DEEP NEURAL NETWORK TECHNOLOGY IS BEING USED IN SMART GRIDS TO DETECT INSTANCES OF ELECTRICITY THEFT

[1]VENNAPUSA SREEEVANI,[2]H.MADHUSUDHANA RAO
[1]Student,[2]Assistant professor, MCA,M.Phill,(Ph.D)
Department of CSE

ABSTRACT:

Theft of power is a worldwide issue that has a detrimental impact not just on utility providers but also on individuals who consume electricity. It creates potentially dangerous situations involving electricity and contributes to the already high cost of energy for consumers. Additionally, it destabilises the economic growth of utility corporations. The development of smart grids plays a significant part in the detection of power theft because smart grids create huge amounts of data, which includes data on customer use. This data, when combined with methods such as machine learning and deep learning, may be used to identify electricity theft. This article presents the technique for detecting theft that makes use of extensive information in both the time and frequency domains in a classification approach that is based on a deep neural network. The methods of data interpolation and the development of synthetic data allow us to solve flaws in the dataset, such as the presence of issues with class imbalance and missing data. Experiments are done in the combined and reduced feature space using principal component analysis, and lastly, a minimal redundancy maximum relevance strategy is included in order to validate the most essential features. We begin by analysing and comparing the contribution of features from the time domain and then go on to the frequency domain. We increase the performance of the power theft detection system by optimising the hyperparameters using a Bayesian optimizer. Additionally, we make use of an adaptive moment estimation optimizer in order to conduct out tests with varying values of critical parameters in order to discover the ideal settings that result in the highest level of accuracy. In the last part of this section, we demonstrate the competitiveness of our technique by evaluating it with several alternative methods using the same dataset. During the validation process, we were able to get an area under the curve (AUC) of 97%, which is 1% better than the best AUC achieved by previous research, and an accuracy of 91.8%, which is the second-highest score on the benchmark.

Deep neural network, power theft, machine learning, minimal redundancy maximum relevance, principal component analysis, and smart grids are some of the terms that are included in this index.

I.INTRODUCTION:

ELECTRICITY theft is a problem that affects utility companies worldwide. More than $96 billion is lost by utility companies worldwide due to Non-Technical Losses (NTLs) every year, of which electricity theft is the major contributor [1]. In sub-Saharan Africa, 50% of generated energy is stolen, as reported by World Bank [2]. The ultimate goal of electricity thieves is to consume energy without being billed by utility companies [3], or pay the bills amounting to less than the consumed amount [4]. As a result, utility companies suffer a huge revenue loss due to electricity theft. [5] reports that in 2015, India lost $16.2 billion, Brazil lost $10.5 billion and Russia lost $5.1 billion. It is estimated that approximately $1.31 billion (R20 billion) revenue loss incurred by South Africa (through Eskom) per year is due to electricity theft [2]. Apart from revenue loss, electricity theft has a direct negative impact on the stability and reliability of power grids [3]. It can lead to surging electricity, electrical systems overload and public safety threats such as electric shocks [4]. It also has a direct impact on energy tariff increases, which affect all customers [3]. Implementation of smart grids comes with many opportunities to solve the electricity theft problem [4]. Smart grids are usually composed of traditional power grids, smart meters and sensors, computing facilities to monitor and control grids, etc., all connected through the communication network [6]. Smart meters and sensors collect data such as electricity usage, grid status, electricity price, Many Utilities sought to curb electricity theft in traditional grids by examining meters' installation and configurations, checking whether the power line is bypassed, etc. [4]. These methods are expensive, inefficient and cannot detect cyber attacks [4], [7]. Recently, researchers have worked towards detecting electricity theft by utilizing machine learning classification techniques using readily available smart meters data. These theft detection methods have proved to be of relatively lower costs [8]. However, existing classification techniques consider time-domain features and do not regard frequency-domain features, thereby limiting their performance. Regardless of the fact that there is active ongoing research on electricity theft detection, electricity theft is still a problem. The major cause of delay in solving this problem may be that smart grids deployment is realized in developed nations while developing nations are lagging behind [9]. The challenges of deploying smart grids include the lack of communication infrastructure and users' privacy concerns over data reported by the smart meters [10]. However, [10] reports that smart meters are being considered by many developed and developing countries with aims that include solving NTLs. [11] predicted smart grids global market to triple in size between 2017 and 2023, with the following key regions leading smart grids deployment: North America, Europe and Asia. In this paper, we present an effective electricity theft detection method based on carefully extracted and selected features in Deep Neural Network (DNN)-based classification approach. We show that employing frequency-domain features as opposed to using time-domain features alone enhances classification performance. We use a realistic electricity consumption dataset released by State Grid Corporation of China (SGCC) accessible at [12]. The dataset consists of electricity consumption data taken from January 2014 to October 2016.

The main contributions are as follows:

- Based on the literature, we propose a novel DNN classification-based electricity theft detection method using comprehensive time-domain features. We further propose using frequency-domain features to enhance performance.

- We employ Principal Component Analysis (PCA) to perform classification with reduced feature space and compare the results with classification done with all input features to interpret the results and simplify the future training process.

- We further use the Minimum Redundancy Maximum Relevance (mRMR) scheme to identify the most significant features and validate the importance of frequency-domain features over time-domain features for detecting electricity theft.

- We optimize the hyperparameters of the model for overall improved performance using a Bayesian optimizer. We further employ an adaptive moment estimation (Adam) optimizer to determine the best ranges of values of the other key parameters that can be used to achieve good results with optimal model training speed.

- Lastly, we show 1% improvement in AUC and competitive accuracy of our model in comparison to other data-driven electricity theft detection methods in the literature evaluated on the same dataset. The remainder of this paper is organized as follows. Section II covers the related work done in literature to tackle the electricity theft problem. In Section III, we briefly introduce techniques used in this paper. Section IV covers step by step method taken in this work; which includes dataset analysis and work done to improve its quality and customers' load profile analysis which lead to features extraction and classification. In Section V, we show and discuss the results. We finally conclude the paper in Section VI.

II.LITERATURE SURVEY:

Research on electricity theft detection in smart grids has attracted many researchers to devise methods that mitigate against electricity theft. Methods used in the literature can be broadly categorized into the following three categories: hardware-based, combined hardware and data-based detection methods and data-driven methods. Hardware-based methods [13]–[19] generally require hardware devices such as specialized microcontrollers, sensors and circuits to be installed on power distribution lines. These methods are generally designed to detect electricity theft done by physically tampering with distribution components such as distribution lines and electricity meters. They can not detect cyber attacks. Electricity cyber attack is a form of electricity theft whereby energy consumption data is modified by hacking the electricity meters [7]. For instance, in [13], an electricity meter was re-designed. It used components that include: a Global System for Mobile Communications (GSM) module, a microcontroller, and an Electrically Erasable Programmable Read-Only Memory (EEPROM). A simulation was done and the meter was able to send a Short Message Service (SMS) whenever an illegal load was connected by bypassing the meter. Limited to detecting electricity theft done by physically tampering with distribution components such as distribution lines and electricity meters. Authors in [16] used the GSM module, ARM-cortex M3 processor and other hardware components to solve the electricity theft problem done in the following four ways: bypassing the phase line, bypassing the meter, disconnecting the neutral line, and tampering with the meter to make unauthorized modifications. A prototype was built to test all four possibilities. The GSM module was able to notify with SMS for each theft case. Authors in [17] designed ADE7953 chip-based smart meter which is sensitive to current and voltage tempering, and mechanical tempering. ADE7953 was used to detect overvoltage, dropping voltage, overcurrent, the absence of load and other irregularities in voltage and current. It sent an interrupt signal to the Microcontroller Unit (MCU) which reported tampering status. Mechanical tampering was overcome by connecting a tampering switch to MCU's IO ports so that it can send alarm signals to MCU once tampered with. The design was tested with tampering cases such as connecting the neutral and the phase lines, connecting the meter input and output in reverse mode, and bypassing the phase line to load. The probability of detection failure was 2.13%. Authors in [15] used a step down transformer, voltage divider circuit, microchip and other hardware components to design a circuitry to detect electricity theft by comparing forward current on the main phase line with reverse current on the neutral line. The circuitry was installed before the meter.The design was tested on Proteus software and on actual hardware. When the meter was bypassed, the problem was detected and an alarm sounded. In [14], a circuit to detect electricity theft done by bypassing the meter was designed. The transformers, rectifiers, microcontroller, GSM module and other hardware components were used. The GSM controller notified the operator with SMS when the meter was bypassed. Authors in [18] proposed putting the Radio Frequency Identification (RFID) tags on ammeters and capturing unique data about each ammeter. Ammeters were to be tracked and managed real-time. Electricity theft was to be inspected onsite. Damaged, removed or a tag with a different information from the original one means high possibility that an electricity theft happened. Evaluation based on analysis on cost of deployment. With a case study made on utility company in China, Return on Investment (ROI) was found to be >1. In [19], An Arduino-based real-time electricity theft detector was designed. The following hardware was used: Arduino Uno, GSM module, current sensors and LCD. The Arduino Uno obtained measurements from current sensors which were located one on the secondary side of the transformer and the other on the electric service cap. If the difference between current sensors' measurements exceeded a set threshold, the message would be sent to the operator via a GSM module. The simulation was done using Proteus 8 software and the prototype was built on hardware, which was able to report theft cases when tested. Apart from their inability to detect cyber attacks, these methods are also expensive due to their need for special hardware deployment and maintenance. Combined hardware and data-based electricity theft detection methods [20]–[22] employ the use of hardware, machine learning and/or deep learning techniques to tackle the electricity theft problem. Due to hardware requirements, these methods also pose the challenge of being expensive to deploy and maintain. In [20], a method to measure the total consumption of a neighbourhood and compare the results with the usage reported by the smart meters in that neighbourhood was proposed. A significant difference between smart meters' and transformers' measurements would mean the presence of unfaithful customers in the neighbourhood. To locate the unfaithful customers in the neighbourhood, the authors proposed using a Support Vector Machine (SVM) classifier. The classifier was tested on a dataset of 5000 (all faithful) customers. A maximum detection rate of 94% and a minimum false positive rate of 11% were achieved. Authors in [22] developed a predictive model to calculate TLs. To get NTL, TLs would be subtracted from total distribution network losses. Based on an assumption that distribution transformers and smart meters share data to the utility after every 30 minutes, a smart meter simulator was used to generate data for 30 users in 30 minutes intervals for 6 days. On the simulator, unfaithful users stole electricity by bypassing the meter. Stolen electricity was varied between 1% and 10% of the total consumption. For stolen electricity value over 4%, the detection rate was 100%, which diminished as stolen electricity percentage was decreased. In [21], a method which would use an observer meter that would be installed on a pole away from households and record the total amount of electricity supplied to n households where it is suspected that one or more meters have been tampered with was proposed. The observer meter would have camera surveillance to protect it from being tampered with. A mathematical algorithm that utilizes data from an observer meter and smart meters to detect a smart meter tempered with was developed. A mathematical algorithm was tested with a real-world consumption dataset by increasing the consumption of some meters which were picked randomly. The algorithm was able to detect the meters with altered consumption. Due to high-cost demand in the above categories, many researchers work on data-driven methods to overcome the electricity theft problem. For instance, the authors in [3] designed an electricity theft detection system by employing three algorithms in the pipeline: Synthetic Minority Over-sampling Technique (SMOTE), Kernel function and Principal Component Analysis (KPCA) and SVM. They used SMOTE to generate synthetic data for balancing an unbalanced dataset, KPCA to extract features and SVM for classification. They obtained maximum overall classifier quality characterized by Area Under the Curve (AUC) of 89% on validation. Authors in [4] used wide and deep Convolutional Neural Networks (CNN) model to detect electricity theft. Based on that normal electricity consumption is periodical while stolen electricity consumption data is not periodical, wide was to learn multiple co-occurrences of features for 1-D series data, while deep CNN was used to capture periodicity with data aligned in a 2-D manner by weeks. They varied training and validation data ratios, to obtain maximum AUC value of 79%. By utilizing the same dataset used in [3] and [4], the method we present in this paper achieves AUC results beyond 90% on both validation and testing. In [23], PCA was used to transform original highdimensional consumption data by extracting Principal Components (PCs) which retained the desired variance. An anomaly score parameter that was defined between set minimum and maximum thresholds was introduced. For each test sample, the anomaly score parameter was calculated. If the result was not between the set thresholds, the sample would then be treated as malicious. The true positive rate (TPR) was used to evaluate the method, which hit the best-recorded value of 90.9%. Authors in [24] used One-Class SVM (O-SVM), Cost-Sensitive SVM (CS-SVM), Optimum Path Forest (OPF) and C4.5 tree. From customer consumption data, different features were selected, and the performance of each classifier was analyzed independently on a different set of features, followed by combining all classifiers for the best results. Best results were achieved when all classifiers were combined, with 86.2% accuracy. Authors in [25] employed a combination of CNN and Long Short-Term Memory (LSTM) recurrent neural network deep learning techniques. Seven hidden layers were used, of which four of them were used by CNN and three were utilized by LSTM. This method relied on CNN's automatic feature extraction ability on a given dataset. Features were extracted from 1-D time-series data. On model validation, the maximum accuracy achieved was 89%. The authors in [26] used a combination of Local Outlier Factor (LOF) and k-means clustering to detect electricity theft. They used k-means clustering to analyze the load profiles of customers, and LOF to calculate the anomaly degrees of customers whose load profiles were from their cluster centres. On the evaluation of the method, they attained an AUC of 81.5%. Our model achieves a maximum value of 91.8% accuracy and 97% on validation. In [27], two electricity theft models were developed. The first model is based on Light Gradient Boosting (LGB) classifier. A combination of SMOTE and Edited Nearest Neighbour (ENN) was used to balance the dataset. Feature extraction was done using AlexNet, followed by classification with LGB. This proposed model was named as SMOTEENN-AlexNet-LGB (SALM). The second model is based on the Adaptive Boosting classifier. Conditional Wasserstein Generative Adversarial Network with gradient penalty (CWGAN-GP) was used to generate synthetic data that resembled the minority class data to balance data of the unbalanced classes. Feature extraction was performed using GoogleNet, then classification by AdaBoost followed. The proposed model was named as GAN-NETBoost. The models were evaluated with SGCC data used in this work. SALM and GAN-NetBoost attained an accuracy of 90% and 95%, and AUC of 90.6% and 96% respectively on validation. Although these models were able to achieve impressive results, their consideration of time-domain features alone limited their performance. Our solution shows that adding frequency-domain features on time-domain features improves classification performance.

Artificial Neural Networks (ANNs) are a class of machine learning techniques that have been built to imitate biological human brain mechanisms [28], [29]. They are typically used for extracting patterns or detecting trends that are difficult to be detected by other machine learning techniques [30]. They consist of multiple layers of nodes/neurons which are connected to subsequent layers [29]. A neuron is the basic element of a neural network, which originates from the McCulloch-Pitts neuron, a simplified model of a human brain's neuron [31]. Figure 1 shows a model diagram of a neuron that comprises a layer following the input to the ANN.

III.CONCLUSION:

In this work, the detection of electricity theft in smart grids was investigated using time-domain and frequency-domain features in a DNN-based classification approach. Isolated classification tasks based on the time-domain, frequencydomain and combined domains features were investigated on the same DNN network. Widely accepted performance metrics such as recall, precision, F1-score, accuracy, AUCROC and MCC were used to measure the performance of the model. We observed that classification done with frequency-domain features outperforms classification done with time-domain features, which in turn is outperformed by classification done with features from both domains. The classifier was able to achieve 87.3% accuracy and 93% AUC-ROC when tested. We used PCA for feature reduction. With 7 out of 20 components used, the classifier was able to achieve 85.8% accuracy and 92% AUC-ROC when tested. We further analyzed individual features' contribution to the classification task and confirmed with the mRMR algorithm the importance of frequency-domain features over time-domain features towards a successful classification task. For better performance, a Bayesian optimizer was also used to optimize hyperparameters, which realized accuracy improvement close to 1%, on validation. Adam optimizer was incorporated and optimal values of key parameters were investigated. In comparison with other data-driven methods evaluated on the same dataset, we obtained 97% AUC which is 1% higher than the best AUC in existing works, and 91.8% accuracy, which is the second-best on the benchmark. The method used here utilizes consumption data patterns. Apart from its application in power distribution networks, it can be used in anomaly detection applications in any field. Our work brings a small contribution towards accurately detecting energy theft as we detect theft that only took place over time. We wish to extend our method to detect real-time electricity theft in the future. Since this method was evaluated based on consumption patterns of SGCC customers, it can further be validated against datasets from different areas to ensure its applicability anywhere.

REFERENCES

[1] S. Foster. (Nov. 2, 2021). Non-Technical Losses: A $96 Billion Global Opportunity for Electrical Utilities. [Online]. Available: https://energycentral.com/c/pip/ non-technical-losses-96-billion-globalopportunity-electrical-utilities [2] Q. Louw and P. Bokoro, ''An alternative technique for the detection and mitigation of electricity theft in South Africa,'' SAIEE Afr. Res. J., vol. 110, no. 4, pp. 209–216, Dec. 2019. [3] M. Anwar, N. Javaid, A. Khalid, M. Imran, and M. Shoaib, ''Electricity theft detection using pipeline in machine learning,'' in Proc. Int. Wireless Commun. Mobile Comput. (IWCMC), Jun. 2020, pp. 2138–2142. [4] 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. [5] P. Pickering. (Nov. 1, 2021). E-Meters Offer Multiple Ways to Combat Electricity Theft and Tampering. [Online]. Available: https://www.electronicdesign.com/technologies/meters [6] X. Fang, S. Misra, G. Xue, and D. Yang, ''Smart grid—The new and improved power grid: A survey,'' IEEE Commun. Surveys

Tuts., vol. 14, no. 4, pp. 944–980, 4th Quart., 2012. [7] M. Ismail, M. Shahin, M. F. Shaaban, E. Serpedin, and K. Qaraqe, ‘‘Efficient detection of electricity theft cyber attacks in AMI networks,’’ in Proc. IEEE Wireless Commun. Netw. Conf. (WCNC), Apr. 2018, pp. 1–6. [8] A. Maamar and K. Benahmed, ‘‘Machine learning techniques for energy theft detection in AMI,’’ in Proc. Int. Conf. Softw. Eng. Inf. Manage. (ICSIM), 2018, pp. 57–62. [9] A. Jindal, A. Schaeffer-Filho, A. K. Marnerides, P. Smith, A. Mauthe, and L. Granville, ‘‘Tackling energy theft in smart grids through data-driven analysis,’’ in Proc. Int. Conf. Comput., Netw. Commun. (ICNC), Feb. 2020, pp. 410–414. [10] I. Diahovchenko, M. Kolcun, Z. Čonka, V. Savkiv, and R. Mykhailyshyn, ‘‘Progress and challenges in smart grids: Distributed generation, smart metering, energy storage and smart loads,’’ Iranian J. Sci. Technol., Trans. Electr. Eng., vol. 44, no. 4, pp. 1319–1333, Dec. 2020.