabilistic model. In LR, the final classification is performed using a sigmoid function. Table 5 presents the results of LR. The LR model performs better than SVM using NM data balancing technique. However, it poorly performs for the SMOTE technique as compared to SVM.

TABLE 5. Classification results of LR.

Evaluation metrics	Without handling imbalanced class	SMOTE	NM
Precision	0.45	0.58	0.68
Recall	0.38	0.45	0.64
Accuracy	0.41	0.56	0.67
F1-score	0.29	0.35	0.66
MCC	0.20	0.38	0.42

3) CONVOLUTIONAL NEURAL NETWORK

It is a deep learning model, which is mostly practiced in ETD as a benchmark model [17], [48]. Firstly, CNN uses different convolutional and pooling operations to capture the prominent features from the consumption profiles of consumers. Subsequently, the classification is carried out accordingly. The ETD results of CNN are provided in Table 6.

E. PROPOSED AlexNet-AdaBoost-ABC

The performance analysis of the proposed hybrid model is presented in this section. Fig. 11 and Fig. 12 show the learning process of the model in terms of loss and accuracy. The blue curve denotes training loss and orange curve exhibits testing loss. In the figures, the decrement in loss and increment in accuracy prompt that the hybrid model

TABLE 6. Classification results of CNN.

Evaluation metrics	Imbalanced data distribution	SMOTE	NM
Precision	0.49	0.73	0.76
Recall	0.51	0.75	0.78
Accuracy	0.55	0.76	0.77
F1-score	0.55	0.72	0.79
MCC	0.28	0.55	0.66

TABLE 7. Classification results of AlexNet-ABC.

Evaluation metrics	Imbalanced data distribution	SMOTE	NM
Precision	0.59	0.77	0.93
Recall	0.68	0.82	0.84
Accuracy	0.65	0.80	0.88
F1-score	0.63	0.81	0.88
MCC	0.35	0.62	0.78

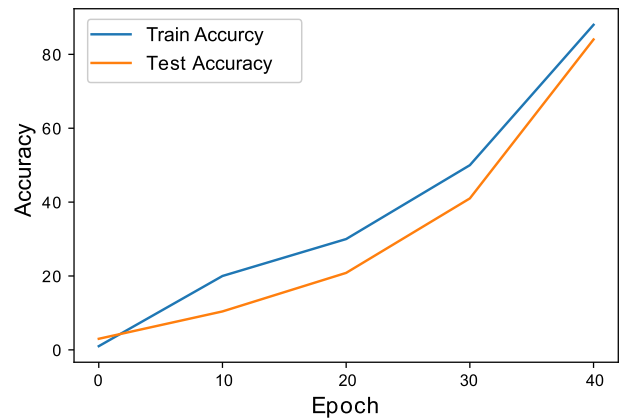


FIGURE 11. Accuracy curve of the proposed model.

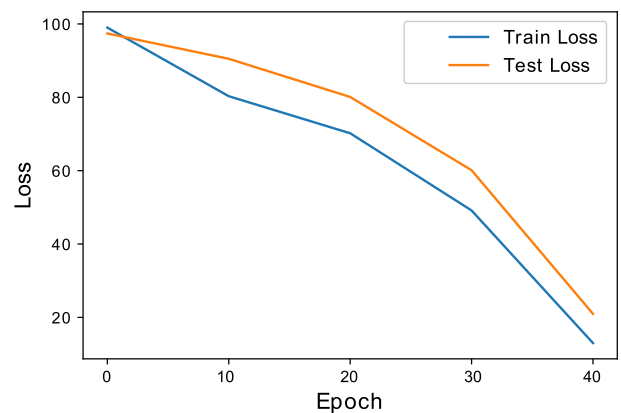


FIGURE 12. Loss curve of the proposed model.

intelligently learns EC patterns. It is seen that the model converges quickly due to the self learning of latent features. Table 7 and Table 8 describe the overall performance results of the AlexNet-ABC and the proposed model, respectively.

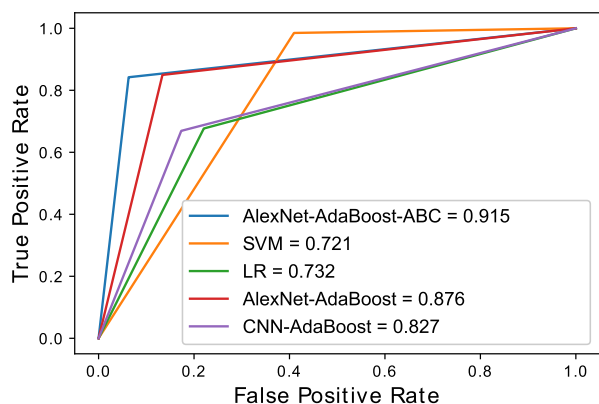


FIGURE 13. ROC curve of the proposed and baseline classifiers.

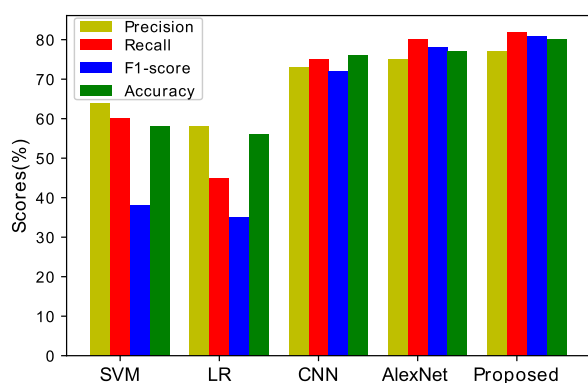


FIGURE 14. Performance of the proposed and baseline models using SMOTE.

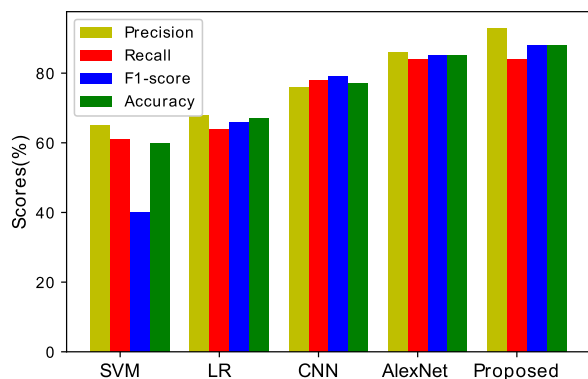


FIGURE 15. Performance of the proposed and baseline models using NM.

Moreover, Fig. 13 illustrates AUC-ROC of the proposed and benchmark models. The proposed model obtains the highest AUC-ROC of 0.91 due to the strong feature learning abilities of AlexNet and the integration of ABC for hyper-parameter tuning. The ABC attempts to find the suitable parameter set for both AlexNet and AdaBoost models. By doing this, the convergence rate of the hybrid model is increased to a greater extent. Moreover, high value of AUC indicates that the model efficiently differentiates between the positive and the negative class samples. Furthermore, Fig. 14 and Fig. 15

TABLE 8. Classification results of the proposed AlexNet-AdaBoost-ABC.

Evaluation metrics	Imbalanced distribution data	SMOTE	NM
Precision	0.55	0.75	0.86
Recall	0.61	0.80	0.84
Accuracy	0.59	0.77	0.88
F1-score	0.58	0.78	0.85
MCC	0.32	0.62	0.78
AUC-ROC	0.56	0.83	0.91

show the performance results of the proposed and baseline models using SMOTE and NM sampling techniques. From the figures, it is seen that the model performs well using NM as compared to SMOTE. The reason is that SMOTE generates overlapping samples after a certain threshold, which leads to overfitting issue. However, NM intelligently removes the overlapping samples from the larger class. That is why it gives better results as compared to SMOTE. Similarly, Fig. 16 and Fig. 17 show the MCC score of the proposed and baseline models using NM and SMOTE sampling techniques, respectively. The proposed model obtains the highest MCC score using NM as compared to SMOTE. This implies that the model successfully maintains the correlation between positive and negative predicted classes. This achievement enables the application of the proposed model in real world scenarios. Moreover, it is important to mention that the basic building block of our proposed ETD model is based on data driven models [37]. These models perform more efficiently when the dataset is huge. In this work, a massive dataset of smart meters is leveraged to train the proposed ETD model. So, the model learns EC patterns more efficiently when a large number of EC records is used. As a result, the proposed ETD model easily deals with the less or high volume of data, which shows the scalability of the proposed ETD model. Furthermore, the training time of the proposed and benchmark models is shown in Fig. 18. The figure illustrates that the computational complexity of the proposed ETD model is greater than the baseline models except SVM. The reason is that the integration of ABC incurs extra time to find the optimal subset of hyper-parameters. However, there is a tradeoff between computational time and detection accuracy. The ETD results of the proposed model are better than its counterparts. So, in ETD, accurate detection of energy thieves is more important instead of computational time. Moreover, SVM has the highest computational overhead because it does not perform well on non-linear and high dimensional data. In SVM,  $n - 1$  ( $n$  represents features) hyperplanes are drawn and only one hyperplane is selected from them having maximum margin from the support vectors. Due to this reason, it has the highest computational overhead. In contrast, LR has the least execution time because it consists of a single layer of neural networks. However, the performance of LR is not satisfactory because it is unable to capture complex non-linearity from the EC data. In addition, CNN consumes less execution time as

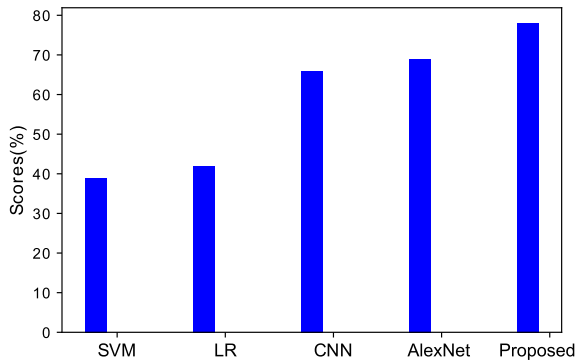


FIGURE 16. MCC score of the proposed model using NM.

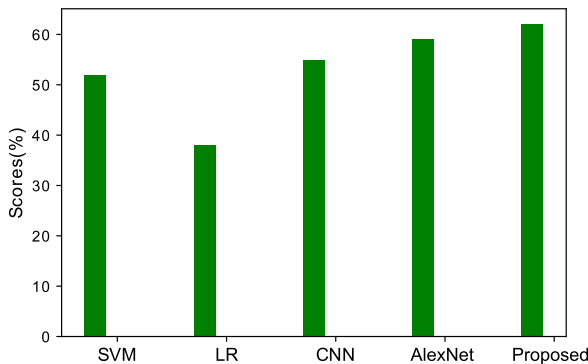


FIGURE 17. MCC score of the proposed model using SMOTE.

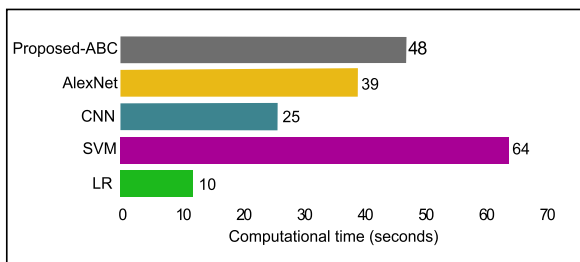


FIGURE 18. Training time of ETD models.

compared to other deep models. However, it underperforms while achieving satisfactory ETD accuracy. Similarly, the computational complexity of the AlexNet model is more than CNN and LR because of extra convolution and max pooling layers.

## VI. CONCLUSION AND FUTURE WORK

In this article, a hybrid deep learning model is introduced for the detection of malicious and non-malicious electricity consumers. The proposed model combines AlexNet, AdaBoost and ABC optimization algorithm. The Alexnet model is employed to cope with the high dimensionality problem. Whereas, the AdaBoost model is utilized to differentiate between honest and dishonest electricity consumers. Moreover, the problem of imbalanced data distribution is handled using an undersampling technique, named as NM.

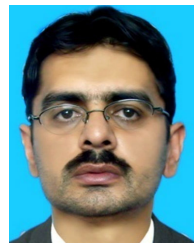
The NM technique intelligently equalizes the proportion of classes by minimizing the majority class samples. Furthermore, ABC optimization algorithm is exploited for hyper-parameters’ tuning of the proposed ETD model. The efficacy of the proposed model is ensured on a realistic smart meters’ dataset that is taken from SGCC, a well-known electric utility in China. Afterwards, the appropriate performance metrics are opted to measure the effectiveness of the proposed ETD model. The simulation results exhibit that the proposed ETD model outperforms the benchmark models. Our model achieves 88%, 86%, 84%, 85%, 78% and 91% accuracy, precision, recall, F1-score, MCC and AUC-ROC scores, respectively. Besides, the proposed model has some shortcomings. The model is trained on only EC dataset without considering other non-sequential factors, which limits its performance towards identifying the accurate location of the energy thieves. Moreover, the low-sampled EC data is considered, which affects the ability of the proposed model while capturing more granular information about energy theft. So, in the future, we will consider high sampling EC data and other non-sequential data for the accurate identification of energy thieves.

## REFERENCES

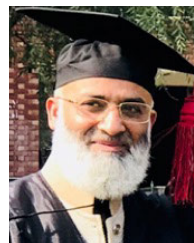
- [1] J. A. Meira, P. Glauner, R. State, P. Valtchev, L. Dolberg, F. Bettinger, and D. Duarte, “Distilling provider-independent data for general detection of non-technical losses,” in *Proc. IEEE Power Energy Conf. Illinois (PECI)*, Feb. 2017, pp. 1–5.
- [2] K. M. Ghori, R. A. Abbasi, M. Awais, M. Imran, 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.
- [3] Z. Hussain, S. Memon, R. Shah, Z. A. Bhutto, and M. Aljawarneh, “Methods and techniques of electricity thieving in Pakistan,” *J. Power Energy Eng.*, vol. 4, no. 9, pp. 1–10, 2016.
- [4] 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, Oct. 2019, Art. no. 106458.
- [5] 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.
- [6] 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.
- [7] C.-H. Lo and N. Ansari, “CONSUMER: A novel hybrid intrusion detection system for distribution networks in smart grid,” *IEEE Trans. Emerg. Topics Comput.*, vol. 1, no. 1, pp. 33–44, Jun. 2013.
- [8] S. Amin, G. A. Schwartz, and H. Tembine, “Incentives and security in electricity distribution networks,” in *Proc. Int. Conf. Decis. Game Theory Secur.* Berlin, Germany: Springer, 2012, pp. 264–280.
- [9] S. K. Gunturi and D. Sarkar, “Ensemble machine learning models for the detection of energy theft,” *Electr. Power Syst. Res.*, vol. 192, Mar. 2021, Art. no. 106904.
- [10] J. Pereira and F. Saraiva, “Convolutional neural network applied to detect electricity theft: A comparative study on unbalanced data handling techniques,” *Int. J. Electr. Power Energy Syst.*, vol. 131, Oct. 2021, Art. no. 107085.
- [11] A. Ullah, N. Javaid, A. S. Yahaya, T. Sultana, F. A. Al-Zahrani, and F. Zaman, “A hybrid deep neural network for electricity theft detection using intelligent antenna-based smart meters,” *Wireless Commun. Mobile Comput.*, vol. 2021, pp. 20–25, Aug. 2021.
- [12] 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.



- [13] F. Shehzad, N. Javaid, A. Almogren, A. Ahmed, S. M. Gulfam, and A. Radwan, "A robust hybrid deep learning model for detection of non-technical losses to secure smart grids," *IEEE Access*, vol. 9, pp. 128663–128678, 2021.
- [14] M. Adil, N. Javaid, U. Qasim, I. Ullah, M. Shafiq, and J.-G. Choi, "LSTM and bat-based RUSBoost approach for electricity theft detection," *Appl. Sci.*, vol. 10, no. 12, p. 4378, Jun. 2020.
- [15] Z. A. Khan, M. Adil, N. Javaid, M. N. Saqib, M. Shafiq, and J.-G. Choi, "Electricity theft detection using supervised learning techniques on smart meter data," *Sustainability*, vol. 12, no. 19, p. 8023, Sep. 2020.
- [16] 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.
- [17] 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.
- [18] 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.
- [19] 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.
- [20] Z. Aslam, N. Javaid, A. Ahmad, A. Ahmed, and S. M. Gulfam, "A combined deep learning and ensemble learning methodology to avoid electricity theft in smart grids," *Energies*, vol. 13, no. 21, p. 5599, Oct. 2020.
- [21] J. L. Viegas, P. R. Esteves, and S. M. Vieira, "Clustering-based novelty detection for identification of non-technical losses," *Int. J. Elect. Power Energy Syst.*, vol. 101, pp. 301–310, Oct. 2018.
- [22] 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.
- [23] Y. Huang and Q. Xu, "Electricity theft detection based on stacked sparse denoising autoencoder," *Int. J. Electr. Power Energy Syst.*, vol. 125, Feb. 2021, Art. no. 106448.
- [24] 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.
- [25] 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.
- [26] 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.
- [27] 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.
- [28] 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.
- [29] J. Brownlee. (Mar. 17, 2021). *SMOTE for Imbalanced Classification With Python*. [Online]. Available: <https://machinelearningmastery.com/smote-oversampling-for-imbalanced-classification/>
- [30] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, no. 1, pp. 321–357, 2002.
- [31] SGCC. *State Grid Corporation of China*. Accessed: Feb. 10, 2022. [Online]. Available: <http://www.sgcc.com.cn>
- [32] M. Asif, A. Ullah, S. Munawar, B. Kabir, A. Khan, and N. Javaid, "Alexnet-AdaBoost-ABC based hybrid neural network for electricity theft detection in smart grids," in *Proc. Conf. Complex, Intell., Softw. Intensive Syst. Cham, Switzerland: Springer*, 2021.
- [33] 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.
- [34] J. Wei. (Jul. 3, 2019). *11AlexNet: The Architecture that Challenged CNNs*. [Online]. Available: <https://towardsdatascience.com/alexnet-the-architecture-that-challenged-cnns>
- [35] S. Bitam, M. Batouche, and E.-G. Talbi, "A survey on bee colony algorithms," in *Proc. IEEE Int. Symp. Parallel Distrib. Process., Workshops Phd Forum (IPDPSW)*, Apr. 2010, pp. 1–8.
- [36] G. Zhu and S. Kwong, "Gbest-guided artificial bee colony algorithm for numerical function optimization," *Appl. Math. Comput.*, vol. 217, no. 7, pp. 3166–3173, Dec. 2010.
- [37] N. Javaid, "A PLSTM, AlexNet and ESNN based ensemble learning model for detecting electricity theft in smart grids," *IEEE Access*, vol. 9, pp. 162935–162950, 2021.
- [38] Kaggle. (2022). *Kaggle: Your Machine Learning and Data Science Community*. Accessed: Jan. 12, 2022. [Online]. Available: <https://www.kaggle.com/>
- [39] Y. Li and Z. Yang, "Application of EOS-ELM with binary Jaya-based feature selection to real-time transient stability assessment using PMU data," *IEEE Access*, vol. 5, pp. 23092–23101, 2017.
- [40] D. Karaboga and B. Akay, "A comparative study of artificial bee colony algorithm," *Appl. Math. Comput.*, vol. 214, no. 1, pp. 108–132, Aug. 2009.
- [41] S. Anam, "Multimodal optimization by using hybrid of artificial bee colony algorithm and BFGS algorithm," *J. Phys., Conf. Ser.*, vol. 893, Oct. 2017, Art. no. 012068.
- [42] R. Kumar, P. Kumar, and Y. Kumar, "Integrating big data driven sentiments polarity and ABC-optimized LSTM for time series forecasting," *Multimedia Tools Appl.*, pp. 1–20, May 2021.
- [43] T. Ozcan and A. Basturk, "Transfer learning-based convolutional neural networks with heuristic optimization for hand gesture recognition," *Neural Comput. Appl.*, vol. 31, no. 12, pp. 8955–8970, Dec. 2019.
- [44] J. C. Bansal, H. Sharma, A. Nagar, and K. V. Arya, "Balanced artificial bee colony algorithm," *Int. J. Artif. Intell. Soft Comput.*, vol. 3, no. 3, pp. 222–243, 2013.
- [45] N. Javaid, N. Jan, and M. U. Javed, "An adaptive synthesis to handle imbalanced big data with deep Siamese network for electricity theft detection in smart grids," *J. Parallel Distrib. Comput.*, vol. 153, pp. 44–52, Jul. 2021.
- [46] S. Ray. (Sep. 13, 2017). *Understanding Support Vector Machine (SVM) Algorithm From Examples*. [Online]. Available: <https://www.analyticsvidhya.com>
- [47] Ayush Pant. (Jun. 22, 2019). *Introduction to Logistic Regression*. [Online]. Available: <https://towardsdatascience.com/>
- [48] L. Zheng, Y. Yang, and Q. Tian, "SIFT meets CNN: A decade survey of instance retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 5, pp. 1224–1244, May 2018.



**ASHRAF ULLAH** is currently pursuing the Ph.D. degree in computer science with the Communications Over Sensors (ComSens) Research Laboratory, COMSATS University Islamabad, Islamabad Campus, under the supervision of Prof. Nadeem Javaid. His research interests include cloud computing, smart grids, electricity load forecasting, and electricity theft detection.



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



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



**MUHAMMAD UMAR JAVED** received the bachelor's and master's degrees in electrical engineering from Government College University Lahore, Lahore, Pakistan, in 2014 and 2018, respectively. He is currently pursuing the Ph.D. degree in computer science with the Communications Over Sensors (ComSens) Research Laboratory, COMSATS University Islamabad, Islamabad Campus, under the supervision of Prof. Nadeem Javaid. His research interests include smart grids, electric vehicles, and blockchain.



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

...