frontiers | Frontiers in Energy Research

TYPE Original Research
PUBLISHED 28 December 2022
DOI 10.3389/fenrg.2022.1043593

Check for updates

OPEN ACCESS

EDITED BY
Kashif Mehmood,
Southeast University, China

REVIEWED BY
Sheeraz Iqbal,
University of Azad Jammu and Kashmir,
Pakistan
Muhammad Asad,
Dalhousie University, Canada
Omar Abu Arqub,
Al-Balqa Applied University, Jordan

*CORRESPONDENCE
Naveed Ishtiaq Chaudhary,
chaudni@yuntech.edu.tw

SPECIALTY SECTION
This article was submitted to Smart Grids, a section of the journal Frontiers in Energy Research

RECEIVED 13 September 2022
ACCEPTED 17 October 2022
PUBLISHED 28 December 2022

CITATION
Munawar S, Khan ZA, Chaudhary NI, Javaid N, Raja MAZ, Milyani AH and Azhari AA (2022), Novel FDIs-based data manipulation and its detection in smart meters' electricity theft scenarios.
Front. Energy Res. 10:1043593.
doi: 10.3389/fenrg.2022.1043593

Novel FDIs-based data manipulation and its detection in smart meters' electricity theft scenarios

Shoaib Munawar[1], Zeshan Aslam Khan[1], Naveed Ishtiaq Chaudhary[2]*, Nadeem Javaid[3], Muhammad Asif Zahoor Raja[2], Ahmad H. Milyani[4] and Abdullah Ahmed Azhari[5]

[1]Department of Electrical and Computer Engineering, International Islamic University, Islamabad, Pakistan, [2]Future Technology Research Center, National Yunlin University of Science and Technology, Yunlin, Taiwan, [3]Department of Computer Science, COMSATS University Islamabad, Islamabad, Pakistan, [4]Department of Electrical and Computer Engineering, King Abdulaziz University, Jeddah, Saudi Arabia, [5]The Applied College, King Abdulaziz University, Jeddah, Saudi Arabia

Non-technical loss is a serious issue around the globe. Consumers manipulate their smart meter (SM) data to under-report their readings for financial benefit. Various manipulation techniques are used. This paper highlights novel false data injection (FDIs) techniques, which are used to manipulate the smart meter data. These techniques are introduced in comparison to six theft cases. Furthermore, various features are engineered to analyze the variance, complexity, and distribution of the manipulated data. The variance and complexity are created in data distribution when FDIs and theft cases are used to poison SM data, which is investigated through skewness and kurtosis analysis. Furthermore, to tackle the data imbalance issue, the proximity weighted synthetic oversampling (ProWsyn) technique is used. Moreover, a hybrid attentionLSTMInception is introduced, which is an integration of attention layers, LSTM, and inception blocks to tackle data dimensionality, misclassification, and high false positive rate issues. The proposed hybrid model outperforms the traditional theft detectors and achieves an accuracy of

Abbreviations: *ADASYN*, adaptive synthetic; *ADASYNENN*, adaptive synthetic edited nearest neighbor neural network; *AMI*, advanced metering infrastructure; *ANFIS*, adaptive neural fuzzy inference system; *ANN*, artificial neural network; *AUC*, area under the curve; *DSN*, deep siamese network; *DWMCNN*, day week month convolutional neural network; *ETD*, electricity theft detection; *FDI*, false data injection; *FIS*, fuzzy interface system; *FPR*, false positive rate; *FRESH*, feature extraction and scalable hypothesis; *GRU*, gated recurrent unit; *KNN*, K-nearest neighbor; *LLE*, locally linear embedded; *NAN*, neighborhood area network; *NCA*, neighborhood component analysis; *NTLs*, non technical losses; *PCA*, principal component analysis; *ProWsyn*, proximity weighted synthetic oversampling; *RE*, reconstruction error; *RESNet*, residual network; *SAGAN*, self attention generative adverserial neural network; *SCADA*, supervisory control and data acquisition; *SMs*, smart meters; *SMOTE*, synthetic minority oversampling technique; *TPR*, true positive rate; $1-$ *DCNN*, 1 dimensional convolutional neural network.

0.95%, precision 0.97%, recall 0.94%, F1 score 0.96%, and area under-the-curve (AUC) score 0.98%.

KEYWORDS

inception module, AttLSTM, false data injection, electricity theft detection, time series data

1 Introduction

Smart grids are serially interconnected networks having resilient features of unity power factor, self-healing, system monitoring, load balancing, and two-way communication. The communication channel is a delicate part of the power network. Network stability is interrupted when it is interfaced with a wrong information flow (Rawat and Bajracharya, 2015). Various types of cellular technologies, wireless sensor protocols, WLAN, and WAN are used for the purpose of communication in smart grids (Parikh et al., 2010), (Djennadi et al., 2021a). The world is moving towards the development of smart grids for efficient and reliable smart energy where different types of energy-production sources are integrated for optimal and reliable operations (Cai et al., 2017), (Djennadi et al., 2021b). The survival of societies is based on economic growth and an uninterrupted electrical energy supply. Losses are of two types: technical losses (TLs) and non technical losses (NTLs) (Jeyaraj et al., 2020), (Guo et al., 2018). An almost as large portion of the losses in the electrical system are NTLs. TLs are the inherent losses of the electrical power system, whereas NTLs occur due to the problems of double tapping, by-passing of smart meters (SMs), and tampering with SM readings, etc., in order to under-report the consumed electrical energy (Buzau et al., 2019), (Somefun et al., 2019). The deployment of smart grids can easily regulate customers' consumption behavior. Detection of NTLs secures the smart grids against anomalies and optimal flow of energy is managed (Rodriguez et al., 2017), (Arqub, 2018).

Advanced metering infrastructure (AMI) is an intelligent infrastructure to detect NTLs, however, it is a hardware-based architecture with multiple architectural flaws. False Data injections (FDIs) of novel nature are used to manipulate in SMs data, which are difficult to investigate and detect by AMI architecture. FDIs are novel techniques, which are used to manipulate the data of SM readings to gain illegal financial benefit. AMI collects the data with the help of a neighborhood area network (NAN). NAN is a useful architecture designed to manage energy in order to forecast short-term load and to investigate the optimal energy scheduling by the utility providers (UPs) (Depuru et al., 2011), (Arqub, 2020). Traditional grids use supervisory control and data acquisition (SCADA) in order to monitor grid operations and ensure security (Yasakethu and Jiang, 2013), (Sweis et al., 2022). Conventional machine learning techniques are used for the detection of NTLs, however, techniques such as support vector machine (SVM), random forest (RF), and 1D-convolutional neural network (1D-CNN) have low detection accuracy in classification scenarios (Glauner et al., 2016). Henceforth, classifiers with high detection and low false positive rate (FPR) are required to mitigate the problem of misclassification.

1.1 Motivation

Electricity theft is extant worldwide. Utility providers look for problems in their consumers' premises due to NTLs. Consumers opt for various electricity theft techniques in order to under-report their consumption. Some of these techniques are (Rawat and Bajracharya, 2015) tampering with the data with shunt devices (Parikh et al., 2010), double tapping of SMs, and (Djennadi et al., 2021a) electronic faults. These traditional approaches capture the behavior of NTLs where various hand-drafted mechanisms are developed due to a lack of clear mathematical formulations. Developing such solutions for each individual theft case is very expensive and time-consuming due to their relience on expert knowledge. In order to tackle such issues we propose a deep-learning based architecture that self-learns features of the observed data and automatically detects NTLs. Such architecture is operated in less time in order to mitigate the need for experts and excessive costs. Moreover, false data injection techniques (FDIs) are introduced in this paper, which can be used in real-time applications to manipulate SM readings. These manipulating techniques are highly intensive in nature and they can manipulate the data accordingly to the consumer's choice. So highlighting the detection of such intensive techniques improves the detection scenarios and minimizes the chances of theft. Consideration of such FDIs in detection scenarios minimizes NTLs, and manipulated patterns found with attributes of such theft traces can easily be identified as theft. Moreover, an efficient model should be used to detect and segregate fraudulent and benign consumers in such scenarios with minimal FPR. Minimal FPR is an effective parameter and minimizes excessive on-site costs for verification of fraudulent consumers.

2 Literature review

This section provides an overview of the existing literature related to electricity theft detection (ETD) in smart grids.

In (Takiddin et al., 2020), an ensemble detector is proposed, which is a combination of deep auto encoders with attention-gated recurrent units (GRU) and feed-forward neural networks. Similarly, (Kocaman and Tümen, 2020) proposes an LSTM classifier for the detection of malicious customers. Data selection, normalization, and weights updating mechanisms are used as preprocessing mechanisms in both of the proposed solutions. Architectures of the LSTM classifier contain LSTM cells, dropout layers, relu activation function, and softmax classifier. Precision, accuracy, and recall matrix is used to evaluate the performance of the proposed models.

Study in (Li et al., 2019) uses a convolutional neural network and random forest (CNN-RF) as a novel hybrid classifier for the detection of NTLs. CNN is used as a down-sampler to extract key features of the time series data. The featured data is inputted to RF for further classification in order to identify anomalous consumers. Similarly (Javaid et al., 2021a), uses adaptive synthesis (ADASYN) for the provision of balanced data. A hybrid module of CNN and LSTM is proposed to detect ill-intent within consumers' profiles. CNN is used to extract abstract features from weekly time series data, however, LSTM is trained on the inputted data of the CNN. Integration of CNN-LSTM is named deep siamese network (DSN), which segregates honest and thieving customers. In (Pereira and Saraiva, 2021), data augmentation techniques are analysed to evaluate the performance of various minority over-sampler techniques. A pool of data augmentation techniques enlisting cost-sensitive learning, random over-sampling, K-medoids over-sampling, cluster-based over-sampling, and synthetic minority over-sampling technique (SMOTE) are used to balance the imbalanced data. The balance data is inputted to CNN. The performance of CNN is evaluated on each of the data augmentation techniques, respectively. Furthermore, CNN is used as a binary classifier for data classification, and area under the curve (AUC) is used as a performance matrix to evaluate the classifier's performance. Literature in (Blazakis et al., 2020) uses an adaptive neural fuzzy inference system (ANFIS), which is a combination of artificial neural network (ANN) and fuzzy set theory in order to investigate NTLs. ANFIS utilizes back propagation learning of ANN and sugeno fuzzy inference system (FIS) to detect maliciousness in time series data of SMs. To maximize the efficiency of the classifier, neighborhood component analysis (NCA) is used to select the optimal ranking of the important features such as mean, medium, entropy, and load factor. Furthermore, accuracy, precision, F1 score, and AUC score are used to evaluate the performance of the classifier.

Similarly, in (Himeur et al., 2021) an ensemble model based on genetic optimization is developed to detect anomaly. SMOTE, a data over-sampling technique, is used to balance the data distribution. Afterward, features of the anomalous consumers are extracted using principal component analysis (PCA) along with the data dimensionality reduction. The abstract information of customers' behavior is extracted using AdaBoost technique and architectural optimization of the deep neural network is analysed through genetic algorithms. Moreover (Hussain et al., 2021), presents a novel supervised learning solution, which is an integration of catboost and SMOTETOMEK algorithms. Data preprocessing is tackled by K-nearest neighbor (KNN) in order to fill missing values, while data augmentation is carried out using SMOTETOMEK in order to mitigate biasness towards a majority class. Furthermore, to extract key features of highly dense time series data feature extraction and scalable hypothesis (FRESH) is used. The extracted data is inputted into catboost classifier for classification and a tree-SHAP algorithm is used as a decision-maker for theft identification. Study in (Cheng et al., 2021) proposes RF based classifier for the detection of an anomaly in a time series data. To reduce heavily dense time series data K-means method is used, whereas, a neural network of day, week, and month convolutional neural network (DWMCNN) is used to analyse the SMs' consumption data and to extract key features. To evaluate the performance AUC score is used as a performance metric. In order to segregate the honest and fraudulent consumers (Javaid et al., 2021b) proposes two supervised learning models. One of the models is an integration of self-attention generative adversarial network (SAGAN) and CNN. Important features of the time series data are extracted using the locally linear embedding technique (LLE) technique and to tackle the class imbalance issue adaptive synthetic edited nearest neighbor (ADASYNENN) is utilized. Furthermore, an ensemble model ERNET is developed, which consists of an efficient net residual network (ResNet) and gated recurrent unit (GRU). ResNet and GRU hybrid model is used as a second classifier to detect NTLs. Robust learning rate and data imbalance issues are tackled with root mean square propagation (RMSProP) and SMOTE edited the nearest neighbor, respectively.

Various proposed solutions have been presented in the literature, however, slow computations in RNN, the need for bulk training data in the case of CNN, performance declination in AFNIS due to the provision of less training data, and non-availability of intrinsic evaluation metric for SAGAN, we propose the AttentionLSTMInception model to overcome all these issues. Moreover, the Attention layer memorizes the large sequence of data. LSTM has more additional units which can hold information longer. An additional number of parameters such as learning rate, input and output biases, updating of weights, and backpropagation make the model more flexible. The inception module is added for better utilization of the computing resources in order to avoid excessive computational load. These are deeper networks, which are used for dimensionality reduction with stacked convolutions. Furthermore, the proposed hybrid model utilizes the attributes of long-term memorization of information and backpropagation of LSTM, data filtering for dimensionality reduction of CNN, and cognitive attention towards the prominent features of the attention layers, we are integrating them to introduce a novel hybrid model AttenLSTMInception for the detection of NTLs. The proposed hybrid model tackles issues of long-term memory dependencies, vanishing grading, under-fitting, over-fitting, and high FPR.

2.1 Paper organization

The rest of the paper is organized as follows. **Section 3** provides a list of contributions and their mapped solutions. **Section 4** determines the importance of feature engineering. **Section 5** and **Section 6** provide a detailed study of the system model and its workings, respectively. **Section 7** highlights performance evaluation and **section 8** is simulations results. Finally, a conclusion is drawn in **section 9**.

3 List of contributions

The contributions of the study are enlisted as follows.

- Diversity and dense variability in data distribution confuse the classification scenario and require special filtering mechanisms, which are tackled in this paper.
- Novel false data injection techniques (FDIs) are investigated, which manipulate the SMs data extremely and remained still undetectable in literature.
- A problem of high FPR due to extensive misclassification is tackled, which causes financial overburdens.
- To tackle data reductionality issues, inception, attention, and filtering mechanisms are introduced to hybridize the existing classifying architectures.
- In order to retain long-term memorization, the inputted data is overlapped through segmented attributes of sliding windows to adopt cognitive learning of the data.
- Data synthesizing through ineffective balancing techniques mimic resembled, overlapped, and replicated data, which is tackled by introducing a novel proximity-weighted synthetic oversampling (ProWsyn) technique.

3.1 Dataset

SMs installed on consumer premises record the electricity consumption for the consumed energy. Consumed energy is recorded in the form of time series data. In this paper, a realistic dataset, named as state grid corporation of China (SGCC) is used which contains 42,372 consumers. We are considering 6 months of data from 1500 benign consumers only for data classification and manipulation due to the limited resources of our machine (Punmiya and Choe, 2019). Our machine specifications are intel(R) core (TM) M-5y10c, CPU@ 0.80 GHz 1.00GHz, RAM 4 GB. Moreover, the simulator is google CoLab. The dataset contains a few missing readings, which are due to the mal-operation and malfunctioning of the sensors deployed over the installed SMs. Such erroneous readings create ambiguity over the classification scenario and ultimately result in a low detection rate. A straightforward approach to eliminating such readings disrupts the time series data's sequence and integrity. Considering optimal data filling techniques and operating such techniques over the perspective rows provide refined and complete consumption data of each consumer. A 24-h time series data for every consumer is recorded by an SM. A unique consumer ID is assigned to each consumer. A label is indexed for the identification of honest and fraudulent consumption. A binary representation of 0 and 1 is used where 0 represents benign class data and 1 represents fraudulent class data. Due to the rarity of theft class data, we are proposing false data injection techniques (FDIs) to manipulate the benign class data in order to synthesize fraudulent class data. FDIs are proposed in comparison to theft cases (Sha et al., 2022), which are shown in **Eqs. 1**, **2**. Moreover, the dataset is online available at: https://github.com/henryRDlab/ElectricityTheftDetection.

3.2 Data preprocessing

Electricity consumption time series data is a series of numeric values, which is monitored by the installed SMs on the consumers' premises. Such time series data contain missing values and outliers due to the mal-operation and malfunctioning of the deployed SMs. Filling in the missing values and removing the outliers are necessary steps. A simple Imputer technique is used to fill in the missing values and to remove the outliers. To fill in the missing values, a mean-based strategy is operated row-wise. Furthermore, data normalization is carried out to normalize the data into a specific range. The normalized data is the input data, which is then transformed and scaled to carry out further operations.

3.3 Data augmentation

The problem of skewness towards the majority class by the classifier is a serious issue, which needs proper attention. To tackle the data imbalance issue, synthetic data is synthesized by oversampling minority class data. Weight value-based approaches transform the data into equal distribution, however, most of the techniques synthesize inappropriate data, which ultimately results in a poor distribution of the classes. To overcome such problems, this paper proposes a proximity-weighted synthetic oversampling technique (ProWsyn) (Islam and Belhaouari, 2022). ProWsyn targets the minority class samples to balance the data. The proximity information of

TABLE 1 Mapping of limitations and proposed solutions.

Sr	Limitation Identified	Solution Number	Proposed Solution	Validations
L1	Misclassification due to the dense variability of the distributed data	S1	Addition of Inception module for filtering abstract features	V1: Table 2
L2	Lack of theft class data samples	S2	Synthesizing through novel FDIs	V2: Eq.1
L3	High FPR	S3	Hybrid model architecture to tackle extensive misclassification	V3: Figure 6
L4	Problem of short term information memorization	S4	Data segmentation and overlapping	V4: Figure 1
L5	Imbalance data and model's skewness towards the majority class	S5	ProWsyn data resampling technique	V5: Algorithm 2

each sample is measured based on the distance from the decision boundary. Distance-based proximity helps to generate the effective weights for the minority class samples. Such effective weights of the minority samples normalize the data distribution, which mitigates the skewness of the model towards the majority class samples. The data is balanced and synthetic samples are generated. ProWsyn is a clustering-based technique, which operates in two steps.

- In the first step, the distance between the residing position of the sample and the decision boundary is monitored for each of the minority samples. All the samples are partitioned (P) upon the splitting.
- In the second step, the partitioned data samples are assigned a proximity level (L).

The proximity level is directly proportional to the distance. A smaller proximity level gives more important samples, whereas, a greater proximity level gives less important samples. Algorithm 2 shows the operating mechanism of the ProWsyn technique.

In step 1 of Algorithm 2, input parameters are defined. Step 2 considers new sampling based on EU. New samples are synthesized and considered if EU of the corresponding sample is less than the corresponding cluster and weight of the sample is updated accordingly. However, if the EU is greater it is ignored. Finally, in step 3, the number of honest consumers and fraudulent consumers is balanced.

$$\begin{cases} FDI_1 = \dfrac{mean\,(E)\,{*}random\,(0.1-0.9)}{E} \\ whereE\ >\ 1\ \leq\ mean \\ FDI_2 = \sqrt{(mean(E))}\ {*}\ random\,(0.1-0.9) \\ FDI_3 = \sqrt{(E)}\ {*}\ random\,(0.1-0.9) \\ FDI_4 = mean\,(E) - (\gamma) \\ where\ \gamma\ is\ a\ constant\ consumption\ and \\ \gamma\ \leq\ mean \\ FDI_5 = E - \gamma_i \\ where\ i =\ 0, \ldots, E_{max} \\ FDI_6 = E\,(t-d) = 0\ if\ t\ <\ d\ and\ 1\ if\ t\ \geq\ d \\ where\ t, d\ is\ time\ and\ difference,\ respectively. \end{cases} \tag{1}$$

```
Step 1: Defining fraudulent and honest consumers:
Input: Honest Consumers $H_{E_c}$, Fradulent Consumers $F_{E_c}$ Sample $S_i$, Euclidean distance EU, Decision Boundary $DB$, Weight W

Step 2: Introducing FDIs:
$F_{E_c} > H_{E_c}$;
$S_i$ if EU is geater ignore $S_i$;
update W;
consider $S_i$ if EU is less;
skip: and go to next sample
Step 3: Balancing :
$F_{E_c} = H_{E_c}$
STOP
Output:Target (Proximity $S_i$ having EU >), Skip (Proximity $S_i$ having EU <)
```

Algorithm 1. Data Augmentation using proWsyn Technique.

$$\begin{cases} T1\,(E_t) = E_t\ {*}\ random\,(0.1, 0.9) \\ T2\,(E_t) = E_t\ {*}\ E_t\,(E_t = random\,(0.1, 0.9)) \\ T3\,(E_t) = E_t\ {*}\ random\,[0, 1] \\ T4\,(E_t) = mean\,(E_t)\ {*}\ random\,(0.1, 1.0) \\ T5\,(E_t) = mean\,(E_t) \end{cases} \tag{2}$$

4 Feature engineering

The data distribution analysis is presented in Table 2. Effective classification is based on the data's nature. Complex data is very difficult to be learned and classified by weak models. Such complexity is based on the variance among the data samples that need special attention before deploying any model to tackle the classification problem. Various types of features are engineered, which include min, max, standard

TABLE 2 Data distribution analysis.

Data Manipulation Scheme	Kurtosis	Skewness
FDIs	6	46
Theft Cases	1	10

deviation, mean, root mean square error, skew, kurtosis, quantile, and rolling mean. Mean, min, max and standard deviation are basically the stochastical features, whereas, root mean square error, skew, kurtosis, quantile, and rolling mean are the static features based on the dynamics of the time series data. Stochastical features show the randomness and variations in the data, which helps to know the complexity of the distributed data. Whereas, the root mean square responds to the provision of the actual information of variations and distribution in the data. Skewness factor (S_k) judges the symmetry and resemblance of the data. In literature, it is represented as three-point plotting. One point is a central point and the other two lies on the left and the right of the central point, respectively. A symmetric distribution is the same to the left and right of the central point. Mathematically it can be represented as:

$$\left\{ S_k = \frac{\sum_{j=1}^{M} (Wj - W^{-})^3 / (M)}{q^3}, \right. \tag{3}$$

The kurtosis parameter helps to investigate the problems associated with the outliers and the data's distribution. It shows the difference between each and every point within the data whether it is symmetric or un-symmetric. Mathematically, it can be represented as:

$$\left\{ Kurtosis = \frac{\sum_{j=1}^{M} (Wj - W^{-})^4 / (M)}{q^4} - 3, \right. \tag{4}$$

Where W^{-} is the mean, q is the standard deviation and M is the number of data samples. Positive kurtosis represents a heavy-tailed distribution, whereas, negative kurtosis is a light tailed distribution. Normal data distribution has a zero kurtosis. Quantile concludes the shape of the distribution. It distributes the observations in the same number of samples based on the probability distribution.Rolling mean (R_m) is a computing window, which computes the mean on a piece of the data slab. The rolling window rolls on continuous time series data and computes for a subset. The computed subset is the rolling average for that specific slab of the data. It basically accesses the stability within the data distribution. Mathematically, it is represented as (Blanca et al., 2013):

$$\left\{ R_m = \frac{E_t + E_{t-1} + E_{t-2} + \cdots + E_{t-n} + 1}{M}, \right. \tag{5}$$

4.1 Data manipulation

Novel FDI techniques are proposed in comparison to six theft cases for data manipulation (Pamir et al., 2022a).

- FDI 1 under-reports the consumption by manipulating the SM's data as shown in **Figure 1A**. The total consumption is aggregated into a mean. A random number is multiplied by the aggregated mean, which ranges between (0.1–0.9). The product is divided by a number greater than 1 and less than a number equal to the aggregated mean, which vanishes the consumed energy reading and limits it to a zero reading.
- FDI2 targets the mean and a random number's product, which is squarely rooted in order to inject false reading by manipulating SM's consumption data. This data subjectively minimizes the consumption of energy almost by 1/2 of the total consumed energy as shown in **Figure 1B**.
- FDI3 is the periodic bulk manipulation of the total consumed energy over monthly and weekly based. It is a specific defined time period manipulation. The square-rooted consumption is multiplied by a random number ranging between (0.1–0.9) in order to get more financial benefits as shown in **Figure 2A**.
- FDI4 is a two-phase manipulation. One is mean-based manipulation and the second one is a constant numeric number subtraction-based manipulation. The mutual difference between both strategies the SM's consumption data under-reports the original consumption as shown in **Figure 2B**.
- FDI5 is the manipulation of the SM's data during off-peak and off-peak hours. A γ factor is a difference-based manipulation variable, which is represented by a simple numeric number. The variable is subtracted from the recorded readings to under-report the consumed energy as shown in **Figure 3A**.
- FDI6 is a unit-step function-based manipulation at the consumer's end. It manipulates the consumption with a choice to operate it at any time stamp or periodically. It can steal 100% of the consumed energy in the extreme. However, in the case of equilibrium, a 50% of the theft is expected. During such modes of manipulation, the consumption is limited to 0 or 1 where 1 shows the original consumption and 0 shows the manipulated consumption as shown in **Figure 3B**.

4.2 Model's architecture

The input data is segmented into various data subsets in form of slabs through a dynamic sliding window. The dynamic

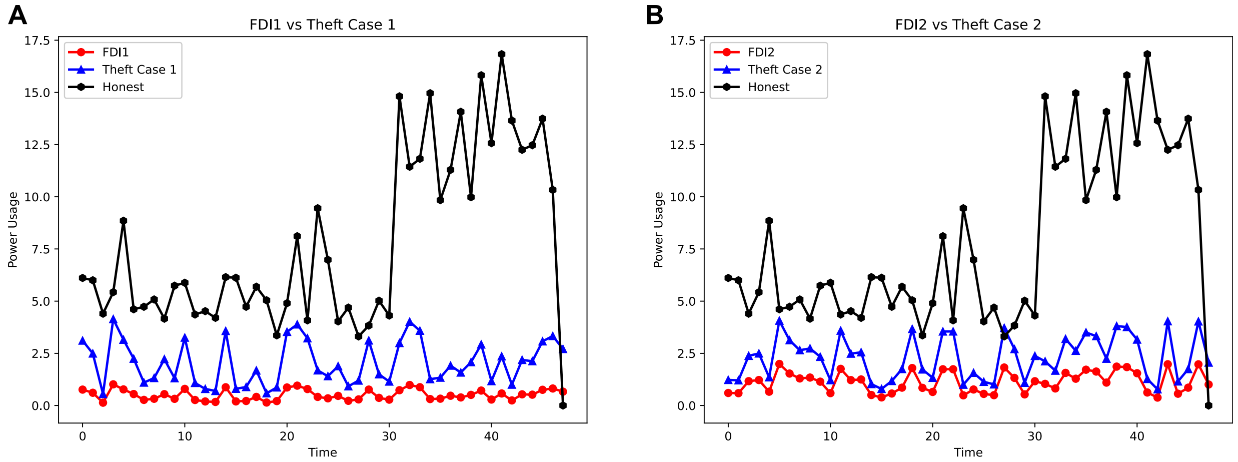

FIGURE 1
(A) Theft Case 1 vs. FDI 1. **(B)** Theft Case 2 vs. FDI 2.

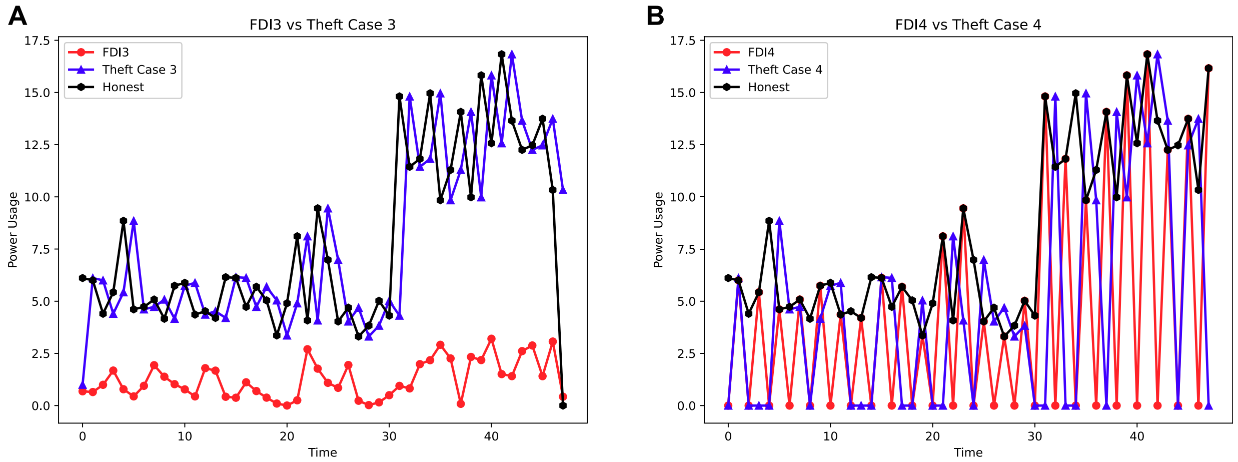

FIGURE 2
(A) Theft Case 3 vs. FDI 3. **(B)** Theft Case 4 vs. FDI 4.

sliding widow overlaps the input data by 50%. Data's subsets contain resizing strategy over k = 20 where 10 previous and 10 next records are buffered. Every next sliding slab selects the data starting from the data point residing on the 10th index of the previous slab. The data is resized in similar fashion until the very end of the array is reached. The same phenomenon is repeated consecutively for the oncoming next slab. The 50% overlapping of the data is a linear traversal of the data, which minimizes the complexity of the dense time series data and finds an optimized data resizing strategy for the input data. The developed hybrid model is a delicately structured architecture, which is a multivariate model and inspired by the long-term short-term memory and fully convolutional network (LSTM-FCN). In order to retain recurrent information of the time series data the modules are integrated in parallel where the LSTM module is connected to an inception time network with additional layers of attention (Abbasimehr and Paki, 2022). Novel FDI techniques are proposed in comparison to six theft cases for data manipulation (Dua et al., 2022). AttenLSTMInception model is a multivariate resolution feature of the time series data. The ultimate goal is to capture and analyse the variance between the classes' data. In order to retain the information LSTM-Inception model contains two residual blocks. Information propagation between the residual blocks is initiated by an ultimate short linear connection where inputs are added to the next block. Such schematics mitigate the vanishing gradient problem due to the direct flow of the gradient. Stacking the inception modules, the first inception component is named

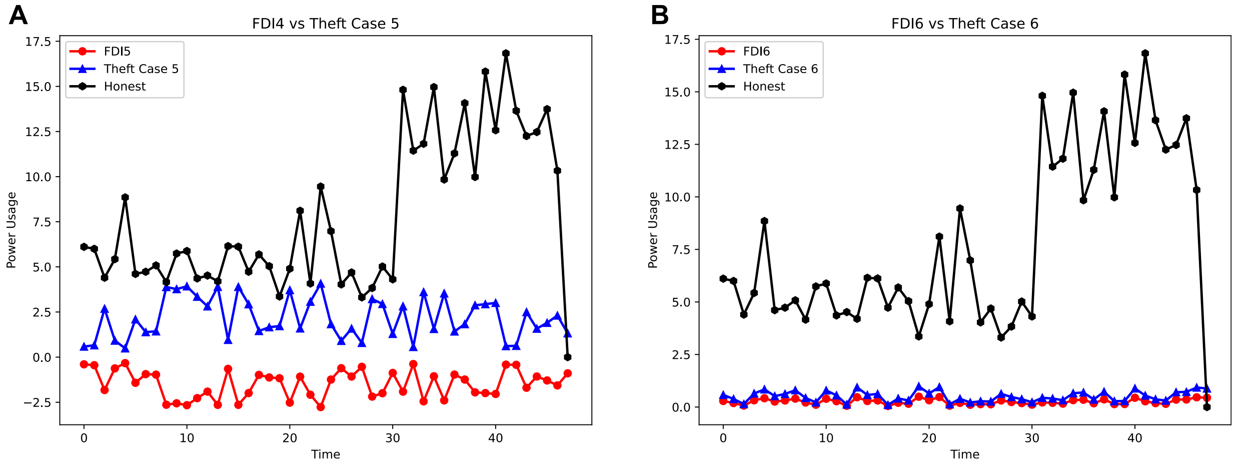

FIGURE 3
(A) Theft Case 5 vs. FDI 5. **(B)** Theft case 6 vs. FDI 6.

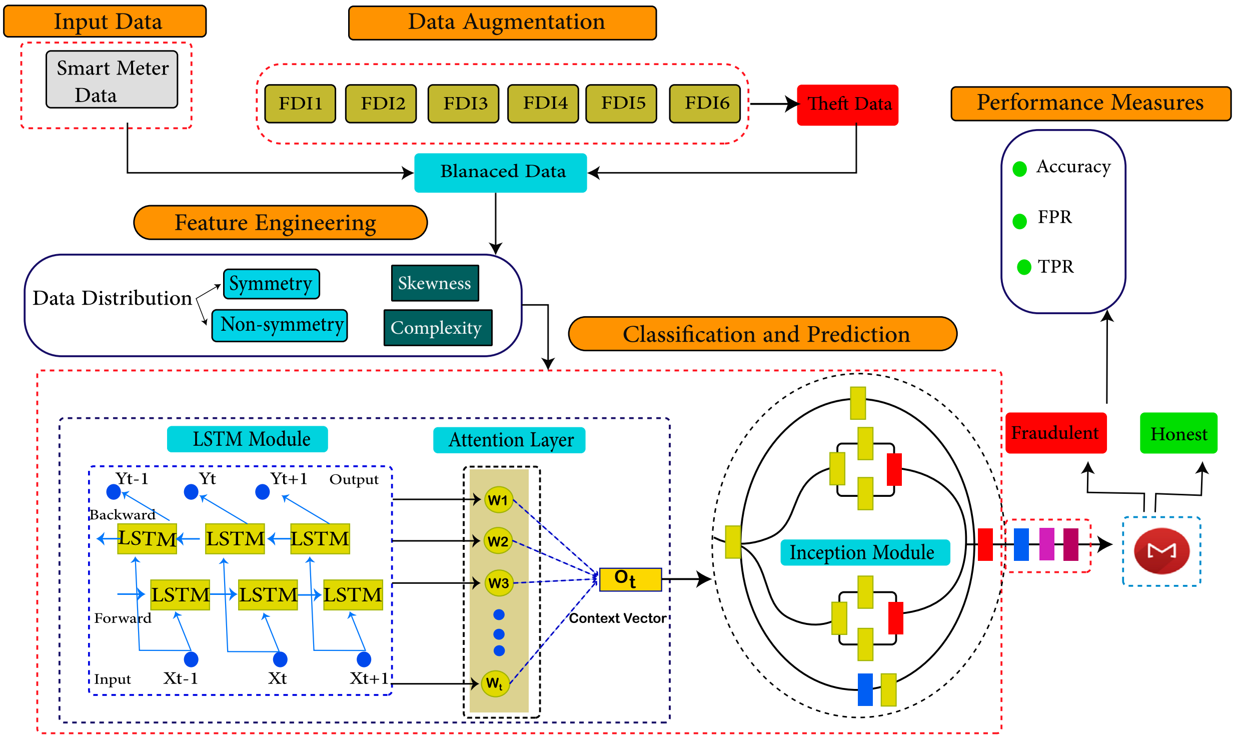

FIGURE 4
The proposed system model.

the bottleneck layer, which performs sliding operation over the data. Such layers reduce the data's dimensionality due to the sliding operation of the filters. Integrating networks in such scenarios mitigate the over-fitting issue, model's complexity, and complex dimensionality. It is necessary to mention that the bottleneck technique maximizes filter length in terms of pulling, which helps in reducing the computational complexity. The max-pooling generates sequential attributed data, which is concatenated with the inception modules' output. The hierarchical latent features are extracted *via* stacking and backpropagation mechanisms. The global pooled output of the inception module and AttenLSTM block are concatenated, which

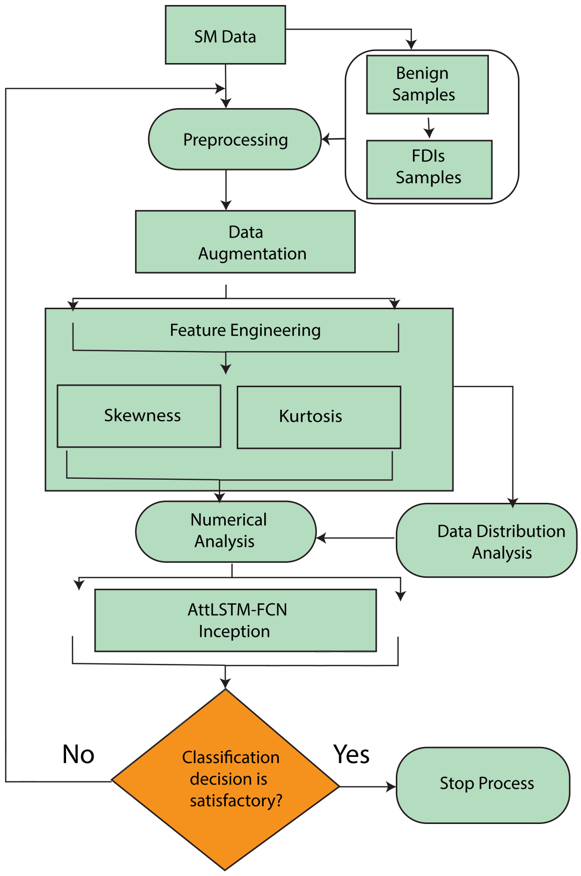

FIGURE 5
Working of flowchart.

```
Step 1: Defining fraudulent and honest consumers:
Input: Honest Consumers $H_{E_c}$, Fradulent Consumers $F_{E_c}$
Step 2: Introducing FDIs:
$FDI1=\frac{mean(E)*random(0.1-0.9)}{E}$;
$FDI2=\sqrt{(}mean(E_c))*random(0.1-0.9))$;
$FDI3=\sqrt{(}E_c)*random(0.1-0.9)$;
$FDI4=mean(E)-(\gamma)$;
$FDI5=E-\gamma_i$;
$FDI6=E(t-d)=0$ if $t<d$ and $1$ if $t>=c$;
Step 3:  Data Augmentation and Concatenation:
$Concat(\ FFDI1 + FDI2 + FDI3 + FDI4$
  $+ FDI5 + FDI6\ )$
$F_{E_c}=FDI_i+\ldots+FDI_n$: where $i=1,\ldots,6$.
$E_{C_T}=H_{E_c} + F_{E_c}$
Step 4:  Data Equilibrium:
$H_{E_c} = F_{E_c}$;
$F_{E_c} > H_{E_c}$; apply proWsyn to $H_{E_c}$.
$H_{E_c} = F_{E_c}$.
Step 5:  Feature Engineering:
$E_{C_T}=H_{E_c} + F_{E_c}$
$skewness(mean(E_{C_T}))$
$kurtosis(mean(E_{C_T}))$
Step 6:  Classification
Output:$E_c\ \varepsilon\ F_{X_c}$;
$E_c\ \varepsilon\ H_{X_c}$
```

Algorithm 2. AttenLSTMInception based Electricity Theft Detection Scheme.

is connected to the inception layer and classification operator function.

5 Proposed system model

The limitations with the proposed solutions are presented in Table 1. While the system model in Figure 4 represents our proposed solution for the aforementioned limitations. It is divided into five sections (1) Data preprocessing (2) Data manipulation (3) Data augmentation (4) Feature engineering and (5) Classification. The data distribution analysis is presented in Table 2.

- Initially in section (1), data is preprocessed where the missing values and outliers are filled and removed by the simple Imputer technique, respectively. A row-wise operation is carried out on the data to tackle such issues.
- In data manipulation section (2), consumers are defined based on their provided SM's readings, which are labeled with a binary representation of 0 and 1, where 0 stands for honest consumers and 1 for fraudulent consumers. Honest consumer data is manipulated in order to synthesize fraudulent consumers' data by applying FDI techniques. Data is synthesized due to the rare availability of the theft class data. Synthesized data by such FDI techniques show fraudulent consumers' data. The defined FDI techniques result in six variants for each benign sample.
- In section (3), data balancing is required in order to mitigate the Model's biasness and skewness towards a majority class. Dense skewness poisons the model's classification, which tends to increase the false positive rate (FPR). A data augmentation technique is required to mitigate such issues (Ullah et al., 2021), (Asif et al., 2021a), (Asif et al., 2021b), (Kabir et al., 2021). ProWsyn based data augmentation strategy is applied in the proposed work to balance fraudulent and benign class samples.
- In section (4), the balanced data is observed by the feature engineering module where the data's nature and distribution

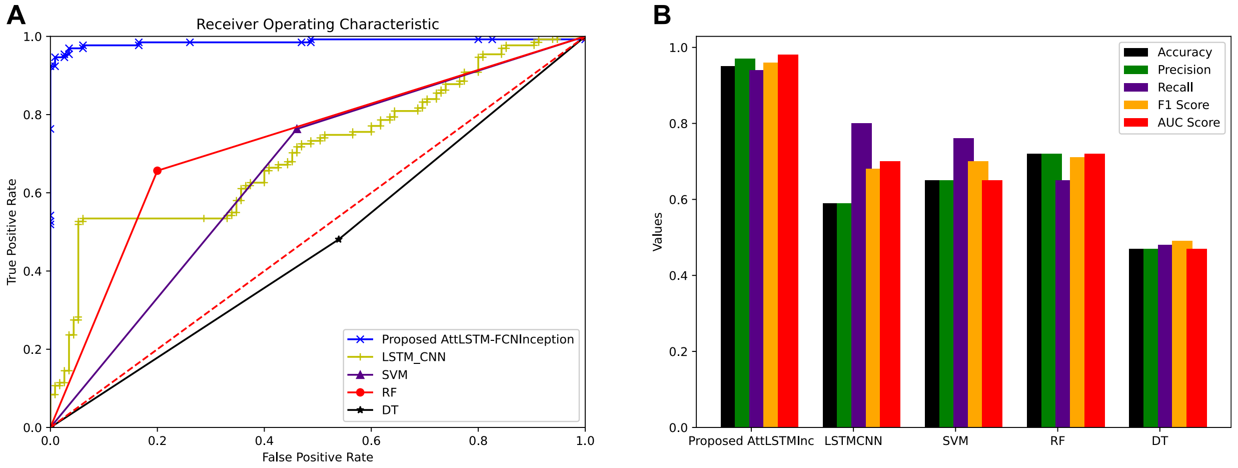

FIGURE 6
(A) Performance analysis of the benchmark and proposed model **(B)** Performance comparison.

are studied. Stochastical features, which contain mean, min, max, and standard deviation are generated to study the data's distribution. In addition, the skewness factors, kurtosis, quantile, root mean square and rolling features are engineered, which shows the distribution symmetry and its deviation. Such investigating factors result in deciding the model's complexity and deepness for the classification scenario. Highly skewed, defused and un-symmetric data needs a heavily featured classifying model for effective class segregation and classification.

- In section (5), to classify the samples effectively a hybrid model AttLSTM-FCNInception model is adopted, which is an integration of attention layers (Pustokhin et al., 2020), LSTM module (Yu et al., 2019) and Inception (Yang et al., 2019). Two of the Inception modules and attention layers are integrated into LSTM. The model is fed with the affine preprocessed data, which is suitable to tackle complex and un-symmetric data. **Algorithm 1** defines a summary of the whole system model.

6 Working of the system model

The working of the whole classification scenario is defined in **Figure 5**.

- Initially in step 1, the SMs' time series data is analyzed and benign samples are considered only due to non-availability of the theft class samples.
- In step 2, the benign class data is manipulated by six FDIs, and six new variants are synthesized for a single benign sample. Such variants for a single benign sample disrupt the data balancing, which requires balancing techniques to balance the data.
- In step 3, a ProWsyn minority class oversampling technique is opted to balance the data. Each sample is considered on a proximity basis where EU distance is measured by assigning weights to the samples. The nearest sample of the cluster to the decision boundary is weighted greater, whereas, the sample with a large EU distance from the perspective cluster is weighted less. The assigned weights help to mitigate the issues of misclassification and high FPR.
- In step 4, various features are engineered in order to investigate the complexity and distribution of the data. Two major mean-based synthesized features are targeted to investigate the complexity and distribution of the data. Kurtosis and skewness are the mean-based engineered features, which visualize the data's symmetry and far-tailed numeric outliers.
- In step 5, in order to enhance the data memorization, a sliding window segments the data with a 50% overlap, which carries the previous and next step information segments of the input data. Such translation of the available information flows back and forth, which increases the memorization capability of the model.
- In step 6, the segmented data is fed to a hybrid AttenLSTMInception model for classification. The fed data is classified and fraudulent consumers are detected with a low FPR, effectively.

7 Performance evaluation

ETD is a binary classification problem where benign and fraudulent classes are represented as positive and negative, respectively. In a binary classification scenario, the positive class

TABLE 3 Performance comparison of the proposed and existing models.

Classifier	Accuracy	Precision	Recall	F1 Score	AUC Score
Proposed AttLSTMInception	0.95	0.97	0.94	0.96	0.98
LSTMCNN	0.59	0.59	0.80	0.68	0.70
SVM	0.65	0.65	0.76	0.70	0.65
RF	0.72	0.72	0.65	0.71	0.72
DT	0.47	0.47	0.48	0.49	0.47

is labeled as 0 and the negative is labeled as 1. Precision, detection rate (DR), accuracy, AUC, and F1 score are used to evaluate the performance of the model. AUC is the area under the curve with two distinguishing parameters, TPR and FPR. TPR is the detection sensitivity of the model and FPR is the specificity. A comparative investigation between the accurate identification of true positive samples and true negative samples constructs AUC. Four parametric attributes are collectively mapped to measure the sensitivity and specificity of the model. Sensitivity is DR and specificity is the FPR of the model. Mathematically, it can be represented as (Jones and Athanasiou., 2005):

$$DR = \frac{TP}{TP + FN}, \tag{6}$$

$$FPR = \frac{TN}{TN + FN}, \tag{7}$$

Where TP, TN, FP, and FN represent true positive, true negative, false positive, and false negative, respectively. TP, TN, FP, and FN are the confusion matrix attributes, which investigate binary classification.

8 Simulation results

In order to compare the proposed AttLSTMInception model with the existing models DT, RF, SVM, and LSTMCNN, a comparative analysis is shown in Figure 6A, Figure 6B. Accuracy, precision, recall, F1 score, and AUC are the performance parameters, which are considered to investigate the performance of the models. The results in Table 3 show that the proposed model outperforms the rest of the models. The effective performance of the proposed model is due to the attention and inception modules. The attention module mimics cognitive attention, which focuses on the prominent and important features rather than non-useful data. The inception module adds the properties of efficient computations and dimensionality reduction by using multiple data filtering sizes. The addition of the inception module tackles the problem of over-fitting and computational complexity. RF (Nguyen and Phan, 2021), SVM (Lin et al., 2021), DT, and LSTMCNN (Hasan et al., 2019) perform very badly. They cannot perform on complex time series data and cause overfitting issues. Furthermore, the performance of the proposed model is enhanced by using dropout regularization and adam optimization. Figure 6 shows the AUC of various models against the proposed model. The proposed model outperforms the rest of the models. Initially, the proposed model classifies the time series data of the honest and fraudulent consumers with zero FPR, however, at an AUC score of 0.92 a minimal FPR is reported. The slight change in reporting FPR is due to the increased data complexity. LSTMCNN performs efficiently with a slight FPR, however, it reduces its performance over the increased complexity of the data. Figures 6A,B shows that the low FPR is achieved by the proposed model as compared to other models, which means that the fraudulent and honest consumers are accurately classified. Similarly, the AUC score of the conventional machine learning techniques SVM (Pamir et al., 2022b), RF, and DT (Munawar et al., 2021) is very bad and reports high FPR. Figure 6 shows the accuracy, precision, recall, F1 scores, and AUC scores of the models. It can be seen that the proposed model outperforms the rest of the models in each of the performance parameters.

9 Conclusion and future work

In this paper, novel FDIs techniques are proposed in comparison to theft cases. The proposed FDIs manipulate the data severely as compared to the theft cases. The variations and complexity in data distribution caused by the proposed FDIs and theft cases are investigated through data distribution techniques. The analysis shows that the proposed FDIs are severe in nature while manipulating data of SMs' as compared to theft cases. FDIs observe minimal skewness and complexity in data distribution as compared to the theft cases data. Furthermore, six variants are synthesized for each of the honest consumers. A novel data balancing technique, ProWsyn is used to balance the data. Moreover, the attLSTMInception model is proposed, which is an integration of LSTM, attention layers, and inception modules. The proposed model outperforms the rest of the existing models and achieves an accuracy of 0.95%, precision 0.97%, recall 0.94%, F1 score 0.96%, and AUC score 0.98%. In future work, we will investigate the extraction of abstract features for dimensionality reduction and the addition of more memory modules for long-term dependencies of the data in our proposed model to reduce FPR furthermore.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Materials, further inquiries can be directed to the corresponding author.

Author contributions

Conceptualization, SM and NJ; methodology, SM; software, SM; validation, NJ; writing—original draft preparation, SM; writing—review and editing, NJ, ZK, NC, and MR; supervision, ZK and NC; project administration, MR, AM, and AA; funding acquisition, AM and AA. All authors have read and agreed to the published version of the manuscript.

Funding

The Deanship of Scientific Research (DSR) at King Abdulaziz University (KAU), Jeddah, Saudi Arabia has funded this project, under grant no. (KEP-MSc: 122-135-1443).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The handling editor, KM declared a past co-authorship with the author(s), NC and AM.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Abbasimehr, H., and Paki, R. (2022). Improving time series forecasting using LSTM and attention models. *J. Ambient. Intell. Humaniz. Comput.* 13 (1), 673–691. doi:10.1007/s12652-020-02761-x

Arqub, O. A. (2020). Numerical simulation of time-fractional partial differential equations arising in fluid flows via reproducing Kernel method. *Int. J. Numer. Methods Heat. Fluid Flow.* 30, 4711–4733. doi:10.1108/hff-10-2017-0394

Arqub, O. A. (2018). Numerical solutions for the Robin time-fractional partial differential equations of heat and fluid flows based on the reproducing kernel algorithm. *Int. J. Numer. Methods Heat. Fluid Flow.* 28, 828–856. doi:10.1108/hff-07-2016-0278

Asif, M., Kabir, B., Ullah, A., Munawar, S., and Javaid, N. (2021). "Towards energy efficient smart grids: Data augmentation through BiWGAN, feature extraction and classification using hybrid 2DCNN and BiLSTM," in *International conference on innovative mobile and internet services in ubiquitous computing* (Cham: Springer), 108–119.

Asif, M., Ullah, A., Munawar, S., Kabir, B., Khan, A., and Javaid, N. (2021). "Alexnet-AdaBoost-ABC based hybrid neural network for electricity theft detection in smart grids," in *Conference on complex, intelligent, and software intensive systems* (Cham: Springer), 249–258.

Blanca, M. J., Arnau, J., Lopez-Montiel, D., and Bendayan, R. (2013). Skewness and kurtosis in real data samples. *Methodol. (Gott).* 9 (2), 78–84. doi:10.1027/1614-2241/a000057

Blazakis, K. V., Kapetanakis, T. N., and Stavrakakis, G. S. (2020). Effective electricity theft detection in power distribution grids using an adaptive neuro fuzzy inference system. *Energies* 13 (12), 3110. doi:10.3390/en13123110

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 Trans. Power Syst.* 35 (2), 1254–1263. doi:10.1109/tpwrs.2019.2943115

Cai, Y., Li, Y., Cao, Y., Li, W., and Zeng, X. (2017). Modeling and impact analysis of interdependent characteristics on cascading failures in smart grids. *Int. J. Electr. Power & Energy Syst.* 89, 106–114. doi:10.1016/j.ijepes.2017.01.010

Cheng, G., Zhang, Z., Li, Q., Li, Y., and Jin, W. (2021). Energy theft detection in an edge data center using deep learning. *Math. Problems Eng.* 2021, 1–12. doi:10.1155/2021/9938475

Depuru, S., Reddy, S. S., Wang, L., and Devabhaktuni, V. (2011). Electricity theft: Overview, issues, prevention and a smart meter based approach to control theft. *Energy policy* 39 (2), 1007–1015. doi:10.1016/j.enpol.2010.11.037

Djennadi, S., Shawagfeh, N., and Arqub, O. A. (2021). A fractional Tikhonov regularization method for an inverse backward and source problems in the time-space fractional diffusion equations. *Chaos, Solit. Fractals* 150, 111127. doi:10.1016/j.chaos.2021.111127

Djennadi, S., Shawagfeh, N., Osman, M. S., Gómez-Aguilar, J. F., Arqub, O. A., and Abu Arqub, O. (2021). The Tikhonov regularization method for the inverse source problem of time fractional heat equation in the view of ABC-fractional technique. *Phys. Scr.* 96 (9), 094006. doi:10.1088/1402-4896/ac0867

Dua, N., Singh, S. N., Vijay, B. S., Kumar Challa, S., and Challa, S. K. (2022). Inception inspired CNN-GRU hybrid network for human activity recognition. *Multimed. Tools Appl.*, 1–35. doi:10.1007/s11042-021-11885-x

Glauner, P., Augusto Meira, J., Valtchev, P., Radu, S., and Franck, B. (2016). *The challenge of non-technical loss detection using artificial intelligence: A survey*. arXiv preprint arXiv:1606.00626.

Guo, Y., Yang, Z., Feng, S., and Hu, J. (2018). "Complex power system status monitoring and evaluation using big data platform and machine learning algorithms: A review and a case study," in *Complexity 2018.*

Hasan, M., Toma, R. N., Nahid, A-A., Islam, M. M., and Kim, J-M. (2019). Electricity theft detection in smart grid systems: A CNN-LSTM based approach. *Energies* 1217, 3310. doi:10.3390/en12173310

Himeur, Y., Ghanem, K., Abdullah, A., Bensaali, F., and Amira, A. (2021). Artificial intelligence based anomaly detection of energy consumption in buildings: A review, current trends and new perspectives. *Appl. Energy* 287 (2021), 116601. doi:10.1016/j.apenergy.2021.116601

Hussain, S., Mustafa, M. W., Jumani, T. A., Khan Baloch, S., Alotaibi, H., Khan, I., et al. (2021). A novel feature engineered-CatBoost-based supervised machine learning framework for electricity theft detection. *Energy Rep.* 7 (2021), 4425–4436. doi:10.1016/j.egyr.2021.07.008

Islam, A., and Belhaouari, S. B. (2022). *Atiq ur rahman, and halima bensmail. "K nearest neighbor OveRsampling approach: An open source python package for data augmentation*. Software Impacts, 100272.

Javaid, N., Gul, H., Baig, S., Shehzad, F., Xia, C., Guan, L., et al. (2021). Using GANCNN and ERNET for detection of non technical losses to secure smart grids. *IEEE Access* 9 (2021), 98679–98700. doi:10.1109/access.2021.3092645

Javaid, N., Jan, N., and Umar Javed, M. (2021). An adaptive synthesis to handle imbalanced big data with deep siamese network for electricity theft detection in smart grids. *J. Parallel Distributed Comput.* 153, 44–52. doi:10.1016/j.jpdc.2021.03.002

Jeyaraj, P. R., Edward, P. R. S. N., Kathiresan, A. C., and Siva, P. A. (2020). Smart grid security enhancement by detection and classification of non-technical losses employing deep learning algorithm. *Int. Trans. Electr. Energ. Syst.* 30 (9), e12521. doi:10.1002/2050-7038.12521

Jones, C. M., and Athanasiou., T. (2005). Summary receiver operating characteristic curve analysis techniques in the evaluation of diagnostic tests. *Ann. Thorac. Surg.* 79 (1), 16–20. doi:10.1016/j.athoracsur.2004.09.040

Kabir, B., Ullah, A., Munawar, S., Asif, M., and Javaid, N. (2021). “Detection of non-technical losses using MLP-GRU based neural network to secure smart grids,” in *Conference on complex, intelligent, and software intensive systems* (Cham: Springer), 383–394.

Kocaman, B. T., and Tümen, V. (2020). Detection of electricity theft using data processing and LSTM method in distribution systems. *Sādhanā* 451, 286–10. doi:10.1007/s12046-020-01512-0, no.

Li, S., Han, Y., Yao, X., Song, Y., Wang, J., and Zhao, Q. (2019). Electricity theft detection in power grids with deep learning and random forests. *J. Electr. Comput. Eng.* 2019, 1–12. doi:10.1155/2019/4136874

Lin, G., Feng, X., Guo, W., Cui, X., Liu, S., Jin, W., et al. (2021). Electricity theft detection based on stacked autoencoder and the undersampling and resampling based random forest algorithm. *IEEE Access* 9 (2021), 124044–124058. doi:10.1109/access.2021.3110510

Munawar, S., Asif, M., Kabir, B., Ullah, A., and Javaid, N. (2021). “Electricity theft detection in smart meters using a hybrid Bi-directional GRU Bi-directional LSTM model,” in *Conference on complex, intelligent, and software intensive systems* (Cham: Springer), 297–308.

Nguyen, T. H. T., and Phan, Q. B. (2021). “Electricity theft detection in power grid with a hybrid convolutional neural network-support vector machine model,” in *The 5th international conference on future networks & distributed systems*, 24–30.

Pamir, N. J., Javaid, S., Asif, M., Umar Javed, M., Adamu, S. Y., Aslam, S., et al. (2022). Synthetic theft attacks and long short term memory-based preprocessing for electricity theft detection using gated recurrent unit. *Energies* 15 (8), 2778. doi:10.3390/en15082778

Pamir, N. J., Qasim, U., Adamu, S. Y., Alkhammash, E. H., and Hadjouni, M. (2022). *Non-technical losses detection using autoencoder and bidirectional gated recurrent unit to secure smart grids*. IEEE Access.

Parikh, P. P., Kanabar, M. G., and Sidhu, T. S. (2010). “Opportunities and challenges of wireless communication technologies for smart grid applications,” in *IEEE PES general meeting* (IEEE), 1–7.

Pereira, J., and Saraiva, F. (2021). Convolutional neural network applied to detect electricity theft: A comparative study on unbalanced data handling techniques. *Int. J. Electr. Power & Energy Syst.* 131, 107085. doi:10.1016/j.ijepes.2021.107085

Punmiya, R., and Choe, S. (2019). Energy theft detection using gradient boosting theft detector with feature engineering-based preprocessing. *IEEE Trans. Smart Grid* 10 (2), 2326–2329. doi:10.1109/tsg.2019.2892595

Pustokhin, D. A., Pustokhina, I. V., Dinh, P. N., Van Phan, S., Gia, N. N., Joshi, G. P., et al. (2020). An effective deep residual network based class attention layer with bidirectional LSTM for diagnosis and classification of COVID-19. *J. Appl. Statistics*, 1–18. doi:10.1080/02664763.2020.1849057

Rawat, D. B., and Bajracharya, C. (2015). “Cyber security for smart grid systems: Status, challenges and perspectives,” in *SoutheastCon 2015*, 1–6.

Rodriguez, V., Esther, J. D. S., Oregi, I., Bilbao, M. N., and Gil-Lopez, S. (2017). Detection of non-technical losses in smart meter data based on load curve profiling and time series analysis. *Energy* 137, 118–128. doi:10.1016/j.energy.2017.07.008

Sha, Y., Faber, J., Gou, S., Liu, B., Li, W., Schramm, S., et al. (2022). An acoustic signal cavitation detection framework based on XGBoost with adaptive selection feature engineering. *Measurement* 192 (2022), 110897. doi:10.1016/j.measurement.2022.110897

Somefun, T. E., Awosope, C. O. A., and Chiagoro, A. (2019). Smart prepaid energy metering system to detect energy theft with facility for real time monitoring. *Int. J. Electr. Comput. Eng.* 9 (5), 4184. doi:10.11591/ijece.v9i5.pp4184-4191

Sweis, H., Shawagfeh, N., and Arqub, O. A. (2022). Fractional crossover delay differential equations of Mittag-Leffler kernel: Existence, uniqueness, and numerical solutions using the Galerkin algorithm based on shifted Legendre polynomials. *Results Phys.* 41, 105891. doi:10.1016/j.rinp.2022.105891

Takiddin, A., Ismail, M., Zafar, U., and Serpedin, E. (2020). Robust electricity theft detection against data poisoning attacks in smart grids. *IEEE Trans. Smart Grid* 12 (3), 2675–2684. doi:10.1109/tsg.2020.3047864

Ullah, A., Munawar, S., Asif, M., Kabir, B., and Javaid, N. (2021). “Synthetic theft attacks implementation for data balancing and a gated recurrent unit based electricity theft detection in smart grids,” in *Conference on complex, intelligent, and software intensive systems* (Cham: Springer), 395–405.

Yang, S., Lin, G., Jiang, Q., and Lin, W. (2019). A dilated inception network for visual saliency prediction. *IEEE Trans. Multimed.* 22 (8), 2163–2176. doi:10.1109/tmm.2019.2947352

Yasakethu, S. L. P., and Jiang, J. (2013). “Intrusion detection via machine learning for SCADA system protection,” in *1st international symposium for ICS & SCADA cyber security research 2013 (ICS-csr 2013)*, 1, 101–105.

Yu, Y., Si, X., Hu, C., and Zhang, J. (2019). A review of recurrent neural networks: LSTM cells and network architectures. *Neural Comput.* 31 (7), 1235–1270. doi:10.1162/neco_a_01199