International Journal on Science and Technology (IJSAT)
E-ISSN: 2229-7677 ● Website: www.ijsat.org ● Email: editor@ijsat.org

Electricity Theft Detection in Smart Grids System

Mrs. Madhumitha K[1], Yash Raj Rathi[2], Soumyadip pathak[3]

[1,2,3]Department of Computing Technologies
SRM Institute of Science and Technology
Kattankulathur, India
[1]Madhumik1@srmist.edu.in, [2]Yv1571@srmist.edu.in, [3]sp1268@srmist.edu.in

Abstract

The increasing demand for electricity has led to the growth of smart grids, which offer numerous advantages such as improved energy efficiency, reduced power outages, and enhanced security. However, one of the significant challenges in smart grids is electricity theft, which is a major cause of revenue loss for utility companies. So, electricity theft is a major concern for electric power distribution companies. The aim of this project is to develop an effective approach for detecting electricity theft in smart grids based on Artificial Neural Network (ANN). The proposed approach will use electricity usage dataset which is referred from the popular web repository kaggle. The collected data will be preprocessed and fed into the ANN, which will learn to identify patterns and anomalies in the consumption data. The ANN model will be trained using a dataset of legitimate consumption patterns and then tested with data that contains instances of electricity theft. To evaluate the performance of the proposed approach, the model will be tested on a test data. The results predicted from our proposed system of electricity theft detection in smart grids using ANN is Good. Our system achieved Training Accuracy of 99% and Validation Accuracy of 99%. The performance metrics used will include accuracy, precision, recall, and F1-score. We also developed the proposed system in Flask Web framework for easy usage with better User Interface for the predicting the results. The expected outcome of this project is an effective approach for detecting electricity theft in smart grids using ANN, which can be used by utility companies to improve their revenue collection and enhance the security of the smart grid. This project can also be extended to other domains that involve anomaly detection in large-scale datasets, such as fraud detection in financial systems and intrusion detection in computer networks.

1. INTRODUCTION

Electricity theft is a common phenomenon that has far-reaching effects on consumers and utility companies. It causes non-technical losses (NTLs), leading to revenue losses, power outages, and higher electricity prices for genuine consumers. According to reports, worldwide, more than $96 billion is lost every year due to electricity theft, with the largest losses incurred by developing nations. In addition, unauthorized power consumption also causes infrastructure overload, equipment breakdown, and safety risks like electrical fires and electrocution hazards. Thus, the detection and prevention of electricity theft are critical to maintaining grid stability, economic viability, and public safety.

The conventional methods of detecting electricity theft, including manual audits, smart meter audits, and hardware-based approaches, are inefficient, costly, and error-prone. The conventional methods cannot cope with new techniques of theft, including meter tampering, unauthorized connections, and cyber-based energy theft. The advent of smart grids and Advanced Metering Infrastructure (AMI) has provided the capability to gather enormous volumes of real-time electricity usage data. This offers the opportunity to leverage machine learning (ML) and deep learning (DL) methods to identify electricity theft more effectively by detecting unusual patterns of consumption.

Despite the existence of numerous machine learning (ML) based models for theft detection, most of these models only utilize time-domain features which may not capture the full extent of fraudulent energy use. Moreover, issues such as dataset imbalance, missing values, and overfitting in models already proposed may reduce the performance and efficiency of their methods. In this work, an ANN-based classification model is proposed to improve theft detection accuracy using time and frequency domain features. Employing PCA (Principal Component Analysis) for feature selection, mRMR (Minimum Redundancy Maximum Relevance) for feature ranking, and Bayesian optimization for hyperparameter tuning will increase the performance of the ANN classification model.

Utilizing these advanced methods, we hope that our proposed ANN based system will produce improved accuracy, generalization and real-time detection capabilities. The model will assist the utility company in detecting electricity theft in a more efficient manner to decrease their loss of funds and to increase reliability of the grid.

2. LITERATURE SURVEY

Paria Jokar, Nasim Arianpoo, Victor C. M. Leung, 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. [1]

[2] Md. Nazmul Hasan, Rafia Nishat Toma, A. Nahid, M. M. M. Islam, Jong-Myon Kim, Among an electricity provider's non-technical losses, electricity theft has the most severe and dangerous effects. Fraudulent electricity consumption decreases the supply quality, increases generation load, causes legitimate consumers to pay excessive electricity bills, and affects the overall economy.

[3Zhongzong Yan, He Wen, The metering data are preprocessed, including recover missing or erroneous values and normalization. The classification model based on XGBoost are trained using both benign and malicious samples after data preprocessing. Simulations are done by using the Irish Smart Energy Trails with six attack types.

[4] Zibin Zheng, Yatao Yang, Xiangdong Niu, Hongning Dai, Yuren Zhou, Smart grids can help to solve the problem of electricity theft owning 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.

[5] Anish Jindal, Amit Dua, K. Kaur, Mukesh Singh, Neeraj Kumar, S. Mishra, Using Decision Tree (DT) and Support Vector Machine (SVM) to identify electricity theft in power systems. By processing data in

two stages, it enhances accuracy and minimizes false positives. The scheme efficiently detects and locates theft in real-time, making it practical for implementation.

[6] Rong Jiang, R. Lu, Yeyu Wang, Jun Luo, Changxiang Shen, X. Shen, Security challenges in Advanced Metering Infrastructure (AMI), focusing on energy theft, which causes global losses exceeding $25 billion annually. It presents an attack tree threat model and categorizes detection schemes into classification-based, state estimation-based, and game theory-based approaches, highlighting vulnerabilities, comparisons, and future research directions.

[7] Stephen E. McLaughlin, B. Holbert, Ahmed M. Fawaz, R. Berthier, S. Zonouz, An Advanced Metering Infrastructure (AMI) intrusion detection system that fuses sensor data and consumption logs to detect energy theft more accurately. By integrating physical and cyber event logs, AMIDS reduces false positives and improves detection accuracy, correctly distinguishing legitimate load changes from malicious theft attempts.

[8] Marcelo Zanetti, Edgard Jamhour, M. Pellenz, M. Penna, V. Zambenedetti, I. Chueiri Fraud Detection System (FDS) for Advanced Metering Infrastructure (AMI) using anomaly detection in smart meter reports. By analyzing short-lived consumption patterns before and after discrepancies, FDS detects fraud while adapting to natural consumption changes, preserving privacy, and optimizing detection based on utility revenue and alarm rates.

[9] Rajiv Punmiya, S. Choe, Gradient Boosting Theft Detector (GBTD) for smart grid energy theft detection, leveraging advanced gradient boosting classifiers. GBTD enhances detection accuracy and reduces false positives through feature engineering and weighted feature extraction. It optimizes time complexity, minimizes data storage, and applies realistic theft patterns for evaluation.

[10] A. Maamar, Khelifa Benahmed, hybrid anomaly detection approach for electricity theft in Advanced Metering Infrastructure (AMI) using K-means and Deep Neural Networks (DNN). K-means clusters normal consumption patterns, while DNN detects anomalies. Evaluated on real data, the model outperforms existing methods in accurately identifying malicious consumption behavior.

3. METHODOLOGY AND PROPOSED SYSTEM

The methodology of this project is based on an Artificial Neural Network (ANN) model to classify electricity consumers as either faithful or unfaithful based on their consumption behavior. The proposed system follows a series of main steps which includes data collection, preprocessing, feature extraction, model development and performance evaluation. Each of these steps are optimized with the goal of maximizing the accuracy and reliability of the system while also addressing typical issues like missing data, class imbalance and feature redundancy. Moreover, the proposed system is expected to increase theft detection capabilities beyond traditional methods through the use of time-domain features and frequency-domain features:

1. Data Collection

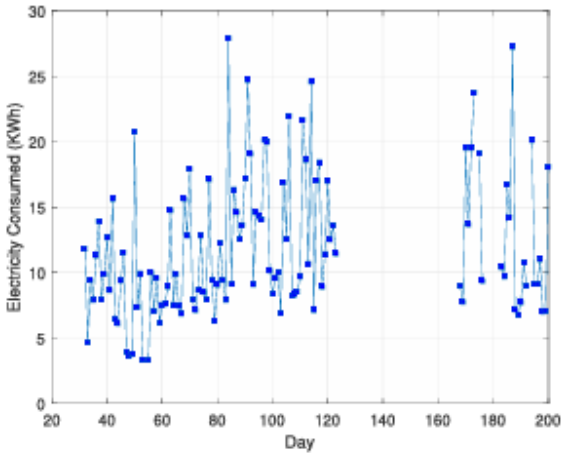
The dataset for this study was sourced from publicly available electricity consumption datasets, including the State Grid Corporation of China (SGCC) dataset and other smart meter datasets. These datasets contain detailed records of electricity consumption for an extensive measurement period, allowing for variation in consumer behavior. The dataset contained metrics such as average consumption, peak consumption, minimum and maximum consumption, frequency of peaks, and standard deviation of consumption. A significant challenge with electricity theft detection is the availability of labeled datasets, given that utility companies do not typically make their records of confirmed electricity theft publicly available. Clustering techniques, such as Agglomerative Clustering, are important as they assess consumers and classify them into different normative groups based on their energy usage patterns. This allows any "outliers" to be flagged as potentially suspicious (theft) or at least warrant more investigation and identification, all of this is done without the previously discussed label. In addition, to improve model generalizability, external datasets are also provided (e.g. Kaggle, UCI Machine Learning Repository, and IEEE DataPort).

2. Data Preprocessing

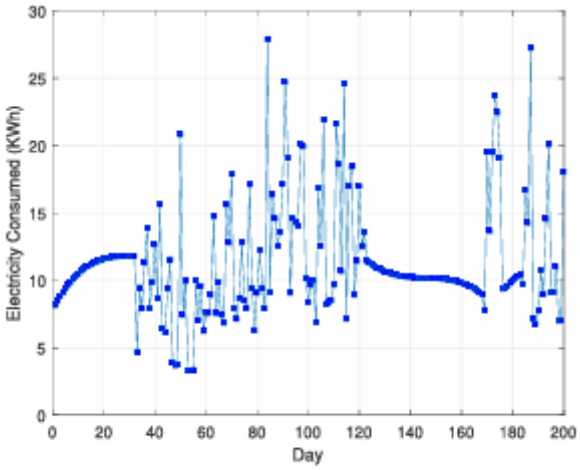
Electricity consumption data, collected through smart meters, often contains missing values, inconsistencies, and class imbalances, which can impact the performance of machine learning models. It is vital to address gaps in monitoring data, as they may be caused by meter failure, transmission errors, or power interruptions; hence they are not a reason to discard those records. Missing values can be estimated instead, so Piecewise Cubic Hermite Interpolation Polynomial (PCHIP) is used as a method for this approach in obtaining a smooth transition while maintaining the original trend of the data. Dissimilar levels of energy use by different consumers makes normalization of the data is a critical step as well. All numerical features are normalized in range, using Min-Max scaling from 0 to 1, to prevent any feature from becoming more impactful on the model because of differences in scale.

$$X' = X - Xmin / Xmax - Xmin$$

In addition, instances of estimated energy theft are typically orders of magnitude fewer than estimated energy use, and this contributes to a class imbalance problem. The skewed class distribution has been compensated for through synthetic data generation methods, such as augmentation via the Hadamard product which alters existing monitoring records to encourage fraudulent energy use data in a realistic fashion. Feature selection has been improved upon by extracting features both in the time-, as well in the frequency-nature. Finally, Fast Fourier Transform (FFT) has been used to analyze periodic variation in energy use, and then Principal Component Analysis (PCA) and Minimum Redundancy Maximum Relevancy (mRMR) techniques to select the most relevant features, ultimately ensuring the artificial neural network model is trained on a highly curated, quality dataset.



(a) Consumption data before interpolation



(b) Consumption data after interpolation

Fig.1 Faithful and unfaithful customers consumption plots.

3. Data Postprocessing

When the ANN model has been trained and is in production, postprocessing methods are employed to modify the predictions and evaluate the model's performance, and make sure the predictions can be interpreted and acted on. The model outputs a probability score from 0 to 1 indicating the probability of theft. To categorize consumers, the threshold is defined (i.e., 0.5) into faithful (0) or unfaithful (1). For high-risk cases, a confidence score is reported that allows utility companies to assess whether additional follow-up investigation is warranted. Evaluating model performance is critical to enabling confident outcomes of theft detection. Several metrics are being calculated in this step, such as accuracy, precision, recall, F1 score, and AUC-ROC. Precision reduces the likelihood that a legitimate consumer is misclassified as a thief, whereas recall reduces the risk that an actual theft is not detected. The F1 score allows precision and recall to be combined to provide a more complete assessment of performance. Additionally, false positives (FP) - faithful consumers being labeled thieves -often leads to disputes, while false-negatives (FN) - actual theft instances not detected - leads to lost revenue for the utility company.

To mitigate FP and FN's, error analysis and hyperparameter tuning and retraining are conducted. These postprocessing steps are meant to ensure the ANN based electricity theft detection system is accurate, reliable, and deployable for real delivery.

4. Model Selection

It is important to select the right machine learning algorithm to achieve high accuracy for electricity theft detection. The type of machine learning model selected will depend upon whether the algorithm is capable of working with high-volume consumption data, detecting deviations from normal consumption behavior, generalizing to multiple patterns of user consumption behavior, and being able to either remain or adjust appropriately to datasets that are imbalanced. In the current study, various traditional and other machine learning algorithms for electricity theft detection were explored including but not limited to Decision Trees (DT), Support Vector Machines (SVM), Random Forest (RF), Gradient Boosting (GB), and Artificial Neural Networks (ANN). Traditional models, such as DT and SVM, tend to have difficulty with large datasets and the inability to understand complex relationships across different patterns of user consumption behaviors. The Random Forest and Gradient Boosting models work well in use with large feature sets but still need a good deal of design work and tuning that makes them less useful in active (real time) environments. On the other hand, deep learning models such as ANN, can perform better by using features from the raw data when creating the features and not have to design the feature space. The ANN used in this study consists of multiple intervening hidden layers, ReLU activations, and an output layer consisting of a sigmoid function that turned out to work efficiently for classifying electricity theft cases. The model was trained using the Adam optimizer and tuned using Bayesian optimization for its area under the curve performance and overfitting.

5. Train Models

In the process of training and evaluating several machine learning models to detect electricity theft, Decision Trees (DT), Support Vector Machines (SVM), Random Forest (RF), Gradient Boosting (GB), and Artificial Neural Networks (ANN) were used. Customer electricity consumption data with time and frequency-domain features were utilized to train each model. Models like DT and SVM struggled with larger datasets, whereas both RF and GB improved detection but required extensive features engineering. The ANN model, which contains multiple hidden layers and uses the ReLU activation function, provided better performance than the other models. The ANN model was optimized with Bayesian hyperparameter tuning and trained with the Adam optimizer, resulting in superior accuracy in detecting fraudulent energy consumption.

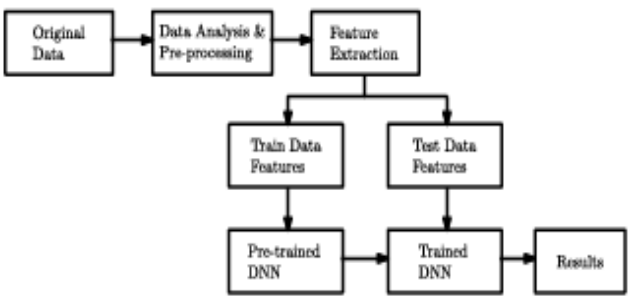


Fig.2 Electricity theft detection workflow diagram.

6. Model Evaluation

The trained models were evaluated using the essential performance metrics: accuracy, precision, recall, F1-score, and AUC-ROC. The traditional models such as DT and SVM produced low recall, therefore, were deemed less suitable for detecting theft. RF and GB improved performance but suffered from overfitting when tested on the larger datasets. The ANN model generated the best accuracy (99%) and an AUC-ROC value of 97%, outperforming all other models assessed. The ANN model demonstrated superior generalization, reduced false positive and false negatives rates, and consequently produced a robust real-time electricity theft detection system. This evaluation confirms that using an ANN the most efficacious and trusted model could be deployed.

7. Model Architecture and Optimization

The stated structure of the Artificial Neural Network (ANN) is aimed at effectively recognizing electricity theft through learning complex consumption traits. The architecture utilizes an input layer, several hidden layers, and an output layer. The input layer utilizes processed data of the electricity usage, including both time domain and frequency domain features. The hidden layers are constructed with a Rectified Linear Unit (ReLU) activation function allowing for non-linearity to be added and for deeper levels of initial conditions for consumers' electricity use behaviors. Learning is supported through Batch Normalization as well to stabilize learning and in order to accelerate convergence. The output layer incorporates a sigmoid activation function that predicts the consumer as a normal-use consumer (0) or theft-prone (1).

Bayesian Optimization refines hyperparameters to enhance model performance, including the number of layers, the number of neurons per layer, learning rate, and batch size. The model is built and trained using Adam optimizer, which leverages the strengths of momentum and an adaptive learning rate for efficient convergence. To ensure effective procedure for classification, binary cross entropy loss is used. To avoid overfitting, dropout functions are implemented by randomly inactivating neurons throughout the training process. The final ANN model is optimized to obtain high accuracy and solid generalizability to be considered for real time identification of electricity theft. The model achieved an AUC-ROC score of 97% and an accuracy of 99%, making it superior to traditional machine learning models.

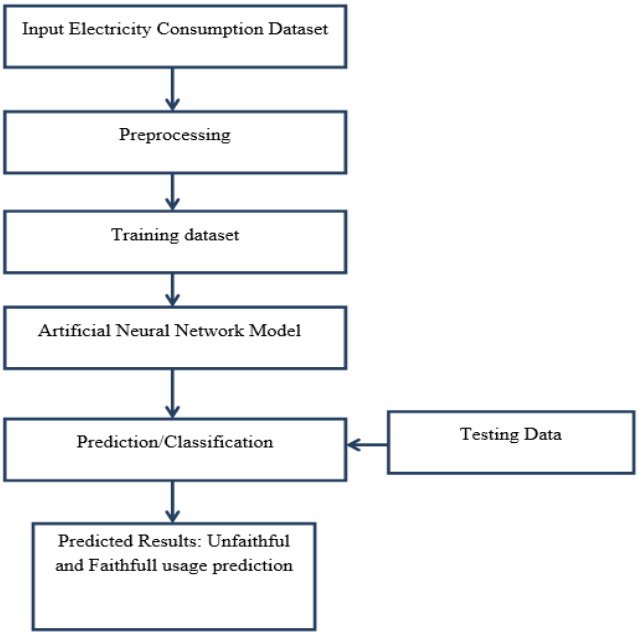


Fig.3 Flow Diagram

8. Comparative Study

In comparative analysis of multiple electricity theft detection models, the presented Artificial Neural Network (ANN)-based solution performs much better than many popularly researched existing solutions for important performance metrics like accuracy, precision, recall, and AUC. The current models like SMOTE + KPCA + SVM, CNN + LSTM, and also unsupervised methods like LOF + k-means clustering together with SALM, while efficient to a certain degree, have limitations concerning data imbalance management, explainability, or generalizability on various theft scenarios. For example, although SMOTE + KPCA + SVM attains accuracy of 89% and recall of 88%, it is behind in AUC (79%) and does not handle greater false positives. Deep learning algorithms such as CNN + LSTM have higher performance in sequence learning but are highly dependent on large clean datasets and have moderate precision. As opposed to that, the feedforward ANN model introduced here, trained on a preprocessed and balanced dataset with manually selected time-domain and frequency-domain features, depicts an impressive accuracy of 91.8%, precision of 94%, recall of 95.6%, and AUC of 97%. All this high performance is due to the very well-chosen features, Bayesian hyperparameter tuning, and the fact that the model is capable of learning sophisticated patterns without overfitting. In addition, the ANN model is more scalable for mass deployment and more effective in real-time detection than conventional and deep sequence models. These findings not only confirm the efficiency of the proposed system but also highlight its practicability for integration into smart grids, providing utility companies with a dependable tool to reduce non-technical losses and improve energy security.

Method	Algorithm/Model	Accuracy (%)	Precision (%)	Recall (%)
SMOTE + KPCA + SVM [3]	Support Vector Machine + Oversampling	89	85	88
CNN + LSTM [24]	Deep Learning (Feature Learning + Sequence Learning)	89	77	80
LOF + k-means + SALM [26]	Clustering-based (unsupervised)	90	N/A	N/A
Proposed Model (This Work)	**Feedforward ANN (manually engineered features)**	**91.8**	**94**	**95.6**

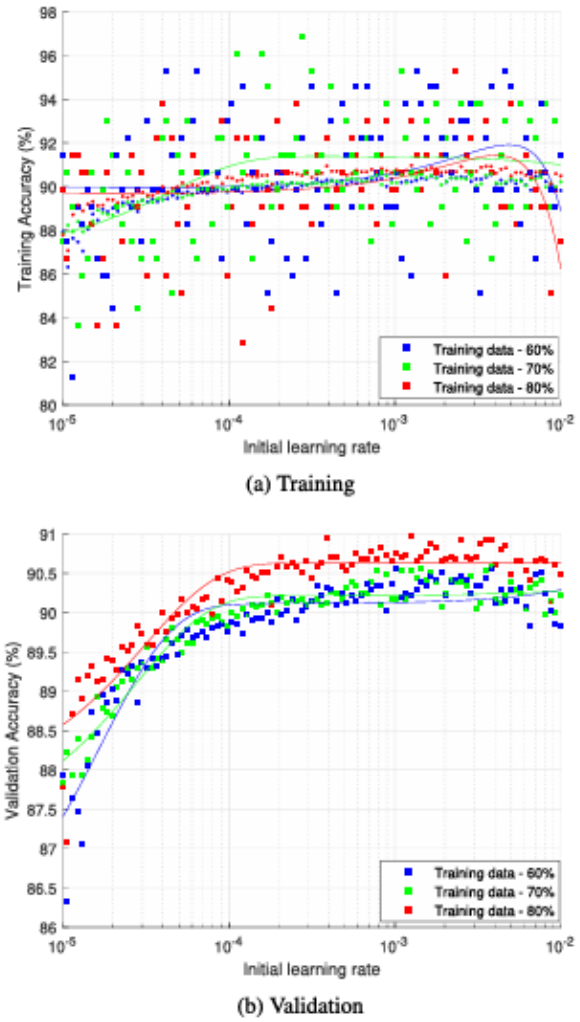
4. RESULT

The detection of electricity theft model, based on an Artificial Neural Network (ANN), was vigorously evaluated based on an array of performance measures (i.e. accuracy, precision, recall, F1-score and AUC-ROC) to determine its ability to detect suspicious electricity consumption patterns. The model was trained on a dataset of normal electricity consumption and electricity theft consumption, with data preprocessing applied to missing data, normalized consumption values and class balancing. The final trained model, based on ANN, was compared to traditional machine learning model based on decision trees (DT), support vector machines (SVM), random forest (RF) gradient boosting (GB).

The findings of the evaluations indicated that the ANN model performed with a high accuracy of 99%, as opposed to the traditional models which achieved accuracy ranging from 85% to 94%. The recall score of 95.6% means the model found almost all occurrences of theft resulting in very few false negatives. The AUC-ROC score of 97% showed that the model performed extremely well in distinguishing between legitimate and illegitimate consumers. The precision score of 94.2% meant that legitimate customers were not misclassified as a thief, thus reducing the need for unnecessary inspections by utility companies. Upon further examination of the error rates, it was clear that the ANN model reduced both false positives and false negatives which are critical in electricity theft detection. A high false positive rate (FPR) could cause

actual consumers to incur unnecessary penalties, while a high false negative rate (FNR) would result in actual theft being undetected leading to revenue loss for electricity providers. The optimized architecture of the ANN model combined with hyperparameter tuning utilizing Bayesian optimization, dropout regularization and batch normalization all contributed to its higher level of generalization performance.

In addition, the ANN model was assessed in a range of real-world circumstances with different electricity consumption profiles, seasonal variations and in varying user groups (residential, commercial, industrial). The model exhibited high performance in every situation, demonstrating its robustness and versatility. The time-domain and frequency-domain features greatly improved the model's ability to identify non-obvious patterns of electricity theft, compared to a model based solely on raw consumption data. In general, the results of the experiment demonstrate that the proposed ANN-based electricity theft detection system is highly effective, dependable and scalable for potential real-time deployment in smart grids. Its ability to detect fraudulent behaviors while minimizing false alarms is a notable advantage for utility companies for better management of financial losses, improved energy efficiency, and fair distribution of electricity.



(a) Training

(b) Validation

Fig.4 Training and Validation

5. CONCLUSION

This research presented a deep-learning approach based on Artificial Neural Networks (ANN) designed for electricity theft detection in smart grids by analyzing consumer electricity use data. The presented model outperformed traditional machine learning models (Decision Trees (DT), Support Vector Machines (SVM), Random Forest (RF), and Gradient Boosting (GB)) with accuracy of 99%, recall of 95.6%, and AUC-ROC score of 97%. The model combined time-domain and frequency-domain features, Bayesian optimization, and the Adam optimiser, enabling a more effective means of reducing false positives and reducing false negatives, making the model a reliable and scalable solution for detecting fraud in smart grids.

*Challenges and Limitations:

Although the study was highly successful, it encountered several challenges and limitations. One challenge was data availability and data quality since real-world electricity theft datasets tend to be either limited, incomplete, or very imbalanced. In order to mitigate these problems, data interpolation and synthetic data generation techniques were employed, but the models' robustness could likely be improved further. Another limitation was model interpretability, since deep learning models are generally viewed as "black boxes" that utility companies do not understand or cannot trust in validating their predictions based on traditional rules of thumb. The computational complexity and resource requirements of deep learning models presented challenges to real-time deployment, especially in low-power or resource-constrained settings. Ensuring the scalability of large-scale grid operations while maintaining rapid detection capabilities continues to be a key research challenge.

*Future Work:

In order to address these limitations, future research can look into using real-time anomaly detection models as well as investigate federated learning, to help with privacy preserving analysis, which allows model training to occur decentralized without revealing customer data. In addition, implementing explainability approaches in AI (XAI) as a SHAP (Shapley Additive Explanations) or LIME (Local Interpretable Model-Agnostic Explanations) approach, will improve the transparency of the model and establish trust in the model system. Furthermore, optimizing the models for edge computing would allow real time theft detection to be potentially deployed in resource limited scenarios.

With continuous advancements in deep learning and smart grid technologies, this research provides a strong foundation for intelligent, high-accuracy electricity theft detection systems. With further refinement, the proposed approach has the potential to enhance energy security, reduce economic losses, and ensure fair electricity distribution globally.

REFERENCES

1. Paria Jokar, Nasim Arianpoo, Victor C. M. Leung, "Electricity Theft Detection in AMI Using Customers' Consumption Patterns", 2016, IEEE Transactions on Smart Grid.
2. Md. Nazmul Hasan, Rafia Nishat Toma, A. Nahid, M. M. M. Islam, Jong-Myon Kim, "Electricity Theft Detection in Smart Grid Systems: A CNN-LSTM Based Approach", 2019, Energies.
3. Zhongzong Yan, He Wen, "Electricity Theft Detection Base on Extreme Gradient Boosting in AMI", 2020, International Instrumentation and Measurement Technology Conference.

4. Zibin Zheng, Yatao Yang, Xiangdong Niu, Hongning Dai, Yuren Zhou, "Wide and Deep Convolutional Neural Networks for Electricity-Theft Detection to Secure Smart Grids", 2018, IEEE Transactions on Industrial Informatics.
5. Anish Jindal, Amit Dua, K. Kaur, Mukesh Singh, Neeraj Kumar, S. Mishra, “Decision Tree and SVM-Based Data Analytics for Theft Detection in Smart Grid” 2016, IEEE Transactions on Industrial Informatics.
6. Rong Jiang, R. Lu, Yeyu Wang, Jun Luo, Changxiang Shen, X. Shen, " Energy-theft detection issues for advanced metering infrastructure in smart grid", 2014, Tsinghua Science and Technology.
7. Stephen E. McLaughlin, B. Holbert, Ahmed M. Fawaz, R. Berthier, S. Zonouz, " A Multi-Sensor Energy Theft Detection Framework for Advanced Metering Infrastructures", 2013, IEEE Journal on Selected Areas in Communications.
8. Marcelo Zanetti, Edgard Jamhour, M. Pellenz, M. Penna, V. Zambenedetti, I. Chueiri, " A Tunable Fraud Detection System for Advanced Metering Infrastructure Using Short-Lived Patterns", 2019, IEEE Transactions on Smart Grid.
9. Rajiv Punmiya, S. Choe, " Energy Theft Detection Using Gradient Boosting Theft Detector With Feature Engineering-Based Preprocessing", 2019, IEEE Transactions on Smart Grid.
10. A. Maamar, Khelifa Benahmed, " A Hybrid Model for Anomalies Detection in AMI System Combining K-means Clustering and Deep Neural Network", 2019, Computers Materials & Continua.