Retraction

Retraction: An Efficient Power Theft Detection Using Mean-Shift Clustering and Deep Learning in Smart Grid (IOP Conf. Ser.: Mater. Sci. Eng. 983 012003)

Published 09 November 2021

This article has been retracted by IOP Publishing following an allegation that this article contains text overlap from multiple unreferenced sources [1, 2, 3]. IOP Publishing has investigated and agree the article constitutes plagiarism. IOP Publishing also expresses concern regarding a number of nonsensical phrases used in the article, which suggests the article may have been created at least partly by artificial intelligence or translation software.

If the authors wish to contest this retraction they are advised to contact researchintegrity@ioppublishing.org.

1. Jeyaranjani J, Devaraj D. "Machine Learning Algorithm for Efficient Power Theft Detection using Smart Meter Data" in International Journal of Engineering & Technology, vol. 7.3, pp. 900-904, 2018, https://www.sciencepubco.com/index.php/ijet/article/view/19585

2. Aswini. R. and Keerthiha. V., "IoT Based Smart Energy Theft Detection and Monitoring System for Smart Home," 2020 International Conference on System, Computation, Automation and Networking (ICSCAN), 2020, pp. 1-6, https://ieeexplore.ieee.org/document/9262411

3. Hasan M, Toma RN, Nahid AA, Islam MM and Kim JM (2019) “Electricity Theft Detection in Smart Grid Systems: A CNN-LSTM Based Approach” in Energies 2019, 12(17), 3310; https://doi.org/10.3390/en12173310

Retraction published: 09 November 2021
An Efficient Power Theft Detection Using Mean-Shift Clustering and Deep Learning in Smart Grid

G. Johncy¹, A.Anisha Felise²

¹ Assistant Professor, St.Xavier’s Catholic College of Engineering, Nagercoil, India.
² PG Student, St. Xavier’s Catholic College of Engineering, Nagercoil, India.

Abstract – Energy theft constitutes a major concern for the utility operators in these modern smart homes. The task of detecting and reducing the energy losses has been highly challenging due to insufficient inspection techniques. Energy distribution is comprised of Technical and Non-Technical Losses (NTL). Energy theft generates a major share of Non-Technical Losses which also prompts budgetary misfortunes for the service organizations. The data in the modern smart meters are transmitted in wireless mode. Therefore the smart homes are vulnerable to energy theft. Many new technologies have been adopted to resolve the issue of energy theft in Advance Metering Infrastructure (AMI) in Smart Grids. Consumption pattern must be derived to identify illegal energy consumers. Computational method has been derived to analyze and identify energy consumption patterns based on data mining techniques. Machine learning technique improves the got client energy utilization readings and guides them on contrasting irregularities in use. Deep Learning method as Convolution Neural Network is implemented on the activity of order methods on client energy utilization and illegal use of electricity and the amount of consumption by energy theft.

Keywords: Smart Grid, Non-Technical Loss, Mean shift Clustering, Convolutional Neural Network.

1. Introduction
The electrical grid is comprised of segments, for example, transmission lines, substation furthermore, transformers where the energy is delivered from the power plant to the individual or a business [5][6]. The modern Grid is provided with the digitized two way communication between the consumers and the energy suppliers. A substantial energy loss occurs during the energy distributions. The loss may be either Technical Loss or Non-Technical Loss. Technical Losses are caused by power dissipation in the component. Non-Technical Losses are more specifically due to the energy theft. Generally, techniques for power utilization are i.e consuming all power from matrix legally; ii. Pilfering all power from network illegally; iii. Pilfering a bit of family unit power and expending the rest legitimately.[1][2]

Compared to the traditional meters which can be physically tampered, smart meters may not be tampered physically but the energy theft can happen due to cyber-attacks. An efficient way to resolve this issue to analyze the consumption pattern of the energy in the grid [4].

As of late, digital altering has become an approach to malignantly change the utilization information examples of the smart meter which leads to financial loss to the utility companies. Despite the fact that such power misfortune is brought about by few customers, it diminishes the benefit and energy
effectiveness of the utility organizations, making issues, for example, load-shedding, mechanical routine hampering, and expansion. “NTL costs a huge amount of money for both developed and developing countries such as USA, UK, Brazil, Malaysia, and India. For example, these losses in the power sectors of India and Brazil cost approximately US $44.5 and US $3 billion per year respectively” [6].

To mitigate this issue, a few procedures have been actualized through productive examination to recognize and subside the theft issue. Analysts arranged energy theft location frameworks into three essential gatherings, for example, state-based, game hypothesis based and order based identification. The step by step power usage examples of clients can be analyzed by utilizing AI calculations to set up characterization models, which incorporate choice trees, arbitrary backwoods (RF), uphold vector machines (SVM), neural organizations (NN, etc.[7][8].

The intermittent portrayal of intensity utilization since the consumption of power by the users is more huge in the field of ETD measure of information [9]. Due to the amount of data handled by the service organizations' AI based arrangement has accomplished huge consideration in fraud data detection. Day by day utilization of information is applied to distinguish theft pattern which protects the data protection of the shoppers. Means Shift algorithm was applied to identify peculiarities and anomalies in the gathered information by bunching and Convolutional Neural Network (CNN) is applied to classify the user profile [10].

In this paper, we propose another technique by consolidating a profound learning model with conventional AI models to detect the electricity theft. In particular, we utilize the Convolutional Neural Network (CNN) with Mean shift clustering model to accomplish better exactness and versatility. The strategy has two stages containing complete four stages. In the initial step, we collected the smart meter data, and then preprocess the crude information of records to eliminate insignificant elements. Also, Principal Component Analysis and Mean Shift clustering are applied to find the trustworthiness of customers. Thirdly, a CNN is prepared with these subsets to identify the Energy burglary. Finally, validation is performed on the testing subsets.

2.Related work
Power burglary location can be treated as unique case of anomaly identification or misrepresentation recognition. As an exceptional time series, electricity utilization records have pulled in a lot attention in the recent years. The everyday power utilization patterns of clients can be dissected by utilizing highlight based AI calculations to set up grouping models. Jindal et al. proposed [7] scheme dependent on decision tree and support vector machine is able to distinguish and find ongoing power theft at each level in power transmission and distribution. This plan lessens bogus positives generally and demonstrates its viability in the genuine situations. Results demonstrate that the plan distinguishes fake buyers with an exactness of 92.5% and a bogus positive rate as low as5.12%.

Depuru et al.[8] reported a neural organization (NN) model with a various leveled model for upgraded assessment of the order effectiveness and an encoding method can recognize illicit customers with better proficiency and quicker arrangement of information.

Zheng et al. [10] proposed that the irregular power use examples can be caught by wide and profound convolutional neural organization. Rajiv et al. proposed that Machine Learning calculation Gradient Boosting Theft Detector with include designing declines False Positive Rate and furthermore diminishes client information extra room just as preparing time [11].

Shuan et al.[12] reported a crossover convolutional neural organization irregular timberland model for
automatic power robbery identification shows that the location model beats other methods as far as exactness and effectiveness.

Nazmul et al.[13] proposed an electricity theft detection system is CNN-based LSTM model for keen lattice information grouping on the grounds that the force utilization mark is time-arrangement information and an information pre-handling calculation was executed to process the missing occasions in the dataset, in view of the neighborhood estimates comparative with the missing information point. The engineered minority over-examining procedure was actualized to produce new information focuses so as to address the class awkwardness issue. Furthermore, the tally of power robbery clients was moderately low, which might have made the model to distinguish burglary clients with great exactness.

Zheng et al. examine the fundamental structure of the onlooker meters and the smart meters and a connection based identification strategy utilizing MIC is given to measure the relationship between the altered burden profiles and the NTL. CFSFDP-based strategy improves the location precision and soundness[14]. Buzau et al. investigates the raw daily energy consumption history and incorporates non-consecutive information, for example, its contracted force or geological data. The results show that the data mining techniques to understand the problem. One strategy is the Maximum Information Coefficient (MIC), which can find the connections between the non-specialized misfortune (NTL) and a particular influence conduct of the buyer. MIC can be utilized to exactly recognize burglaries that seem ordinary in shapes [15].

Existing System
The past exploration takes a shot at electricity theft detection centers around the client power use profile information for theft identification. The specific are where there is dissimilarity in supplied power and billed power. All the customers belonging to that area are considered to be suspects. The drawback of the work discussed previously is that power theft identification has been carried out based on the assumption that the customers are suspected to be fraud. This case could identify the potential customer as the fraudulent customers. This provides the motivation for this research work to include the bogus customer power utilization information into the genuine force utilization information. The machine learning algorithms are used to analyze the data that literally cluster and deep learning algorithm is used classifies the Customer. The customer’s data are discriminated as genuine and fraud based on their usage pattern.

A. System Model
This clever thought of Smart Energy Theft System for the shrewd home is more proficient and dependable contrasted with past techniques. Because of a non-meddling strategy for information assortment, the energy observing framework was actualized in a genuine house in Singapore. The gathered information incorporates Time arrangement information power utilization from a non-controlled genuine house environment

Keen Homes are made through execution of Internet of Things (IoT) and smart meters. So as to screen and control the Advanced Metering Infrastructure (AMI), Energy Management System (EMS) was a basic reconciliation of the framework foundation. Request Side Management System (DSMS) is incorporated as an element of EMS. Its usefulness centers principally around dealing with the interest reaction and burdens. It gathers the interest data to direct the ideal force use, for example, executing load-moving to empower the utilization of power markets during top and off-top hours. It permits clients to advantageously direct their shrewd machines inside the home territory by utilizing cell phones. Further developed and created frameworks could additionally break down the information gathered and settle on its own choice for the keen homes to work in a smart and energy-effective manner.
technique dependent on clients' utilization designs. Energy burglary has become a difficult issue in the keen matrix network. It has caused monstrous misfortunes for some nations that surpass billions of dollars.

3. Proposed system

Electricity theft is the criminal act of taking power through unlawful strategies without paying for the devoured power tax which brings about income misfortune for the utility supplier. The proposed model is shown in Figure 1.1

![Figure 1.1. Proposed Model for Power theft identification](image)

A. Data Preprocessing

The collected Smart meter data, the customer profile data of a specific area is chosen. All the customers of the selected region are genuine profiled. The percentage of trustworthiness of the customers is identified by using machine learning algorithm. For this purpose, unsupervised Mean Shift clustering algorithm is applied. The clustering is performed to select the customer’s profile which is more genuine in their power usage pattern by grouping them in clusters. The customers are clustered based on their power usage readings obtained from their smart meter. For this purpose, a sample customer’s smart meter reading for every 30 minutes in (KWH) for 28 days is considered.

B. Principal Component Analysis

At First, it needs to perform mean standardization at highlight scaling, so the highlights, for example, power utilization have zero mean and should have a practically identical scope of qualities using

\[ \mu_j = \frac{1}{m} \sum_{i=1}^{m} X_j(i) \]

where \( X_j \) is the \( j \)th consumer data vector and the average electricity consumption of each consumer. Then we replace each \( (i) \) with \((i) - \mu_j\).

For datasets dimensionality reduction next step is to apply the technique principal component analysis which increasing interpretability but at the same time minimizing information loss. PCA obtains eigenvalues and eigenvectors which represent the characteristics and relationship of the data.
C. Mean Shift Clustering
Mean shift algorithm is a non-parametric grouping technique which doesn't need prior data on the amount of cluster, and doesn't compel the state of the cluster. For every data point, mean shift algorithm characterizes a window around it and registers the mean of the information point. By then it shift the point of convergence of the window to the mean and rehashes the figuring till it combines.

Mean shift ALGORITHM
Input: Bandwidth \( h \), profile function \( g(x) \), and data matrix \( X = [x_1, \ldots, x_n] \), \( n \geq 2 \)
Output: Estimated cluster.
begin
for \( i = 1 \) to \( n \) do
  Initialize: \( j \leftarrow 1; y_j \leftarrow x_i \)
The underlying center community \( y_1 \) is picked to be one of the watched information
repeat
  Evaluate the mean shift vector
  \[
  m(y_j) = \frac{\sum_{i=1}^{n} x_i g \left( \frac{1}{n} y_j - x_i \right)}{\sum_{i=1}^{n} g \left( \frac{1}{n} y_j - x_i \right)} - y_j
  \]
  Do the cluster center update
  \[
  y_{i+1} = m(y_j) + y_j
  \]
  Find the nearest data point
  \[
  y_i = \arg\min_{x} \left\{ x \in \{x_1, \ldots, x_n\} \mid \|y_{i+1} - x\| < \|y_j - x\| \right\}
  \]
  Until convergence happens,
  \[
  y_{i+1} = y_j \text{ for some } j \in N
  \]
Consolidation the assessed cluster communities that are nearer than \( h \) optional: Estimate cluster that pull in modest number of information focus end

Mean shift calculation ensure the assembly of the isolated arrangement. MS calculation appoints the processed MS Vector to the closet data point in the informational collection in every cycle. The quantity of emphases is exceptionally needy to the halting measures in the mean shift calculation however in proposed mean shift calculation we show that the quantity of cycles is constantly limited over the quantity of tests. In mean shift calculation, we need to consolidate the assessed group places that are nearer than the transmission capacity. We have to kill the assessed cluster community that draws in barely any number of focuses.

CNN Model
A CNN is a subclass of neural organization proposed, which is propelled by the working rule of utilizing the human visual cortex for object acknowledgment. CNNs were intended for recognizing objects, just as their classes, in a picture is appeared in Figure 1.2. CNN fluctuates from conventional AI calculations concerning incorporate element extraction, where CNN removes includes internationally through different stacked layers.
Figure 1.2 Convolutional Neural Networks

In general, CNN design involves a few convolution layers and pooling layers. These layers are trailed by one or more fully connected (FC) layers. The convolutional neural organization is the chief structure square of a Convolutional Neural Network. Convolution is a numerical activity that follows up on two sets of information.

\[ Y = X \ast F \rightarrow y[i] = \sum_{j=-\infty}^{+\infty} x[i-j]F[j] \]

On account of CNNs, the two arrangements of data are the info information and a convolution channel, which is additionally called the portion. The convolution activity is performed by sliding the bit over the entire data, which delivers an element map. Practically speaking, various channels are used to play out different convolutions to create particular component maps. These element maps are at long last consolidated to define the last yield from the convolution layer. Initiation capacities are utilized after the convolution activity to acquaint non-linearity with the model. Distinctive enactment capacities, for example, straight capacity, sigmoid, and tanh are utilized, yet the corrected direct unit (ReLU) was utilized in the proposed CNN since it can prepare the model quicker and guarantee close worldwide weight improvement. The ReLU initiation work is characterized as follows:

\[ f(x_i) = \max(0, x_i) \]

The pooling layer shows up close to the convolution layer. These layer down-examples each component guide to decrease their measurements, which thus diminishes over fitting and preparing time. The maximum pooling is generally utilized in CNNs which just chooses the most extreme incentive in the pooling window. The FC layer is fundamentally a completely associated fake neural organization. Basically, in a CNN, the convolution and pooling layers remove low-level highlights, for example, edges, lines, ears, eyes, and legs, and the completely associated layer plays out the order task dependent on these low-level highlights. The initiation work utilized in this last arrangement layer is normally a SoftMax work, which allocates a likelihood incentive to each class which includes to 1

4. Result and Discussion
The first step for the electricity theft detection is to input the consumer electricity data set. It contains the field like consumer no, date, area no, daily power consumption and hourly consumption. The customer profiles are assigned into four cluster 0, 1, 2 and 3. The usage pattern of finalized cluster head originated as the source for the anomaly customer detection. The customers profile which are close to the cluster head is identified as genuine profile. The genuine profiles are separated and their energy consumption data are taken for further investigation. The trustworthiness of the customer is verified by using the clustering algorithm is shown in Figure 1.3
The clustered result is given to the input of CNN classification. The CNN is built to classify the customer’s profile. The three types of bogus data along with the actual data are considered to train the network. After required number of iterations, the network is trained to predict any new customer profile to genuine or fraud shown in the Figure 1.4. From the 2000 data sample 695 is detected as fraud, but by using CNN the predicted data is 574. Thus this model produces Accuracy rate of 85.9%, detection rate of 83%, False positive rate is 5% and it shown that the time taken to execute is 5ms/steps. Recall is 0.897% model F1 is 0.84. For CNN model if we increase the dataset it will produce more accurate result.

5. Conclusion
This project proposes an efficient way of using deep learning algorithm to address the power theft problem. In shrewd framework information examination framework, it is important to know the continuous power utilization information to estimate the specific future interest of power and plan in like manner. The identification of power burglary will likewise expand its help for load gauging that allows the utilities to precisely foresee the force interest for future explicit to singular client. The information produced through this analytics, increment information on client utilization design and the requirement of power for the future

Acknowledgement
G Johncy was graduated in Computer Science and Engineering in 2005 from Anna University-Chennai, India. In 2008, she received M.E degree in Computer Science and Engineering from Anna
University- Chennai, India. She is in teaching profession for the past Twelve years. Currently she is working as an Assistant Professor in the Department of Computer Science at St. Xavier’s Catholic College of Engineering, India. Her areas of interest are Network Security, Image Processing and Smart grid Security. She is a Life member of ISTE.

A.Anisha Felise was graduated in Computer Science and Engineering in 2018 from Anna University- Chennai, India. In 2020, she received M.E degree in Computer Science and Engineering from Anna University- Chennai, India.

REFERENCES

[1] Depuru SS, Wang L, Devabhaktuni V, Green RC 2013 High performance computing for detection of electricity theft, International Journal of Electrical Power & Energy Systems 7 21-30.
[2] Depuru SS, Wang L and Devabhaktuni V 2011 Electricity theft: Overview, issues, prevention and a smart meter based approach to control theft. Energy policy 39 1007-13.
[3] Li W, Logenthiran T, Phan VT and Woo WL 2019 A novel smart energy theft system (SETS) for IoT-based smart home. IEEE Internet of Things Journal 6 5531-9.
[4] Glauner P, Boechat A, Dolberg L, State R, Bettinger F, Rangoni Y and Duarte D 2016 Large-scale detection of non-technical losses in imbalanced data sets. IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT) 6 1-5.
[5] Depuru, Soma ShekaraSreenadh Reddy 2012 Modeling, detection, and prevention of electricity theft for enhanced performance and security of power grid. The University of Toledo.
[6] Prasad J, Samikannu R 2017 Overview, issues and prevention of energy theft in smart grids and virtual power plants in Indian context. Energy Policy 110 365-74.
[7] Jindal A, Dua A, Kaur K, Singh M, Kumar N and Missura S 2016 Decision tree and SVM-based data analytics for theft detection in smart grid. IEEE Transactions on Industrial Informatics 12 1005-16.
[8] Depuru SS, Wang L, Devabhaktuni V and Nelapati P 2011 A hybrid neural network model and encoding technique for enhanced classification of energy consumption data. In2011 IEEE Power and Energy Society General Meeting 24 1-8.
[9] Giani A, Bitar E, Garcia M, McQueen M, Khargonekar P and Poola K. Smart grid data integrity attacks: characterizations and countermeasures. In2011 IEEE International Conference on Smart Grid Communications (smartGridComm) 17 232-237.
[10] ZhengZ, YangY, NiuX, DaiHN and ZhouY 2017 Wide and deep convolutional neural networks for electricity-theft detection to secure smart grids. IEEE Transactions on Industrial Informatics 14 1606-15.
[11] Punmiya R, Choe S 2019 Energy theft detection using gradient boosting theft detector with feature engineering-based preprocessing IEEE Transactions on Smart Grid 10 2326-9.
[12] Li S, Han Y, Yao X, Yingchen S, Wang J and Zhao Q 2019 Electricity Theft Detection in Power Grids with Deep Learning and Random Forests Journal of Electrical and Computer Engineering Oct 1 2019.
[13] Hasan M, Toma RN, Nahid AA, Islam MM and Kim JM 2019 Electricity theft detection in smart grid systems: a CNN-LSTM based approach Energies 12 3310.
[14] Zheng K, Chen Q, Wang Y, Kang C and Xia Q 2018 A novel combined data-driven approach for electricity theft detection. IEEE Transactions on Industrial Informatics 15 1809-19.
[15] Buzau MM, Tejedor-Aguilera J, Cruz-Romero P and Gomez-Exposito A 2019 Hybrid deep neural networks for detection of non-technical losses in electricity smart meters. IEEE Transactions on Power Systems 35 1254-63.
