IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received October 3, 2021, accepted October 7, 2021, date of publication October 13, 2021, date of current version October 22, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3119575

A Novel Method CNN-LSTM Ensembler Based on Black Widow and Blue Monkey Optimizer for Electricity Theft Detection

ABDULWAHAB ALI ALMAZROI[1] AND NASIR AYUB[2]
[1]Department of Information Technology, College of Computing and Information Technology at Khulais, University of Jeddah, Jeddah 21959, Saudi Arabia
[2]Department of Computer Science, Federal Urdu University of Arts, Science and Technology, Islamabad 44000, Pakistan

Corresponding author: Nasir Ayub (nasirayub@fuuast.edu.pk)

ABSTRACT Enhanced metering infrastructure is a key component of the electrical system, offering many advantages, including load management and demand response. However, several additional energy theft channels are introduced by the automation of the metering system. With data analysis techniques, adapting the smart grid significantly reduces energy theft loss. In this article, we proposed deep learning methods for the identification of power theft. A three-stage technique has been devised, which includes selection, extraction, and classification of features. In the selection phase, the average hybrid feature importance determines the most important features and high priority. The feature extraction technique employs the ZFNET method to remove the unwanted features. For the detection of electric fraud, we have applied Long Short Term Memory method embedded in Convolutional Neural Network technique (CNN-LSTM). Meta-heuristic techniques, including Black Widow Optimization (BWO) and Blue Monkey Optimization (BMO), are used to calculate optimized values for the hyperparameters of CNN-LSTM. The tuning of hyperparameters of the classifier helps in better training on data. After extensive simulation, our proposed methods CNN-LSTM-BMO and CNN-LSTM-BWO achieved an accuracy of 91% and 93%. Our proposed methods outperform all the existing compared schemes. The performance of our models has attained high accuracy and low error rate. Furthermore, the statistical analysis also shows the superiority of the proposed methods.

INDEX TERMS Optimization techniques, smart grid, deep learning methods, ensembler, black widow optimizer, CNN.

NOMENCLATURE

m_k	Linear Interpolation value.
NaN	Not a number.
STD	Standard deviation.
N'	Normalized value.
m	Input.
T	Filter/Kernel.
H	Weight Functions.
G	Weight Functions.
φ	Softmax Activation.
x_i	Input data to activation layer.
W	Weight.
X	Function Matrix.
h_t	Hidden State.
X_t	Current Input.
$Tan\ h$	Sigmoid function.
C_t	Multiplication outcome.
F_t	Forgot Gate.
$X1\ X2$	BWO Parents.
$Y1\ Y2$	BWO Descendants.
$rate$	Power of Monkey.
W_{leader}	Leaders Weight.
W_t	Monkeys Weight.
X_{best}	Monkeys Leader position.
$Rand$	Random Number.
$Rate_{ch}$	Children Weight.
$W_{ch(leader)}$	Weight of younger leader.
$W_{ch(i)}$	Random Weight of children.
X_{ch}	Child Position.
$W_{ch(best)}$	Children Leader Position.

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

I. INTRODUCTION

Nowadays, electricity has become a necessary component of our everyday lives. Electricity is produced and then transmitted over vast networks from big power plants to customers, with the loss occurring during both the production and transmission stages [1]. Due to the shortage of energy resources and the high cost of acquiring them, every country's social and economic development must include safe and effective use provisions for its energy resources.

Smart Grid (SG) technology has shown itself as a credible alternative for monitoring energy in the future. The SG system is a type of energy system that comprises of a power distribution system and computers that are used to manage and regulate electricity use using an automated control system that keeps track of all system users [2]. When new digital technology is integrated with the electrical grid, utilities and customers can monitor, track, and predict energy usage. The collector (device) transmits an internet-based energy usage reading to the operational hub and the power transmission company, then bills the client. Simultaneously, the utility gathers user readings through a wireless network via sporadic warnings from nearby consumers. Their main objective is to reduce energy losses and deliver consistent, dependable, and cost-effective power [3]. A smart meter is an electronic version of a conventional meter that monitors and reports on usage. Smart meters are vulnerable to a variety of security flaws. These are low-cost consumer devices that are widely utilized. They are designed to last longer to save cost on replacement. Due to its microchip, nonvolatile storage, networking capabilities, and ability to sustain overall consumer energy production are critical components of smart grid architecture [4]. Power loss is the difference in the amount of energy generated as well as the amount of electricity provided to customers. Smart meters are critical for measuring energy loss in smart grids. Advanced energy meters get data from customers' load meters and then compute hourly power usage. This meter provides additional information to the retailing business to facilitate appropriate administration, accounting, and two-way communication between the retailer and its consumers [5]. Energy consumption may be reduced by disconnecting and reconnecting the power source from any remote location.

Smart meter security flaws that lead to Non-Technical Loss (NTL) fraud were classified into three categories: network penetration, meter modification, and measurement interruption. Meter tampering tactics include altering meter firmware, password cracking, and key spoofing. Network intrusion attacks include data injection attacks and communication interception. Finally, measurement disruption involves partial or full meter bypassing or the placement of powerful magnets near meters. Electricity thieves result in a substantial loss of revenue for the electric company. The two kinds of electrical loss are NTL and Technical Loss (TL). TL is the abbreviation for the power loss produced by resistance in transmission networks [6]. Calculating TL is a time-consuming procedure, and determining the moment of failure and calculating the amount of energy lost is challenging. While it cannot be eliminated, we may significantly decrease it by changing current systems. To correctly identify and rectify NTL, the TL must be estimated in the first place. A utility company doesnot may have access to the network topology or cable impedance measurements needed to estimate TL.

NTL refers to the difference between complete loss and TL. NTL is mainly caused by billing delays, irregularities, theft, malfunctioning energy meters, fraud, and outstanding bills. NTL often entails bypassing electric meters, altering them, or hacking them [7]. Energy theft will result in higher energy rates, a heavy load on the grids, a loss of income for the energy supplier, and a decrease in profit, as well as higher costs for all users and other issues including offloading, disruption of business schedules, and inflation. Controlling energy theft is a significant issue for the nations listed above in economic development [8]. Only real-time fraud detection can reduce electricity theft. Power theft is a form of NTL that deprives the power industry of profits, thus damaging the country's economy.

A. PROBLEM STATEMENT

Several methods for detecting and minimizing theft have been used. More analysis is necessary to address the issues of Electricity Theft Detection (ETD) adequately and overcome the constraints of inadequate theft detection owing to unbalanced data and the limited capacity of Machine Learning (ML) algorithms. We discovered that just a few publications in the current literature had addressed the impact of unbalanced data in their system models [9]. The authors in the literature solved the class imbalance issue by using Adaysn; however, this produces overfitting and repeats the samples of the closest neighbor, which will not represent theft cases of real-world [10]. Several past surveys suggest that smart grid data gathering will help identify energy theft. This strategy has the following drawbacks: Linear Regression (LR) and Support Vector Machine (SVM) have low performance in identifying energy theft [11]. The CNN-LSTM hybrid structure is also utilized to forecast energy theft detection [12]. The CNN-LSTM model performed well in each theft detection scenario.

In contrast, raw data with no preprocessing was used to predict an electrical theft [13]. Many parameters of the sensors, such as level of noise, the scale of sensor power, and so on, might affect data quality. However, using an adequate preprocessing technique is essential. On energy consumption data, these researchers [14] constructed a regression model employing hybrid structures such as a Classification algorithm i.e, CNN-LSTM. The power consumption pattern dataset was used to address the classification problem, and the CNN-LSTM hybrid model was combined with a preprocessing approach. This led us to develop a hybrid structure that analyses customers' irregular consumption patterns to detect power theft. These existing methods offer good outcomes, but as indicated in Table 2, they have limitations.

We proposed the following solutions for the problems identified in the above articles:

TABLE 1. Limitations of existing system.

Ref	Methods	Limitations
[1]–[3]	Conventional ETD	It includes manual methods like humanly checking, for which we need to hire an inspection team with many members. There is a massive inconsistency in the manual checking.
[4]	Game-Based Theory	This method has a low identification rate but a high rate of False Positives (FPR).
[9]	Missing Values Data	Erroneous data are used in some situations, which reduce classification precision.
[10]–[13]	State-Based Solution	This solution is costly since it involves the installation of new machinery.
[13], [14]	ML Techniques	The critical issue with previous ML approaches was dealing with unbalanced data. This unbalance issue is left unaddressed in conventional models. Synthetic Minority Over-sampling TEchnique (SMOTE) and Random Under Sampling (RUS) approaches result in information loss and overfitting.
[15]	Traditional ML Techniques	On large datasets, traditional approaches such as Logistic Regression (LR) and SVM performed poorly.

- The accuracy problem is resolved by using the Deep Neural Network (DNN) method, including the tuning of parameters by the novel optimization techniques.
- The un-balancing of the data is handled using the SMOTE algorithm.
- Tuning the parameters performed in our solution has reduced the issue of data overfitting.
- As the missing values reduce the classifier training and classification precision, we have applied to preprocess step to adjust the missing values.
- We have used AI techniques to eliminate the hardware issues and efficiently detect electricity fraud in light of the hardware-based issues.

II. RELATED WORK

The literature review related to electricity theft detection can be split into two approaches: hardware-based and data-driven-based. These approaches are discussed below:

A. HARDWARE-BASED APPROACH

In this approach, researchers focus on developing special metering devices and facilities for quickly detecting power. Smart meters with RF tags and anti-tamping sensors are examples of these systems [16]. These hardware-based implementations have several drawbacks, including hardware component vulnerability, malfunction due to environmental factors, and difficulty sustaining these machines (replacing batteries, etc.).

B. DATA-DRIVEN BASED APPROACH

Due to difficulties and limitations of hardware base solutions, data-driven solutions have drawn attention recently. Previously, researchers suggested various methods for identifying electricity fraud from the electricity theft data. Weixian *et al.* proposed a three-layer architecture [17]. Souza *et al.* developed a framework for defense against cyber-assaults based on Phasor Measurement Units (PMU). Furthermore, the author developed a dynamic matrix pencil method to detect electricity theft [18]. Also, rule base and rough set models were applied in [19] to identify the electricity thieves. In [19], using statistical techniques, conventional theft detection was addressed. Further, Fuzzy networks and rough collections were used to compare irregular activity in standard meter readings [20]. In [21], the author used Light GBM, XGBoost, and Cat Boost Learning (CBL) to detect NTL. The concept of smart grids heralds a new age of detecting energy theft. In certain instances, smart meter data were used for further implementation.

Using smart meter data, ML-based classification has received much coverage recently. Daily electricity usage is used to detect theft trends to protect consumers' privacy [22]. Inconsistencies and abnormalities in the obtained data were regularly detected using the SVM [23]. Clustering was used as a primary and secondary stage in some algorithms, making it more suitable for modeling and identifying energy consumption profiles. For anomaly detection, some researchers used research areas like intrusion detection [24]. Furthermore, the author introduced a new Intrusion Detection System (IDS) and Distributed Intelligent Energy Theft (DIET) attack and to defend the AMI system. The suggested IDS can monitor the system in the background and detect potential threats. This IDS is strengthened and dependable as a result of these features.

The author aims to study and transform fine-grained smart-meter data into usable information that may be utilized in consumer behavior modeling and distribution system operations to model complex customer behavior in [25]. Individual load forecasting and customer aggregate are two of the most prominent work. Other works covered include pattern detection, load profiling, personalized pricing design, household behavior coding, and identification of socio-demographic information [26].

Anomaly detection has received a lot of attention from smart grids because it may assist enhance security and protection in smart metering networks by enhancing control reliability and identifying frauds [27]. In smart grids, techniques such as SVM, clustering, and classification were utilised to detect anomalies. To identify power theft, a Rough set and Decision Tree (DT) were employed in [28]. A rule-based approach was employed to detect NTLs in [29]. Other techniques, such as C4.5, DT, Optimum Path Forest (OPF), and one-class SVM [30], were integrated to identify normal users and power thieves.

Many approaches, including DT with SVM [31], Genetic Algorithm (GA) with SVM [32], fuzzy logic with SVM [33], Online Sequential ELM (OS-ELM) [34], and Extreme Learning Machine (ELM) [35], were used to enhance the accuracy of categorising normal and theft users.

Internet access increases the possibility of hacking on smart grid networks. Wei *et al.* suggested a defense architecture for reducing cyber-attacks. According to [36], smart grid remote monitoring services may be connected to the safety thread to make data safer and more private. The Euclidean distance to the cluster core was often used as an unsupervised tool for fuzzy classification [37]. In [38], a wavelet-based approach based on an Artificial Neural Network (ANN) was used to analyze and classify dishonest consumers. An ANN-SVM hybrid platform improves ETD performance in smart grid networks [39]. Load profiles have emerged as a viable and cost-effective method of detecting fraudulent users. Researchers also used various pattern recognition approaches to create load profiling tools based on locally recorded patterns.

Furthermore, deep learning has significant success in visual processing and computer vision. Deep learning techniques simplify feature extraction and classification of data derived from the smart grid due to their ability to process and monitor massive data. CNN is used in [40] to detect energy theft. Hybrid deep learning approaches have recently been used for load forecasting. In [41], a CNN-LSTM combination was utilised for short-term load forecasting. This model has a better performance than others. This model was also used to forecast energy prices and household power demand. This model is more effective and efficient than other methods. The authors in [42] present an LSTM-based evaluation technique for precisely and promptly assessing the system's stability state. Furthermore, the suggested technique outperformed more traditional evaluation methods that depended on shallow learning. Authors utilized the CNN-LSTM based model for non-intrusive load decomposition [42] and electricity theft detection using smart meter data [43]. They addressed the problem of detecting electricity theft with high accuracy, however, they didn't deal with the model's computational complexity. The summary of some literature reviews is shown in Table 2.

III. PROPOSED SYSTEM

In this article, we utilized a new DNN based CNN integrated with LSTM and tuned with the novel optimization technique Black Widow Optimization (BWO) and Blue Monkey optimization (BMO). The CNN algorithm is capable of automatically extracting features from a given data set. However, the LSTM algorithm works well in our case as it produces better results when dealing with sequential data. The combination of these algorithms is used in various applications, including text extraction from images, text extraction from Natural Language Processing (NLP), videos, and Sentiment Analysis (SA) [41]. In this article, we will solve a binary classification problem using the CNN-LSTM method.

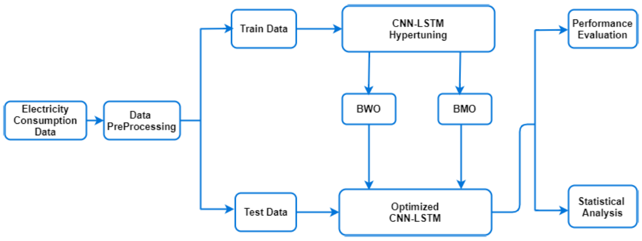


FIGURE 1. Proposed system model.

We use seven hidden layers in this article, four of which will perform CNN operations. Each CNN layer consists of 20 feature sets. The remaining layers would execute the LSTM process. The first, second, and third layers of the LSTM contain 10, 5, and 100 neurons, respectively. The input is first loaded into the model, which will perform data pre-processing and after interpolation and normalization on the data; if the processed data under-sampling is satisfied, it moves on to the next feature extraction phase; otherwise, preprocessing is repeated. We extracted features through ZFNet. After the extraction of the feature, the value is optimized through BWO and BMO, and data is passed toward the next phase of classification. The classification of data is performed through CNN-LSTM. After predicting class, we used different performance matrix to improve performance, i.e., MAPE, RMSE, MSE, F1Score, precision, and recall. Finally, statistical analysis is performed on the achieved result through different techniques like Pearson's test, Spearman's test, Kendell's test, etc. Figures 1 and 2 show the proposed model's block diagram and flow chart.

A. INPUT DATA DESCRIPTION

The research is based on the State Grid [40] collection of real consumption data from customers conducted by the Chinese government. The description of the dataset is shown in Table 3. The dataset used in this article comprises 9655 energy consumption data of consumers gathered for one year. Our main discovery here is that normal and abnormal users generate distinct patterns of energy usage. The energy consumption of two customers is shown in Figure 3; one is an electrical theft user, while the other is a legitimate energy consumer. According to the consumption pattern, an abnormal or electricity theft consumer has an even more fluctuating pattern than a genuine user.

Data on electricity consumption is typically collected using smart meters or different sensors equipped at the client-side. After that, the data is combined and routed to a prime hub via a communications network. this situation, smart meters may malfunction, detectors might fail, network delays may occur, and database servers could fail.

Consumption datasets are unavoidable. In this dataset, we also observed some outliers. If the missing values are simply omitted, the dataset gets smaller significantly, creating

TABLE 2. Existing ETD literature methods.

Methodology (s)	Aims and objectives	Sources (s) of Information/Achievement(s)	Drawback(s)
Hardware Devices [15], [16]	Electricity Theft Detection	Radio frequency-based identification of theft.	Vulnerability, failure due to a fault sometimes in hardware. Replacing batteries etc.
3-layer framework [17]	Cyber Attack detection and prevention in smart grid	Create an X2 detector capable of detecting random data injection, wrong data injection, and Dos attacks.	This detector can be breached if false data is injected.
PMU based security system [18]	Protecting system from security attacks	Detection of unobserved attacks	State data problems, Asynchronous meter reading, and taking harmful action of grid operators can compromise data.
Complex Matrix Approach [19]	Smart Load Monitoring	Load monitoring, Noise reduction, less data for identifying signals, easy method for extracting poles.	Correct correlation is a must for classification, Accuracy issues.
Rough set models [19], [20]	Fraud Detection and Electricity Price Forecasting	Fraud detection through high and low voltage consumption, Price forecasting through RNN.	Some fraudulent users were identified as non-fraudulent, Also show some user who is neither fraudulent nor genuine users, and also in case of price forecasting, it creates conflicts.
Statistical Techniques [21], [22]	Conventional Theft Detection	Credit card detection, telecommunication fraud detection, intrusion, and account defaulting.	Costly, Difficult to implement and maintain.
NTL Analysis [23]–[27]	Detection of abnormalities	Accuracy and time performance was excellent.	The result was not optimum.

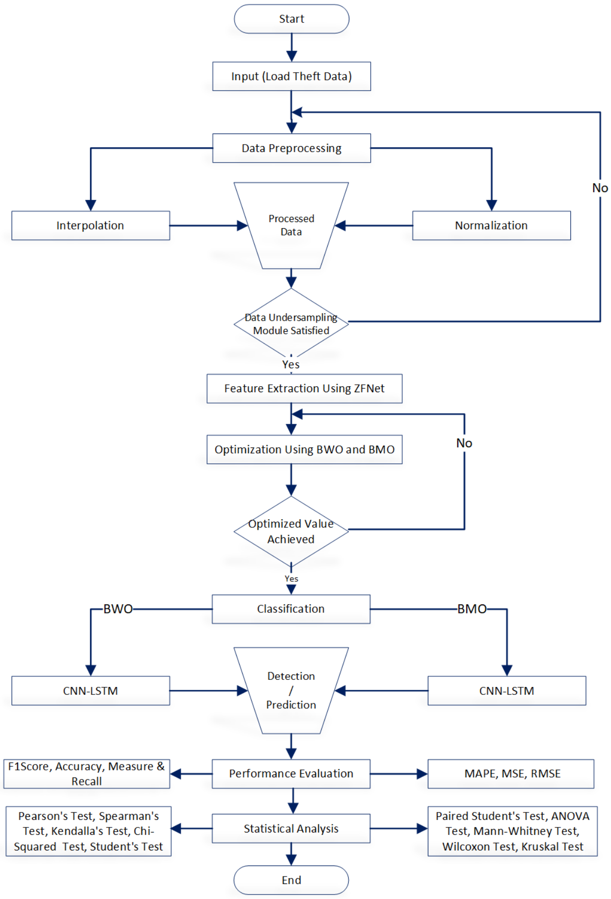

Start
Input (Load Theft Data)
Data Preprocessing
Interpolation
Processed Data
Normalization
No
Data Undersampling Module Satisfied
Yes
Feature Extraction Using ZFNet
Optimization Using BWO and BMO
No
Optimized Value Achieved
Yes
Classification
BWO
BMO
CNN-LSTM
Detection / Prediction
CNN-LSTM
F1Score, Accuracy, Measure & Recall
Performance Evaluation
MAPE, MSE, RMSE
Pearson's Test, Spearman's Test, Kendalla's Test, Chi-Squared Test, Student's Test
Statistical Analysis
Paired Student's Test, ANOVA Test, Mann-Whitney Test, Wilcoxon Test, Kruskal Test
End

FIGURE 2. Flow chart of proposed system.

reliable analysis difficult. We proposed a data preprocessing algorithm to avoid downsizing in the dataset.

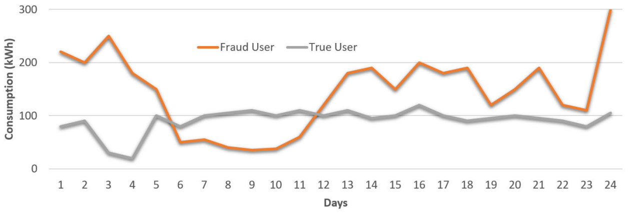


FIGURE 3. Power consumption trend.

IV. PREPROCESSING OF DATA

The interpolation method is used to preprocess the data, which helps in achieving accuracy. Equation 1 explains the interpolation method [41].

$$\begin{cases} \dfrac{m_{m+1} + m_{k-1})}{2} & \text{if } m_k \in NaN, m_{k-1} \text{ and } m_{k+1} \notin \text{NaN} \\ 0 & \text{if } m_k \in NaN, m_{k-1} \text{ and } m_{k+1} \in \text{NaN} \\ m_k & m_k \in \text{NaN} \end{cases} \tag{1}$$

The 3-sigma rule is then applied to the raw data to eliminate outliers [8], [41]. These outliers are aware that non-working days have the highest energy consumption. We use Equation 2 to restore these values by the 3-Sigma rule of thumb.

$$f\ (m_k) = \begin{cases} avg(m) + 2\,\text{std}(m) & \text{if } m_k > \text{avg}(m) + 2std(m) \\ m_k & \text{else} \end{cases} \tag{2}$$

The average of m is represented by *avg(m)*, and the standard deviation of m is defined by *STD (m)* in equation 2. This method is effective in dealing with outliers.

To normalize the data between the 1 and 0 scales, we used the Min-Max scaling approach, interpolation, and

Algorithm 1 Electricity Theft Detection Model

```
Require: Electricity Theft Data (DTA);
/* Separating the Data into target and features */
X: Features DTA;
Y: Target DTA;
/* Preproccesing using interpolation and normalization methods */
Preprocessing(DTA);
/* Data Balancing of data */
SMOTEAlgorithm (DTA);
/* Data Splitting into Training and Testing */
x_test, x_train, y_test, y_train = split(X,Y);
/* Feature Extraction using ZFNET */
Selected_Feature = ZFNet(x_train, y_train);
if (SelectedFeaturesImportance > Threashold) then
    Reserve Feature;
else
    Drop Feature
end if
/* (Optimization using BWO and BMO) */
CNN Parameters Setting;
BMO_Optimized_Value = BMO (CNN_LSTM_Parameters);
BWO_Optimized_Value = BWO (CNN_LSTM_Parameters);
if (BMO_Optimized_Value and BWO_Optimized_Value == Satisfied) then
    CNNModel(BMO_BWO_Optimized_Value);
else
    Repeat Parameter Setting;
end if
Prediction using CNN−LSTM;
Compare predictions with y_test;
Applying Performance evaluation metric;
Statistical Analysis of proposed algorithm and state of the art methods;
```

TABLE 3. Dataset description and details.

Description of Data	Value
Data collection time frame	1 Jan-31 Dec
Total Consumer	9956
Normal User	8562
Abnormal or Thief Consumer	1394

the 3-sigma law. It is required because neural networks perform poorly when the results are inconsistent. By giving the data a consistent scale, data normalization helps the training phase of DL models. Equation 3 is used to normalize the data [41].

$$N' = \frac{N - \mathrm{Min}(N)}{\mathrm{Max}(N) - \mathrm{Min}(N)} \tag{3}$$

N' represents the normalized value. The consistency of the data used to train ML algorithms affects the efficiency of the algorithms. Data pre-processing improves the precision and reliability of the data used in these models.

A. DATA BALANCING

In the dataset, the number of average energy users outnumbers the number of theft users. The ratio of data mismatch is a significant problem in ETD that must be addressed immediately; else, because the classifier would be skewed towards the categories that contain more data, the classifier's performance will be low [42].

SMOTEBoost and SMOTE, assisting in navigating the imbalanced collection of results. In this article, we also used a sampling-based method. Strategies based on sub-sampling or oversampling the unbalanced data set are used to reduce the quantity difference between the two data categories. To avoid overfitting, the entries of the majority class are rejected automatically to minimize the occurrence ratio of the majority class. Although random removal may eliminate crucial data, which may or may not be a fair sample representation, this method reduces the dataset, which is statistically beneficial. Because the model was built using test data, the information it provides may be less accurate. It seeks to balance class representation by removing instances of the majority class arbitrarily. In the case of two distinct classes that are similar, we eliminate all instances of the majority class to maximize the spacing between the two classes. This facilitates the classification process.

Most under-sampling techniques use methods based on near-neighbors to reduce the problem of data loss. The following is a basic overview of the process of some near-neighbor ways [41], [42]:

- Step 1: The process first determines the dissimilarities between the majority and minority class instances. In this case, an under-sampling of the majority class is required.
- Step 2: The majority class N instances with the smallest distances from the minority class are then chosen.
- Step 3: If the minority class has k instances, the next process is k*n instances within the majority class.

B. ZFNET FEATURE EXTRACTION

ZFNet is the updated 5-layer version of CNN. A 7 × 7 filer and a reduced stride value are used in layer-1. Softmax is the final layer of ZFNet. It's used in learning how to isolate and disseminate features. The representation spaces created by all layer filters are presented in detail in this article using ZFNET for feature extraction. Using a deconvolution network, all of a layer's activations are utilized to remove the associated features [44]. Convolutional and pooling layers are used. In the last dense layer, the Softmax is used as the activation mechanism. The ZFNET modules' multi-pooling layers outperform the competition when it comes to significant data advancements. We'll examine the input image that maximizes the filter's activation and see what features each filter captures. In the convolutional approach,

TABLE 4. Hyper-parameters of ZFNET.

Parameters	Values	Description
Epoches	5	No of iterations
Batch Size	80	Training samples for each iteration
Optimizer	Adam	Learning rate
Dropout	0.001	Over fitting Resolving rate
Learning Rate	0.01	Tuning of parameters

spreading the kernel over the full inputs gives a function chart. After multiple feature mapping processes, the kernel function combines the final output of the convolution layer [37].

$$k = m \times T \rightarrow k[s] = \sum_{d=-\infty}^{+\infty} \times [s-d]T[d] \tag{4}$$

In Equation 4, the input is m, and the filter is T, also known as the kernel. It is possible to compute the failure by multiplying the number of times a particular filter is activated by the number of times the input image is initialized to be a random image. The activation function Relu is used in the model to create nonlinearity by acting as an activation function [39]:

$$\text{Relu}(m) = \max(0, m) \tag{5}$$

After the dropout layer operations, the essential features are visualized using a dense layer. To avoid over-fitting, the learning rate is set at 0.001, and the dropout rate is set at 0.01. This method can be used as a Softmax activation mechanism for the final dense layer [45].

$$P\left(k = s \mid \varphi^{(d)}\right) = \frac{\neg\varphi^{(d)}}{\sum_{S=0}^{l} \neg\varphi_l^{(d)}} \tag{6}$$

If H *and* G are the functions and weight matrices, then s in the above equation is calculated as follows [39], [43]:

$$\varphi = \sum_{d=1}^{k} H_d G_d = H^F G \tag{7}$$

Table 4 shows the values of the ZFNET's hyper-parameters as well as their meanings. These hyper-parameters are the learning rate, the optimizer, the batch size, the number of eras, and the dropout rate. These criteria are critical for achieving the best possible results from the ZFNET module.

C. CLASSIFICATION USING CNN-LSTM OPTIMIZED BY BWO AND BMO

1) MODEL OF CNN

CNN is introduced as a kind of DNN class that was implemented first [38]. This approach is influenced by the Human Visual Cortex, which is used for object detection. In an image, CNN recognizes objects and their class. In terms of feature extraction, it differs from Conventional Machine Learning (CML). It extracts functions globally across a variety of layers.

Several pooling layers, a Fully Connected layer, and a convolution layer are included in this architecture. CNN is built on the foundation of a main convolutional layer.

2) CONVOLUTIONAL LAYER

The convolutional layer collects input data and transfers its outcome to the next phase. Its function is synchronized with the response of neurons in the visual cortex to a specific stimulus. Convolutional neurons only process information that falls inside their receptive field. It is a mathematical approach that involves the use of two sets of data.

The inputs and the kernel, which is a convolution filter, are the two sets of information that make up the framework of CNN. Convolutional operations are applied to the input data by slide the kernel over the complete input, resulting in a function map. Different filters were used to conduct several convolutions on the data to generate distinct feature maps. After extracting distinct maps from the data, these maps are finally combined to create output from the convolutional layer.

3) ACTIVATION LAYER

After convolutional layer processes, activation functions were used to add nonlinearity to the model. There are several activation mechanisms, such as linear functions, sigmoid functions, and tanh functions. However, we can use the Rectified Linear Unit (RELU) since it allows us to train the model faster and ensures close weight global optimization. Equation 8 shows the RELU function [40].

$$F(xi) = \max(0, x\ i) \tag{8}$$

4) POOLING LAYERS

Pooling layers reduce data dimensionality by merging the outputs of neuron clusters in one layer into a single neuron in the next layer. It appears next to the convolutional layer. It reduces over-fitting and training time by reducing dimensions. Max pool function is commonly used in CNN's to pick the highest value in the pooling window.

5) FULLY CONNECTED LAYER

This layer connects a neuron from one layer to a neuron from every other layer. CNN extracts pooling layers and low-level convolutional characteristics such as points, and the FC layer then does classification based on the retrieved features. The Softmax function is employed as the activation function in this last layer, and it assigns a probability to each class ranging from 0 to 1.

a: WEIGHTS

Each CNN/ANN neuron generates an output value by using activation functions to input from the receptive region of the preceding layer. A vector of weights determines the attribute that is added to the input values. W represents the weight, and X represents the function matrix in the Equation 9 [33], [38], [40].

$$\emptyset = \sum_{i=0}^{k} W_i X_i = W^T X \tag{9}$$

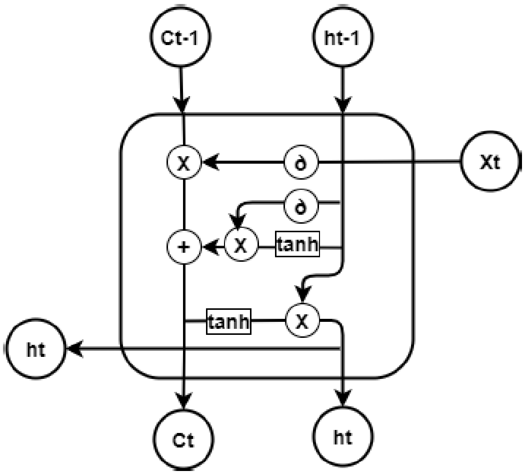


FIGURE 4. LSTM architecture.

D. LSTM ALGORITHM

The LSTM technique is a Recurrent Neural Network (RNN) subclass that was built to overcome the RNN's short-term memory issue. This algorithm is capable of propagating and recalling specific details from the beginning to the end. This method will employ the LSTM's basic form, as depicted in Figure 4 [39], [40]. An LSTM has the same setup as an RNN, but its module has distinct internal components, as shown in Figure 4. This method relies heavily on the cell state, which provides information and the chain. Several units known as gates modify or drop cell state detail. The LSTM is made up of three gates (forgotten, input, and output).

The forgot gate is composed of a sigmoid layer that generates an output set by combining the initial hidden state *(ht-1)* and the current input *(Xt)* (0 to 1). The decision of keeping and discarding the data is made in this layer. 0 indicates that the previous value should be forgotten, while one suggests that the previous detail should be retained. Equation 5 shows the output gate from this gate [40].

$$f_t = \sigma \left(W_f \left[h_{t-1} \right] + b_f \right. \tag{10}$$

The forgot gate then uses *tan h* and sigmoid features to determine what details should be applied to the cell state. Both functions accept (*ht-1*) and (*Xt*) as inputs. The sigmoid function's performance indicated whether or not the present information is significant, and the *Tanh* function reacted quickly to the network by squashing values ranging from +1 to −1. As seen in equation 11, results were then multiplied [21].

$$\begin{aligned} i_t &= \sigma \left(W_i \left[h_{t-1}, x_t \right] + b_i \right. \\ \tilde{C} &= \tan h \left(W_C \cdot \left[h_{t-1}, x_t \right] + b_C \right. \end{aligned} \tag{11}$$

Cell state information is changed after receiving output from the input and forget gates. This is the product of a pointwise multiplication of the contribution of the forgotten state and the present state. If $F_{t=0}$, the multiplication outcome C_t would be 0, indicating that the prior amount has been fully dropped, while if $F_t = 1$, it will be kept. The cell's state is then updated via pointwise addition [22].

$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C}t \tag{12}$$

Finally, the output gate finds the final output. This performance is the next hidden state (h_t). The sigmoid function takes (h_{t-1}) and (X_t) as inputs in this gate, and the current state C_t is passed via *tan h*. Finally, all outputs are multiplied to determine which information is borne by hidden layers [38]–[40].

$$\begin{aligned} S_t &= \sigma \left(W_0 \cdot \left[h_{t-1}, x_t \right] + b_0 \right) \\ h_t &= S_t \times \tan h \left(C_t \right) \end{aligned} \tag{13}$$

That is why CNN–LSTM is being used. Figure 5 shows CNN layers preceding LSTM layers, which are highly effective and resilient for detecting energy theft by classification in smart grids.

1) BWO

BWO is an optimization algorithm introduced in [45] that focuses on the black spider's mating behavior. It has a good output in the discovery and exploitation phases and has a quick convergence pace, and also it avoids the local optima problem.

The steps of BWO are given below:

a: INITIAL POPULATION

BWO chromosomes include widow, similar to the chromosomes of Particle Swarm Optimization (PSO). Each black widow spider represents a value of the problem variable. The steps of BWO are shown in Figure 6. The formula shown in equation 11 can be used to calculate widow fitness [38]:

$$Fitness = f(widow) \tag{14}$$

b: PROCREATE

To replicate an array named alpha, the widow array should be generated with random numbers. The offspring is generated with α, where *x1* and *x2* are parents and *y1* and *y2* are descendants in the following equations [38], [44].

$$\begin{aligned} y_1 &= \alpha \times x_1 + (1 - \alpha) \times x_2 \\ y_2 &= \alpha \times x_2 + (1 - \alpha) \times x_1 \end{aligned} \tag{15}$$

This technique will be performed Nvariable twice, although random numbers must not be repeated. Next, the children and mothers will be sorted into an array according to their health value.

c: CANNIBALISM

Cannibalism is classified into three groups. Sexual Cannibalism occurs when a female black widow spider eats her spouse, and sibling cannibalism occurs when a powerful black widow spider eats their weak siblings. Child cannibalism occurs when a child spider eats her mother. In this algorithm, we will set a cannibalism value that will be used to evaluate survivor

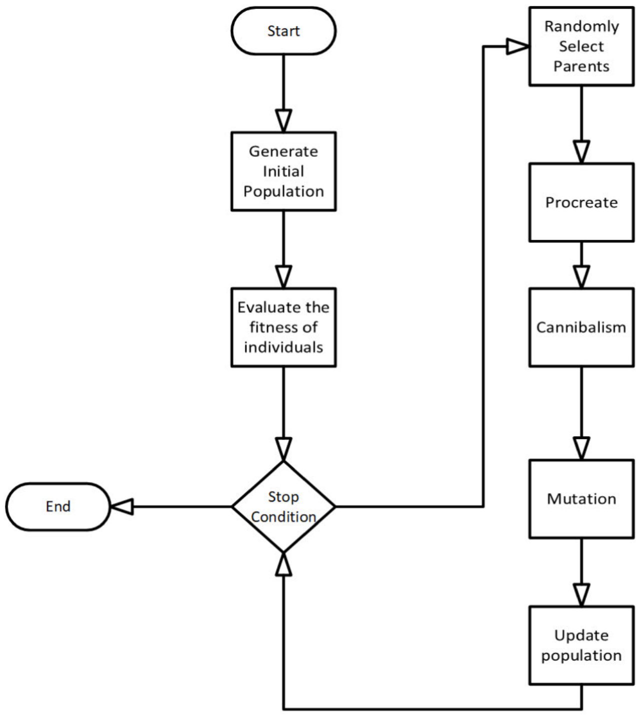


FIGURE 5. Steps of BWO.

spiders. We also use fitness values to determine whether a spider is strong or weak.

d: MUTATION

Mute pop is chosen randomly in this step. The mutation rate can be used to calculate mute pop.

e: CONVERGENCE

Like many other algorithms, this algorithm will take into account three-stop conditions: a) A predetermined no of iterations, b) several changes in the fitness value, and c) meeting the defined limit of accuracy.

f: PARAMETER SETTINGS

Specific parameters help in producing better performance. These parameters include the rate of procreation, the rate of Cannibalism, and the rate of mutation.

E. BMO

BMO is a novel algorithm inspired by natural blue monkey swarms [45]. This technique determines the number of males in a group. There is only one male present in the group of blue monkeys other than the season of breeding. The monkeys are divided into groups, searching for good places for living and food over long distances. The younger male monkeys should leave the female group as soon as possible to become more successful. They will enter into a challenge with the dominant male monkey of another family. If the young monkey is successfully defeating the dominant male monkey, he will become the family leader. These challenges select a good leader who can live with females and young monkeys and provide better food and accommodation. There are generally many children, women, and only one male in the group of blue monkeys.

1) UPDATING POSITIONS

The updating of positions in the group for every monkey depends on the best monkey's position. This can be done with the help of equations 14 and 15 [46].

$$\begin{aligned} \text{Rate}_{\,i+1} &= (\,0.\,7 \times \text{ Rate}_{\,i}) + (W_{\text{leader}} - W_i) \\ &\quad \times \text{rand} \times (X_{\text{best}} - X_i) \\ X_{i+1} &= X_i + \text{ Rate}_{\,i+1} \times \text{ rand} \end{aligned} \tag{16}$$

In the above equations, *rate* represents the power of the monkey, W_{leader} is the leader's weight, and W_i is the monkey's weight where all weights of the monkeys are a random number from 4-6. X is monkey position, X_{best} is monkey's leader position, and *rand* is a random number (0, 1).

For updating children, equation 16 can be utilized [47].

$$\begin{aligned} &Rate^{ch_{(i+1)}} \\ &\quad = \left(0.7 \times Rate^{ch_{(i)}}\right) \\ &\qquad + \left(W^{\,ch_{(leader)}} - W^{ch_{(i)}}\right) \times \text{rand} \times X^{ch_{(best)} - X^{ch_{(i)}}}\Big) \end{aligned} \tag{17}$$

$$\begin{aligned} &X^{ch_{(i+1)}} \\ &\quad = X^{ch_{(i)}} + Rate^{ch_{(i+1)}} \times \text{ rand} \end{aligned} \tag{18}$$

$Rate_{ch}$ represent children weight, $W_{ch(leader)}$ is the weight of younger leader in children, $W_{ch(i)}$ is the random weight of the children from (4, 6). X_{ch} represents the kid position, $W_{ch(best)}$ represents the leader's position in children, and rand represents a random integer (0, 1). The position will be updated in each iteration of the algorithm. In our work, we will use this algorithm for tuning CNN-LSTM.

F. CLASSIFICATION WITH ENSEMBLER

The CNN-LSTM is tuned with the BWO and BMO to perform ETD classification. BWO and BMO calculate the best values for CNN LSTM parameters, as shown in Figure 7. The optimization techniques determine the best suitable value for the classifier's parameters, based on which classifier performs better.

V. EXPERIMENTAL RESULTS

The results of the proposed model implementation, as measured by performance metrics, are presented in this section. The following system requirements were used to implement our proposed model: Core (i7), RAM (16GB), Processor (4.8 GHz), as well as Anaconda (Spyder) as an IDE and Python as a programming language. The explanation of the simulation's outcomes is given below:

A. PERFORMANCE EVALUATION

The efficiency of our suggested model was assessed using evaluation and error metrics. F-score, accuracy, recall, and

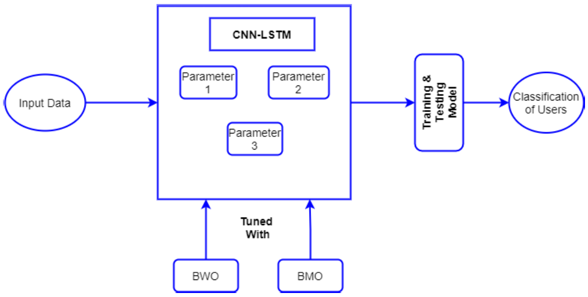


FIGURE 6. Ensembler view of proposed system.

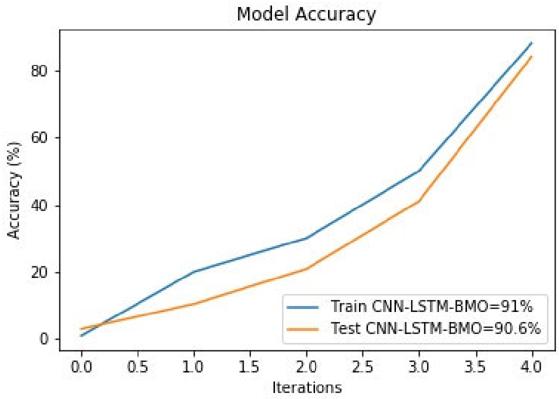


FIGURE 7. Accuracy VS iteration of CNN-LSTM-BMO.

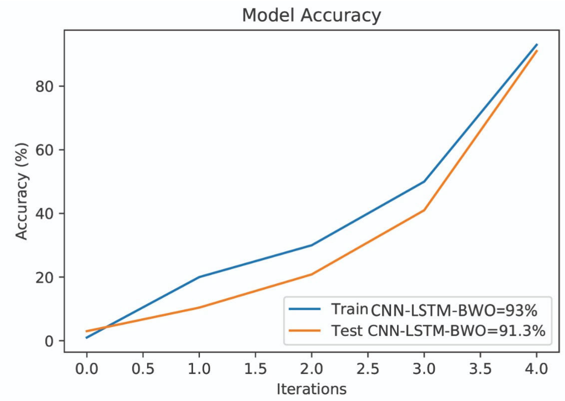


FIGURE 8. Accuracy VS iteration of CNN-LSTM-BWO.

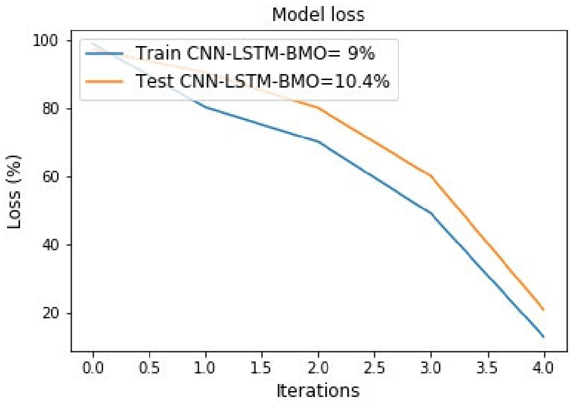


FIGURE 9. Loss VS iteration of CNN-LSTM-BMO.

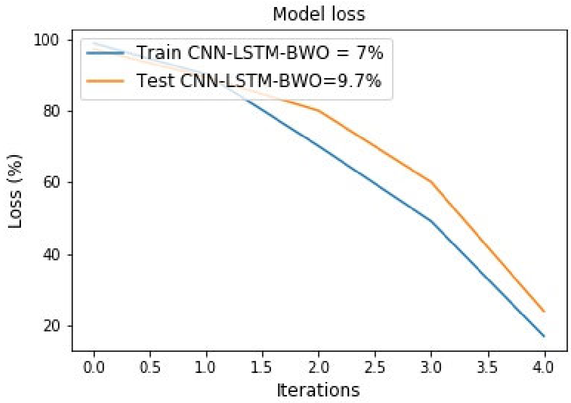
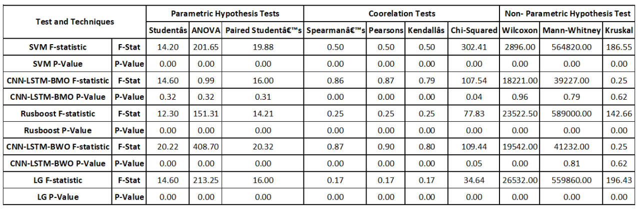


FIGURE 10. Loss VS iteration of CNN-LSTM-BWO.



Test and Techniques		Parametric Hypothesis Tests			Coorelation Tests				Non- Parametric Hypothesis Test		
		Studentâs	ANOVA	Paired Studentâ€™s	Spearmanâ€™s	Pearsons	Kendallâs	Chi-Squared	Wilcoxon	Mann-Whitney	Kruskal
SVM F-statistic	F-Stat	14.20	201.65	19.88	0.50	0.50	0.50	302.41	2896.00	564820.00	186.55
SVM P-Value	P-Value	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
CNN-LSTM-BMO F-statistic	F-Stat	14.60	0.99	16.00	0.86	0.87	0.79	107.54	18221.00	39227.00	0.25
CNN-LSTM-BMO P-Value	P-Value	0.32	0.32	0.31	0.00	0.00	0.00	0.04	0.96	0.79	0.62
Rusboost F-statistic	F-Stat	12.30	151.31	14.21	0.25	0.25	0.25	77.83	23522.50	589000.00	142.66
Rusboost P-Value	P-Value	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
CNN-LSTM-BWO F-statistic	F-Stat	20.22	408.70	20.32	0.87	0.90	0.80	109.44	19542.00	41232.00	0.25
CNN-LSTM-BWO P-Value	P-Value	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.00	0.81	0.62
LG F-statistic	F-Stat	14.60	213.25	16.00	0.17	0.17	0.17	34.64	26532.00	559860.00	196.43
LG P-Value	P-Value	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

FIGURE 11. Performance metrics evaluation of proposed model vs. benchmark algorithm.

precision are among the evaluation criteria, whereas RMSE, MSE, and MAPE are performance error metrics. Our suggested methods beat benchmark methods, including the most significant value and lowest error rate in MAPE, MSE, and RMSE performance measures. Figure 8 illustrates the accuracy of CNN-LSTM-BMO, which is 91% on training data and 90.6% on testing data with a 0.4% variance. Figure 9 shows the accuracy of CNN-LSTM-BWO is 93% on training data and 91.3% on testing data with a 1.7% variation.

Figure 10 illustrates the loss of CNN-LSTM-BMO, which is 9% on training data and 10.4% on testing data with a 1.4% variance, and Figure 11 represents the loss of CNN-LSTM-BWO is 7% on training data and 9.7% on testing data with a 1.4% variation.

The formulas for performance error metrics and performance evaluation metrics are given below [48]:

$$
\begin{aligned}
Precision &= \frac{Pos\_True}{(Pos\_False + Pos\_True)} \\
F1 &= \frac{2 * Precision * Recall}{Precision + Recall} \\
Recall &= \frac{Pos\_True}{Pos\_True + Neg\_False} \\
Accuracy &= \frac{Pos\_True+Neg\_True}{Pos\_True+Neg\_True+Pos\_False+Neg\_False}
\end{aligned}
\tag{19}
$$

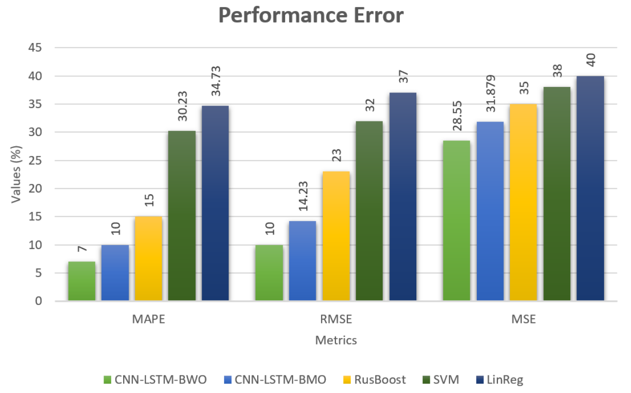


FIGURE 12. Statistical analysis of proposed model VS benchmark algorithm.

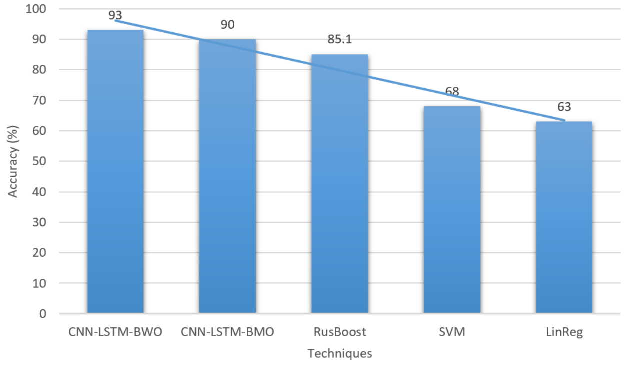


FIGURE 13. Accuracy of proposed techniques VS benchmark algorithm.

TABLE 5. Tabular view of performance evaluation metrics values.

Techniques	Evaluation Metrics (%)				Performance Error Metrics (%)		
	F1-Score	Accuracy	Precision	Recall	MAPE	RMSE	MSE
CNN-LSTM-BWO	92	93	92.32	94.02	7	10	28.55
CNN-LSTM-BMO	87	90	89	92.87	10	14.23	31.879
RusBoost	86.2	85.1	87.43	88	15	23	35
SVM	71	68	65	72	30.23	32	38
LinReg	61	63	67	65	34.73	37	40

$$\begin{aligned} \text{MSE} &= \frac{1}{N}\sum_{i=1}^{N}(\text{Actual-Predicted})^2 \\ MAPE &= \frac{1}{N}\sum_{i=1}^{N}\left|\frac{\text{Actual-Predicted}}{\text{Actual}} * 100\right| \\ RMSE &= \sqrt{\frac{\sum(\text{ Predicted-Actual })^2}{N}} \end{aligned} \tag{20}$$

Figures 12 and 13 illustrate the values of the evaluation and performance error matrices, demonstrating that the error values of our suggested approaches are minimal when compared to the other techniques. Figure 12 clearly shows that the CNN-LSTM-BWO and CNN-LSTM-BMO have the maximum accuracy of 93 and 90 percent, respectively. Furthermore, the CNN-LSTM-BWO and CNN-LSTM-BMO had the lowest MAPE error of 7% and 10%, respectively. These values demonstrate the superiority of our proposed approaches. Table 5 shows the statistical analysis of the proposed approaches and benchmark techniques. Table 6 displays the performance values in tabular format. The CNN-LSTM-BWO and CNN-LSTM-BMO have the lowest MAPE error and highest accuracy. Figure 14 depicts the accuracy of our techniques.

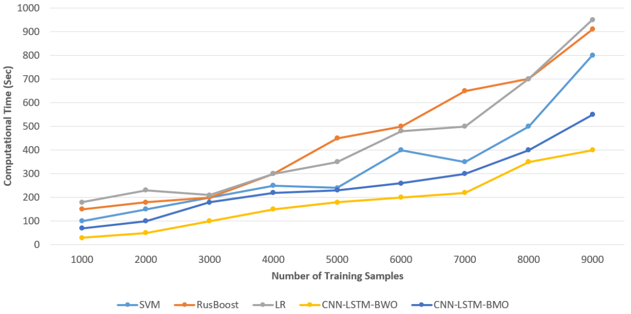


FIGURE 14. Computational complexity of proposed techniques VS benchmark algorithm.

Figure 14 shows the computational cost of the proposed model and the state-of-the-art methods. The reason behind the low complexity and better classification are that our proposed method is better tuned with BWO and BMO algorithms, which results in better classification. We can see that the computational complexity of our proposed model is less than the other methods. As the number of training samples increases, our proposed model complexity is increasing gradually but less than other methods.

VI. CONCLUSION AND FUTUREWORK

In this work, we have applied novel optimization techniques BMO and BWO to the deep learning model CNN embedded LSTM (CNN-LSTM) for electricity theft detection. We have mainly focused on the extraction of electricity usage patterns in the dataset. For better accuracy in detecting electricity thieves, we have proposed a model consisting of data preprocessing, feature engineering/pattern extraction, optimization of classification technique, training/testing, classification/detection, performance evaluation, and statistical analysis of classifiers. In a pre-processing step, we have applied interpolation and normalization to the dataset. Furthermore, we have also used SMOTE method for downsampling the data, i.e., equal the number of electricity thieves' data and normal data. Afterward, we have applied ZFNet for feature extraction/pattern extraction. After cleaning the data and extracting the features, the data is sent to the classifier for training purposes. The hyper-parameters of the proposed classifier CNN-LSTM are tuned by two optimization techniques, i.e., BMO and BWO. The tuned classifier is then trained and tested on the cleaned data. Our proposed model CNN-LSTM-BWO and CNN-LSTM-BMO accuracy and recall are 93%, 90%. and 94.02% and 92.87% Furthermore, the performance error MAPE, RMSE and MSE of CNN-LSTM-BWO and CNN-LSTM-BMO is 7%, 10%, 28.55% and 10%, 14.23%, 31.87%. Our proposed model accuracy is 5 to 8% better than state-of-the-art algorithms. Furthermore, the error rate is also 4% less than the other algorithms.

We have also performed statistical analysis to verify the superiority of our proposed techniques and to clarify the performance of the proposed classifier and state-of-the-art. The performance evaluation and statistical analysis results show that our proposed model has high accuracy and precision and the lowest error rate. The significant findings of our model are to classify the theft user and normal user accurately and within less time. Our model is more scaleable to a large amount of data. As its computational complexity is less than the state-of-the-art algorithms.

In the future, we will apply more novel optimization methods, hybrid ML, and deep learning techniques to handle a huge amount of data for classification/detection.

REFERENCES

[1] S.-Y. Chen, S.-F. Song, L.-X. Li, and J. Shen, "Survey on smart grid technology," *Power Syst. Technol.* vol. 33, no. 8, pp. 1–7, 2009.

[2] O. M. Butt, M. Zulqarnain, and T. M. Butt, "Recent advancement in smart grid technology: Future prospects in the electrical power network," *Ain Shams Eng. J.*, vol. 12, no. 1, pp. 687–695, Mar. 2021.

[3] M. Jayachandran, C. R. Reddy, S. Padmanaban, and A. H. Milyani, "Operational planning steps in smart electric power delivery system," *Sci. Rep.*, vol. 11, no. 1, pp. 1–21, 2021.

[4] S. Pealy and M. A. Matin, "A survey on threats and countermeasures in smart meter," in *Proc. IEEE Int. Conf. Commun., Netw. Satell. (Comnetsat)*, Dec. 2020, pp. 417–422.

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

[6] T. N. Mahbub, S. S. Hossain, R. A. Akash, S. M. S. Reza, and Z. Tasnim, "Implementing fuzzy analytical hierarchy process (FAHP) to measure malicious behaviour of codes in smart meter," in *Proc. 2nd Int. Conf. Robot., Electr. Signal Process. Techn. (ICREST)*, Jan. 2021, pp. 90–94.

[7] B. K. Sovacool, "The political economy of energy poverty: A review of key challenges," *Energy Sustain. Develop.*, vol. 16, no. 3, pp. 272–282, 2012.

[8] 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. 2015.

[9] N. Ayub, K. Aurangzeb, M. Awais, and U. Ali, "Electricity theft detection using CNN-GRU and manta ray foraging optimization algorithm," in *Proc. IEEE 23rd Int. Multitopic Conf. (INMIC)*, Nov. 2020, pp. 1–6.

[10] 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, 2020.

[11] 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, 2019.

[12] 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.

[13] S. K. Singh, R. Bose, and A. Joshi, "Energy theft detection for AMI using principal component analysis based reconstructed data," *IET Cyber-Phys. Syst., Theory Appl.*, vol. 4, no. 2, pp. 179–185, 2019.

[14] M. A. de Souza, J. L. R. Pereira, O. G. de Alves, B. C. de Oliveira, I. D. Melo, and P. A. N. Garcia, "Detection and identification of energy theft in advanced metering infrastructures," *Electr. Power Syst. Res.*, vol. 182, May 2020, Art. no. 106258.

[15] W. Xiong, Y. Cai, L. Wang, Y. Zhang, Q. Ai, Z. Li, Y. Wang, and S. Yin, "Study on energy theft detection based on customers' consumption pattern," in *Proc. 11th Int. Conf. Appl. Energy*, vol. 3, 2019, pp. 1–4.

[16] A. A. Chauhan, "Non-technical losses in power system and monitoring of electricity theft over low-tension poles," in *Proc. 2nd Int. Conf. Adv. Comput. Commun. Eng.*, May 2015, pp. 280–284.

[17] 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.

[18] C. A. C. Montañez and W. Hurst, "A machine learning approach for detecting unemployment using the smart metering infrastructure," *IEEE Access*, vol. 8, pp. 22525–22536, 2020.

[19] 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.

[20] S.-C. Yip, K. Wong, W.-P. Hew, M.-T. Gan, R. C.-W. Phan, and S.-W. Tan, "Detection of energy theft and defective smart meters in smart grids using linear regression," *Int. J. Electr. Power Energy Syst.*, vol. 91, pp. 230–240, Oct. 2017.

[21] 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.

[22] T. Zhang, R. Gao, and S. Sun, "Theories, applications and trends of non-technical losses in power utilities using machine learning," in *Proc. 2nd IEEE Adv. Inf. Manage., Commun., Electron. Automat. Control Conf. (IMCEC)*, May 2018, pp. 2324–2329.

[23] F. Kaytez, "A hybrid approach based on autoregressive integrated moving average and least-square support vector machine for long-term forecasting of net electricity consumption," *Energy*, vol. 197, Apr. 2020, Art. no. 117200.

[24] M. Chakraborty, "Advanced monitoring based intrusion detection system for distributed and intelligent energy theft: DIET attack in advanced metering infrastructure," in *Transactions on Computational Science*. Berlin, Germany: Springer, 2018, pp. 77–97.

[25] Y. Wang, Q. Chen, and C. Kang, *Smart Meter Data Analytics: Electricity Consumer Behavior Modeling, Aggregation, and Forecasting*. Singapore: Springer, 2020.

[26] M. S. Sheikh, J. Wang, and A. Regan, "A game theory-based controller approach for identifying incidents caused by aberrant lane changing behavior," *Phys. A, Stat. Mech. Appl.*, vol. 580, Oct. 2021 Art. no. 126162.

[27] T. Kotsiopoulos, P. Sarigiannidis, D. Ioannidis, and D. Tzovaras, "Machine learning and deep learning in smart manufacturing: The smart grid paradigm," *Comput. Sci. Rev.*, vol. 40, May 2021, Art. no. 100341.

[28] 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.

[29] Z. Qu, H. Li, Y. Wang, J. Zhang, A. Abu-Siada, and Y. Yao, "Detection of electricity theft behavior based on improved synthetic minority oversampling technique and random forest classifier," *Energies*, vol. 13, no. 8, p. 2039, 2020.

[30] R. M. R. Barros, E. G. da Costa, and J. F. Araujo, "Evaluation of classifiers for non-technical loss identification in electric power systems," *Int. J. Electr. Power Energy Syst.*, vol. 132, Nov. 2021, Art. no. 107173.

[31] Z. Qu, H. Liu, Z. Wang, J. Xu, P. Zhang, and H. Zeng, "A combined genetic optimization with AdaBoost ensemble model for anomaly detection in buildings electricity consumption," *Energy Buildings*, vol. 248, Oct. 2021, Art. no. 111193.

[32] 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.

[33] 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.

[34] Y. Li, R. Qiu, and S. Jing, "Intrusion detection system using online sequence extreme learning machine (OS-ELM) in advanced metering infrastructure of smart grid," *PLoS ONE*, vol. 13, no. 2, 2018, Art. no. e0192216.

[35] Z. Yan and H. Wen, "Electricity theft detection base on extreme gradient boosting in AMI," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–9, 2021.

[36] L. Guo, L. Gao, B. Cai, Y. Qu, Y. Zhou, and S. Yu, "A covert electricity-theft cyber-attack against machine learning-based detection models," *IEEE Trans. Ind. Informat.*, early access, Jun. 16, 2021, doi: 10.1109/TII.2021.3089976.

[37] N. Somu, M. R. G. Raman, and K. Ramamritham, "A hybrid model for building energy consumption forecasting using long short term memory networks," *Appl. Energy*, vol. 261, Mar. 2020, Art. no. 114131.

[38] D. Lehri and A. Choudhary, "A survey of energy theft detection approaches in smart meters," in *Intelligent Energy Management Technologies*. Singapore: Springer, 2021, pp. 9–24.

[39] U. Ugurlu, I. Oksuz, and O. Tas, "Electricity price forecasting using recurrent neural networks," *Energies*, vol. 11, no. 5, p. 1255, 2018.

[40] A. Ullah, N. Javaid, O. Samuel, M. Imran, and M. Shoaib, "CNN and GRU based deep neural network for electricity theft detection to secure smart grid," in *Proc. Int. Wireless Commun. Mobile Comput. (IWCMC)*, Jun. 2020, pp. 1598–1602.

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

[42] X. Zhou, J. Feng, and Y. Li, "Non-intrusive load decomposition based on CNN-LSTM hybrid deep learning model," *Energy Rep.*, vol. 7, pp. 5762–5771, Nov. 2021.

[43] R. U. Madhure, R. Raman, and S. K. Singh, "CNN-LSTM based electricity theft detector in advanced metering infrastructure," in *Proc. 11th Int. Conf. Comput., Commun. Netw. Technol. (ICCCNT)*, Jul. 2020, pp. 1–6.

[44] K. Aurangzeb, N. Ayub, and M. Alhussein, "Aspect based multi-labeling using SVM based ensembler," *IEEE Access*, vol. 9, pp. 26026–26040, 2021.

[45] S. Pande, N. K. Rathore, and A. Purohit, "A survey and analysis of extreme machine learning models and its techniques," Shri Govindram Seksaria Inst. Technol. Sci., Indore, India, Tech. Rep. rs-599856/v1, 2021.

[46] S. Guo, Q. Bai, and X. Zhou, "Foreign object detection of transmission lines based on faster R-CNN," in *Information Science and Applications*. Singapore: Springer, 2020, pp. 269–275.

[47] V. Hayyolalam and A. A. P. Kazem, "Black widow optimization algorithm: A novel meta-heuristic approach for solving engineering optimization problems," *Eng. Appl. Artif. Intell.* vol. 87, Jan. 2020, Art. no. 103249.

[48] P. Mukilan and W. Semunigus, "Human object detection: An enhanced black widow optimization algorithm with deep convolution neural network," *Neural Comput. Appl.*, vol. 33, pp. 1–12, Aug. 2021.

ABDULWAHAB ALI ALMAZROI received the M.Sc. degree in computer science from the University of Science, Malaysia, and the Ph.D. degree in computer science from Flinders University, Australia. He is currently an Assistant Professor with the Department of Information Technology, College of Computing and Information Technology at Khulais, University of Jeddah, Saudi Arabia. His research interests include parallel computing, cloud computing, wireless communication, and data mining.

NASIR AYUB received the M.S. degree in software engineering from COMSATS University Islamabad, Islamabad. He is currently pursuing the Ph.D. degree with the School of Electrical Engineering and Computer Science (SEECS), National University of Science and Technology (NUST), Islamabad. He worked as a Research Associate with the Communications Over Sensors (ComSens) Research Group, Department of Computer Science, COMSATS University Islamabad. Currently, he is working as a Lecturer with the Department of Computer Science, Federal Urdu University Islamabad, Pakistan. His research interests include energy management, natural language processing, and data science.