



Article Title: A Deep Neural Network Based on Bayesian Optimizer for Electricity Theft Detection

A Deep Neural Network Based on Bayesian Optimizer for Electricity Theft Detection

Ashna S Kumar¹, Sujitha R²

¹Student, Department of Master of Computer Application, Narayanaguru College of Engineering, Manjalumoodu, Anna University

²Asst. Professor, Department of Master of Computer Application, Narayanaguru College of Engineering, Manjalumoodu, Anna University

ABSTRACT

Electricity theft is a global problem that negatively affects both utility companies and electricity users. It destabilizes the economic development of utility companies, causes electric hazards and impacts the high cost of energy for users. The development of smart grids plays an important role in electricity theft detection since they generate massive data that includes customer consumption data which, through machine learning and deep learning techniques, can be utilized to detect electricity theft. This paper introduces the theft detection method which uses comprehensive features in time and frequency domains in a deep neural network-based classification approach. It addresses dataset weaknesses such as missing data and class imbalance problems through data interpolation and synthetic data generation processes. It analyse and compare the contribution of features from both time and frequency domains, run experiments in combined and reduced feature space using principal component analysis and finally incorporate minimum redundancy maximum relevance(mrml) scheme for validating the most important features. It improves the electricity theft detection performance by optimizing hyper parameters using a Bayesian optimizer and they employ an adaptive moment estimation optimizer to carry out experiments using different values of key parameters to determine the optimal settings that achieve the best accuracy. On validation, had obtained 97% area under the curve (AUC), which is 1% higher than the best AUC in existing works, and 91.8% accuracy, which is the second-best on the benchmark.

Keywords: Electricity theft, deep learning, machine learning, smart grids, frequency domains, time domains, Bayesian Optimizer, Minimum redundancy Maximum relevance

1 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. In sub-Saharan Africa, 50% of generated energy is stolen, as reported by World Bank.

The ultimate goal of electricity thieves is to consume energy without being billed by utility companies, or pay the bills amounting to less than the consumed amount. As a result, utility companies suffer a huge revenue loss due to electricity theft. Reports that in 2015, India lost \$16.2 billion, Brazil lost \$10.5 billion and Russia lost \$5.1 billion.



Article Title: A Deep Neural Network Based on Bayesian Optimizer for Electricity Theft Detection

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.

Apart from revenue loss, electricity theft has a direct negative impact on the stability and reliability of power grids. It can lead to surging electricity, electrical systems overload and public safety threats such as electric shocks. It also has a direct impact on energy tariff increases, which affect all customers. Implementation of smart grids comes with many opportunities to solve the electricity theft problem. 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, etc. [6].

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. In this paper, it presents 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. It further propose using frequency-domain features to enhance performance. It employs 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. It further uses 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. It optimizes the hyper parameters of the model for overall improved performance using a Bayesian optimizer. It further employs 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.



Article Title: A Deep Neural Network Based on Bayesian Optimizer for Electricity Theft Detection

2 Recent Works

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 cannot 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. It 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.

[1] An Alternative technique for the detection and mitigation of electricity theft in South Africa, Quentin Louw, 2019. Electricity theft and illegal connection by ground surface conductors is a pervasive problem in South Africa. The impact this phenomenon has is not only limited to revenue loss and equipment damage, but also presents a life threatening hazard. Although the issues of non-technical losses have been researched for decades, no universal solution has been presented, due to the complexity of the problem. This paper sets out to propose an alternative strategy using ZSC to detect and mitigate illegal connections specifically because of bare conductors lying on the ground surface scenario. For decades the issue of NTL and electricity theft has been the focus of continued research. Most of the focus has been on dealing with the point of supply at the customer end, rather than also considering the consequential safety effect the illicit acts present. With the advent and deployment of smart metering technology systems, which primarily focus on revenue improvement, energy balancing can be applied to detect NTL due to perceived electricity theft. This philosophy however has limitations, as it depends on accurate network data (data mining) and demand load management patterns, and furthermore cannot detect and isolate the primary supply node, should these illegal connections occur. The only way to deal with these illegal connections, using this approach, is to dispatch engineering teams to investigate and remove the illegal



Article Title: A Deep Neural Network Based on Bayesian Optimizer for Electricity Theft Detection

connections which are problematic as the acts of removal initiate volatile environments for the teams to operate under, and this therefore further escalates the safety risk. The alternative proposed ZSC philosophy does not consider load demand patterns in the decision-making algorithm and can therefore be applied as a detection and mitigation strategy. The presence of ZSC due to ground surface conductors can be measured in the star-point of the primary supply node. With the proposed IED (Intelligent Electronic Device) development installed, dispatch alarms can isolate the affected node automatically, so mitigating the safety aspect and NTL losses. Simulation and experimental results show the validity of this technique as well as its dependence on seasonal change of the soil resistivity.

[2] Electricity Theft Detection using Pipeline in Machine Learning, Mubbashra Anwar, 2020. Electricity theft is the primary cause of electrical power loss that significantly affects the revenue loss and the quality of electrical power. Nevertheless, the existing methods for the detection of this criminal behaviour of theft are diversified and complicated since the imbalanced nature of the dataset, and high dimensionality of time-series data make it challenging to extract meaningful information. In this work, a pipeline is proposed to detect electricity theft in SG. The proposed pipeline is made up of SMOTE, KPCA and SVM. The imbalanced class issue is resolved using SMOTE, KPCA is used for feature extraction and SVM for the classification of electricity theft. It is the most efficient and simplest technique that is able to classify the fraudulent and non-fraudulent consumers accurately. SMOTE algorithm uses K-nearest neighbour (K-NN) approach to generate synthetic data. Minority instances are used to generate the data. To introduce the synthetic data, SMOTE takes nearest neighbours from feature vectors, computes the distance between them, multiplies the difference by number (0,1), and adds to feature space. SMOTE is used in the proposed pipeline to balance the dataset. PCA is a statistical technique. PCA is used for data analysis. The rate of loss of information during dimensions' reduction is minimum in PCA. It identifies useful patterns from the dataset and retains spatial attributes of data. High dimensional electricity consumption data is reconstructed by reducing the dimensions and extracting underlying consumption trends. Data transformed by the PCA has maximum variance. SVM is flexible and a powerful supervised machine learning model. It is used for both classification and regression problem. SVM works by creating one or more hyperplanes that separate the data clusters. Hyperplane is a decision plane which divides objects of different classes. The uses of kernel in SVM make it equivalent to feed forward neural network. The kernel trick converts a low dimensional space into high dimensional space that makes it flexible, powerful and accurate. Besides, various performance metrics are used for the evaluation of binary classification problems, such as: ROC curve, precision, recall, F1-score, MCC, and MAP are used to evaluate the performance of the proposed model. The model has achieved the precision of 0.85, recall of 0.88, and AUC-ROC of 0.89. Furthermore, the comparison with other benchmarks, such as: logistic regression (LR), decision tree (DT), RF, CNN, and LSTM has shown that the proposed model is superior in the prediction rate. Real-time power consumption data is used to train and test the proposed



Article Title: A Deep Neural Network Based on Bayesian Optimizer for Electricity Theft Detection

model. The comparison section shows that the proposed model gives better results than other machine learning and deep learning techniques. The proposed method is general and can be applied to any field to detect the anomaly. However, our contribution is just a small step towards the goal of accurate detection of NTLs.

[3] Wide and Deep Convolutional Neural Networks for Electricity-Theft Detection to Secure Smart Grids, Zibin Zheng, 2017. Electricity theft can be harmful to power grid suppliers and cause economic losses. Integrating information flows with energy flows, smart grids can help to solve the problem of electricity theft owing to the availability of massive data generated from smart grids. The data analysis on the data of smart grids is helpful in detecting electricity theft because of the abnormal electricity consumption pattern of energy thieves. However, the existing methods have poor detection accuracy of electricity theft since most of them were conducted on one dimensional (1-D) electricity consumption data and failed to capture the periodicity of electricity consumption. In this paper, we originally propose a novel electricity-theft detection method based on Wide & Deep Convolutional Neural Networks (CNN) model to address the above concerns. In particular, Wide & Deep CNN model consists of two components: the Wide component and the Deep CNN component. The Deep CNN component can accurately identify the non-periodicity of electricity-theft and the periodicity of normal electricity usage based on two dimensional (2-D) electricity consumption data. Meanwhile, the Wide component can capture the global features of 1-D electricity consumption data. The Wide component is a fully-connected layer of neural networks and it learns the global knowledge from the 1-D electricity consumption data. According to the preliminary analysis, the electricity consumption of customers fluctuates from time to time while the normal electricity usage reveals the periodicity and the electricity consumption of energy thieves is less periodic or non-periodic. The electricity consumption of one customer is essentially one dimensional (1-D) time series data. Our preliminary analytical results reveal the periodicity of the normal electricity usage and the non-periodicity of the electricity theft. However, it is difficult to identify the periodicity or the non- periodicity of electricity usage from the 1-D electricity consumption data since the electricity consumption in every day fluctuates in a relatively independent way. Nevertheless, our preliminary results also imply that we can easily identify the abnormal electricity usage if we analyse the electricity consumption by aligning the consumption data of several weeks together.

[4] Efficient Detection of Electricity Theft Cyber Attacks in AMI Networks, A. Muhammad Ismail, 2018. Advanced metering infrastructure (AMI) networks are vulnerable against electricity theft cyber-attacks. While this could be true against physical attacks, the AMI network opens the door for new types of electricity theft attacks, namely cyber-attacks. Electricity theft cyber-attacks against smart meters target the integrity of energy consumption data where a hacker maliciously manipulates the values of energy consumption to reduce the electricity bills. Different from the existing research that exploits shallow machine learning architectures for electricity theft detection; this paper proposes a deep neural network (DNN)-



Article Title: A Deep Neural Network Based on Bayesian Optimizer for Electricity Theft Detection

based customer- specific detector that can efficiently thwart such cyber-attacks. One effective approach applies machine learning techniques on energy consumption load profiles of the customers to detect electricity theft. In this context, a classifier is trained offline on honest and malicious sample energy consumption data to learn their respective patterns. The classifier then is used to detect online any malicious behaviour. Since an on-site inspection is triggered once a malicious attack is detected by the classifier, false alarm is considered to be a limiting performance indicator for electricity theft detectors. Further performance improvement can be achieved if a more complex architecture is adopted for the machine learning technique underlying the electricity theft detector. The proposed DNN based detector implements a sequential grid search analysis in its learning stage to appropriately fine tune its hyper-parameters, hence, improving the detection performance. Extensive test studies are carried out based on publicly available real energy consumption data of 5000 customers and the detector's performance is investigated against a mixture of different types of electricity theft cyber-attacks. A customer- specific DNN-based electricity theft detector is proposed. The developed detector makes use of deep (up to 8 hidden layers) and wide (up to 1000 neurons per hidden layer) architectures to learn complex representations in energy consumption patterns of the customers, leading to better detection performance when compared with other detectors based on shallow architectures. Through a sequential grid search analysis on the detector's hyper-parameters, the proposed DNN based detector achieves a detection rate up to 93% and a false alarm rate as low as 2.3%, which presents an 80% reduction in false alarm rate compared with a benchmark detector that is based on shallow machine learning architecture.

[5] Electricity Theft Detection in AMI Using Customers' Consumption Patterns, Paria Jekar, 2015. As one of the key components of the smart grid, advanced metering infrastructure brings many potential advantages such as load management and demand response. However, computerizing the metering system also introduces numerous new vectors for energy theft. In this paper, we present a consumption pattern-based energy theft detector (CPBETD) that employs novel techniques to overcome the problems associated with existing classification-based ETDSs. In CPBETD, the total consumption of each neighbourhood is measured by transformer meters, and is compared with the total amount of usage reported by the smart meters. If at this level a nontechnical loss (NTL) is detected, customers in the area with abnormal patterns will be selected as suspicious users. For each customer, a multiclass support vector machine (SVM) is trained using historic data of the user as well as a synthetic attack dataset. The classifier is then used to decide whether a new sample is normal or malicious. The main contributions to this paper are as follows.

- 1) We design a novel algorithm for detecting energy theft attacks against AMI. CPBETD employs transformer meters along with monitoring of abnormalities in customers' consumption patterns to provide a cost-effective and high-performance solution for energy theft detection. Through application of appropriate clustering techniques and transformer meters, unlike existing classification-based methods, CPBETD is robust



Article Title: A Deep Neural Network Based on Bayesian Optimizer for Electricity Theft Detection

against contamination attacks and no malicious changes in consumption patterns, and therefore, achieves a higher DR and a lower FPR.

- 2) We address the problem of imbalanced data and zero-day attacks by generating a synthetic attack dataset, benefiting from the fact that theft patterns are predictable. Through extensive experiments we show that this significantly improves the DR and enables the detection of a wide range of attack types.
- 3) We study the effect of sampling rate on detection performance, and show that compared to existing methods, CPBETD provides a higher performance with a lower sampling rate. Hence, it has a smaller effect on customers' privacy. Application of appropriate classification and clustering techniques, as well as concurrent use of transformer meters and anomaly detectors, make the algorithm robust against no malicious changes in usage pattern, and provide a high and adjustable performance with a low sampling rate. Therefore, the proposed method does not invade customers' privacy. Extensive experiments on a real dataset of 5000 customers show a high performance for the proposed method.

[6] Electricity Theft Detection Using Smart Meter Data, Sanujit Sahoo, 2020. Electricity theft is a major concern for the utilities. With the advent of smart meters, the frequency of collecting household energy consumption data has increased, making it possible for advanced data analysis, which was not possible earlier. We have proposed a temperature dependent predictive model which uses smart meter data and data from distribution transformer to detect electricity theft in an area. In this work, we have fine-tuned the predictive model for calculating technical loss for a branch in the distribution network by incorporating the temperature dependency of resistances in a distribution network. One effective way of estimating non-technical losses in the distribution network is by correctly estimating technical losses in the network and then subtracting it from the total loss in the network. The new model performed better than the constant resistance model and gave better power theft detection rates. In addition, we tested the proposed predictive models on distribution circuits, which were approximating them as linear circuits. The performance of our models on these circuits was also very good which implies that they can be used to detect electricity thefts using data from actual smart meters. The model was tested for varying amounts of power thefts and also for different types of circuit approximations. The results are encouraging and the model can be used for real world application.

[7] Towards Sustainable Energy Efficiency with Intelligent Electricity Theft Detection in Smart Grids Emphasising Enhanced Neural Networks, Abdulaziz Aldegheishem, 2021. In smart grids, electricity theft is the most significant challenge. It cannot be identified easily since existing methods are dependent on specific devices. Also, the methods lack in extracting meaningful information from high-dimensional electricity consumption data and increase the false positive rate that limits their performance. Moreover, imbalanced data is a hurdle in accurate electricity theft detection (ETD) using data driven methods. To address this problem,



Article Title: A Deep Neural Network Based on Bayesian Optimizer for Electricity Theft Detection

sampling techniques are used in the literature. However, the traditional sampling techniques generate insufficient and unrealistic data that degrade the ETD rate. In this work, two novel ETD models are developed. In the first model, we present a combination of two different sampling approaches, such as SMOTE and edited nearest neighbour (ENN), known as SMOTEENN, to balance the dataset. Thereafter, a light gradient boosting (LGB) model is used for classification. The proposed model is named as SMOTEENN AlexNet LGB (SALM) for ETD. In the second model, a novel approach is proposed for ETD that consists of a Generative Adversarial Network (GAN), Google Net, and Adaptive Boosting (AdaBoost), named as GAN-NET Boost. In the model, conditional Wasserstein GAN (CWGAN) with gradient penalty (CWGAN-GP) is used to balance the data by synthesizing the fake electricity consumption profiles of the minority class. GAN has gained much attention for anomaly detection. A hybrid sampling approach, i.e., synthetic minority oversampling technique with edited nearest neighbour, is introduced in the first model. Furthermore, Alex Net is used for dimensionality reduction and extracting useful information from electricity consumption data. Finally, a light gradient boosting model is used for classification purpose. In the second model, conditional Wasserstein generative adversarial network with gradient penalty is used to capture the real distribution of the electricity consumption data. It is constructed by adding auxiliary provisional information to generate more realistic data for the minority class. Moreover, Google Net architecture is employed to reduce the dataset's dimensionality. Finally, adaptive boosting is used for classification of honest and suspicious consumers. The proposed models' performance is evaluated using different performance metrics like precision, recall, accuracy, F1-score, etc. The simulation results prove that the proposed models outperform the existing techniques, such as support vector machine, extreme gradient boosting, convolution neural network, etc., in terms of efficient ETD.

[8] Electricity Theft Detection in AMI Based on Clustering and Local Outlier Factor, Yanlin Peng, 2021. As one of the key components of smart grid, advanced metering infrastructure (AMI) provides an immense number of data, making technologies such as data mining more suitable for electricity theft detection. However, due to the unbalanced dataset in the field of electricity theft, many AI-based methods such as deep learning are prone to under-fitting. In this study, we proposed a CLOF based method for electricity theft detection in AMI. By combining k-means and LOF together, this method utilizes LOF to calculate the anomaly degree of outlier Candidates selected by k-means. And a detection framework for practical application is designed. Numerical experiments based on realistic dataset from SGCC with 7 attack types shows that, the proposed method exhibit excellent performance in all attack types except type 1. Thus, our method outperforms other approaches in detecting type MIX which is closer to the real scene. Considering the fact there is no one-fit-all solution to handle all sorts of attack types, the CLOF method is of high value in practical application. However, there are also some limitations in the proposed method. First, the proposed method only analyses electricity consumption data alone, which may contain limited information. In addition to meter



Article Title: A Deep Neural Network Based on Bayesian Optimizer for Electricity Theft Detection

reading data, the other information such as climatic factors (temperature), regional factors, and some electric factors (current and voltage) is worth being studied in the future. Second, our method does not specialize in detecting linear FDI (type 1), which is adopted by most physical attacks. Therefore, it is worthwhile for us to investigate how to supplement the detection for linear FDI in next step. Then, customers whose load profiles are far from the cluster centres are selected as outlier candidates. After that, the LOF is utilized to calculate the anomaly degrees of outlier candidates. Corresponding framework for practical application is then designed. Finally, numerical experiments based on realistic dataset show the good performance of the presented method.

3 Techniques Used

In this work, it gives a summary of the main techniques used, which are: Deep Neural Networks (DNNs), Principal Component Analysis (PCA), and Minimum Redundancy Maximum Relevance (mRMR)

3.1 Principal Component Analysis

PCA is used to extract important information from a data table of inter-correlated features/variables that represent observations. This extracted information is represented as a new set of Orthogonal Variables known as Principle Components (Pcs). In this work, PCA uses a Singular Value Decomposition (SVD) Algorithm. The Singular Value Decomposition (SVD) of a matrix is a factorization of that matrix into three matrices. It has some interesting algebraic properties and conveys important geometrical and theoretical insights about Linear Transformations. In the SVD ($A=U\Sigma V^T$), V is the eigenvector of the Covariance Matrix, while the eigenvalues of it (λ) are hidden in Singular Values (σ). The main goals achieved with PCA are as follows: Extraction of most important information from data/feature table, thereby compressing and simplifying dataset description, and Analysis of observations and variables' structure. PCA was used to transform original high-dimensional consumption data by extracting Principal Components (PCs) which retained the desired variance. PCA is used to extract important information from a data table of inter-correlated features/variables that represent observations. This extracted information is represented as a new set of orthogonal variables known as Principal Components (PCs). In this work, PCA uses a Singular Value Decomposition (SVD) algorithm [43] which works in the following manner: for input feature matrix X , SVD decomposes it into three matrices, i.e., $X = PQR$, such that V is the normalized eigen vectors of the matrix XX^T , Q is the normalized eigen vectors of matrix $X^T X$, and E is a diagonal matrix of eigen values of matrix XX^T , and R is the normalized eigen vectors of matrix $X^T X$. When PCA is applied to a matrix X of size $m \times n$, n PCs are obtained, which are ordered in descending order with respect to their variances [23]. A PC at position p and its variance is obtained by evaluating $\frac{1}{n} \sum_{i=1}^n X_{ip}^2$. The main goals achieved with PCA are as follows: Extraction of most important information from data/feature table, thereby compressing and simplifying dataset description, and Analysis of observations and variables' structure. For dimensionality



Article Title: A Deep Neural Network Based on Bayesian Optimizer for Electricity Theft Detection

reduction purposes, the first r PCs that retain acceptable variance can accurately represent feature matrix X in a reduced r -dimensional subspace.

3.2 Deep Neural Networks

Artificial Neural Networks (ANNs) are a class of machine learning techniques that have been built to imitate biological human brain mechanisms. They are typically used for extracting patterns or detecting trends that are difficult to be detected by other machine learning. They consist of multiple layers of nodes/neurons which are connected to subsequent layers. 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 Figure 1 shows a model diagram of a neuron that comprises a layer following the input to the ANN. The Deep Neural Networks (DNNs) concept originates from research on ANNs. DNNs are characterized by two or more hidden layers. They are able to learn more complex and abstract features than shallow ANNs.

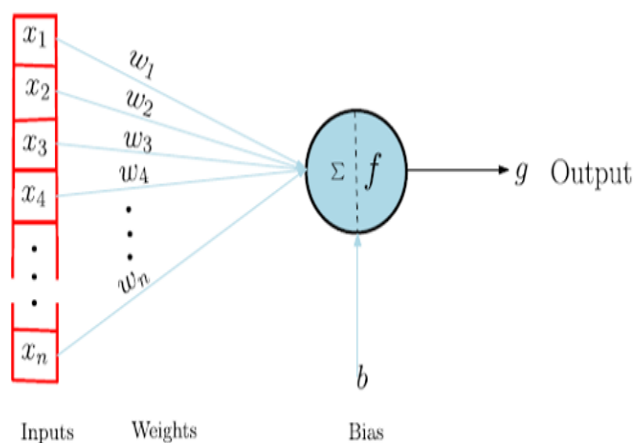
It consists of an activation function f , which takes a weighted sum of the real number input signal and gives real number output y given by Equation (1).

$$y = f(\sum_{i=1}^n w_i x_i + b); \quad (1)$$

where $x_i \in x$, $w_i \in w$, x is input vector, w is weights vector and b is the bias. Neural network nodes mimic the brain's neurons, while connection weights mimic connections between neurons, which are unique for each connection. A neural network stores information in the form of weights and bias.

3.3 Deep Neural Network Training

A large dataset and high computational abilities are the major requirements in training the DNN since weight updates require multiple iterations. DNN training process is concerned with adjusting the weights between the neurons. Through the training process, the DNN learns information from the data. Learning can be in the following major four ways: supervised, semi-supervised, and unsupervised or reinforcement.





Article Title: A Deep Neural Network Based on Bayesian Optimizer for Electricity Theft Detection

Figure 1: First Hidden Layer

3.4 Minimum Redundancy Maximum Relevance

An [mRMR] is a feature selection scheme that selects features that have a high correlation with the response variable and low correlation with themselves. It ranks features based on mutual information of a feature and a response variable, and pairwise mutual information of features. Mutual information between variables **A** and **B** is given by **MI(A;B)**

$$MI(A, B) = \sum_{a \in A} \sum_{b \in B} p(a, b) \log \frac{p(a, b)}{p(a)p(b)}. \quad (2)$$

For all features fX_i , maximum relevance R_l is implemented using mean value of their mutual information with an output class O .i.e.,

$$R_l = \frac{1}{|X|} \sum MI(O, X_i). \quad (3)$$

Minimum redundancy R_d helps to select features that are mutually maximally dissimilar. It is given

$$R_d = \frac{1}{|X|^2} \sum MI(X_i, X_j). \quad (4)$$

Where $X_i, X_j \in X$. mRMR feature selection goal is achieved by optimizing relevance and redundancy in the following ; manner: $\max(R_l - R_d)$.

3.5 Bayesian Optimizer

Bayesian Optimization to find the best transferred parameters. It is an approach for Model-based Global Optimization of Black-box Function. Traditionally, to find the best parameters, all parameters need to be evaluated, which is time-consuming. To address this issue, Gaussian Process-based Bayesian Optimization Approach issued to get the optimal parameters quickly. Genetic Algorithm and Bayesian Optimization can be used if the input space is an Euclidean space.

4 Proposed Work

An effective electricity theft detection based on Deep Neural Network (DNN) Based Classification Approach is proposed. By employing frequency-domain features as opposed to using time-domain features alone enhances classification performance. Secondly, proposed a Minimum Redundancy Maximum Relevance (Mrmr) scheme to identify the most significant features. It selects features (time domain and frequency domain) that have a high correlation with the response variable and low correlation with themselves. Finally, a Fully Connected



Article Title: A Deep Neural Network Based on Bayesian Optimizer for Electricity Theft Detection
Feed-forward Deep Neural Network (DNN) with Genetic Optimizer architecture is used for the classification process

5 System Architecture

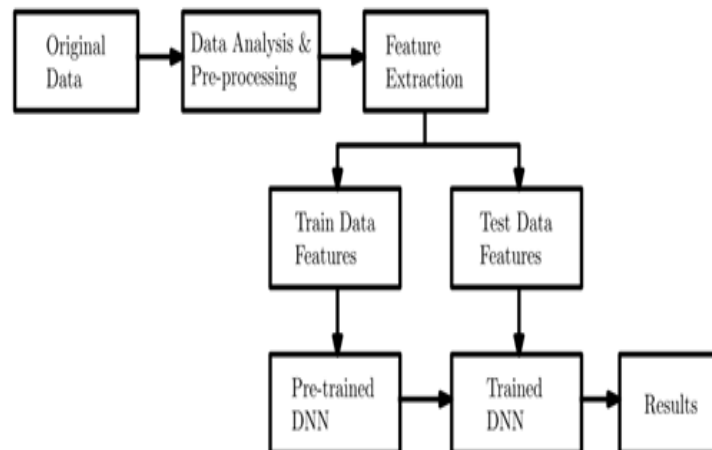


Figure 2: System Architecture

6 Modules

6.1 Data Analysis and Pre-Processing

i.e., given a dataset A of size N , $a_i \in A$; where a_i is the i_i observation of A in this sub-section, we present the dataset used and its quality improvement by identifying and removing observations that had no consumption data. In this work, an observation refers to a single instance/record in the dataset, for the duration of measured consumption. And $1 \leq i \leq N$. We show customers' load profiles analysis. We further present data interpolation and synthetic data generation details that have been undertaken. Using data analysis methods, we found approximately 5.45% of observations in this dataset to either have only null values, or zeros, or a combination of both, for the whole duration of 1034 days. These observations were regarded as empty observations.

6.2 Feature Extraction

Electricity consumption data used in this project is univariate time-series data. A univariate measurement is a single measurement frequently taken over time. For solving classification problems, data can be represented by its features (properties), which can then be fed as input to the classifier. Data is classified based on the similarity between features given a dataset of different samples. In this work, time-domain and frequency domain features were extracted and used as input to a deep neural network for classification. Classification performance comparison between time-domain, frequency domain and combined features from both domains was carried out.



Article Title: A Deep Neural Network Based on Bayesian Optimizer for Electricity Theft Detection

6.3 Classification

6.3.1 Network architecture

A fully connected feed-forward DNN architecture shown in Figure, was used for the classification process. In order to avoid network under fitting and overfitting, the following rule of thumb methods was considered in the design of hidden layers of a deep neural network classifier:

1. Number of hidden neurons should be between the size of the input layer and size of the output layer.
2. Number of hidden neurons should be approximated to the summation of 2, 3 size of input layer and size of the output layer.
3. Number of hidden neurons should be less than twice the size of the input layer. Rectified Linear unit (ReLU) activation function was used in the hidden neurons because of its better convergence property in comparison to other activation functions

6.3.2 Training

The maximum number of training iterations was limited to 1000. The classification approach was split into four parts. In the first part, only time-domain features were used for classification. In the second part, only frequency-domain features were used. The third part comprised of combined features from both domains, while in the last part, classification was performed in reduced feature space by incorporating PCA.

6.3.3 Performance Metrics

Based on true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) obtained from a confusion matrix; we used the following performance metrics to evaluate the classifier's performance: Recall/True Positive rate (TPR), Precision/Positive Predictive Value (PPV), Score, Matthews Correlation Coefficient (MCC), Accuracy and Area under the Curve of Receiver Operator Characteristic (AUC-ROC) curve. We briefly introduce performance metrics used as follows.

- **Recall / True Positive Rate (TPR):** is the measure of the fraction of positive examples that are correctly labelled. It is given by:

$$TPR = \frac{TP}{TP + FN} \quad (5)$$

- **Precision / Positive Predictive Value (PPV):** is the measure of the fraction of examples classified as positive that are truly positive. It is given by:

$$PPV = \frac{TP}{TP + FP} \quad (6)$$

- **F1-Score:** shows the balance between precision and recall. It is given by:



Article Title: A Deep Neural Network Based on Bayesian Optimizer for Electricity Theft Detection

$$F1-Score = 2 * \frac{TPR * PPV}{TPR + PPV}. \quad (7)$$

- **Accuracy:** shows the fraction of predictions classified correctly by the model. It is given by:

$$\begin{aligned} Accuracy &= \frac{\text{Number of correct predictions}}{\text{Total number of predictions}} \\ &= \frac{TP + TN}{TP + TN + FP + FN}. \end{aligned} \quad (8)$$

- **Matthews Correlation Coefficient (MCC):** a single digit that measures a binary classifier's performance. Its value ranges from -1 to $+1$, with values closer to $+1$ signifying good performance, while values closer to -1 signify bad performance. MCC is given by:

$$MCC = \frac{TP * TN - FP * FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}. \quad (9)$$

- **Area Under the Curve (AUC):** measures the classifier's overall quality. Larger AUC values indicate better classifier performance.

7 Result and Discussion

7.1 Validation Results before Synthetic Data Generation

When there was an imbalance in the number of observations between two classes, the classifier performed badly on the class with a relatively lower number of observations. The classifier was trained with features extracted from an original dataset with no augmented synthetic data. 80% of the data was used for training while 20% was used for validation. The third column of Table 4 shows the validation results. For the faithful customer's class, validation results are much better than the unfaithful class. Compared with validation results in combined domains before the incorporation of PCA, there was no significant change in the recall, precision and F1-score for faithful customers' class since the difference in corresponding values was within 1% margin. However, for the unfaithful class, which was the minority class, validation results in terms of recall, precision and F1-score were not good at all before balancing the classes. A significant improvement was obtained after balancing the classes. This shows that the sensitivity of the classifier to the minority class was not as good as its sensitivity to the opposite class. The subsequent subsections show the results which were obtained after augmenting synthetic data to the original dataset to balance classes.

7.2 Different Domains Features' Contribution Analysis

To ensure the reliability and robustness of the method introduced in this work, we present experimental results based on widely-accepted performance metrics summarized. To simplify the analysis, classification performance between time-domain, frequency-domain and



Article Title: A Deep Neural Network Based on Bayesian Optimizer for Electricity Theft Detection

combined features from both domains is graphically presented, it can be seen that the classification process taken with time-domain features gave impressive validation and test results for both faithful and unfaithful customers classes. An experiment done with frequency-domain features alone showed improved results. The best results were obtained when all features from both domains were combined. For example, on validation, accuracy was 87.5%, which improved to 89.9%, and final 91.1% when the experiment was done with time-domain features, frequency-domain features and all features from both domains respectively. The red trend line in Figure 8 graphs portrays significant improvement on results obtained from experiments done with time domain features, frequency-domain features and all features from both domains. This improvement can be explained by a bar chart of predictors presented in order of their prominence, which has been produced through the mRMR scheme. There are more frequency-domain features to the left of the bar chart (i.e., features with the best scores) than time-domain features, with mean frequency achieving the highest predictor score. We confirmed the exactness of features' ranking through the mRMR scheme by doing classification tasks using top 3, middle 3 and bottom 3 features on the same network classification accuracy and AUC-ROC results. Comparing the results, we observed that accuracy and AUC-ROC are best for the top 3 features and worst for the bottom 3 features, as expected. MCC was determined on the last experiment when all features were combined. Its values were found to be 0.84 and 0.75 on validation and test respectively, which are closer to +1 than -1. AUC-ROC values were found to be 97% and 93% on validation and test respectively. These results portray a satisfactory overall classification task.

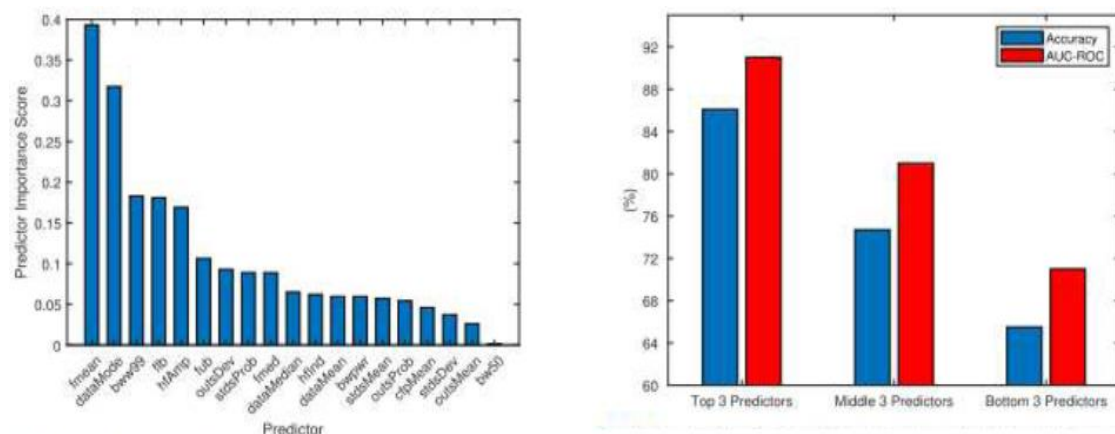


Figure 3: Classification Results Comparison of Features Ordered by mRMR Scheme

7.3 Analysis of Components Reduction with Principle of Component Analysis

When PCA was incorporated with the component reduction criterion of leaving enough components to explain 95% variance, seven components were left, having the following percentages contributions to total variance: 35.84%, 27.02%, 15.55%, 7.69%, 4.87%, 3.30% and 1.81%. 2-D bi plots of original features contribution to each of the components in the



Article Title: A Deep Neural Network Based on Bayesian Optimizer for Electricity Theft Detection

principal components space. Frequency-domain feature vectors are labelled with 's', while time-domain features are labelled with 't'. The contribution of each feature to the principal component is signified by that feature's vector direction and length. This was also confirmed by features importance scores analysis based on the mRMR scheme. The last two columns show both validation and test results obtained after components reduction with PCA.

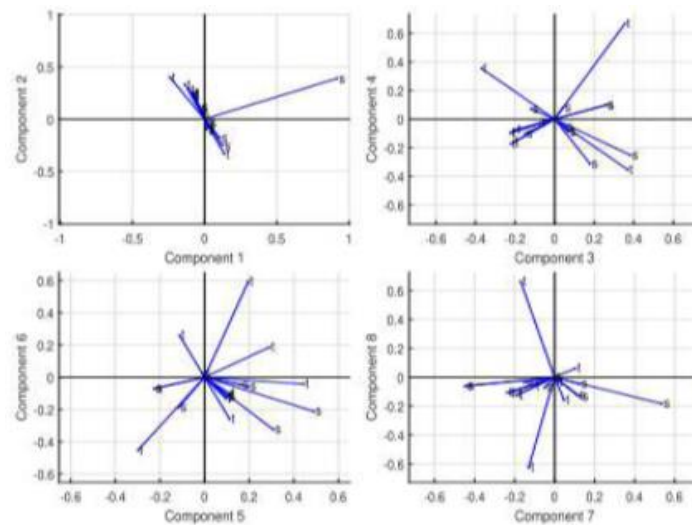
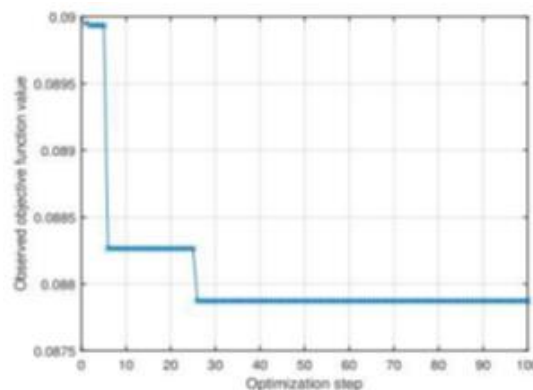


Figure 4: Graphical Display

7.4 Hyper parameters Optimization Results

The hyper parameters optimization procedure is stipulated, observed objective function values vs optimization steps. The best hyper parameters combination was obtained at the 26th optimization step and remained unchanged till the 100th step. An improved classification network architecture constructed with optimized hyper parameters achieved maximum validation and test accuracies of 91.8% and 88.1% respectively, which are 0.7% and 0.8% higher than an optimized.





Article Title: A Deep Neural Network Based on Bayesian Optimizer for Electricity Theft Detection

Figure 5: Objective Function Value

Table 1: Optimized hyper parameter values

Parameter	Value
Fully-connected Layers	[41 21]
Regularization strength	5.6882×10^{-7}
Activation function	Sigmoid

8 Conclusion

In this work, the Detection of Electricity Theft in Smart Grids was investigated using Time-domain and Frequency-domain Features in a Deep Neural Network (DNN)-based Classification Approach. Isolated Classification Tasks based on the time-domain, frequency domain and combined domains features were investigated on the same DNN network. Widely accepted performance metrics such as Recall, Precision, F1-score, Accuracy, AUC-ROC and MCC were used to measure the performance of the model. The classifier was able to achieve 87.3% accuracy and 93% AUC-ROC when tested.

Reference

1. Q. Louw; P. Bokoro, Year: 2019, "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.
2. Z. Zheng; Y. Yang; X. Niu; H.-N. Dai; Y. Zhou, Year: 2018, "Wide and deep convolutional neural networks for electricity-theft detection to secure smart grids", IEEE Transactions. International. Information., Vol: 14, no: 4, pp. 1606 – 1615.
3. S. B. Yousaf; M. Jamil; M. Z. U. Rehman; A. Hassan; S. O. G. Syed, Year: 2016, "Prototype development to detect electric theft using PIC18F452 microcontroller", Indian J. Sci. Technology., Vol: 9, no: 46, pp. 1 5.
4. X. Fang; S. Misra; G. Xue; D. Yang, Year: 2012, "Smart grid —The new and improved power grid: A survey", IEEE Communication. Surveys Tuts., Vol: 14, no: 4, pp. 944 – 980.
5. P. Jokar; N. Arianpoo; V. C. M. Leung, Year: 2015, "Electricity theft detection in AMI using customers' consumption patterns", IEEE TransactionsSmart Grid, Vol: 7, no: 1, pp. 216 – 226.
6. I. Diahovchenko; M. Kolcun; Z. Čonka; V. Savkiv; R. Mykhailyshyn, Year: 2020, "Progress and challenges in smart grids: Distributed generation, smart metering, energy storage and smart loads", Iranian J. Sci. Technol., Transactions. Electrical. Engineering. Vol: 44, no: 4, pp. 1319 – 1333.
7. W. Han; Y. Xiao, Year: 2017, "A novel detector to detect colluded non-technical loss frauds in smart grid", Computer. Network., Vol: 117, pp. 19 – 31.



Article Title: A Deep Neural Network Based on Bayesian Optimizer for Electricity Theft Detection

8. D. O. Dike, U. A. Obiora, E. C. Nwokorie, and B. C. Dike, Year: 2015, “Minimizing household electricity theft in Nigeria using GSM based prepaid meter”, Amer. J. Eng. Res., Vol: 4, no: 1, pp. 59 – 69.
9. P. Dhokane; M. Sanap; P. Anpat; J. Ghuge; P. Talole, Year: 2017, “Power theft detection & intimate energy meter information through SMS with auto power cut off”, International. J. Current Res. Embedded System. VLSI Technology., Vol: 2, no: 1, pp. 1 – 8.
10. Y. Peng, Y. Yang, Y. Xu, Y. Xue, R. Song, J. Kang, and H. Zhao, Year: 2021, “Electricity theft detection in AMI based on clustering and local outlier factor”, IEEE Access, Vol: 9, pp. 107250 – 107259.
11. B. Vega-Márquez; I. Nepomuceno-Chamorro; N. Jurado-Campos; C. Rubio-Escudero, Year: 2020, “Deep learning techniques to improve the performance of olive oil classification”, Frontiers Chem., Vol: 7, pp. 929.
12. M. K. Uçar; M. Nour; H. Sindi; K. Polat, Year: 2020, “The effect of training and testing process on machine learning in biomedical datasets”, Mathematical. Problems Eng., Vol: 2020, pp. 1 – 17.
13. M. Billah; S. Waheed, Year: 2020, “Minimum redundancy maximum relevance (mRMR) based feature selection from endoscopic images for automatic gastrointestinal polyp detection”, Multimedia Tools Applications., Vol: 79, nos. 3334, pp. 23633 – 23643.
14. M. N. Hasan; R. N. Toma; A.-A. Nahid; M. M. M. Islam; J.-M. Kim, Year: 2019, “Electricity theft detection in smart grid systems: A CNN-LSTM based approach”, Energies, Vol: 12, no: 17, pp. 3310.
15. A. Aldegheishem; M. Anwar; N. Javaid; N. Alrajeh; M. Shaq; H. Ahmed, Year: 2021, “Towards sustainable energy efficiency with intelligent electricity theft detection in smart grids emphasising enhanced neural networks”, IEEE Access, Vol: 9, pp. 25036 – 25061.