JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 2368 - 2372
https://publishoa.com
ISSN: 1309-3452

Use of Deep Learning and Random Forest Algorithms to Predict Electricity Theft in Power Grids

Dr. T.K.Shaik Shavali [1], Mohammed Khateeb Ahmed [2], Syed Fahad [3], Mohammed Ashfaq Zaheer [4]

[1] Professor, Department of Computer Science and Engineering, Lords Institute of Engineering and Technology, Hyderabad, Telangana, India.
[2, 3, 4] Research Scholar, Department of Computer Science and Engineering, Lords Institute of Engineering and Technology, Hyderabad, Telangana, India.

Email : *[1] drskshavali@gmail.com*

ABSTRACT

There are several factors that contribute to distributing infrastructure easily understood losses (NTLs), but since it has a substantial impact on electricity grid performance and value is electricity thefts. Convolutional neural networks (CNNs) and random forests (RFs) are used in this article to support utilities firms address the issues of ineffective power inspections and unpredictable power usage. Convolution and down sampling are used in this approach to teach a convolutional neural network (CNN) [14] how to distinguish between the various times of day and days of the week in huge and constantly changing data from smart metres. To avoid over fitting, an absent / if these factors is introduced, and the training algorithm technique is used to adjust network parameters during training. Based on these attributes, a random forest (RF) is utilized [7] to detect if an electrical thief is lurking in the house. RF parameters for the hybrid model are found using the grid search engine. It is concluded that the suggested classification algorithm surpasses existing techniques in terms of effectiveness and precision by carrying out experiments on real datasets.
Index Terms— Electrical theft, power grid, CNN, Random Forest algorithm.

Index Terms— Electrical theft, power grid, CNN, Random Forest algorithm.

I. INTRODUCTION

Power providers around the world confront a major challenge in reducing the amount of energy lost in electricity transmission lines. Transmission losses (TLs) and nonprofessional losses (NTLs) are sometimes associated with energy losses. The transmission lines and transformers, as well as other parts of the power system, are responsible for the TL, which is the change in the total loses and TLs. The NTL, on the other hand, is mostly due to electricity theft. Actually, the majority of fraudulent activity is committed physically, such as by tapping into power lines or cutting into metres to alter the readings [3]. Electricity companies may suffer income reductions as a result of these fraudulent activities involving electricity. Energy theft costs roughly $4.5 billion a year in US alone [4]. Electricity theft is expected to cost utilities throughout the world $20 billion a year [5]. Electricity theft behaviour might potentially have an impact on the power system's safety. For example, if the electrical systems are overloaded due to electricity theft, fires could result. Therefore, accurate detection of electricity theft is critical to the safety and stability of the power grid.

AMI in smart grids allows us to collect enormous volumes of dataset obtained at a particular intensity through smart appliances, which is useful in detecting electricity theft [6, 7]. Although the AMI network has its advantages, it also has its disadvantages. Various digital techniques and cyber attacks can be used to carry out these attacks on the AMI. Human examination of unauthorised line diversion, comparison of malicious metre records with benign records, and examination of faulty equipment or hardware are the key methods of detecting power theft. When all metres in a system are verified, these procedures are exceedingly time-consuming and expensive. In addition, these manual methods are vulnerable to cyber attacks. Many ways have been proposed in recent years to address the issues raised above. There are three basic categories of these methods: state-based, game theory-based, and artificial intelligence-based [8].

Wireless sensors and power conditioning [12] are crucial components of state-based detection [9–11]. However, these approaches could identify electricity theft, but they require real-time acquisition of network topologies and extra physical measures, which may not be completely possible, It is possible to deduce distinct distribution of regular and delinquent behaviour from the gaming equilibrium using game-based detection algorithms. According to [13], they are able to reduce energy theft at a fair cost. Though it's still a problem to figure out the value each actor (such as suppliers, authorities, and fraudsters) provides, some examples of artificial intelligence includes ml algorithms. In [14] machine learning algorithms are further divided into categorization and grouping models. In spite of being cutting-edge and noteworthy, the machine learning detection algorithms described above aren't good enough for practical use. Due to their limited capacity for responsible for considerable, most of these techniques necessitate manual feature extraction. Aside from the confidence interval and greatest and lowest values of consumption statistics included in traditional hand-designed features Smart metre data's 2D characteristics cannot be extracted manually because the process is complicated and time consuming. In [7], a comparison between different neural network architectures is shown, including deep neural networks (CNNs), long-short-term memory (LSTM) recursive neural networking (RNNs), and stackable auto - encoder. Synthetic data is used to study the detecting' performances, but this makes it impossible to compare the detectors' performance to shallow systems. A deep neural network-based (DNN) customer-specific detector described by the researchers in [9] can effectively block such cyber attacks. Valuable and discriminatory features can be generated from raw data using the CNN in recent years [12]. In order to detect electricity theft, CNN utilised to extract features from high-resolution smart metre data. [13] created and used a large and deep convolutional neural network (CNN) approach for the study of smart grid power thefts.

The back propagation algorithm is used to train the max - pooling classification layer in a basic CNN, just as it is in a conventional single convolutional back propagation neural networking (SLFN) [2]. SLFN's generalisation performance will degrade if it is over trained during the back propagation procedure, because of this. The back propagation algorithm, on the other hand, is based on empirical organisational resilience, which is appropriate to cultural minima of training mistakes. As previously stated, the CNN is not really the best choice for classification due to the limitation of the softmax function, but it has showed remarkable promise in the extracting the features. As a result, it's critical to identify a classifier that not only matches the soft - max classifiers in terms of competence but also makes full advantage of the newly acquired capabilities. Packing and various feature selections are two effective machine learning approaches that can overcome the softmax classifier's limitations in most classifiers. A new convolutional neural network-random forest (CNN-RF) algorithm for detecting electricity theft is derived from these studies. As a critical aspect in the efficiency of the power theft classification algorithm, a CNN is configured to automatically extract numerous capacities of consumers' consumption pattern from smart metre data. The RF is utilised as a replacement for the softmax classifier in order to increase the detection performance. All of the consumers of the electrical provider in Iceland and Londoners have provided data for this algorithm to be received training on.

II. RELATED WORK

Electrical theft: overview, challenges, and a smart meter-based solution to preventing theft are all addressed in this report

Utility providers in poor nations have had a tough time catching those responsible for theft due to non-technical loss (NTL) in the transmission of electrical power. NTL is largely made up of cases of electricity theft. As a result of these outages, customers are forced to pay higher prices and the generating station has to work harder. The reasons why people steal electricity are examined in this research. A variety of strategies for detecting and estimating the theft are presented in light of these negative consequences. A smart metre, external monitoring stations, harmonics generation, and filtering circuits are all proposed [5-12] in this work. The goal of this project is to eliminate unlawful energy users while also conserving and properly utilising available resources. Smart metres can also provide information on a wide range of other characteristics relating to current power consumption on-the-fly. In order to calculate NTL in the distribution system, an external control station uses data from the distributing feeder's sending end data. [11] Harmful harmonic generators are activated at a particular feeder when NTL levels are high enough, so that unlawful consumers' appliances can be destroyed. An example of a cost-benefit analysis is provided for the proposed system's adoption in India.

Analysis of nontechnical power utility loss using a machine-learning algorithm

Using the cutting-edge computational method known as extreme learning machine, the authors in this work provide a new way to utilities' nontechnical losses (NTL) analysis (ELM). In both advanced and emerging economies, nontechnical losses account for a large percentage of electricity losses. With this ELM-based technique, anomalous behaviour that has been linked with NTL activities can be discovered by looking at client load profiles. It is possible to mine past kWh consumption data for processes involved in the consumer behaviour using this methodology [13], which uses data mined. Classifier categories are derived from the data, and these can be used to determine whether or not any new, noteworthy behaviour are the consequence of consumption irregularities. Classification performance and accuracy are both increased by using ELM and asynchronous sequential-ELM (OS-ELM) methods. ELM's effectiveness and efficiency in NTL analysis are found to be superior to other different classifiers, like SVM method.

Problems with smart grid security and privacy

We are on the cusp of the biggest technological shift in global electrical systems since it became possible to have electricity in the home. Rather of relying on an outdated infrastructure to distribute electricity to our households and businesses, a new technology known as the smart grid is being implemented. Grid modernisation allows customers and utilities to better track, manage and predict their energy usage and costs with this technology.

III. METHODOLOGY

Comparative tests depending on hand - crafted features with no deep learning methods such as SVM, RF, gradient boosting decision tree (GBDT), and logistic regression are carried out to evaluate the accuracy of the CNN-RF model for accuracy (LR). CNN features with an SVM classifier (CNN-SVM) and CNN characteristics with a GBDT (CNN-GBDT) classifier were used to compare their classification performance to those achieved in the prior classification study. Here, the outcomes of the following 5 methods are shown, and then the techniques are evaluated:

Logistic regression: sigmoid activation function is a fundamental model in classification algorithm that is similar to a single layer of neural network Sigmoid functions uses linear regression to derive values ranging between 0 to 1. Then, any value greater than 0.5 would be considered a normal pattern, and any value less than 0.5 would be considered abnormal.

Support vector machine: To solve a nonlinear separable issue, this classifier uses kernel functions to translate the data into feature space and then determine an optimal separate hyper plane for each feature. It is used to anticipate consumer behaviour depending on the handmade characteristics stated above.

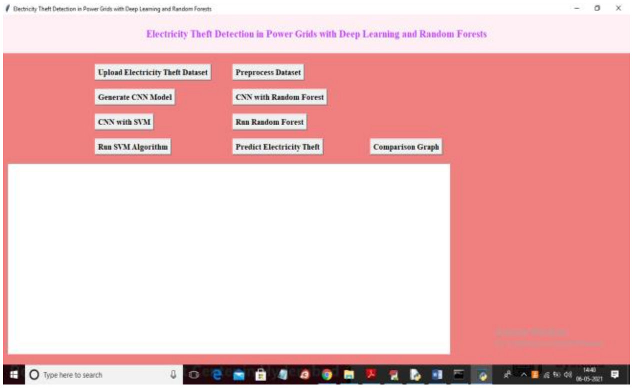
Random forest: When over fitting is effectively controlled, it is possible to get higher performance by combining many decision trees than by using a single one. Additionally, the RF classifier is able to deal with large-dimensional datasets while maintaining a high level of computational efficiency.

Gradient boosting decision tree: Decision trees are used in an iterative process that uses many trees. Random forests and the GBDT both use majority voting for ultimate output, but the GBDT accumulates all results or weights.

Deep learning methods: CNN, CNN-GBDT, and CNN-SVM all use softmax in the final layer, whereas the suggested technique does not.

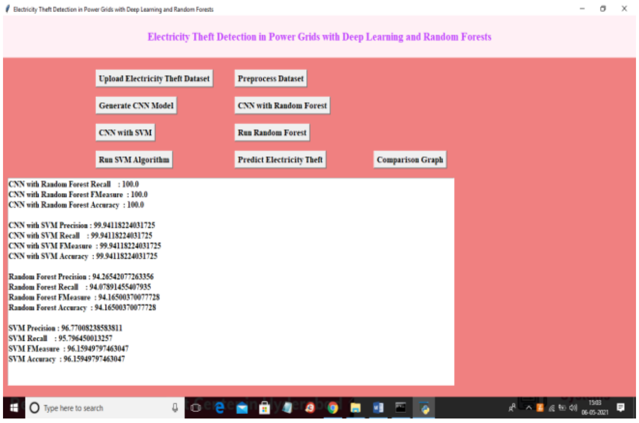
IV. RESULT AND DISCUSSION

Run the project to get below screen


Electricity Theft Detection in Power Grids with Deep Learning and Random Forests
Electricity Theft Detection in Power Grids with Deep Learning and Random Forests
Upload Electricity Theft Dataset
Preprocess Dataset
Generate CNN Model
CNN with Random Forest
CNN with SVM
Run Random Forest
Run SVM Algorithm
Predict Electricity Theft
Comparison Graph
Type here to search

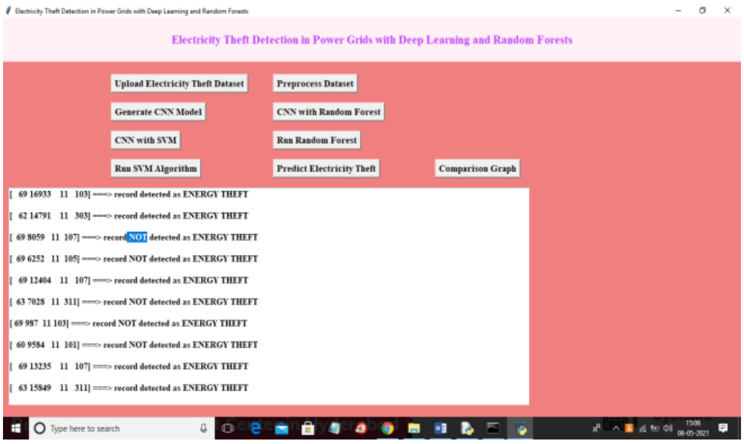
Then you can upload the Electricity Theft Dataset.

The following result displays the accuracy, recall, and Fmeasure for each algorithm.


Electricity Theft Detection in Power Grids with Deep Learning and Random Forests
Electricity Theft Detection in Power Grids with Deep Learning and Random Forests
Upload Electricity Theft Dataset
Preprocess Dataset
Generate CNN Model
CNN with Random Forest
CNN with SVM
Run Random Forest
Run SVM Algorithm
Predict Electricity Theft
Comparison Graph
Type here to search

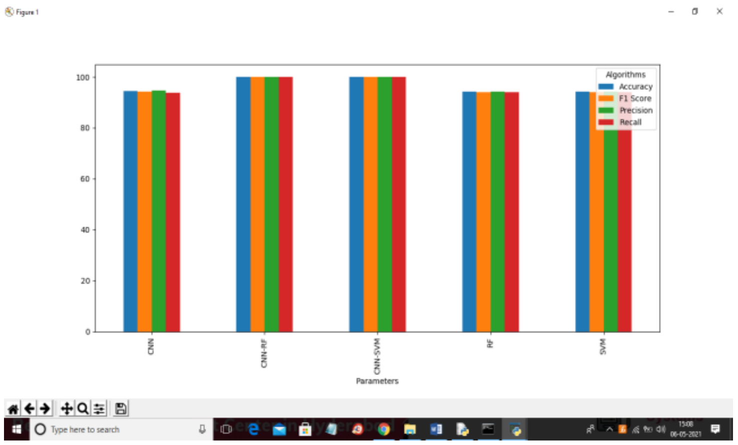
```
CNN with Random Forest Recall    : 100.0
CNN with Random Forest FMeasure  : 100.0
CNN with Random Forest Accuracy  : 100.0

CNN with SVM Precision : 99.94118224031725
CNN with SVM Recall    : 99.94118224031725
CNN with SVM FMeasure  : 99.94118224031725
CNN with SVM Accuracy  : 99.94118224031725

Random Forest Precision : 94.26542077263356
Random Forest Recall    : 94.07891455407935
Random Forest FMeasure  : 94.16500370077728
Random Forest Accuracy  : 94.16500370077728

SVM Precision : 96.77008238583811
SVM Recall    : 95.796450013257
SVM FMeasure  : 96.15949797463047
SVM Accuracy  : 96.15949797463047
```

Now select the 'test.csv' file and afterwards click on the 'Open' tab in order to load the data and receive the following prediction performance.


Electricity Theft Detection in Power Grids with Deep Learning and Random Forests
Electricity Theft Detection in Power Grids with Deep Learning and Random Forests
Upload Electricity Theft Dataset
Preprocess Dataset
Generate CNN Model
CNN with Random Forest
CNN with SVM
Run Random Forest
Run SVM Algorithm
Predict Electricity Theft
Comparison Graph
Type here to search

```
[  69 16933   11   103] ===> record detected as ENERGY THEFT
[  62 14791   11   303] ===> record detected as ENERGY THEFT
[ 69 8059   11  107] ===> record NOT detected as ENERGY THEFT
[ 69 6252   11  105] ===> record NOT detected as ENERGY THEFT
[  69 12404   11   107] ===> record detected as ENERGY THEFT
[ 63 7028   11  311] ===> record NOT detected as ENERGY THEFT
[69 987  11 103] ===> record NOT detected as ENERGY THEFT
[ 60 9584   11  101] ===> record NOT detected as ENERGY THEFT
[  69 13235   11   107] ===> record detected as ENERGY THEFT
[  63 15849   11   311] ===> record detected as ENERGY THEFT
```

Test results are shown in square brackets in the above outcome, and the estimated values are shown in square brackets after the quote marks. To see the following graph



Algorithm names are shown on the x-axis; precision, recall, FSCORE, and accuracy are shown on the y-axis. CNN-RF is 100 percent accurate across all algorithms.

V. CONCLUSION

To identify electricity theft, a new CNN-RF model is described in this study. These smart metre data are examined by the CNN, which serves as an automated analysis tool. In order to reduce the likelihood of over fitting, a dropout's percentage of 0.4 is used in the training phase for a fully connected layer. Data imbalance is addressed by utilising the SMOT method as well. SVM, RF, GBDT, and LR are some of the machine learning and data mining learning techniques that have been performed to same challenge as benchmarks. Two characteristics of the CNN-RF model suggest that it is a promising classification tool for detecting electricity theft: One advantage of the hybrid model is that it can automatically variables with the help, whereas traditional classifiers rely on retrieving good hand-designed features, which is a time-consuming and hard process. As both RF and CNN are widely used and successful classifiers for power theft detecting, a hybrid model incorporates the benefits of both.

Future research will examine how the precision and durability of intelligent power information affect the security of users because the identification of fraudulent activity affects this privacy. Using the hybrid CNN-RF model for additional purposes (such as load forecasting) is an idea worth exploring.

REFERENCES

1. S. S. S. R. Depuru, L. Wang, and V. Devabhaktuni, "Electricity theft: overview, issues, prevention and a smart meter based approach to control theft," Energy Policy, vol. 39, no. 2, pp. 1007–1015, 2011.View at: Publisher Site | Google Scholar

2. J. P. Navani, N. K. Sharma, and S. Sapra, "Technical and non-technical losses in power system and its economic consequence in Indian economy," International Journal of Electronics and Computer Science Engineering, vol. 1, no. 2, pp. 757–761, 2012.View at: Google Scholar

3. S. McLaughlin, B. Holbert, A. Fawaz, R. Berthier, and S. Zonouz, “A multi-sensor energy theft detection framework for advanced metering infrastructures,” IEEE Journal on Selected Areas in Communications, vol. 31, no. 7, pp. 1319–1330, 2013.View at: Publisher Site | Google Scholar

4. P. McDaniel and S. McLaughlin, “Security and privacy challenges in the smart grid,” IEEE Security & Privacy Magazine, vol. 7, no. 3, pp. 75–77, 2009.View at: Publisher Site | Google Scholar

5. T. B. Smith, “Electricity theft: a comparative analysis,” Energy Policy, vol. 32, no. 1, pp. 2067–2076, 2004.View at: Publisher Site | Google Scholar

6. J. I. Guerrero, C. León, I. Monedero, F. Biscarri, and J. Biscarri, “Improving knowledge-based systems with statistical techniques, text mining, and neural networks for non-technical loss detection,” Knowledge-Based Systems, vol. 71, no. 4, pp. 376–388, 2014.View at: Publisher Site | Google Scholar

7. C. C. O. Ramos, A. N. Souza, G. Chiachia, A. X. Falcão, and J. P. Papa, “A novel algorithm for feature selection using harmony search and its application for non-technical losses detection,” Computers & Electrical Engineering, vol. 37, no. 6, pp. 886–894, 2011.View at: Publisher Site | Google Scholar

8. P. Glauner, J. A. Meira, P. Valtchev, R. State, and F. Bettinger, “The challenge of non-technical loss detection using artificial intelligence: a surveyficial intelligence: a survey,” International Journal of Computational Intelligence Systems, vol. 10, no. 1, pp. 760–775, 2017.View at: Publisher Site | Google Scholar

9. S.-C. Huang, Y.-L. Lo, and C.-N. Lu, “Non-technical loss detection using state estimation and analysis of variance,” IEEE Transactions on Power Systems, vol. 28, no. 3, pp. 2959–2966, 2013.View at: Publisher Site | Google Scholar

10. O. Rahmati, H. R. Pourghasemi, and A. M. Melesse, “Application of GIS-based data driven random forest and maximum entropy models for groundwater potential mapping: a case study at Mehran region, Iran,” CATENA, vol. 137, pp. 360–372, 2016.View at: Publisher Site | Google Scholar

11. N. Edison, A. C. Aranha, and J. Coelho, “Probabilistic methodology for technical and non-technical losses estimation in distribution system,” Electric Power Systems Research, vol. 97, no. 11, pp. 93–99, 2013.View at: Publisher Site | Google Scholar

12. J. B. Leite and J. R. S. Mantovani, “Detecting and locating non-technical losses in modern distribution networks,” IEEE Transactions on Smart Grid, vol. 9, no. 2, pp. 1023–1032, 2018.View at: Publisher Site | Google Scholar

13. S. Amin, G. A. Schwartz, A. A. Cardenas, and S. S. Sastry, “Game-theoretic models of electricity theft detection in smart utility networks: providing new capabilities with advanced metering infrastructure,” IEEE Control Systems Magazine, vol. 35, no. 1, pp. 66–81, 2015.View at: Google Scholar

14. A. H. Nizar, Z. Y. Dong, Y. Wang, and A. N. Souza, “Power utility nontechnical loss analysis with extreme learning machine method,” IEEE Transactions on Power Systems, vol. 23, no. 3, pp. 946–955, 2008.View at: Publisher Site | Google Scholar

15. J. Nagi, K. S. Yap, S. K. Tiong, S. K. Ahmed, and M. Mohamad, “Nontechnical loss detection for metered customers in power utility using support vector machines,” IEEE Transactions on Power Delivery, vol. 25, no. 2, pp. 1162–1171, 2010.