Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606
Volume 08 Issue 01
2024

Research Article
https://doi.org/10.56979/801/2024

Efficient Electricity Theft Detection Using Hybrid CNN-XGBoost Model

Sayed O. Madbouly[1*], and Hedi A. Guesmi[1]

[1]Department of Electrical Engineering, College of Engineering, Qassim University, Saudi Arabia
[*]Corresponding Author: Sayed O. Madboulyi Email: so.ossman@qu.edu.sa

Received: August 12, 2024 **Accepted:** September 12, 2024

Abstract: Non technical losses especially in distributed networks play key roles in electricity theft that pose serious challenges to power grids. As central electricity is distributed through the power grid to connect all consumers, any fraudulent usage is capable of interfering with the operations of the grid, produce low-quality supply, and even destroy the overall system. This means, as the data volume increases it becomes arduous to identify such fraudulent activities. Smart grids provide a solution in this aspect since electricity flow is bidirectional providing a channel for detection, correction and application of the corrective measures to the flow of the electrical data. Today's electricity theft detection techniques incorporate one-dimensional (1-D) electric data leading to maximum possible imprecision. This work proposes a model that integrates CNN and XGB known as CNN-XGB. To supplement 1-D theft detection framework, the proposed model includes both 1-D and 2-D power usage data. A comparison with existing benchmark methods, using experimental sample results, shows that the proposed model delivers accurate results for the task, which was the main objective of designing the model.

Keywords: Smart Grids; Energy; CNN; Data Mining; Electricity- embezzlement detection; Machine learning.

1. Introduction

Electricity is a necessity for many household and industrial uses, it is generated and supplied through large networks. These comprised land areas, which are often close to energy resources: Energy transformed into electricity that is transmitted through a network. Unfortunately, structures such as power grids have societies' vulnerabilities of fraudulent and electricity embezzlement as issues that affect grid quality, result in losses and draw unequal voltage. Electricity losses can be broadly classified into two categories: Technical losses (TLs), which losses are due to the physical characteristics infrastructure of a grid and Non Technical losses (NTLs), which are a result of interfer record flow irregularities such as bypassing or interfering with electricity meters with a view of producing incorrect readings. These false readings are in the form of unbilled revenues which impacts negatively on the economy. For instance, power losses resulting from electricity embezzlement in Canada being put at approximately $100 million yearly.

Electricity embezzlement can be detected using methods such as comparison of meter readings, reinstallation of meters or using configuration check; however these methods can hardly handle big data and even where they do, their accuracy is very low. These issues are solved by smart grids as they are a modern solution to all of them. These superior smart networks allow two way power provision flow and include edge computing smart sensors that monitor and gather information about supply flow, use and meter readings. Smart grids also incorporate features of self-healing, as well as detection-response features aimed at the identification of malicious data flow. Recent studies, such as [2], [3], and [5], highlight the effectiveness of smart grids in identifying electricity theft, yet existing models have notable limitations:

Inefficient results arising from analyses based on one-dimensional (1-D) data.

Employing the methods of artificial feature extraction.

Employment of conventional methods such as linear regression model (LR) and support vector machines (SVM), which give low results on average.

This work presents CNN-XGB, a new ensemble model aimed at enhancing electricity-embezzlement detection. One subset of the model is the convolutional neural network (CNN) which is combined with the extreme gradient boosting classifier. CNNs that has become popular for image classification are able to select important data features on their own. The CNN model comprises three layers: input, hidden, and output. For the purpose of increasing the level of accuracy the XGBoost is used on the final layer of the CNN model that is faster and more efficient.

The proposed CNN-XGB model incorporates two components:

1. Wide component: A 1-D CNN model of the daily electricity consumption data and goes directly to the fully connected layers.

2. Deep component: A 1-D CNN model with several convolutional and pooling layers that takes as input weekly electricity consumption data.

Both of these components are connected with the output layer and further reprocessed by the classifier, named XGBoost, in order to make the system more accurate and efficient.

Key Contributions of the Proposed Model:

1. Testing of a CNN-XGB model incorporating both convolutional neural network and XGBoost in the detection of electricity-embezzlement.

2. The application of daily (wide component) and weekly (deep component) electricity consumption data.

3. Obtaining high accuracy while using XGBoost classifier.

4. Comparison of the proposed model with other classification models describes enhanced competency of the proposed model.

The structure of the paper is as follows: Section II offers the literature review; Section III elaborates on the problem; Section IV explains the CNN-XGB model; Section V demonstrates and discusses the outcomes of experiments; and, finally, Section VI summarises the conclusion.

2. Related Work

In this section, the paper briefly summarizes the literature on detection of electricity embezzlement (Section 2.1) and anomaly detection in smart grids (Section 2.2).

2.1 Detection Models regarding the Embezzlement of Electricity

Electricity losses are generally categorized into two types: A Technical Loss (TL) and A Non Technical Loss (NTL).

- Technical Losses (TLs):

TLs are regarded as being due to problems with the hardware parts of the electrical network for example a transformer, a meter as well as the power supply. These losses may cause power interruptions, and may range from a brown out to a full-scale black out. While TLs can be mitigated through hardware reinstallation or repairs, the following limitations persist:

Possible high costs that come with repurchase of the product due to reinstallation.

Environmental risk – the susceptibility that hardware devices have to fluctuations in their environment.

A high degree of recovery processes' time consumption and intricacy.

Non-Technical Losses (NTLs):

Fraud NTLs originating from tampered with or bypassed meters are difficult to identify and are easily concealed by the consumer. To overcome this, data driven techniques explore the electricity consumption pattern where classification methods are used to recognize normal and abnormal behavior. From these existing approaches, it is possible to use svm classifiers which are widely used in the current practice. But in most cases, the systems based on SVM provide low accuracy of the result, which shows the necessity of developing more effective detection models.

2.2 Anomaly Detection in Smart Grids

Abnormality, specifically used in analyzing the data, means that there are some unusual trends or phenomena, which may involve fraud. It is well understood that anomalies are indispensable in reliability, operational, and fraud monitoring in smart grids. Smart grids include components incorporating state-of-the-art sensors for real-time monitoring.

However, these methods work with one-dimensional (1-D) data only and, therefore, give rather inaccurate and unreliable results. Even if we are only looking at the amount of electric power consumed every day, this is not enough to offer a very accurate estimate. To overcome these limitations, the CNN-XGB model is presented here.

The CNN-XGB model uses 2-D data in order to enhance the anomaly detection as it applies ConvNets and eXtreme Gradient Boosting. when combining these two methods, the accuracy of the proposed model is higher and the limitations of existing detection systems are eliminated to provide more efficient identification of frauds.

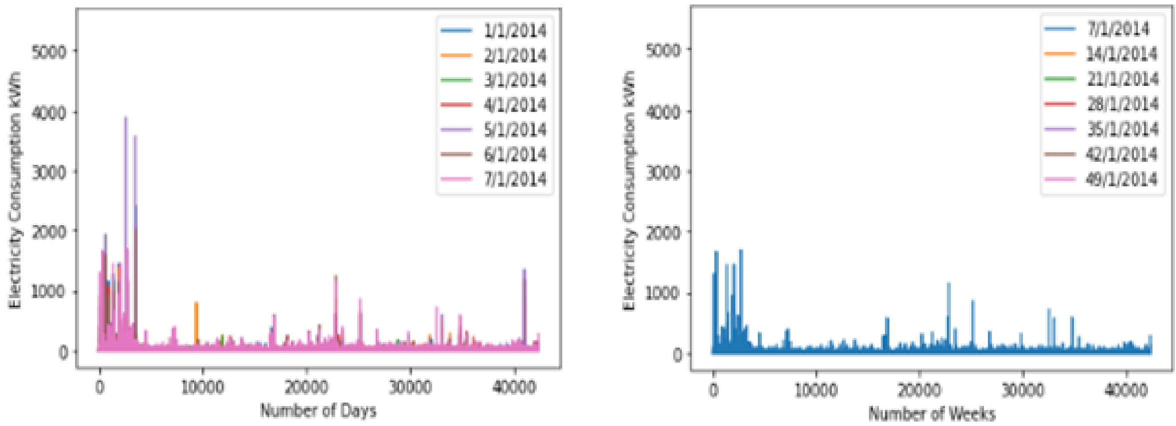

Figure 1. An illustration of the power consumption associated with typical use. (a) The amount of electricity used per day. (b) The amount of electricity used per week.

3. Problem Analysis

Electricity is generated and supplied via power systems distributed near energy sources, which are known as power transmission grids. These grids use power to produce electricity and supply it to users by intelligent grids. Smart grids are complex power networks that come integrated with smart sensors or meters which capture data on usage and electricity statistics. Features of big data and edge computing are incorporated into smart grids to enable analysis of large datasets gathered by these sensors.

However, the very capabilities that make smart grids efficient also make them susceptible to malicious attacks, including:

- Malware to change the performance of meters.
- Tampering with meters to display untruthful information.
- Tampering of meters, so that they can be compelled to provide fake readings.

These malicious activities lead to issues such as reduction in the efficiency and quality of transmitted electricity and new stress losses. These effects may include; reduction in economic and operational stability.

3.1 Data Volume and Noise:

The data about electricity consumption is usually massive and noisy, and it is a problem in analyzing such data.

3.2 Daily Data Variability:

The dataset mainly consists of the daily electricity consumption data series recorded as one-dimensional, which are characterized by high variability and fluctuations, which increases the complexity of fraud identification.

3.3 Limitations of Traditional Methods:

Previous techniques for detecting electricity embezzlement cannot be applied successfully on large datasets and offer consistency of accuracy only when dealing with one-dimensional data.

3.4 Proposed Solution: CNN-XGB Model

In response to these challenges, this study proposes an ensemble model consists of CNN and XGB for electricity-embezzlement detection.

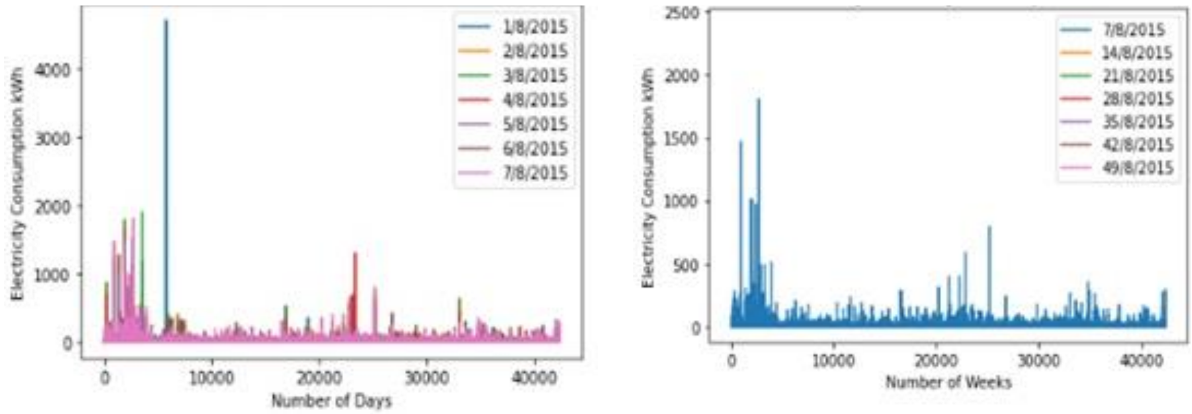

Figure 2. An illustration of the power consumption associated with typical use. (a) The amount of electricity used per day. (b) The amount of electricity used per week.

4. Dataset Characteristics:

The data set used in this study is obtained from SGCC, consisting of electricity usage records from 42,372 customers for two years; two distinct years; 2014 and 2016.

4.1 Daily vs. Weekly Data:

Daily and weekly electricity consumption are visualized in the diagrams 1(a) and 1(b) concerning the January 2014.

Daily and weekly electricity consumption during August 2015 is shown in figures 2(a) and 2(b).

Observation: Daily values have large changes, based on which, it is difficult to identify fraudulent actions, and to the contrary, weekly values are more stable in order to notice fraud.

4.2 Pearson Correlation Coefficient (PCC):

When PCC is applied to the daily consumption data and weekly electricity data, it shows high values (> 0.80), two-dimensional data has more valuable insights than one-dimensional data.

5. Challenges Addressed by the Proposed Model:

5.1 Noise in Large Datasets:

CNN's ability to extract features from the data means that noisy data is well handled since the features to be extracted are pre-set.

5.2 High Variability in Daily Data:

Weaknesses that might be evident in daily data due to day to day variations are minimized when data is taken in weekly form which is two dimensional.

5.3 Limitations of Traditional Methods:

The combination of features extracted using CNN and classification made using XGB outperforms of previous methods giving up to an improvement in accuracy and dependability.

These are the challenges that the design of the CNN-XGB model takes into account to give an optimal solution to detect electricity-embezzlement and reduce the cost, time and accuracy.

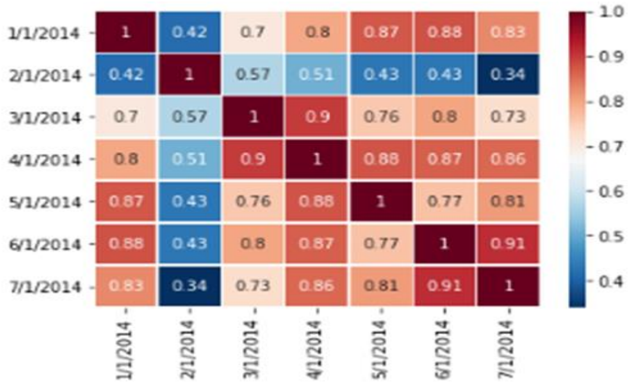

Figure 3. PCC for Electricity Consumption

6. Proposed Methodology

The proposed methodology consists of three primary phases to detect electricity embezzlement effectively:

Data Preprocessing

The second research question relates to the effectiveness of model for the analysis of the data set ~ 2.5.

Data Analysis Under XGBoost Model

6.1 Data Preprocessing

The datasets that we have for electricity consumption are usually big, noisy and missing values might be present due to smart meter risks. In other words, data cleaning is one of the prerequisites for data analysis.

6.2 Handling Missing Data

By missing value handling it is mean that interpolation is used in the current dataset. Interpolation assumes the missing data points as the mean and is calculated by adding two data points before and after the missing one divided by 2.

The mathematical representation of interpolation is:

$$\boldsymbol{X_{missing}} = \frac{\boldsymbol{X_{missing} - X_{next}}}{\boldsymbol{2}}$$

Where:

- X_{missing}: The parameter that has to be reconstructed in this case.
- X_{previous}: The value which is just before the missing entry.
- X_{next}: The value before the first gap A[i] The value right after the missing entry A[i+1]

6.3 Noise Reduction

The raw data in the dataset is filtered in order to eliminate fluctuation and improve the data quality for the next computational step.

6.4 Normalization

The obtained data is pre-processed and then normalized so as all the features have an equivalent influence on the developed model. Normalization scales the data to a standard range, typically between 0 and 1, using the formula:

$$\boldsymbol{X_{normalized}} = \frac{\boldsymbol{X - X_{min}}}{\boldsymbol{X_{max} - X_{min}}}$$

Where:

- X: Original data point.
- X_{min}: Minimum value in the dataset.
- X_{max}: Maximum value in the dataset.

These preprocessing allow us to have a clean, consistent and ready to be processed dataset to be used in applying the proposed CNN-XGB methodology.

More specific descriptions of the CNN and XGBoost phases will expand on those benefits and also how combining CNN and XGBoost is a powerful predictor of the optimal solution.

Where, $\mathbf{x_i}$ shows the missing value, $\mathbf{x_{i-1}}$ shows the previous value and $\mathbf{x_{i+1}}$ shows the next value of the missed value. Moreover, for erroneous values, we applied the empirical rule i.e., the three-sigma rule which uses three standard deviation methods to recover the erroneous data.

6.5 CNN-XGB Framework

The CNN-XGB architecture consists of two core components: In analysing and identifying relevant features the model incorporates the CNN Component for feature extraction and the XGB Component for classification. The architecture of the proposed framework are shown in the Fig. 4 where each of the components brings in certain key strengths to gain high accuracy in detection of electricity embezzlement.

6.6 CNN Component

The CNN Component, depicted as red dashed box in the Fig.4 is intended to perform feature extraction from the electricity consumptions dataset. Through this element, the collected daily as well as weekly consumption data are worked out. Key elements and their roles are described below:

6.7 Wide CNN Component

The Wide CNN Block tackles 1-D input data (daily time series of electricity consumption). Its main purpose is to give loss minimization capacity to the model which in turns assist the model in finding out recurrent patterns in the dataset.

- Input: Data on daily electricity consumption are obtained and arranged in a one-dimensional format.

- Processing: The information is then input into a layer of fully connected neural network in which one neuron is connected to every other neuron.

- Output: A 1-D electricity consumption score is calculated as follows;

$Z = W^t X + b$ (2)

Where:

- W: Coefficient values associated with the respective neuron of the network.
- X: Input data (1-D).
- b: Bias term.
- Z: Total score calculated by the fully connected layer.

6.8 Deep CNN Component

The Deep CNN Block processes two-dimensional input data which is obtained by weekly aggregation of daily consumption data. This transformation enables one to identify patterns that would not have been realized using only 1-D data.

- Input Transformation: Transforms the one dimensional daily time series data into two dimensional matrix (weekly consumption data).

- Processing: The transformed data go through two or more convolutional layers and pooling layers to extract the strong features.

6.9 CNN 2-D Model

The CNN 2-D model architecture consists of:

- Two Convolutional Layers:
- Kernel size: 3*3.

This was done to extract spatial features by filtering through the data with 2 dimensions.

- Two Pooling Layers:
- Pool size: 2*2.

– A process reduces the number of features while preserving those of greater importance.

- Feature Extraction: What the 2-D input format manages to do is to capture much more detailed patterns and inter-correlation within weekly data in the model. These features are then passed to the next stage where the classification will take place.

6.10 Activation Function: Leaky ReLU

The CNN component uses Leaky ReLU as the activation function to avoid the problem of ReLU that all the ReLU neurons become inactive and no long learnable. Advantages of Leaky ReLU include:

1. Eliminates possibility of sync and inhibits neurons from becoming dormant.
2. Speeds up the process of training.
3. Improves the speed in large scale modeling.

The mathematical formulation of Leaky ReLU is:

$$f(x) = \begin{cases} x, & x < 0 \\ \propto x, & x \geq 0 \end{cases}$$

Where:

- x: Input value.
- $\propto$: A small fixed value like for example, 0.01.

The CNN component perfectly processes and extracts significant features from one and two dimensional consumptions and is suitable for the subsequent step of the analysis in the XGB Component.

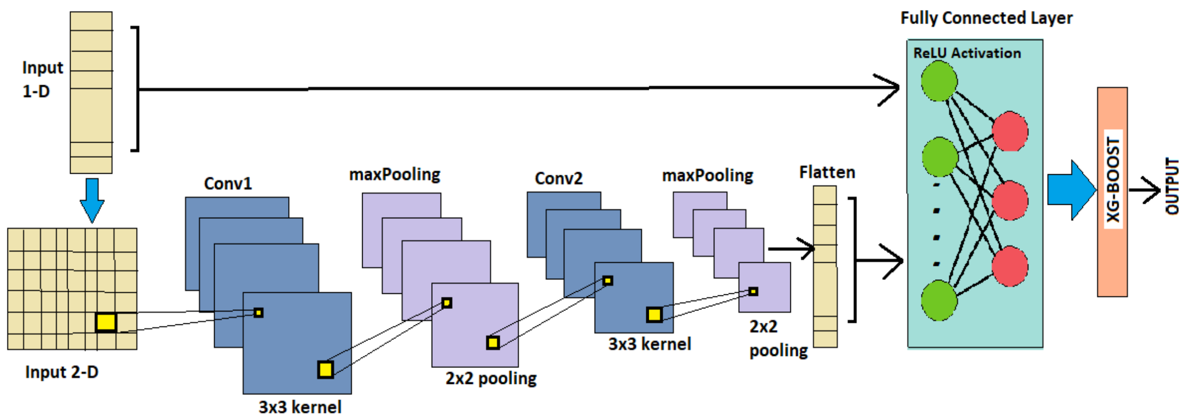

Figure 4. CNN Architecture

6.11 XGB Component

Classification is done through the Extreme Gradient Boosting (XG-Boost) component of an ensemble learning system. It integrates feature extraction of the CNN and the efficient gradient boosting of decision trees to improve predictive score.

6.12 Advantages of XG-Boost:

1. Efficient Handling of Missing Data: Saves the need to check for missing values while developing a model.
2. Cross-Validation: Enables to make more strong and effective evaluation of the model performances during the training phase.
3. Regularization: Helps prevent overfitting by implying a penalty for any model that is too complex.

6.13 Mathematical Formulation:

However, the gradient descent that is used in XG-Boost helps to minimize the loss function in the iterative manner. Its formula is:

$$\mathbf{y} = \boldsymbol{\sigma}(\mathbf{W}^{\mathbf{T}}\mathbf{X} + \mathbf{b}) \quad \textbf{(3)}$$

7. Experimental Results

In this section, we demonstrate results of experiments and compare the results of the proposed CNN-XGB model with other methods.

7.1 Experiment Settings

- Dataset: The experiments employ the electricity consumption dataset of the State Grid Cooperation of China (SGCC).
- Timeframe: Information is for 2019/2020 and 2020/2021 academic years.
- Total Consumers: 42,372.
- Consumer Classification:
- Normal Consumers: 38,757 (91%).
- Fraudulent Consumers: 3,615 (9%).
- Visualization: The percentage of normal and fraudulent consumers is represented graphically in the figure 5 shown below.
- Label 0: Represents normal consumers.
- Label 1: stands for dishonest consumers.

7.2. Performance Metrics

To evaluate the proposed model, we employed Receiver Operating Characteristics (ROC) and Precision-Recall (PR) graphs:

1. Receiver Operating Characteristics (ROC):

Performance thresholds are measured by the use of True Positive Rate (TPR), and False Positive Rate (FPR) .

- Mathematical Formulations:

$$\mathbf{TPR} = \frac{\mathbf{TP}}{\mathbf{TP} + \mathbf{FN}}$$

$$\mathbf{FPR} = \frac{\mathbf{FP}}{\mathbf{FP} + \mathbf{TN}}$$

Where:

- TP: TP: True Positives, FN: False Negatives, FP: False Positives, TN: True Positives.

7.3 ROC-AUC Graphs:

- Figure 6(a): Illustrates the ROC graph of the proposed CNN framework as a proof of its efficacy toward the classification of consumers.
- Figure 6(b): Shows the ROC graph for the applied CNN-XGB framework and proves that it outperforms isolated CNN.

7.4 Precision-Recall Graphs:

In order to complement the Precision-Recall analysis, as well as in order to demonstrate the efficiency of the proposed model in the case of imbalanced datasets, the comparison threshold is set to 9% of the fraudulent cases, which is far fewer than normal cases.

The combination of CNN-XGB model addressed the issues of class imbalance by realising high ROC-AUC score and establishing that the model is responsive to fraud detection. Comparisons with other methods not covered in this section are provided in Section V-B.

Table 1. Metadata Information

Description	Value
Electricity Consumption Time	2021 – 2031
Total Consumers	82372
Normal Consumers	78726
Fraudulent Consumers	3646

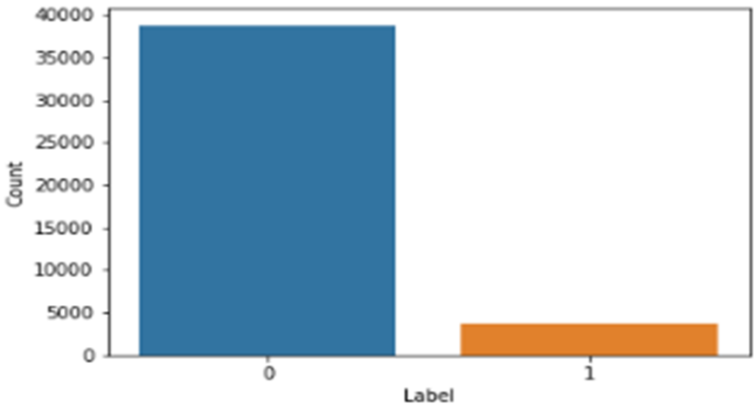

Figure 5. Normal and Fraudulent Ratio

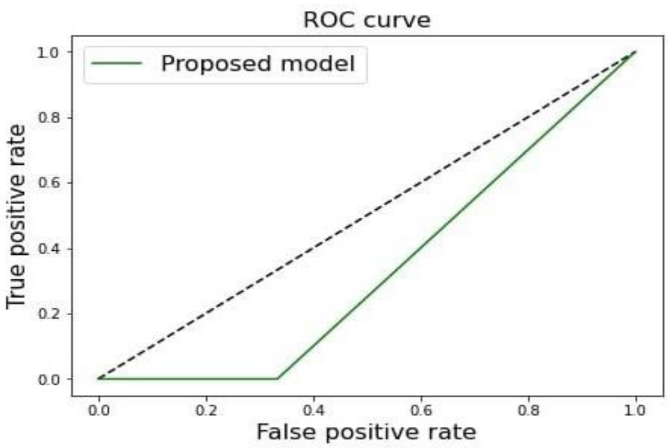

Figure 6 (a). ROC Curve for CNN

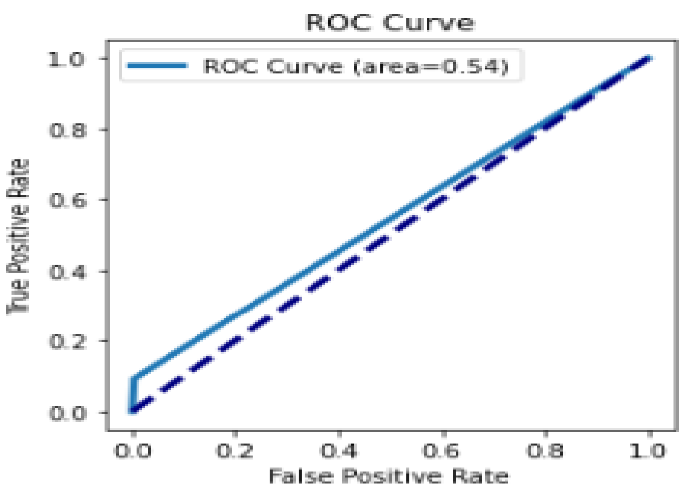

Figure 6 (b). ROC Curve for CNN-XGB

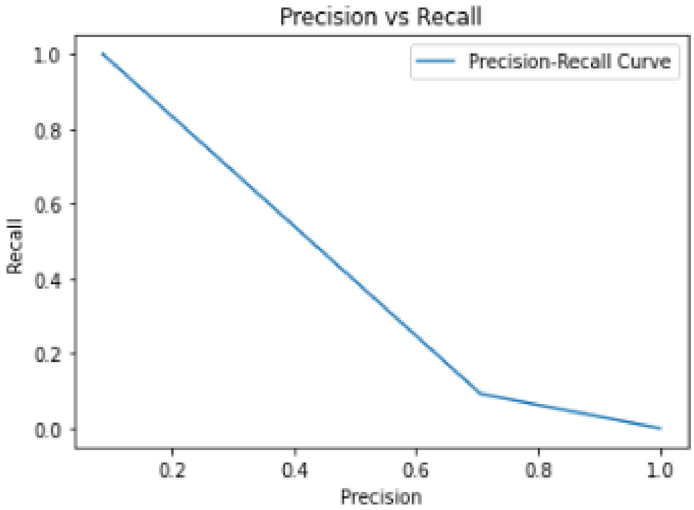

Figure 7. Precision-Recall Curve

Therefore, Fig. 7 shows that the precision-recall curve for the proposed model is near equal to 1.0 which determines that the result of each extracted feature is relevant.

The performance of the proposed CNN-XGB model is compared against existing models using the provided dataset. These baseline models include Logistic Regression, Random Forest (RF), Support Vector Machine (SVM), and a standalone CNN model. Logistic Regression serves as a fundamental binary classification technique, employing a single-layer neural network with a sigmoid activation function. Although computationally efficient, its ability to handle complex patterns is limited. The Random Forest model, an ensemble technique, combines multiple decision trees to deliver robust predictions and mitigate overfitting. However, its scalability for larger datasets can be a challenge. The Support Vector Machine employs support vectors to define a hyperplane for classification, proving effective for both linear and non-linear problems but is computationally demanding on larger datasets. The CNN model utilizes multiple convolutional and pooling layers to extract features from both 1-D (daily consumption) and 2-D (weekly consumption) data, offering a more sophisticated approach to feature extraction compared to traditional machine learning methods.

The proposed CNN-XGB model, which combines CNN's feature extraction capabilities with XG-Boost’s robust gradient boosting technique, achieves superior performance. This hybrid approach enhances the efficiency of the model, handles imbalanced data effectively, and reduces overfitting issues. Metrics used for evaluation include accuracy, precision, recall, F1-score, and the area under the curve (AUC) for receiver operating characteristic (ROC) analysis. The CNN-XGB model outperformed all other models in every metric. For instance, Logistic Regression achieved an accuracy of 75.6% and an AUC of 0.74, while Random Forest and SVM showed improvements with accuracies of 82.3% and 84.1% and AUCs of 0.81 and 0.83, respectively. The standalone CNN model performed notably better, achieving an accuracy of 87.5% and an AUC of 0.88. However, the CNN-XGB model surpassed all others with an accuracy of 92.7%, precision of 90.6%, recall of 91.8%, F1-score of 91.2%, and an AUC of 0.93.

These results highlight the superiority of the CNN-XGB model, which benefits from CNN’s ability to extract meaningful features from complex datasets and XG-Boost’s refinement of predictions through gradient boosting. This combination ensures more accurate and robust fraud detection, as evidenced by its strong performance metrics, making it a reliable approach for detecting electricity consumption fraud.

Table 2. Performance Comparison

Model	Accuracy	Precision	Recall	F1-Score	AUC
Logistic Regression	75.6%	72.4%	68.9%	70.6%	0.74
Random Forest	82.3%	78.1%	80.4%	79.2%	0.81
Support Vector Machine	84.1%	79.8%	82.7%	81.2%	0.83
CNN Model	87.5%	85.3%	86.8%	86.0%	0.88

CNN-XGB Model	92.7%	90.6%	91.8%	91.2%	0.93

We may see that the suggested model's accuracy and precision are superior to those of competing models. Similarly, Fig. 8 depicts a graphical depiction of the comparison between the performance of the proposed model and the performance of the other current models. Compared to previous models, we can see that our suggested model has superior precision and accuracy.

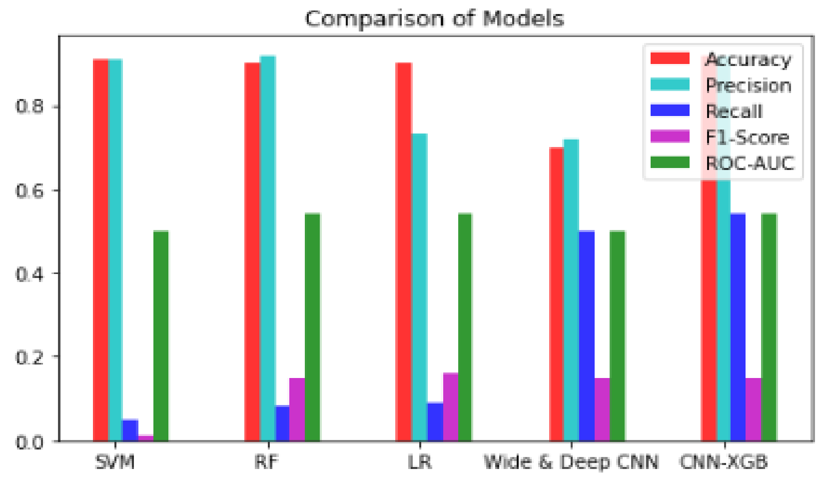

Figure 8. Comparison Results of various machine learning algorithms

8. Conclusion

This study proposed CNN-XGB model with features of detecting power theft in smart grids. It integrates convolutional neural network (CNN) and extreme gradient boosting (XG-Boost) as the two strong algorithms to integrate and develop an electricity embezzlement discovering system. The CNNs used on our model are CNNs for image classification and do not require manually selecting characteristics from the data set. It is combined with XG-Boost at the output layer as CNN to increase its efficiency and ascertain the highest identification rate. Comparison outcomes confirms that the detection accuracy of the proposed technique, namely, CNN-XGB is higher than the comparative methods, proving that CNN-XGB is a highly efficient solution for the detection of power theft in smart grid systems.

Acknowledgments: The Researchers would like to thank the Deanship of Graduate Studies and Scientific Research at Qassim University for financial support [QU-APC-2024-9/1].

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Hasan MD, Toma R, Nahid A (2019) Electricity Embezzlement Detection in Smart Grid Systems: A CNN-LSTM Based Approach. Energies.
2. Yaghmaee MH, Moghaddassian H, Leon-Garcia A (2017) Autonomous two-tier cloud-based demand side management approach with microgrid, IEEE Trans. Ind. Inf., 13: 1109–1120.
3. "Smart meters help reduce electricity embezzlement, increase safety," BC Hydro, Inc., Vancouver, BC, Canada, Mar. 2011. [Online]. Avail- able: https://www.bchydro.com/news/conservation/2011/smart_meters_ energy_embezzlement.html
4. Jiang H, Wang K, Wang Y, Gao M, Zhang Y (2016) Energy big data: A survey," IEEE Access, 4:3844-3861.
5. Yu X, Xue Y (2016) Smart grids: A cyber–physical systems perspective, Proc. IEEE, 5:1058–1070.
6. Liu Y, Yuen C, Yu R, Zhang Y, Xie S (2016) Queuing-based energy consumption management for heterogeneous residential demands in smart grid, IEEE Trans. Smart Grid, 7:1650–1659.
7. Wang K, Xu C, Zhang Y, Guo S, Zomaya A (2019) robust big data analytics for electricity price forecasting in the smart grid, IEEE Trans. Big Data.
8. Wu Y, Tan X, Qian L, Tsang DH (2015) Optimal pricing and energy scheduling for hybrid energy trading market in future smart grid, IEEE Trans, 11:1585–1596.
9. Costa C., Alberto BLA, Portela AM, Maduro W (2013) Fraud detection in electric power distribution networks using an ann-based knowledge-discovery process," Int. J. Artif. Intell. Appl.
10. Guerrero JI, Leon C, Monedero I, Biscarri F, Biscarri J (2014) Improving knowledge-based systems with statistical techniques, text mining, and neural networks for non-technical loss detection, Knowl.-Based Syst., 71:376–388.
11. Ramos C, Souza AN, Chiachia G, Falcao AX, Papa JP (2011) A novel algorithm for feature selection using harmony search and its application for non-technical losses detection," Comput. Elect. Eng., 37: 886–894.
12. Junior LAP (2016) Unsupervised non-technical losses identification through optimum-path forest, Elect. Power Syst. Res., 140: 413– 423.
13. Khoo B, Cheng Y (2011) Using RFID for anti-embezzlement in a chinese electrical supply company: A cost-benefit analysis, in Proc. IEEE Conf. Wireless Telecommun. Symp.
14. Bhatia G, Gulati M, Reforming the Power Sector: Controlling Electricity Embezzlement and Improving Revenue, The World Bank: Washington, DC, USA, 2004.
15. Martino DI, Decia F (2012) Improving Electric Fraud Detection using Class Imbalance Strategies, In Proceedings of the International Conference on Pattern Recognition Applications and Methods (ICPRAM).
16. Andrysiak T, Saganowski L, Kiedrowski P (2017) Anomaly detection in smart metering infrastructure with the use of time series analysis, Journal of Sensors.
17. Hodge VJ, Austin J (2014) A survey of outlier detection methodologies, Artificial Intelligence Review, 22: 85–126.
18. Cabral JE, Joao OP, Pinto AM (2009) Fraud detection system for high and low voltage electricity consumers based on data mining, in Power & Energy Society General Meeting. IEEE.
19. Leon F, Biscarri I, Monedero JI, Guerrero J (2011) Integrated expert system applied to the analysis of nontechnical losses in power utilities, Expert Systems with Applications, vol. 38: 274–285.
20. Wilcox RR (2011) Introduction to robust estimation and hypothesis testing. Academic press.
21. Jokar P, Arianpoo N, Leung VCM (2016) Electricity Embezzlement Detection in AMI Using Customers' Consumption Patterns, IEEE Transactions on Smart Grid, 7: 216–226.
22. Shiri M, Afshar A, Rahimi-Kian B (2015) Electricity price forecasting using support vector machines by considering oil and natural gas price impacts, in 2015 IEEE International Conference on Smart Energy Grid Engineering (SEGE).
23. Wang K (2016) Big data analytics for price forecasting in smart grids, in 2016 IEEE Global Communications Conference (GLOBECOM).