Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 12, No. 1, December : 2023

IDENTIFICATION OF POWER THEFT IN SMART GRID USING DEEP NEURAL NETWORK

Dr. Praveena Chaturvedi Professor Department of Computer Science, Kanya Gurukul Campus Dehradun-248001(Campus of Gurukul Kangri Deemed to be University, Haridwar, Uttrakhand, India

Abstract

The Power stealing is a worldwide issue that has a detrimental impact on both utility providers and electrical customers. It destabilises utility companies, economic development and leads to electric risks having an influence on the in-height rate of current for patrons. The expansion of shrewd networks is crucial in the detection of power theft because they create vast amounts of data, including consumer usage data, which may be used to identify electricity stealing using machine learning and bottomless knowledge techniques. This research describes a method for detection of theft that employs extensive information in the while and regularity domains in a deep neural network-based classification approach. We fix dataset problems including missing information and class imbalance concerns using data approximation and artificial information creation techniques.We assess and analyse the contributions of features from the frequency and temporal domains, perform tests in merged and decreased feature space utilising the analysis of principal components, and subsequently use the lowest redundancy maximum relevance strategy for verifying the most pertinent features.We improve power theft detection performance by optimising hyperparameters with a combining an adaptive moment estimate optimizer and a Probabilistic optimizer to run tests with varying values of critical parameters to identify the ideal settings that produce the greatest accuracy.

Keywords: Electricity, Deep Neural Network, CNN, ANN

Introduction

Electricity robbery is a worldwide issue that impacts utility providers. Every year, usefulness businesses lose more than $96 billion owing to Non-Technical Losses (NTLs), with energy theft being the greatest important cause.Giving to the World Bank, 50% of generated electricity in Sub-Saharan Africa is stolen[1].

The main aim of electricity thieves is to consume energy without being charged by utility providers, or to pay mandibles that are less than the quantity spent. As an outcome, utility companies face significant revenue losses as a result of power theft. According to sources, India lost $18.2 billion in 2020, Brazil lost $12.5 billion, and Russia lost $6.1 billion. Electricity theft is estimated to cost South Africa approximately $1.31 billion in revenue each year[2]. Aside from revenue loss, electricity theft has a shortest negative impact on power grid stability and reliability. It can lead to surging electricity, electrical systems overload and public safety hazards such as electric shocks. It also has a straight influence on energy tariff increases, which affect all customers[3].

The rising need for electricity has fueled the development of smart grids, which provide several benefits such as increased energy efficiency, fewer power interruptions, and higher security. However, authority theft is a serious concern in smart grids and a substantial source of income loss for utility corporations. Power robbery is thus a big problem for influence supply firms. The aim of this research is to provide an efficient method for detecting power theft in clever nets using Artificial Neural Networks (ANN). The suggested method would make use of a power use dataset obtained from the renowned web repository kaggle[18]. Pre-processed data will be put into the ANN, which will learn to spot patterns and abnormalities in the consumption data.

The ANN model will be trained on a dataset of lawful usage patterns before being evaluated on data including cases of energy theft.The model will be evaluated using test data to appraise the recital of the suggested strategy. The projected outcomes from our suggested method of power theft detection in smart grids utilising ANN are favourable. Our method obtained 99% Training Accuracy and 99%

Validation Accuracy. Accuracy, precision, recall, and F1-score will be employed as performance measurements.We have built the suggested system on the Flask Web framework for ease of use and a better User Interface for forecasting results.

Review of Literature

Theft of electricity and illicit ground surface conductor connections are a widespread problem in South Africa, according to SAIEE researchers. This marvel not only grounds income loss and equipment damage, but it also poses a life-threatening concern. Despite decades of research into non-technical losses, no universal explanation has been given due to the problem's complexity. This research studies the usage of zero-sequence current-based detection as a mitigation approach for dealing with unauthorised ground surface conductor connections. The validity of this approach, as well as its influence on seasonal changes in soil resistivity, is demonstrated by simulation and experimental data. In this research, Zibin Zheng[4] intend to develop a unique power theft uncovering approach to overcome the aforementioned difficulties. They propose a Wide & Deep Convolutional Neural Networks (CNN) model to learn power usage data and identify electricity thieves. Deep & Wide CNN model is made up of two parts: a Wide component with a fully-connected layer of neural networks and a Deep CNN component with numerous convolutional coatings, a pooling coating, then a fully-connected layer[12]. In essence, the Wide component can acquire global information, In contrast, the Deep CNN element can understand the periodicity of data on power usage. The advantages associated with Wide & Deep CNN are combined in this approach. components, resulting in high performance in power theft detection[5][6].

Some publications investigated ETD approaches, which utilise smart metre consumption data to identify deceptive users[7]. Academics are concerned about the observation of customer load profiles for signs of electricity theft in conservative control networks. Angelos et al. used five parameters, including maximum consumption, mean consumption, a summary of the inspection notes, standard deviation,and neighbourhood mean consumption, to provide a typical form of power consumption for each user[9]. For gathering consumers with similar characteristics, K-means fuzzy clustering was obtained. Customers with plenty of parking near the cluster centres were deemed to be potential cheaters[10].

P. Dhokane and M. Sanap, mentioned that electricity theft through unauthorised connections is a substantial source of non-technical loss contribution. These influences are often linked to South African supply networks' low voltage networks. Socioeconomic conditions are the primary cause of these occurrences, and a collective strategy involving political, economic, and engineering interaction is required to mature answers that address all stakeholder needs though also addressing the safety of the population living in these communities where these illegal connections occur[14].

EXISTING MODEL:

They describe an effective power theft detection technique in the existing system that is based on carefully collected and chosen characteristics in a Deep Neural Network (DNN)-based classification methodology. We demonstrate that utilising frequency-domain features rather than time-domain features alone improves classification performance[8][10]. The previous approach relied on a realistic power usage dataset made available by the Government Net Company of Porcelain (SGCC). To understand the findings and ease future training, the current system used Principal Component Analysis (PCA) to conduct classification with reduced ear planetary and compare the results with classification done with all input characteristics[11].

The old approach relied on the Lowest Joblessness All-out Significance (mRMR) scheme to determine the maximum significant characteristics and justify the importance of frequency-domain data over time-domain features in identifying electricity theft.Whereas prior system models produced outstanding results, their reliance on time-domain properties alone restricted their efficacy.

For training, conventional system model DNN-based approaches require enormous volumes of labelled data. In the situation of control theft detection, this container be a trouble since acquiring labelled data can be difficult and time-consuming. The present system DNN models are computationally costly to train and can take a long time, especially for big datasets. This might make it difficult to adjust fast to new data or changes in the smart grid.
The current system overfitting may arise in DNN models when the model gets overly specialised to the training data and performs badly on fresh, unknown data. This can be a concern in the detection of power theft since it might lead to missed thefts or false alarms.Existing system DNN models are frequently regarded as black-box models, implying that the model's decision-making process might be difficult to explain[13][14]. This can make understanding the aspects that lead to the detection of energy theft difficult, as well as explaining the results to stakeholders or authorities.

PROPOSED METHODOLOGY:
The current system adversarial attacks on DNN models are possible, in which an attacker manipulates the input data to cause the model to generate inaccurate predictions. This can be a serious issue in the detection of power theft since it allows a bad actor to avoid detection.
Our suggested method for perceiving energy robbery in clever networks using fake neural nets (ANN) comprises of three steps: data analysis and pre-processing, feature extraction, and classification. The suggested approach makes advantage of the Kaggle-referenced power usage dataset[. Pre-processing of the obtained data will involve data cleansing, normalisation, and feature extraction. This step is crucial because it guarantees that the data is in a format that the ANN model can learn from. The dataset contains no labels for loyal or unfaithful usage. So we'll start by labelling the dataset via Agglomerative clustering.
The suggested system comprises the development of Clustering (to detect Electricity Theft (Target value)). As in our previous study (base), we used agglomerative clustering with a cluster value of 3.

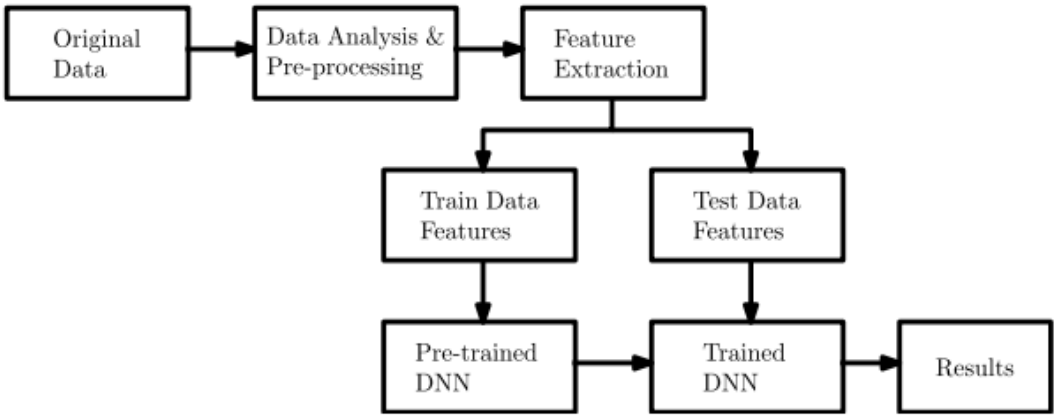


Fig. 1. Proposed Architecture

After that, the suggested system was trained using the Artificial Neural Network (ANN). A big dataset of labelled power use data will be used to train the ANN model. The programme will learn to recognise patterns and abnormalities in data that suggest cases of electricity theft. The model's performance will be measured using several metrics like as accuracy, precision, recall, and F1-score.
Great accuracy: ANN replicas take been established to identify power theft with great accuracy. This is due to the fact that ANN models may learn complicated patterns and correlations in consumption data that are difficult to detect using standard statistical approaches.
Robustness: ANN mockups can arrangement with deafening and missing data, which is frequently encountered in real-world smart grid deployments. This increases the robustness of ANN models and lessens the likelihood of mistakes and false positives.

Adaptability: ANN models are capable of adapting to changes in the keen grid, such as new forms of theft or shifts in consumption patterns. This makes ANN models more adaptable to the dynamic nature of smart grids.
Speed: Because ANN models can handle huge volumes of data fast, they are perfect for sensing energy theft in real time. This can assist utility firms in responding fast and taking corrective measures to reduce revenue losses.
Automation: ANN models may be taught to identify energy theft automatically, removing the need for physical inspection and lowering utility companies' burden. This can result in substantial cost reductions and enhanced efficiency.
This method makes use of univariate time-series data on electricity use. It is a single measurement that is conducted regularly across time. Data can be represented by its features (properties) for classification issues, that is given as input to the classifier, as shown in figure1. Based on collection of distinct samples, data is categorised built on the comparison of characteristics. Time-domain and frequency-domain information were retrieved and utilised as input to a deep neural network for classification in this work.

IMPLEMENTATION

Dataset: For the first module's training and testing purposes, we constructed a mechanism to obtain

5.1 Data Collection and Dataset

In the first module, we do the process of data collection. This is the first real step towards the real development of a machine learning model, collecting data. This is a critical step that will cascade in how good the model will be, the more and better data that we get; the better our model will perform. There are several techniques to collect the data, like web scraping, manual interventions and etc. We have given the dataset in the model folder. The dataset is referred from the popular web repository Kaggle[18]. The dataset comprises a total of 3,510,433 individual data points. It consists of 9 columns with the following descriptions:
LCLid: Identification number for each entry.
day: Date in the format dd/mm/yyyy.
Energy median: Median value of energy.
Energy mean: Mean value of energy.
Energy max: Maximum value of energy.
Energy count: Count of energy values.
Energy std: Standard deviation of energy.
Energy sum: Sum of energy values.
Energy min: Minimum value of energy.
Each row in the dataset represents a unique data entry, capturing information about energy measurements on specific days.We will be using Python language for this. First we will import the necessary libraries such as keras for building the main model, sklearn for splitting the training and test data, PIL for converting the images into array of numbers and other libraries such as pandas, numpy ,matplotlib and tensorflow.

5.2 Clustering (To find Electricity Theft (Target value)) :

To find the Electricity Theft using Agglomerative clustering with cluster value = 3 as from our other analysis (base on mean energy).Then the data set is split into train and test. 80% train data and 20% test data.The second step is to choose a neural network to represent the classification function. For classification problems, it is composed of:
A scaling layer.
A perceptron layer.
A probabilistic layer.
For the scaling layer, the minimum and maximum scaling methods are set.

We set one perceptron layer, with 3 neurons as a first guess, having the logistic activation function.Some major differences between them are biological neural network does parallel processing whereas the Artificial neural network does series processing also in the former one processing is slower (in millisecond) while in the latter one processing is faster (in a nanosecond).

5.3 Architecture of ANN

A neural network comprises multiple layers, and each layer serves a distinct function. As the complexity of the model grows, the number of layers also increases, leading to its designation as a multi-layer perceptron. This architecture allows for the network to learn and extract hierarchical representations of the input data, enabling more intricate and sophisticated pattern recognition. Working of ANN, In the initial stage, the input layer receives the input information, which is then transmitted to the hidden layers. The interconnections between these layers assign random weights to each input during the network's initialization phase. Additionally, a bias is introduced to each input neuron. Subsequently, the weighted sum, a combination of the weights and bias, undergoes processing through the activation function. The activation function determines which nodes should activate for feature extraction, ultimately leading to the computation of the output.

This entire process, known as Forward Propagation, involves feeding information through the layers and calculating the output. Following this, the model compares the output with the desired output, calculating the error. In the subsequent Backward Propagation phase, the weights are adjusted to minimize the error. This iterative process continues for a specified number of epochs. As a result, the model's weights are updated, and predictions can be made based on the updated weights.

5.4 Model selection:

Within this module, we aim to develop and train an Artificial Neural Network (ANN) model utilizing the extracted features and their corresponding labels, which indicate faithfulness or unfaithfulness. To optimize the network's generalization performance, we employ order selection to determine the ideal complexity of the neural network. This involves finding the number of neurons that minimize the error for the selection instances.

Additionally, we employ input selection, also known as feature selection, to identify the most effective set of input variables that enhance generalization. In this case, a genetic algorithm has been implemented for input selection. However, after evaluation, it was determined that the algorithm did not reduce the selection error value. Consequently, we decide to retain all input variables in the model. In Kera's, models are constructed as a sequential sequence of layers, where each layer is added sequentially. The input features are specified when creating the first layer using the input dim argument. In this case, the input dim value will be set to 8, indicating the number of input dimensions or features in the input data. This ensures that the model is properly configured to handle input data with 8 features.We compile the model and apply it using fit function. The batch size is 64. Then we will plot the graphs for accuracy and loss. We got average training accuracy of 99%.

5.5 Analyze and Prediction:

This module will extract relevant features from the processed data to be used as inputs for the ANN. In the actual dataset, we chose only 7 features :

energy_median : energy medium value
energy_mean : energy medium value
energy_max : Energy max value
energy_count : Energy count value
energy_std : Energy std value
energy_sum : Energy sum value
energy_min : Energy min value
Target: Unfaithful and Faithfull

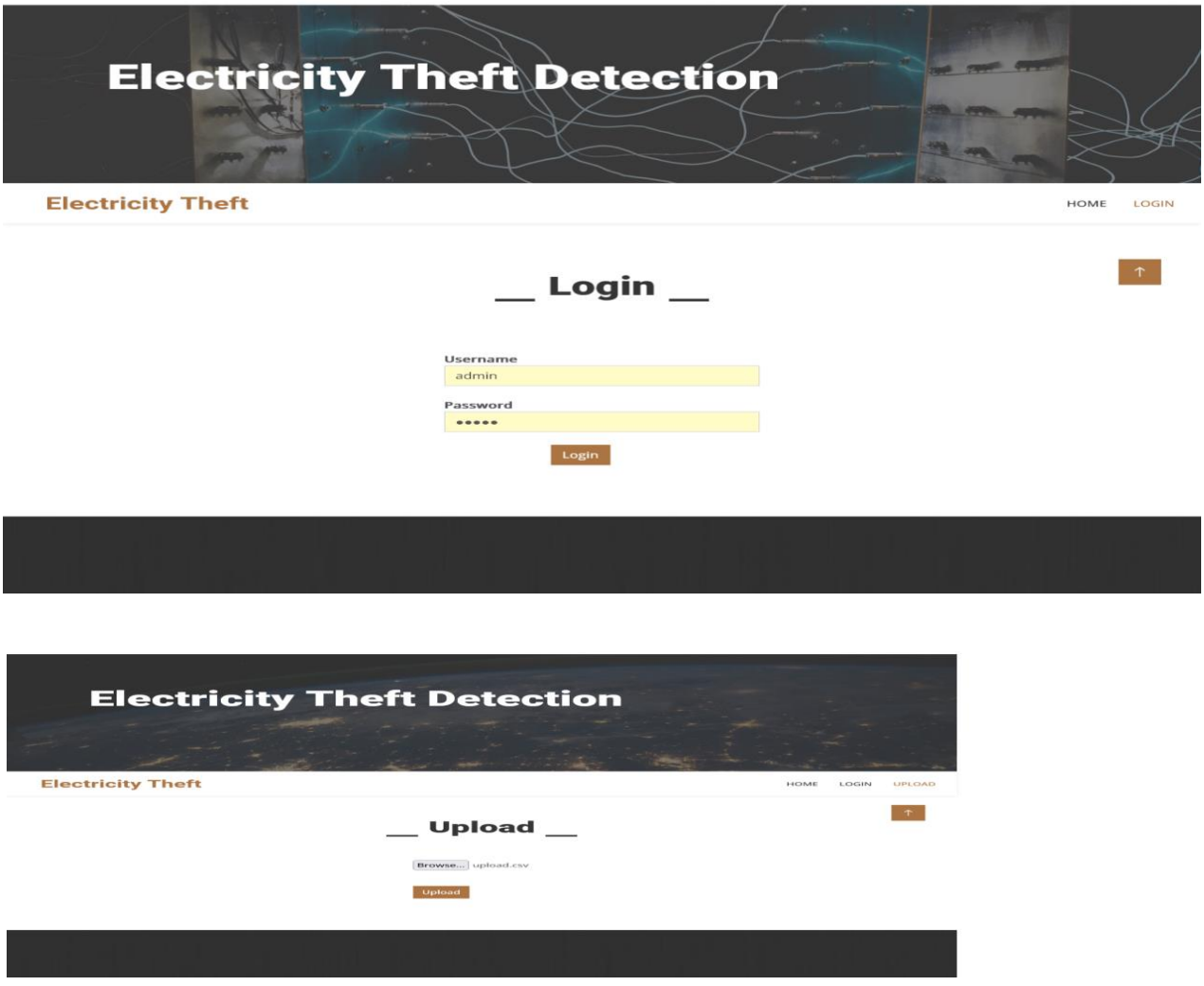
In this we evaluate the performance of the developed ANN model on a test dataset to determine its accuracy and effectiveness. We got an accuracy of 99% on test set. When we have reached a level of

confidence with our trained and tested model and are ready to transition it to a production-ready environment, the initial step involves saving it to a file using a library like pickle.

RESULTS

Applying the test data the result shown in the Fig 2 , Depicting the energy levels at different stages. Based on prediction value, the dataset retrieves the system with pitch and energy during testing and delivers the indicated theft.

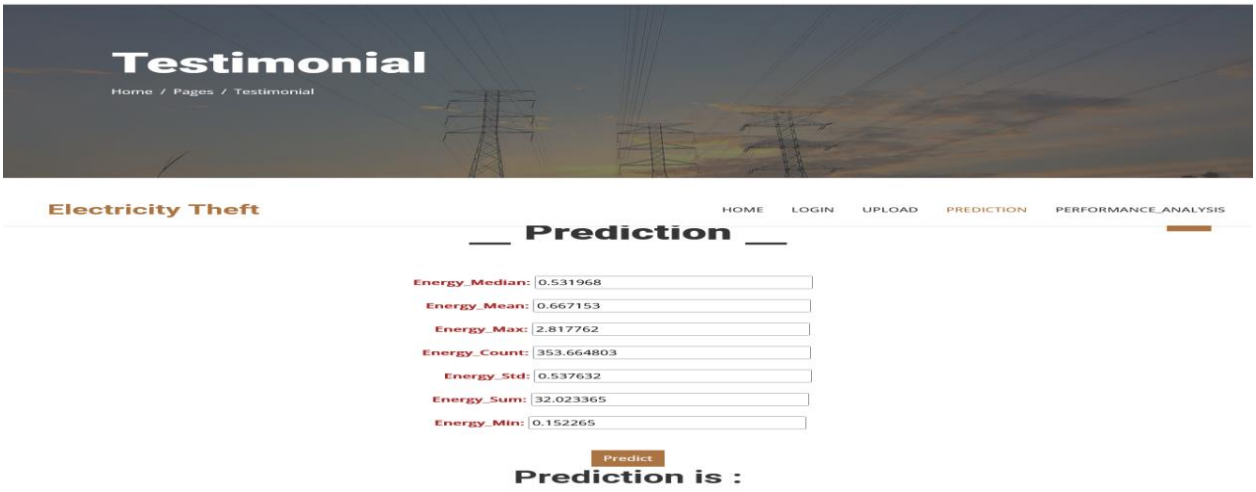
Screen shots of the system



__ Preview __

ID	Day	Energy_median	Energy_mean	Energy_max	Energy_count	Energy_std	Energy_sum	Energy_min
1	15-12-2022	0.4850	0.432045	0.868	22	0.239146	9.505000	0.072

2	16-12-2022	0.1415	0.296167	1.116	48	0.281471	14.216000	0.031
3	17-12-2022	0.1015	0.189812	0.685	48	0.188405	9.111000	0.064
4	17-12-2022	0.1140	0.218979	0.676	48	0.202919	10.511000	0.065
5	18-12-2022	0.1910	0.325979	0.798	48	0.259205	15.647000	0.066
6	19-12-2022	0.2180	0.357500	1.077	48	0.287597	17.160000	0.066



Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 12, No. 1, December : 2023

Fig.2. Result

The classification of electricity, which has several practical uses, is one of the biggest issues in the world. This section tests the proposed system in order to gather and analyse the results that show the system's effectiveness. Various studies have made use of the power usage dataset. In these experiments, the electrical signals classifier configuration is tested, the best BM model configuration (of two chosen layers) is applied, accuracy and loss are measured using the deep CNN and BM models, and accuracy and loss results from the CNN and BM models are then compared to accuracy and loss results from the CNN model alone.

CONCLUSIONS

In this paper, using time-domain and frequency-domain power theft was detected in smart grids in a DNN-based classification technique. Isolated classification tasks based on time-domain, frequency-domain, and mixed-domain features were also examined on the same DNN network. The performance of the model was measured using widely known performance metrics recall, precision, F1-score, accuracy, AUCROC. It was discovered that classification using frequency-domain features outperforms classification using time-domain features, which in turn outperforms classification using both domains. When tested, the classifier achieved 88.3% accuracy and 92% AUC-ROC. For feature reduction, we employed PCA. When evaluated, the classifier achieved 84.8% accuracy and 90% AUC-ROC using 7 out of 20 components. We next examined individual features' contributions to the classification job and validated the relevance of frequency-domain features over time-domain features in a successful classification task using the mRMR method.

REFERENCES

[1] 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.
[2] 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-global opportunity-electrical-utilities
[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] Annu Sharma, Shwetank Arya, Praveena Chaturvedi, “A Novel Image Compression Based Method for Multispectral Fingerprint Biometric System”,Procedia Computer Science,Volume 171, 2020, pp 1698-1707, https://doi.org/10.1016/j.procs.2020.04.182.
[9] 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.
[10] Annu Sharma, Arvind Selwal, Shwetank Arya, “ Multispectral Image Fusion System Based on Wavelet Transformation for Secure Human Recognition”, International Journal of Advanced Science and Technology, vol 28(2019),pp811-820. http://sersc.org/journals/index.php/IJAST/article/view/2667
[11] 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.
[12] 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.
[13] M. Jaganmohan. (Mar. 3, 2022). Global Smart Grid Market Size by Region 2017–2023. [Online]. Available: https://www.statista.com/statistics/246154/global-smart-grid-marketsize-by-region/
[14] Z. Zheng, Y. Yang, X. Niu, H.-N. Dai, and Y. Zhou. (Sep. 30, 2021). Electricity Theft Detection, [Online]. Available: https://github.com/henryRDlab/ ElectricityTheftDetection .
[15] D. O. Dike, U. A. Obiora, E. C. Nwokorie, and B. C. Dike, ‘‘Minimizing household electricity theft in Nigeria using GSM based prepaid meter,’’ Amer. J. Eng. Res., vol. 4, no. 1, pp. 59–69, 2015.
[16] P. Dhokane, M. Sanap, P. Anpat, J. Ghuge, and P. Talole, ‘‘Power theft detection & initimate energy meter information through SMS with auto power cut off,’’ Int. J. Current Res. Embedded Syst. VLSI Technol., vol. 2, no. 1, pp. 1–8, 2017.
[17] S. B. Yousaf, M. Jamil, M. Z. U. Rehman, A. Hassan, and S. O. G. Syed, ‘‘Prototype development to detect electric theft using PIC18F452 microcontroller,’’ Indian J. Sci. Technol., vol. 9, no. 46, pp. 1–5, Dec. 2016.
[18] https://www.kaggle.com/datasets/jayaprakashpondy/electricity-consumption-dataset.