A PLSTM, AlexNet and ESNN based Ensemble Learning Model for Detecting Electricity Theft in Smart Grids

NADEEM JAVAID
Department of Computer Science, COMSATS University Islamabad (CUI), Islamabad 44000, Pakistan
School of Computer Science, University of Technology Sydney (UTS), Ultimo, NSW, 2007, Australia
Correspondence: www.njavaid.com, nadeemjavaid@comsats.edu.pk, nadeem.javaid@uts.edu.au

This work did not receive any funding.

ABSTRACT The problem of electricity theft is exponentially increasing around the globe, which is harmful to the power sectors and consumers. The recent development in the advanced metering infrastructure brings opportunities for experts to identify the electricity thieves in the smart grid community. Many advancements are made in the area of the smart grid for Electricity Theft Detection (ETD), where the data collected from the smart meters is utilized. However, the problems of imbalanced distribution of data and inaccurate classification are not efficiently addressed. Therefore, to overcome the problems, machine learning and deep learning models are proposed for ETD. Initially, to refine the smart meters’ data, pre-processing methods are used. Then, the class imbalance problem is solved through Synthetic Minority Oversampling Tomek Links (ST-Links). It solves the classifier’s biasness problem, which occurs due to imbalanced data. It achieves the benefits of both data oversampling and undersampling. Afterwards, an AlexNet and peephole long short-term memory network based feature extractor with an attention layer is developed to extract the relevant features from electricity consumption profiles that are most suitable to classify honest and theft consumers. After the extraction of suitable features, the classification of consumers is performed by an echo state neural network. Moreover, an evolutionary grey wolf optimization technique is utilized to tune the hyper-parameters of the proposed model. A paired t-test is also applied on the final classification results for a reliable assessment of the proposed model. The simulations are conducted on a realistic smart meters’ dataset of China to check the performance of the proposed model. In addition, different benchmark models are implemented to perform a comparative analysis. Different meaningful performance metrics are considered for the fair evaluation of the proposed model: Matthews Correlation Coefficient (MCC), F1-score, Area Under Curve (AUC), precision and recall. The simulation results depict that the proposed model obtains accuracy, recall, F1-score, AUC, PR-AUC, precision and MCC score of 96.3%, 92.1%, 92.0%, 96.4%, 97.3%, 90.0% and 84.0%, respectively. It is worth mentioning that the application of the proposed solution is quite general. Therefore, it can be used by the power companies to overcome the power losses in the energy sector.

INDEX TERMS AlexNet, Data pre-processing, electricity theft detection, echo state neural network, imbalanced data, peephole LSTM, supervised learning, smart meter.

I. INTRODUCTION

Electricity Theft (ET) is the stealing of electric power from electricity distribution systems. It is performed by meter tampering, bypassing meters, billing irregularities, etc., [1]. Recently, the power sectors have raised concerns about ET because they have to bear huge financial losses and energy inefficiencies. Worldwide, the electric utilities lose $25 billion per year due to ET [2]. The negative financial impact of this act is faced both by the developed and the developing nations. Table 1 shows the statistics of power losses incurred due to ET in different countries. In the USA, the electricity losses exceed $6 billion per year due to ET [3]. Whereas,
the electric utilities of developing countries like India, incur a loss of $4.5 billion every year. This is almost 30% of the total energy produced in India [4]. Likewise, in Pakistan, the commercial losses exceed $0.89 billion per annum due to ET [5]. In literature, many research works are done to resolve the aforementioned challenges. For instance, in [6], Electricity Theft Detection (ETD) is performed using a gradient boosting algorithm. Moreover, a specific theft window is designed to detect the suspicious activity of electricity consumers in the peak hours. However, an efficient and accurate ETD mechanism is still required.

### TABLE 1: Power losses due to ET in different countries

| Countries | Losses (billion)/year |
|-----------|-----------------------|
| USA       | $6                    |
| UK        | $0.23                 |
| Brazil    | $4                    |
| Pakistan  | $0.89                 |
| India     | $4.5                  |

To overcome the ET problem, the researchers have implemented various solutions. The most popular ETD solutions are categorized into state based, game theory based and Machine Learning (ML) based solutions. In state based solutions, hardware devices are involved, i.e., smart meters, sensors, distribution transformers, etc., to detect the ET. In [7], a physically inspired state based solution is presented to detect energy theft in smart grids. A special type of transformer is integrated with the smart meters to investigate the EC patterns of electricity consumers. Moreover, a statistical linear regression algorithm is developed, which perfectly captures the relationship between electricity consumption values and voltage magnitude. The simulation results exhibit that the proposed solution outperforms the baseline models. For game theory based solutions, a game is played between the electricity thieves and the power company. In [8], a game theory based solution is introduced for smart homes to optimize the energy cost in peak hours. Furthermore, the proposed solution establishes the coordination among the smart appliances and mitigates the energy losses to a minimal level. On the other hand, the ML based solutions use the Electricity Consumption (EC) data acquired from smart meters to detect the dishonest electricity consumers. In [9], a Deep Convolution Neural Network (DCNN) is developed to detect the power quality disturbance in power grids. The DCNN model intelligently extracts the potential features from the high dimensional data through convolution operations and performs classification of power quality disturbance using a softmax layer. The performances of the above mentioned solutions are acceptable. However, the existing solutions have their drawbacks, which are itemized as follows.

1) The state based solutions are expensive because they require additional hardware cost to perform ETD [7].
2) The game theory based solutions are complex. They need to define strategies and set of rules for all the players in the game, which make the solutions very difficult to implement and operate [8].
3) In conventional ML models, the overfitting issue occurs because they are trained on high dimensional and complex data [9].
4) The imbalance nature of the dataset leads to model’s biasness, which raises the misclassification rate. The authors of [10] conduct a survey to highlight the challenges of data imbalance.
5) The performance of the classifier is primarily dependent on the input data. In most cases, the available data has missing values, outliers and is unscaled. As a result, the accuracy of the classification models is reduced [11].

A new solution is proposed in this study, which has four main stages: classification, feature extraction, data balancing and data pre-processing. The main contributions of this study are listed as follows.

1) A Synthetic Minority Oversampling Tomek Links (ST-Links) technique is used to resolve the imbalanced data problem. The ST-Links technique achieves the benefits of both data oversampling via Synthetic Minority Oversampling Technique (SMOTE) and under sampling through tomek links method. Moreover, it efficiently extracts the important features from the data to mimic the behavior of real electricity thieves, which significantly improves the performance of the classifier.
2) Essential statistical features, such as mean, median, mode, min and max are calculated to enhance the ETD performance.
3) An AlexNet and Peephole LSTM (APLSTM) based feature extractor along with an attention layer is presented to extract the potential features from the EC data to avoid overfitting.
4) A novel Echo State Neural Network (ESNN) is utilized for the classification of honest and dishonest electricity consumers.
5) An evolutionary Grey Wolf Optimization (GWO) technique is utilized to optimize the hyper-parameters of the proposed model.
6) A paired t-test hypothesis is applied to the final classification results to provide statistical support.
7) Thorough simulations are conducted and the performance of the proposed model is enhanced in terms of ETD. The proposed model is evaluated using suitable performance metrics, such as Precision Recall Area Under the Curve (PR-AUC), Matthews Correlation Coefficient (MCC), AUC, recall and precision.

The remaining part of this paper is organized as follows. The related work is presented in Section II. The proposed system model is discussed in Section III. The simulation results are provided in Section IV. Finally, the conclusion and future work are given in Section V.
II. RELATED WORK

In this section, the details of the work performed in ETD domain is provided. The work done in the existing literature is classified into non-hardware and hardware based solutions. The state based solutions are also known as hardware based solutions [12]. The non-hardware solutions are based on ML and game theory [13]. The ML techniques are further categorized into unsupervised learning (clustering), semi-supervised learning and supervised learning (classification) techniques. In the proposed solution, the supervised techniques are adopted. Moreover, the details of the recent advancements that are made in the ETD domain are studied.

The smart grids are helpful for efficient energy management and provisioning of electricity at lower prices [14], [15]. Moreover, the smart meters’ data helps in detecting the anomalous consumption behaviour of electricity thieves. In [14], a Deep Long Short Term Memory (DLSTM) is developed to forecast the electricity price and electricity load in smart grids. The proposed DLSTM model mitigates the energy losses to a minimal level. The authors of [15] propose a novel framework, which combines a Finite Mixture Model (FMM) and a Genetic Algorithm (GA) to perform ETD. In the proposed work, the FMM model is exploited for the customers’ segmentation and GA is used for the selection of potential features. In [16], electricity thieves are detected using a LSTM model. The LSTM model maintains the temporal correlation between the current and the previous EC time sequences to distinguish the honest consumers’ data from the dishonest consumers’ data. However, the data imbalance issue is not addressed, which increases the misclassification rate. In [17], the authors address the problem of inaccurate detection of Non Technical Loss (NTL) through Support Vector Machine (SVM) and Random Under Sampling (RUS) based Extreme Gradient Boosting (XGBoost). The aim of the study is to achieve high True Positive Rates (TPRs) and low False Positive Rates (FPRs). This methodology is tested on Endesa dataset and validated through AUC and PR-AUC performance measures. In [18], the authors address the problem of high FPR in NTL detection by proposing a new model, which is based on Bidirectional Gated Recurrent Unit (BiGRU) and Synthetic Over Sampling Tomek Links (SOSTLink). For feature extraction, the BiGRU technique is utilized while the SOSTLink technique is used to handle the imbalanced data. This model achieves the lowest FPR when it is compared with SVM, Convolutional Neural Network (CNN) and LSTM models. In [19], Maximum Overlap Discrete Wavelet Packet Transform (MODWPT) with Random Undersampling Boosting (RUSBoost) based scheme is proposed. The MODWPT scheme is used for selecting the features that are relevant to the data. On the other hand, data classification is performed using RUSBoost. However, this model loses important information by randomly reducing the data size that causes the under-fitting issue. In [20], LSTM and Bat optimization based RUSBoost techniques are used to solve the issues of overfitting and imbalanced data in ETD. Moreover, several performance metrics are used to show the effectiveness of the model.

In recent times, most of the researchers modify the internal structure of the ML algorithms for enhancing their performance. For instance, the structure of LSTM is changed using the softplus activation function instead of a tangent hyperbolic activation function [21]. Along with the improved LSTM, Gaussian Mixture Model (GMM) is used for classification. It achieved satisfactory values of accuracy and F1-score. However, high computational time is consumed by the model. In [22], a Smart Energy Theft System (SETs) is proposed for ETD. In the system, smart meters are integrated that digitally transmit the data to reduce the manual work. However, the hardware installation cost problem is not resolved. The authors in [23], [24] use the supervised learning techniques to tackle the class imbalance and the overfitting issues. To resolve the class imbalance issue, SMOTE is used. The ETD model in [24] is developed through a hybrid scheme that combines LSTM and CNN. In [23], the output layer of CNN is replaced by Random Forest (RF) to avoid the overfitting problem. Apart from deep learning techniques, the traditional ML techniques like SVM are also popular for handling the binary classification problems. Likewise, the authors in [25] use SVM for ETD. Besides SVM, a Principle Component Analysis (PCA) model is applied for dimensionality reduction. Although SVM performs better on linearly separable data, it uses a complicated kernel function for non-linear data, which makes it computationally expensive.

In most cases, the EC data is taken from smart meters. However, this data contains outliers and missing values, which raise the misclassification rate. Therefore, the authors in [26] focus on pre-processing to refine the input data and improve the classification accuracy. Moreover, they propose a Gradient Boosting Decision Trees (GBTD) model that helps in selecting the relevant information from the smart meters’ data, which also reduces the execution time. The losses incurred due to ET in Brazil are alarming and the amount of commercial losses faced in 2011 reached $4 billion [27]. To address this problem, the authors utilize the Binary Black Hole Algorithm (BBHA). The algorithm beats the existing optimization techniques, i.e., genetic and particle swarm optimization techniques, in terms of accurate NTL detection and execution time. However, the model is not evaluated using reliable performance metrics, such as recall and precision. A reliable evaluation measure is very necessary for assessing the performance of models in the case of imbalanced data classification problem. To address the NTL problems in Spain, the authors in [28] propose a hybrid model, which consists of LSTM and Multi Layer Perceptron (MLP). In the model, two types of data are utilized to detect the electricity thieves, i.e., sequential data and non-sequential data. The sequential data is processed using the LSTM model and the non-sequential data is learned using MLP. The authors in [29] change the internal structure of CNN to enhance the accuracy of ETD. The enhanced version of CNN is composed of wide and deep components. The former extracts the local features from one-dimensional (1D)
TABLE 2: Overview of supervised ML techniques in ETD

| Objectives | Proposed solutions | Dataset | Performance metrics | Limitations |
|------------|--------------------|---------|---------------------|-------------|
| To detect anomaly in real time data [16] | LSTM | Electricity load diagrams (2011–2014) | Precision, recall, RMSE | Inaccurate classification issue is not addressed |
| To accurately detect NTL [17] | XGBoost, SVM, RUS | Endesa, Spain | PR-AUC, AUC | Loss of information in balancing the data |
| To reduce high FPR [18] | SOSTLink, BiGRU | SGCC | Precision, recall, F1-score, AUC | Improving the system’s robustness is not considered |
| To accurately detect NTL [19] | RUSBoost, MODWPT | Honduras | MCC, F1-Score | Under-fitting problem is not addressed |
| To solve class imbalance problem in ETD [20] | LSTM, Bat based RUSBoost | SGCC | AUC, F1-score, recall and precision | Improving the system’s robustness is not considered |
| To improve the accuracy of ETD [21] | LSTM, GMM | Numenta anomaly benchmark | F1-score | High execution time problem is not addressed |
| To improve the accuracy of ETD in smart homes [22] | SETS | Aeon lab, UK | MAPE | High cost of hardware implementation |
| To solve the imbalanced data and overfitting problems in ETD [24] | CNN, LSTM, SMOTE | SGCC | Precision, recall, F1-score, MCC | High execution time problem is not addressed |
| To increase ETD performance for detecting NTL [23] | CNN, RF, SMOTE | Smart meter data of Ireland and electricity utility of London | Precision, recall, F1-score, AUC | High execution time problem is not addressed |
| To reduce and detect NTL [25] | SVM, PCA | Smart meter data | Accuracy | No reliable metrics are used for model’s evaluation |
| To enhance the classification accuracy and minimize time complexity [26] | GBTD | Smart meter data of Ireland | FPR | Overfitting issue is not addressed |
| To overcome power losses [27] | BBHA | Brazilian electric utility | Accuracy | No suitable metrics are used for model’s evaluation |
| To accurately detect NTL [28] | LSTM, MLP | Endesa, Spain | AUC, PR-AUC | Imbalanced data problem is not handled |
| To accurately detect ET data [29] | Wide and deep CNN | SGCC | AUC | Imbalanced data problem is not handled |
| To reduce NTL [30] | GBM | Endesa, Spain | AUC | Imbalanced data problem is not handled |
| To detect and remove abnormal EC data [31] | SMOTE, Adaboost | State grid Zhejiang electric utility | Accuracy | Overfitting issue is not addressed |

data while the latter derives the global features from the two-dimensional (2D) EC data. In [30], the authors use Gradient Boosting Machine (GBM) to detect the electricity thieves. In [31], the authors use SMOTE based Adaboost algorithm to perform ETD. The boosting algorithm is an ensemble strategy in which multiple weak models are joined to create a strong model to boost the classification performance. The overview of the existing ETD based supervised techniques is given in Table 2, which contains objectives, proposed solutions, datasets, performance metrics and their limitations. With the development of the advanced metering infrastructure, the electric utilities can record a high dimensional data from smart meters that is useful for the researchers to design data driven approaches [23], [24]. However, there are some limitations, which are needed to be addressed.

- The distribution of classes in the data is not balanced. The authors in [18], [23], [24] generate synthetic data to overcome this problem. However, synthetic data does not represent the behavior of real data.
- The features available in the EC dataset are not sufficient for efficient ETD [29].
- The overfitting issue is raised due to the presence of high dimensional, noisy and overlapping features in the EC dataset [25]. Moreover, the ML techniques like Linear Regression (LR) [17] require manual feature extraction and are not suitable for classifying the high dimensional data.
- Limited generalization ability of the traditional ML classifiers adversely affects the ETD process [25].
- Traditional neural networks increase computational complexity due to the excessive learning of weights for each hidden layer [9].
- The selection of suitable hyper-parameters for deep learning models is a challenging task [24], [29].
- Without applying proper statistical tests, the performance results are not reliable [20]–[22], [25].
In this study, a new model is proposed to resolve the aforementioned issues in ETD. The flowchart diagram of the proposed model is depicted in Figure 1. In the figure, the red color depicts the limitations labeled as L.1-L.7 while the blue color highlights the proposed solutions, which are labeled as S.1-S.7.

III. THE PROPOSED SYSTEM MODEL

The proposed model is discussed in this section and is presented in Figure 2. It comprises four components: classification, feature extraction, data balancing and data preprocessing, which are discussed in the following subsections.

A. DATA COLLECTION

The proposed methodology is designed using the real EC data, which is released by State Grid Corporation of China (SGCC) [29]. The dataset details are given in Table 3.

| Description                  | Values       |
|------------------------------|--------------|
| Time window of collected data| 01-01-2014 to 31-10-2016 |
| Total number of consumers    | 42,372       |
| Number of theft consumers    | 3,615        |
| Number of honest consumers   | 38,575       |

It consists of the EC information of forty-two thousand, three hundred and seventy-two consumers. Where 91% consumers are honest and 9% are fraudulent. The information about the fraudulent consumers is a ground reality. The number of honest and theft consumers shows the imbalanced nature of data. Figure 3 shows the EC patterns of two consumers, i.e., the honest consumer and the electricity thief. It shows that the electricity thief has irregular EC patterns and its EC value is less because of meter tampering. Whereas, the honest consumer has normal EC patterns.

B. DATA PRE-PROCESSING

In most cases, the data contains missing values and outliers, and is unscaled. These values occur due to faulty measuring equipment like defective smart meters and sensors. In this regard, the data pre-processing steps are required to be employed. Therefore, in this paper, a series of data pre-processing steps is applied, i.e., data interpolation, removing the outliers and data normalization.

1) Removing the missing values

The missing instances present in the original data mislead the model to inaccurately classify the electricity thieves and honest consumers. Therefore, it is required to fill the missing values in the EC data. For this purpose, the data interpolation method is exploited to handle the missing values. This method takes the mean of the neighboring values and fills the missing value. If the neighboring values are also missing, the missing value is replaced with 0. Equation (1) is used to fill-in the missing values, taken from [29].

\[ f(x_i) = \begin{cases} \frac{x_{i+1} + x_{i-1}}{2}, & \text{if } x_i \in \text{NaN}, x_{i-1} \notin \text{NaN} \\ 0, & \text{if } x_i \notin \text{NaN}, x_{i-1} \in \text{NaN} \\ x_i, & \text{if } x_i \notin \text{NaN} \end{cases} \]

where \( x_i \) represents the EC value of a consumer and \( \text{NaN} \) denotes the values that are non-numeric.

2) Identification and removal of outliers

The outliers are the values present in the dataset, which show abnormal behavior. In this study, a three sigma rule is utilized for removing the outliers from the overall dataset as defined in Equation (2), which is taken from [29].

\[ f(x_i) = \begin{cases} \text{avg}(x) + 2 \text{std}(x), & \text{if } x_i > \text{avg}(x) + 2 \text{std}(x) \\ x_i, & \text{else} \end{cases} \]

where \( \text{avg}(x) \) represents the average EC value of a consumer and the standard deviation is denoted as \( \text{std}(x) \). To remove the outliers, a threshold is set, i.e., a standard deviation of 2. The rule to select the standard deviation is not specified.

3) Data normalization

Data normalization is an important step. In neural networks, it plays a key role to increase the convergence rate and reduce the execution time. The values of different features in the dataset are not the same. Therefore, the min-max normalization technique is used to scale the values. This technique transforms the values of features between 0 and 1. The min-max normalization is done using Equation (3), which is mentioned in [29].

\[ x_{\text{nor}} = \frac{x_i - \text{min}(x)}{\text{max}(x) - \text{min}(x)}, \]

where \( x_{\text{nor}} \) represents the normalized value while the minimum and maximum values of the data are represented as \( \text{min}(x) \) and \( \text{max}(x) \), respectively.

C. FEATURE GENERATION

Initially, the EC dataset is univariant and has only one feature of EC. However, it is important to generate more statistical features for efficient ETD. In this regard, I compute significant statistical features, such as mean, median, mode, min and max from the original EC data to enhance the ETD performance.

D. DATA BALANCING THROUGH ST-LINKS

The major problem in ETD is the imbalanced distribution of honest consumers and electricity thieves in the data. The number of honest consumers is higher than theft consumers. Due to this problem, the classifier gives biased results [32]. The ST-Links technique is proposed to address the imbalanced class issue. Where SMOTE increases the theft class instances after k-nearest neighbours of a similar group are...
found. It draws a line between neighbours and creates new points on that line. The new points represent the theft instances in the dataset. If \((x_1, x_2)\) are the instances of minority class with their nearest neighbours selected as \((x_1', x_2')\), then the data instances synthesized by SMOTE are given in Equation (4) [24]:

\[
X_1, X_2 = (x_1, x_2) + \text{random}(0, 1) \times \Delta,
\]

(4)

where

\[
\Delta = (x_1' - x_1), (x_2' - x_2).
\]

(5)

In Equation (4), \text{random} (0, 1) is a function that generates synthetic random samples in the range 0 to 1 using the neighbours of the minority class. On the other hand, ST-Links removes instances from both classes. This method identifies pairs of the nearest neighbours, which have different groups in the same dataset. It creates a boundary line by identifying the pairs of nearest neighbours of two classes. The generation of new data by the ST-Links technique is shown in Figure 4.

E. FEATURE EXTRACTION USING AN ALEXNET AND PEEPHOLE LSTM BASED HYBRID TECHNIQUE WITH ATTENTION LAYER

Feature extraction is a significant stage in accurate identification of electricity thieves. In this process, more relevant fea-
Nadeem Javaid: Ensemble Learning Model for Detecting Electricity Theft in Smart Grid

Smart meters’ data | Data pre-processing | Data balancing
---|---|---
SGCC | Interpolation, three sigma rule and normalization | ST-Links

\[
\text{ST-Links}
\]

\[
\text{Conv layer 1} \rightarrow \text{Max pool} \rightarrow \text{Conv layer 2} \rightarrow \text{Max pool} \rightarrow \text{Conv layer 3} \rightarrow \text{Max pool} \rightarrow \text{Conv layer 4} \rightarrow \text{Max pool} \rightarrow \text{Conv layer 5} \rightarrow \text{Max pool} \rightarrow \text{FC 1} \rightarrow \text{FC 2} \rightarrow \text{FC 3}
\]

\[
\text{Feature extraction}
\]

\[
\text{Hyper-parameters’ tuning using GWO}
\]

\[
\text{PLSTM}
\]

\[
\text{ESNN}
\]

\[
\text{Output}
\]

\[
\text{Classification}
\]

Limitations addressed:
L.1 Data imbalance issue
L.2 Insufficient features
L.3 Overfitting issue
L.4 Limited generalization ability
L.5 High computational overhead
L.6 Lack of optimal hyper-parameters
L.7 Non reliable classification results

Solutions proposed:
S.1 ST-Links solve data imbalanced issue
S.2 Adding statistical features
S.3 APLSTM avoids overfitting issue
S.4 ESNN increases generalization ability
S.5 ESNN reduces time complexity by updating only output neurons’ weights
S.6 GWO finds optimal hyper-parameters
S.7 Paired t-test hypothesis is utilized for a fair evaluation of the proposed model

Evaluation:
- AUC
- Precision
- Recall
- F1-score
- MCC
- PR-AUC

FIGURE 2: The proposed ETD model
Majority class samples
Normal customer
(kWh)
12
14
18

Therefore, in this paper, an AlexNet [33] and PLSTM based hybrid solution is used along with an attention layer for feature extraction. The architecture of the proposed feature extractor is given in Figure 2. The AlexNet module has three fully connected layers, five pooling layers and five convolutional layers. These layers operate as a sequential model to perform feature extraction. AlexNet extracts relevant features through these layers. Initially, the balanced data is fed to the convolutional layer. Afterwards, the convolutional layers’ output is sent to the pooling layers. In this study, the activation function used to deal with the non-linear data is a Rectified Linear Unit (ReLU) [34], which is given as:

\[
ReLU(x) = \max (0, x).
\]

In Equation (6), \(x\) is the input data. ReLU performs activation if \(x\) is positive, otherwise, it returns 0. The max pooling layer is used in this paper, which computes the highest value for every feature map. The convolutional and pooling layers minimize the data dimensions to avoid the overfitting problem. The max pooling layer is used after each convolutional layer and then three fully connected layers (also called dense layers) are utilized to give the final output. Afterwards, the extracted feature set of AlexNet is passed to PLSTM for extracting more granular temporal correlated features. The PLSTM model is the advanced variant of traditional LSTM in which the concept of peephole connection is introduced. The peephole connections link the value of long-term memory state with every gate at each time step to increase the memorization power of the model. Furthermore, the integration of the attention layer with the PLSTM model guides it to focus only on the relevant information. This process increases the convergence speed of PLSTM to a greater extent. The mathematical representations of PLSTM’s memory gates are given below [28].

\[
i_t = \sigma(W_i[C_{t-1}, h_{t-1}, x_t] + b_i),
\]

(7)

\[
f_t = \sigma(W_f[C_{t-1}, h_{t-1}, x_t] + b_f),
\]

(8)

\[
o_t = \sigma(W_o[C_{t-1}, h_{t-1}, x_t] + b_o),
\]

(9)

\[
c_t = \tanh(W_c[h_{t-1}, x_t] + b_c),
\]

(10)

\[
h_t = o_t * \tanh(c_t).
\]

(12)

Where \(i_t\), \(f_t\) and \(o_t\) denote input gate, forget gate and output gate, respectively. Similarly, \(W_i\), \(W_f\) and \(W_o\) represent the weight matrices of input gate, forget gate and output gate, respectively. \(\sigma\) indicates sigmoid activation function. The original input at time step \(t\) is expressed by \(x_t\), \(C_{t-1}\) and \(h_{t-1}\) denote cell state and hidden state at time step \(t - 1\). \(c\) and \(\hat{c}\) represent the information in cell state at overall and current time step, respectively. Whereas, \(b_i\), \(b_f\), \(b_o\) and \(b_c\) indicate the bias values of input gate, forget gate, output gate and cell state, respectively.
F. CLASSIFICATION USING ECHO STATE NEURAL NETWORK

A novel version of recurrent neural network is utilized, termed as ESNN [35], for the final classification of honest and dishonest electricity consumers. The basic architecture of ESNN is based on recurrent memory networks. The ESNN model is introduced to perform time series prediction and classification tasks. It contains a reservoir where multiple nodes are placed in such a way that they are partially connected with each other to store important time series information. These nodes maintain temporal correlated patterns in their hidden states, which decay over time. The basic structure of the ESNN model consists of three layers: input, reservoir (hidden layer) and output. The extracted feature set of APLSTM is aligned in a 1D vector and is fed to the input layer where random weights are assigned to each input. Afterwards, the feature vector is passed to the hidden layer (reservoir). The nodes in the reservoir capture the necessary temporal patterns from the received input. In particular, they also maintain non-linear variance in EC patterns, which increases the generalization ability of the classification model. The output layer performs the classification of honest and dishonest consumers accordingly. The major advantage of the ESNN model over the deep models is that it updates the weights of only those neurons which are connected with the output layer neurons. By doing this, the computational overhead of the model is reduced and the convergence speed is increased to a greater extent. The mathematical formula used for updating the reservoir [35] nodes is given as follows.

\[
x(n + 1) = 1/(1 + exp(-\alpha \ast (W^{in} u(n + 1) + W^{r} x(n) + W^{back} y(n)) + v))
\]

(13)

Where \(\alpha\) and \(v\) indicate a coefficient and a small proportion of noise, respectively. \(W^{in}\) and \(W^{r}\) represent the weight matrices of input layer and reservoir, respectively. \(W^{back}\) depicts the weight matrix that is sent back to the reservoir for updation. The expression \(u(n + 1)\) represents the input that is given to the network at time step \(n + 1\). The activation of the output layer’s neurons is updated as follows.

\[
y(n + 1) = W^{out}(u(n + 1), x(n + 1), y(t))
\]

(14)

Where \(W^{out}\) depicts the weight matrix of the output layer.

Algorithm 1 describes the feature extraction and classification processes of the proposed model. Lines 1-4 represent the input, output and variables’ initialization of the algorithm. The feature extraction process of AlexNet and PLSTM

| Algorithm 1: Feature extraction and classification through the proposed model |
|---------------------------------|
| **Input:** Balanced dataset, \(X_{data} = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}, \ x, \ y \in \mathbb{R}\) |
| **Output:** \(Det_{model}\) |
| **Variables’ initialization:** \(W^l_{(\cdot)}\) and \(b^l_{(\cdot)}\) for each layer, \(W^{in}, W^{r}, W^{back}, W^{out}, b, context_v\) |
| **x = X_{data}** |
| **Feature extraction:** |
| **while** not converged **do** |
| for each training sample \(x\) **do** |
| **AlexNet layers:** |
| **for each hidden layer \(l = 1, \ldots, L\) do** |
| \(conv^l_i = z_i(W^l_i * x^l_i + b^l_i)\) |
| \(pool^l_i = \max_i \in Rl(\{conv^l_i\})\) |
| \(fully^l_i = z_i(W^l_i + pool^l_i)\) |
| **end** |
| **Peephole LSTM layers:** |
| **for each recurrent layer \(l_p\) do** |
| **for each time step \(t\) do** |
| \(i^l_p = \sigma(W^{l_p}_i [C^l_{t-1}, h^l_{t-1}, x^l_p] + b^l_i)\) |
| \(f^l_p = \sigma(W^{l_p}_f [C^l_{t-1}, h^l_{t-1}, x^l_p] + b^f)\) |
| \(o^l_p = \sigma(W^{l_p}_o [C^l_{t-1}, h^l_{t-1}, x^l_p] + b^o)\) |
| \(c^l_p = tanh(W^{l_p}_c [h^l_{t-1}, x^l_p] + b^c)\) |
| \(h^l_p = o^l_p * tanh(c^l_p)\) |
| **Attention layer:** |
| \(m = score(h^l_p, h_{t-1})\) |
| \(q = exp(m) / \sum_{m} exp(m)\) |
| \(context_{v,t} = \sum_{q} q * h_t\) |
| **end** |
| **Back propagation:** |
| compute: \(\nabla W^l C(x)\) and \(\nabla b^l C(x)\) |
| **end** |
| **Weights and biases’ updates:** |
| \(W^l = W^l - \alpha \sum_x \nabla W^l T(x)\) |
| \(b^l = b^l - \alpha \sum_x \nabla b^l T(x)\) |
| **end** |
| **Classification:** |
| \(x = context_v\) |
| **ESNN layers:** |
| **for epoch = 1 to n do** |
| **Hidden state weights’ updation:** |
| \(x(n + 1) = 1/(1 + exp(-\alpha \ast (W^{in} u(n + 1) + W^{r} x(n) + W^{back} y(n)) + v))\) |
| **Output state weights’ updation:** |
| \(y(n + 1) = W^{out} (u(n + 1), x(n + 1), y(n))\) |
| **Fully connected layer:** |
| \(Det_{model} = \sigma(W * y(t) + b)\) |
| **end** |
models is shown in lines 5-36. Furthermore, lines 37-42 exhibit the final classification of honest and dishonest energy consumers using the ESNN model. In addition, more detailed working of the attention layer and the back propagation process is given in [36].

G. HYPER-PARAMETER OPTIMIZATION USING GREY WOLF OPTIMIZATION

Being inspired from [37], where the authors proposed an ensemble learning and Jaya based solution to predict the post-fault transient stability status of power systems, a GWO and deep learning based ensemble learning solution is proposed to detect the energy thieves in smart grids. The utilization of heuristic techniques have greater influence on the performance of classification algorithms. Moreover, the hyper-parameter tuning is the most important step taken to enhance the performance of deep learning models. In this regard, an evolutionary meta heuristic technique, known as GWO [38], is employed to tune the hyper-parameters of APLSTM and ESNN models and enhance the ETD performance. The GWO technique is based on the working behavior of gray wolves. Moreover, it converges quickly towards a global optimal solution due to random and diverse numbers of individual assignment in exploitation and exploration phases.

IV. SIMULATION RESULTS AND DISCUSSION

The simulation results for the proposed solution is discussed in this section. The proposed solution is compared with other benchmark schemes to demonstrate its effectiveness. Moreover, it is worth mentioning that APLSTM is combined with ESNN after being motivated from the articles [23] and [24] where similar work is done in the context of ETD in smart grids. In these articles, initially, the feature extraction is performed and then the classification is carried out to identify the honest and dishonest electricity consumers. The simulation results of these studies prove that the combined model performs better than the standalone model. Keeping this in view, a combined ensemble model is proposed in this study. Where APLSTM is employed to extract the temporal correlated features from the EC data. Afterwards, the extracted feature set is passed to the ESNN model for final ETD. The classification results proved that the combined APLSTM and ESNN based model is a better solution to detect energy thieves as compared to other baseline models.

A. SIMULATION SETUP

The simulations are performed using open-source libraries of Python programming language, known as Tensorflow and Keras. The simulations are performed to train and validate the model. The data is grouped into 75% and 25% for training and testing the proposed model, respectively.

B. MODEL EVALUATION METRICS

In this regard, more suitable performance metrics are used, i.e., PR-AUC, AUC, precision, MCC, recall and F1-score, to assess the performance of the proposed model. Moreover, similar performance metrics are used in [39] to evaluate the performance of imbalanced classification models. The performance metrics are measured using the confusion matrix, which gives the information about the following:

- False Negative (FN): dishonest users are mistakenly predicted as legitimate.
- False Positive (FP): legitimate users are mistakenly predicted as dishonest users.
- True Negative (TN): dishonest users accurately predicted as dishonest.
- True Positive (TP): legitimate users accurately predicted as legitimate.

Precision [40], recall [41], F1-score [42], [43] and MCC [44] are calculated using Equations (15-17) as:

\[
\text{Precision} = \frac{TP}{TP + FP}, \quad (15)
\]

\[
\text{Recall} = \frac{TP}{TP + FN}, \quad (16)
\]

\[
F1 - score = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}, \quad (17)
\]

Where recall and precision are computed from the values of FN, FP and TP. The precision shows those values that are correctly identified as honest consumers while recall determines the positive class instances that are accurately identified by the model. In Equation (16), the F1-score is calculated, which is a more reliable metric than recall and precision. It balances both recall and precision to provide a single score. The PR-AUC is another useful metric, which is a graphical representation that demonstrates the recall values on the y-axis and the precision values on the x-axis. The outcome of the PR-AUC is between 0 and 1. In all of the above mentioned performance metrics, MCC is more reliable because it takes the correlation of all the four possible confusion matrix outcomes, i.e., FN, FP, TN and TP. It is computed using Equation (18):

\[
MCC = \frac{TP \times TN - FP \times FN}{\sqrt((TP + FP)(TP + FN)(TN + FP)(TN + FN))}. \quad (18)
\]

To get the graphical representation of the evaluation model, the AUC metric is used. It is calculated from TPR and FPR. In ETD, TPR is the number of electricity thieves, which are found guilty while FPR is the number of legitimate energy consumers, which are misclassified as electricity thieves. AUC gives the performance value between 0 and 1. It is calculated using Equation (19), which is taken from [24] as:

\[
AUC = \frac{\text{Recall} + \text{Specificity}}{2}. \quad (19)
\]

In this equation, specificity is a performance metric that shows the percentage of dishonest users that are classified as dishonest by the model. This metric is also known as a True Negative Rate (TNR).
C. RESULTS OF THE PROPOSED MODEL

In this section, the performance of the proposed solution is verified for solving the imbalanced data and the overfitting problems. Moreover, the performance of the model is also evaluated in terms of different performance metrics.

1) Performance on imbalanced dataset

The theft users are remarkably less than the honest users in the SGCC dataset. This makes the model biased towards the majority class during the training process. In this regard, the ST-Links technique is used. For comparison, the performance of the proposed model is compared with different data balancing techniques, i.e., SMOTE [31], Adaptive synthetic (Adasyn) [45] and Near Miss (NM). The Adasyn technique performs relatively better than SMOTE with an improved performance of 85.2% precision, 87.1% recall, 88.3% F1-score, 86.0% AUC, 85.1% PR-AUC and 77.4% MCC, as shown in Figure 5. In SMOTE and Adasyn algorithms, the minority class data is replicated to balance the dataset. However, as a result of the unintelligent replication of the minority class data, an overfitting issue occurs, which causes a high misclassification rate. The aim of using NM is two-fold. Firstly, it creates a boundary line and removes those majority samples that are considered as noise and are far from the decision boundary. Secondly, it reduces the execution time of the model. However, NM causes the loss of information due to random under-sampling, which under-fits the model. Figure 5 shows that the NM technique improves the classifier’s performance by achieving 75.2% precision, 78.4% recall, 80.2% F1-score, 74.6% AUC, 75.3% and 68.1% MCC.

ST-Links is a good solution to balance the dataset. It avoids the loss of information by intelligently selecting the number of instances in the data. Therefore, it prevents the classifier from biasness. It is worth noticing that the performance of ST-Links is better than the other data balancing techniques. Figure 5 shows that the ST-Links technique has improved the model’s performance by securing 90.0% precision, 92.1% recall, 92.0% F1-score, 96.4% AUC, 97.3% PR-AUC and 84% MCC.

I also compare the computational time of the ST-Links technique with other data balancing techniques. ST-Links based ensemble learning model has less execution time than the SMOTE and Adasyn methods. Although it is greater than the NM, however, the NM method removes those samples, which causes noise in the data. Due to the training of the model over less data, NM is beneficial in reducing the computational time of the model. Whereas, SMOTE takes more execution time due to the addition of synthetic samples in the dataset. Moreover, the execution time of Adasyn is the highest because it adds more random values in the data, which are linearly correlated to the main data with less variance. Figure 6 shows the executional time of the four data balancing techniques: SMOTE, Adasyn, NM and ST-Links methods. It is seen that NM has less execution time as compared to Adasyn, SMOTE and ST-Links.

2) Effect of optimization

The accuracy of the classifier is dependent on the hyper-parameters’ values. During the model validation, the GWO technique is used to select the optimal learning rate of ESNN. The scores of MCC, AUC, F1-score, recall, precision and PR-AUC for the ESNN model are depicted in Figure 7. It is observed that without using the optimization technique, the performance of ESNN is satisfactory. It achieved 87.0% precision, 77.2% MCC, 89.2% F1-score, 90.2% AUC, 82.3% PR-AUC and 88.0% recall. On the other hand, the hyper-parameters’ tuning through the GWO enhances the model’s performance by increasing the recall, precision, F1-score, AUC, PR-AUC and MCC by 4.1%, 3.0%, 2.8%, 6.2%, 15% and 6.8%, respectively. Moreover, it is important to mention that the GWO technique selects the best solution by comprehensively assessing the three optimal solutions during the exploration phase [46]. Furthermore, simulation results demonstrate that the aforementioned property of GWO efficiently avoids the problem of premature convergence. It is seen that the selected hyper-parameter set of GWO increases the performance of the proposed model to a greater extent.
3) Model generalization

The term generalization is used to describe the ability of the model to perform on unseen data. In Figure 8, the loss values of APLSTM-ESNN are shown. When a small epoch value is selected for training the model, it effectively trains the model on high dimensional data. Furthermore, as the value of epoch rises to 40, it causes overfitting. It shows 38\% and 41\% loss on training and testing data, respectively. These results show good prediction capability of the proposed model on testing data. Similarly, Figure 9 exhibits the accuracy curve of the model. It is seen that the model’s accuracy gradually improves with each epoch. This implies that the model efficiently captures EC patterns from consumers’ profiles and improves its generalization ability.

In Figure 10, ESNN shows the highest AUC score of 96.4\% and 99.8\% on the testing and training data, respectively. It shows that the model has accurately identified the honest consumers and electricity thieves. Moreover, the model does not show overfitting. Therefore, the proposed model has enough capability to identify the dishonest users even from the unseen data.

D. THE COMPARISON BETWEEN THE PROPOSED AND BENCHMARK MODELS

To show the effectiveness of the proposed model as compared to other benchmark models, I give the performance comparison with CNN-GRU, MLP-LSTM, XGBoost, LR and LSTM-RUSBoost. These state-of-the-art models are used for the classification of electricity thieves in the literature.

1) The results of CNN-GRU model

CNN and GRU are two popular deep learning models, which are used for the identification of electricity thieves in the literature [18], [23]. CNN is used for feature extraction. It has the ability to extract relevant features from the data, which help to perform an accurate ETD. Moreover, GRU performs the final classification. Using the SGCC dataset, I evaluate the CNN-GRU model on the selected performance measures. This model achieves 77.0\% accuracy, 85.4\% precision, 68.0\% recall, 80.1\% F1-score, 84.6\% AUC, 82\% PR-AUC and 64.0\% MCC, as given in Table 4. However, it is a deep learning model where its performance depends on the hyper-parameters’ values and the amount of data used.

2) The results of LR model

LR is used for the ETD in [17]. It uses the principle of neural networks. The results of the LR are given in Table 4. It achieves 73.0\% accuracy, 70.9\% precision, 72.9\% recall, 71.9\% F1-score, 69.4\% PR-AUC and 45.0\% MCC. The AUC of the LR model is given in Figure 11, i.e., 70.3\%. Thus, it is not suitable for binary classification with a large dataset.

3) The results of XGBoost model

In this paper, XGBoost [17] is used as a benchmark model. It is the most popular ML algorithm that won 17 out of 29 competitions in the Kaggle platform in 2015 [45]. The performance of the XGBoost model is evaluated on different performance metrics, which are given in Table 4 and shown in Figure 12. The model achieves 84.1\% accuracy, 86.7\%
precision, 78.7% recall, 82.5% F1-score, 80.8% AUC, 78.1% PR-AUC and 68.4% MCC.

4) The results of MLP-LSTM model
In the hybrid MLP-LSTM model [28], MLP is used for feature extraction. Whereas, LSTM classifies the electricity thieves by capturing the temporal correlations. This model achieves 70.0% accuracy, 79.4% precision, 78% F1-score, 82.3% recall, 68.1% MCC, 80.8% PR-AUC and 82.9% AUC, as given in Table 4.

5) The results of LSTM-RUSBoost model
The authors in [20] use a hybrid of RUSBoost and LSTM for ETD. Where the LSTM model is used for extracting the long term correlations between the consumption patterns. Whereas, the purpose of RUSBoost is two-fold. Firstly, it balances the data using RUS. Secondly, it performs the classification through ensemble learning. In Table 4, the classification results of the LSTM-RUSBoost model are presented. The model achieves 76.9% accuracy, 78.2% precision, 76.0% F1-score, 77.1% recall, 68.3% MCC, 83.3% PR-AUC and 86.0% AUC. However, randomly reducing the size of data through the RUSBoost technique causes the loss of important information, which raises the misclassification rate.

The comparison of the proposed model with the benchmark models is depicted in Figures 12 and 11. It is seen that the proposed model surpasses the existing models in terms of PR-AUC and AUC. The average execution time of the proposed and benchmark schemes is depicted in Figure 14. From the figure, it is seen that the ESNN model has the least training time of 17 seconds because the weight updating process of the model is pretty simple. It updates the weights of only those neurons that participate in the output layer. Due to this property, the proposed model minimizes the computational overhead to a minimal level. Furthermore, it successfully obtains the highest performance results as compared to other baseline models. Therefore, in practical applications, the proposed solution is capable to detect energy thieves accurately within a minimum time. In contrast,
### TABLE 4: The results’ summary for the proposed and benchmark schemes

| Schemes          | PR-AUC | Recall | Precision | F1-score | AUC   | MCC   | Accuracy |
|------------------|--------|--------|-----------|----------|-------|-------|----------|
| CNN-GRU          | 0.820  | 0.680  | 0.854     | 0.801    | 0.846 | 0.640 | 0.770    |
| LR               | 0.694  | 0.729  | 0.709     | 0.719    | 0.703 | 0.450 | 0.730    |
| LSTM-RUSBoost    | 0.833  | 0.771  | 0.782     | 0.760    | 0.860 | 0.683 | 0.769    |
| MLP-LSTM         | 0.808  | 0.823  | 0.794     | 0.780    | 0.829 | 0.681 | 0.700    |
| XGBoost          | 0.781  | 0.787  | 0.867     | 0.825    | 0.808 | 0.684 | 0.841    |
| Proposed         | 0.973  | 0.921  | 0.900     | 0.920    | 0.964 | 0.840 | 0.963    |

### TABLE 5: Mapping of validation results, proposed solutions and problems addressed

| Identified limitations | Proposed solutions | Validations |
|------------------------|--------------------|-------------|
| L.1: The imbalance distribution of classes in the data | S.1: ST-Links is used to solve this problem | V.1: ST-Links has improved the classifier’s results based on handling the dataset, as depicted in Figure 5 |
| L.2: Insufficient features for efficient ETD | S.2: Suitable statistical features, such as mean, median, mode, min and max are calculated from the dataset to enhance the ETD performance | V.2: The final classification results demonstrate the significance of the calculated features, as shown in Table 4 and Figure 13 |
| L.3: The overfitting issue is raised due to the presence of high dimensional, noisy and duplicate features in the EC data | S.3: A peephole LSTM with attention layer and AlexNet based feature extractor is developed to capture the relevant features from the EC data to avoid the overfitting issue | V.3: Figures 7, 10 show that the proposed model avoids overfitting and achieves the best classification performance in the unseen data |
| L.4: Limited generalization ability of the traditional classifiers adversely affect the ETD process | S.4: An ESNN model is employed for the classification of normal and abnormal consumers to improve the ETD performance | V.4: The proposed model achieves more generalized results on both training and testing data, as shown in Figures 8, 9 |
| L.5: Traditional neural networks increase the computational complexity due to the excessive learning of weights for each hidden layer | S.5: ESNN reduces computational overhead by updating the weights of only those neurons that participate in the output layer | V.5: Figure 14 exhibits that the ESNN model has the least training time |
| L.6: Optimal selection of hyper-parameters for deep learning models is a challenging task | S.6: An evolutionary GWO technique is utilized to optimize the hyper-parameters of the proposed solution | V.6: Figure 7 shows that the GWO technique helps to increase the classification results by finding the optimal hyper-parameters’ values |
| L.7: Without applying a proper statistical test, the performance results are not reliable | S.7: A paired t-test is applied on the final classification results to provide statistical support | V.7: The solution is validated by computing p value of t-test, as given in section IV-D6 |

the complex layering structure of the baseline deep models increases the training time to a greater extent. Moreover, the mapping of the addressed problems with the proposed solutions and their respective validation is given in Table 5. The limitations are labeled as L.1-L.7 while the proposed solutions are labeled as S.1-S.7 and the validation of results are labeled as V.1-V.7. Firstly, the model is validated in terms of handling the imbalanced data. The results in Figure 5 show that ST-Links achieved better results than other data balancing techniques, i.e., SMOTE, Adasyn and NM. ST-Links obtained 90% precision, 92.1% recall, 92% F1-score, 96.4% AUC and 84% MCC. Secondly, Figure 13 depicts that the ESNN model performs well with newly generated statistical features. Thirdly, Figures 7, 10 exhibit that the proposed model avoids overfitting and achieves the best classification performance on the unseen data. Fourthly, Figure 10 shows that GWO-ESNN gives improved generalized performance as it achieves the highest score regarding AUC, i.e., 96.4% and 99% on the testing and training data, respectively. Fifthly, Figure 14 illustrates that the ESNN model has the least training time. Sixthly, Figure 7 shows that the GWO technique increases the classification accuracy by selecting the best hyper-parameters’ values for the ESNN model. Seventhly, the classification results of the ESNN model are verified by applying a statistical test, which is known as paired t-test. Finally, benchmark methods are compared with the proposed model, i.e., CNN-GRU, MLP-LSTM, XGBoost, LR and LSTM-RUSBoost. The accuracy performance metric is not suitable to differentiate between the honest and dishonest consumers. It is a misleading metric in a classification problem where the classes of the data are imbalanced. So, more reliable performance metrics are used, i.e., AUC, PR-AUC, precision, F1-score and recall, to assess the proposed model. Using the reliable metrics, it is observed that the performance of the proposed model is better than the existing models. Moreover, it is worth mentioning that the proposed...
ETD model is trained on a realistic dataset of China, which is a massive dataset. In addition, the proposed approach is based on deep neural networks where the models perform efficiently with a huge dataset. In this regard, the proposed ETD approach could be applicable in the scenarios where the number of electricity consumers is high or low. Finally, it is concluded that the proposed ETD approach is scalable.

6) The performance analysis of paired t-test on classification results

The reliance on the final classification accuracy of the proposed model is not justifiable without applying any suitable statistical test. Therefore, a well-known statistical test is used, termed as paired t-test, for the fair evaluation of the final classification results. First of all, I compute 60 independent accuracy values by changing the hyper-parameters’ settings of the APLSTM and ESNN models with different training and testing distributions (e.g., 70:30 and 80:20). Afterwards, the paired t-test is applied on these values to ensure that the obtained accuracy of the classification model does not occur by chance. The test achieves 1.983723 and a p-value of 0.045256, which means that the null hypothesis is rejected. Hence, it is proved that the classification results of the proposed model are strongly reliable without the presence of any randomness. The average accuracy value of the proposed model in 60 different settings is shown in Figure 15.

V. CONCLUSION AND FUTURE WORK

In the proposed study, real EC data of SGCC is used to identify the electricity thieves. Initially, necessary statistical features are computed and data pre-processing is performed...
to refine the input data. Afterwards, the imbalanced data problem is addressed by ST-Links that intelligently balances the data classes. In addition, the relevant features are extracted by APLSTM to improve the ETD performance. The extracted feature set is given to the ESNN model for classifying the electricity thieves. Moreover, the hyper-parameters of the proposed model are tuned through GWO. The hyper-parameters selected through GWO increase the performance of the APLSTM-ESNN model to a greater extent. Furthermore, a statistical test, termed as paired t-test, is applied on the classification results for the reliable assessment of the proposed model. The simulation results depict that the proposed model achieved MCC, AUC, PR-AUC, F1-score, recall, precision and accuracy of 84.0%, 96.4%, 97.3%, 92.0%, 92.1%, 90.0% and 96.3%, respectively. The proposed model performs better than the existing models in terms of classification. Although the proposed model is an ideal solution for efficient ETD, however, there are some abrupt changes in the performance of the proposed model with respect to the input data. Furthermore, the proposed model is trained on the low sampling data, which limits its performance towards capturing the more granular information about EC patterns. Therefore, in future, high sampling ED data and some other factors, such as different usage behavior of consumers, temperature and seasonality will be considered to develop a robust model.

REFERENCES

[1] Messinis, G.M. and Hatziargyriou, N.D., 2018. Review of non-technical loss detection methods. Electric Power Systems Research, 158, pp.250-266.

[2] Jiang, R., Lu, R., Wang, Y., Luo, J., Shen, C. and Shen, X., 2014. Energy-theft detection issues for advanced metering infrastructure in smart grid. Tsinghua Science and Technology, 19(2), pp.105-120.

[3] McDaniel, P. and McLaughlin, S., 2009. Security and privacy challenges in the smart grid. IEEE Security and Privacy, 7(3), pp.75-77.

[4] Razavi, R. and Fleury, M., 2019. Socio-economic predictors of electricity theft in developing countries: An Indian case study. Energy for Sustainable Development, 48, pp.10-21.

[5] Hussain, Z., Menon, S., Shah, R., Bhattu, Z.A. and Aljawarneh, M., 2016. Methods and techniques of electricity thefting in pakistan. Journal of Power and Energy Engineering, 4(09), pp.1-10.

[6] Punnuya, R. and Choe, S., 2021. ToU Pricing-Based Dynamic Electricity Theft Detection in Smart Grid Using Gradient Boosting Classifier. Applied Sciences, 11(1), pp.390-401.

[7] Gao, Y., Foggio, B. and Yu, N., 2019. A physically inspired data-driven model for electricity theft detection using smart meter data. IEEE Transactions on Industrial Informatics, 15(9), pp.5076-5088.

[8] Jamil, A., Alghamdi, T.A., Khan, Z.A., Javaid, S., Haseeb, A., Wadud, Z. and Javaid, N., 2019. An Innovative Home Energy Management Model with Coordination among Appliances using Game Theory. Sustainability, 11(22), pp.6280-6287.

[9] Wang, S. and Chen, H., 2019. A novel deep learning method for the classification of power quality disturbances using deep convolutional neural network. Applied energy, 235, pp.1126-1140.

[10] Kaur, H., Pannu, H.S. and Malhi, A.K., 2019. A systematic review on imbalanced data challenges in machine learning: Applications and solutions. ACM Computing Surveys (CSUR), 52(4), pp.1-36.

[11] Mujeeb, S. and Javaid, N., 2019. ESAENARX and DE-RELM: Novel schemes for big data predictive analytics of electricity load and price. Sustainable Cities and Society, 51, pp.1-8.

[12] Glauner, P., Meira, J.A., Valtchev, P., State, R., and Bettinger, F., 2016. The challenge of non-technical detection using artificial intelligence: A survey. International Journal of Computational Intelligence Systems, pp.760-775.

[13] Viegas, J.L., Vieira, S.M., Melício, R., Mendes, V.M.F. and Sousa, J.M., 2016. Classification of new electricity customers based on surveys and smart metering data. Energy, 107, pp.804-817.

[14] Mujeeb, S., Javaid, N., Ilahi, A., Wadud, Z., Ishmanov, F. and Afzal, M.K., 2019. Deep long short-term memory: A new price and load forecasting scheme for big data in smart cities. Sustainability, 11(44), pp.980-987.

[15] Razavi, R., Gharioum, A., Fleury, M. and Akpan, I.J., 2019. A practical feature-engineering framework for electricity theft detection in smart grids. Applied energy, 235, pp.481-494.

[16] Fenza, G., Gallo, M. and Loia, V., 2019. Drift-aware methodology for electric theft detection in smart grid. IEEE Access, 7, pp.9645-9657.

[17] Buzau, M.M., Tejedor-Aguilera, J., Cruz-Romero, P. and Gómez-Expósito, A., 2018. Detection of non-technical losses using smart meter data and supervised learning. IEEE Transactions on Smart Grid, 10(3), pp.2661-2670.

[18] Gul, H., Javaid, N., Ullah, I., Qamar, A.M., Afzal, M.K. and Joshi, G.P., 2020. Detection of non-technical losses using SOSTLink and bidirectional gated recurrent unit to secure smart meters. Applied Sciences, 10(9), pp.1-26.

[19] Avila, N.F., Figueroa, G. and Chu, C.C., 2018. NTL detection in electric distribution systems using the maximal overlap discrete wavelet-packet transform and random undersampling boosting. IEEE Transactions on Power Systems, 33(6), pp.7171-7180.

[20] Adil, M., Javaid, N., Qasim, U., Ullah, I., Shaqit, M. and Choi, J.G., 2020. LSTM and bat-based eBoost approach for electricity theft detection. Applied Sciences, 10(12), pp.1-21.

[21] Ding, N., Ma, H., Gao, H., Ma, Y. and Tan, G., 2019. Real-time anomaly detection based on long short-Term memory and Gaussian Mixture Model. Computers and Electrical Engineering, 79, pp.1-11.

[22] Li, W., Logenthiran, T., Phan, V.T. and Woo, W.L., 2019. A novel smart energy theft system (SETS) for IoT-based smart home. IEEE Internet of Things Journal, 6(3), pp.5531-5539.

[23] Liu, X., Han, Y., Yao, X., Yingchen, S. and Wang, J., Zhao, Q., 2019. Electricity theft detection in power grids with deep learning and random forests. Journal of Electrical and Computer Engineering, pp.1-12.

[24] Hasan, M., Toma, R.N., Nahid, A.A., Islam, M. and Kim, J.M., 2019. Electricity theft detection in smart grid systems: A CNN-LSTM based approach. Energies, 12(17), pp.1-18.

[25] Jindal, A., Dua, A., Kaur, K., Singh, M., Kumar, N. and Mishra, S., 2016. Decision tree and SVM-based data analytics for theft detection in smart grid. IEEE Transactions on Industrial Informatics, 12(3), pp.1005-1016.

[26] Punnuya, R. and Choe, S., 2019. Energy theft detection using gradient boosting thief detector with feature engineering-based preprocessing. IEEE Transactions on Smart Grid, 10(2), pp.2326-2329.

[27] Ramos, C.C., Rodrigues, D., de Souza, A.N. and Papa, J.P., 2016. On the study of commercial losses in Brazil: a binary black hole algorithm for theft characterization. IEEE Transactions on Smart Grid, 9(2), pp.676-683.

[28] Buzau, M.M., Tejedor-Aguilera, J., Cruz-Romero, P. and Gómez-Expósito, A., 2019. Hybrid deep neural networks for detection of non-technical losses in electricity smart meters. IEEE Transactions on Power Systems, 35(2), pp.1254-1263.

[29] Zheng, Z., Yang, Y., Niu, X., Dai, H.N. and Zhou, Y., 2017. Wide and deep convolutional neural networks for electricity-theft detection to secure smart grids. IEEE Transactions on Industrial Informatics, 14(4), pp.1606-1615.

[30] Coma-Puig, B. and Carmona, J., 2019. Bridging the gap between energy consumption and distribution through non-technical loss detection. Energies, 12(9), pp.1-17.

[31] Qin, H., Zhou, H. and Cao, J., 2020. Imbalanced learning algorithm based intelligent abnormal electricity consumption detection. Neurocomputing, 402, pp.112-123.

[32] Figueroa, G. and Chu, Y.S. Avila, N.ChuCC, 2017. Improved practices in grid management algorithm for NTL detection with imbalanced data. In 2017 IEEE Power Energy Society General Meeting. IEEE, pp.1-5.

[33] Ding, L., Li, H., Hu, C., Zhang, W. and Wang, S., 2018. Alexnet feature extraction and multi-kernel learning for object-oriented classification. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci, 42, pp.277-281.

[34] Pedamonti, D., 2018. Comparison of non-linear activation functions for deep neural networks on MNIST classification task. arXiv preprint arXiv, pp.1-5.

[35] Lin, X., Yang, Z. and Song, Y., 2009. Short-term stock price prediction based on echo state networks. Expert systems with applications, 36(3), pp.1-5.
Nadeem Javaid: Ensemble Learning Model for Detecting Electricity Theft in Smart Grid

[36] Takiddin, A., Ismail, M., Zafar, U. and Serpedin, E., 2020. Robust Electricity Theft Detection Against Data Poisoning Attacks in Smart Grids. IEEE Transactions on Smart Grid, 12(3), pp.2675-2684.

[37] Li, Y. and Yang, Z., 2017. Application of EOS-ELM with binary Jaya-based feature selection to real-time transient stability assessment using PMU data. IEEE Access, 5, pp.23092-23101.

[38] Emary, E., Zawbaa, H.M. and Hassanien, A.E., 2016. Binary grey wolf optimization approaches for feature selection. Neurocomputing, 172, pp.371-381.

[39] Zhang, M., Li, J., Li, Y. and Xu, R., 2021. Deep learning for short-term voltage stability assessment of power systems. IEEE Access, 9, pp.29711-29718.

[40] Hand, D. and Christen, P., 2018. A note on using the F-measure for evaluating record linkage algorithms. Statistics and Computing, 28(3), pp.539-547.

[41] Gu, Y., Cheng, L. and Chang, Z., 2019. Classification of imbalanced data based on MTS-CBPSO method: a case study of financial distress prediction. Journal of Information Processing Systems, 15(3), pp.682-693.

[42] Douzas, G., Bacao, F., Fonseca, J. and Khudinyan, M., 2019. Imbalanced learning in land cover classification: Improving minority classes’ prediction accuracy using the geometric smote algorithm. Remote Sensing, 11(24), pp.30-40.

[43] Smith, A.M., Lampinen, J.M., Wells, G.L., Smalarz, L. and Mackovichova, S., 2019. Deviation from perfect performance measures the diagnostic utility of eyewitness lineups but partial area under the ROC curve does not. Journal of Applied Research in Memory and Cognition, 8(1), pp.50-59.

[44] Greff, K., Srivastava, R.K., Koutnik, J., Steunebrink, B.R. and Schmidhuber, J., 2016. LSTM: A search space odyssey. IEEE transactions on neural networks and learning systems, 28(10), pp.2222-2232.

[45] Khan, Z.A., Adil, M., Javaid, N., Saqib, M.N., Shafiq, M. and Choi, J.G., 2020. Electricity theft detection using supervised learning techniques on smart meter data. Sustainability, 12(19), pp.1-20.

[46] Wang, J.S. and Li, S.X., 2019. An improved grey wolf optimizer based on differential evolution and elimination mechanism. Scientific reports, 9(1), pp.1-21.

Nadeem Javaid (S’08, M’11, SM’16) received the bachelor degree in computer science from Gomal University, Dera Ismail Khan, Pakistan, in 1995, the master 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 an Associate Professor and the Founding Director of the Communications Over Sensors (ComSens) Research Laboratory, Department of Computer Science, COMSATS University Islamabad, Islamabad. He is also working as visiting professor at the School of Computer Science, University of Technology, Sydney, Australia. He has supervised 137 master and 24 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 recipient of the Best University Teacher Award from the Higher Education Commission of Pakistan, in 2016, and the Research Productivity Award from the Pakistan Council for Science and Technology, in 2017. He is also Associate Editor of IEEE Access and Editor of Sustainable Cities and Society journals.

VOLUME 4, 2016

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