

## Research Article

# A Comprehensive Analysis of Supervised Learning Techniques for Electricity Theft Detection

**Farah Aqilah Bohani** , **Azizah Suliman** , **Mulyana Saripuddin** ,  
**Sera Syarmila Sameon** , **Nur Shakirah Md Salleh** , and **Surizal Nazeri** 

*College of Computing & Informatics, Universiti Tenaga Nasional, Putrajaya Campus, Jalan Ikram-Uniten, 43000 Kajang, Selangor, Malaysia*

Correspondence should be addressed to Farah Aqilah Bohani; farah.aqilah@uniten.edu.my

Received 12 April 2021; Accepted 8 July 2021; Published 21 July 2021

Academic Editor: Jit S. Mandeep

Copyright © 2021 Farah Aqilah Bohani et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

There are many methods or algorithms applicable for detecting electricity theft. However, comparative studies on supervised learning methods for electricity theft detection are still insufficient. In this paper, comparisons based on predictive accuracy, recall, precision, AUC, and F1-score of several supervised learning methods such as decision tree (DT), artificial neural network (ANN), deep artificial neural network (DANN), and AdaBoost are presented and their performances are analyzed. A public dataset from the State Grid Corporation of China (SGCC) was used for this study. The dataset consisted of power consumption in kWh unit. Based on the analysis results, the DANN outperforms compared to other supervised learning classifiers such as ANN, AdaBoost, and DT in recall, F1-Score, and AUC. A future research direction is the experiments can be performed on other supervised learning algorithms with different types of datasets and suitable preprocessing methods can be applied to produce better performance.

## 1. Introduction

Electricity loss can be defined as the difference between the energy source that has been injected and the energy that has been delivered to consumers. In a power system, electricity losses generally occur in the processes of generating, transmitting, and distributing electrical energy [1]. Electricity loss can be classified into two categories, namely, technical loss (TL) and nontechnical loss (NTL). TLs involve the component of the electrical system [1, 2] whereas NTL is more related to electricity theft that is caused by tampering of the meter reading device, hacking the electricity meter, stealing (illegal connections), and more [3].

In 2015, Northeast Group reported that the total cost of NTL worldwide was US\$89.3 billion per year. Countries such as India, Brazil, and Russia lost US\$16.2 billion, US\$10.5 billion, and US\$5.1 billion, respectively [4]. NTL can give a negative impact on power companies, which reduces future investments [5]. From this reason, numerous

studies proposed a method to overcome this problem including artificial-intelligence-based, game theory-based, and state-based models [6].

Artificial-intelligence-based approaches consist of machine learning (ML) and deep learning (DL) techniques. Generally, ML is the process of training a machine with an algorithm to handle large data efficiently by predictive analysis. On the other hand, DL [7] is based on artificial neural network (ANN), which is a human brain model and helps to model irrational functions. To model multiple results simultaneously, ANN is extremely flexible as it requires a huge amount of data.

In general, the effectiveness of a ML solution depends on the nature and characteristics of data and the performance of the learning algorithms. There are various real-world application areas, such as text mining [8–16], web mining [17], medical diagnosis [18–21], COVID-19 [22, 23], crime prediction [24–26], and electricity theft [27–30] which used ML algorithms as an effective solution to solve such complex

problems. ML can be mainly classified into three: supervised learning, unsupervised learning, and reinforcement learning [31].

Supervised learning [31] is the most often used ML algorithm for electricity theft detection such as support vector machine (SVM) [27, 28], ANN [32], deep bidirectional recurrent neural network [33], self-attention [34], wide and deep convolutional neural network [29], and more. Supervised learning is able to take input data and labels in a trained model so as to generate predictions. Supervised learning methods have been successfully applied to assist in reducing site inspections cost [1].

To analyze the classification results, accuracy, precision, recall, area under the ROC curve (AUC) [35], and F1-score [36] are calculated to determine the performance of ML models. Besides, to obtain more meaningful performance measures, certain performance criteria, such as the numbers of true positives (TPs), true negatives (TNs), false positives (FPs), and false negatives (FNs), are commonly used in previous studies [29, 37–43]. This article compares and analyzes the predictive accuracy, precision, recall, F1-score, and AUC for four classifiers based on the dataset obtained from the State Grid Corporation of China (SGCC). The classifiers that are compared are decision tree (DT), ANN, deep artificial neural network (DANN), and AdaBoost. The rest of the paper is organized as follows: Section 2 reviews several papers that investigated electricity theft detection issues by means of supervised learning algorithms, and Section 3 briefly explains the supervised learning algorithms used in the research. In Section 4, the dataset is described, while in Section 5, the preprocessing methods of the dataset are explained. Next, Section 6 presents the evaluation metrics. Sections 7 and 8 provide the experimental results and comparative analysis, respectively, for all the comparisons. Finally, the study is concluded in Section 9.

## 2. Related Works

Decision tree (DT) is a supervised learning method that is commonly applied in electricity theft detection [44]. The authors in [44] proposed DT in conjunction with SVM classifiers and compared it against fuzzy classification, rough sets, ANN, SVM, and fuzzy logic coupled with SVM. The results showed that the method successfully produced 92.5% of accuracy and 5.12% of false positive rate. This detector proved its effectiveness in real scenarios. In a different work, AdaBoost associated with SVM (AdaBoost-SVM) was proposed and could efficiently detect electricity theft [45]. AdaBoost-SVM was compared with four conventional ML techniques and two ensemble learning techniques. It was found that the suggested approach performed significantly better in the imbalanced dataset [45].

Besides, ANN can be utilized in detecting electricity theft [46, 47]. Nevertheless, a majority of previous research is found to be less accurate in detecting electricity theft. It is found that extracting artificial features is necessary based on the domain knowledge [29]. Recently, researchers in [29] proposed a wide and deep convolutional neural network (CNN) model. The study aimed to examine the

data of electricity consumption and determine the electricity theft offenders. The Wide and Deep CNN model included a wide component of a fully-connected layer of neural networks and a deep CNN component with multiple convolutional layers, a fully-connected layer, and a pooling layer. This model integrated the benefits of the wide and deep CNN components, which give rise to its useful implementation and good performance in electricity theft detection.

SVM is developed as a supervised machine learning algorithm that has an advantage in performing classification (for nonseparable class) and regression tasks. The concept of SVM is that classes have the capability of separating the hyperplane through support vector [48]. The nonseparable class is categorized via conversion from lower-dimensional space into higher dimensional space by using kernel trick. SVM with linear kernel for classification is known as a linear SVC (support vector classifier), which accelerates performance for large datasets [49]. NTL detection based on SVM is found in several studies in the literature [28, 50]. Due to the parameter tuning problem in SVM that increases implementation time, SVM has been associated with the genetic algorithm (GA), DT, social spider optimization, and fuzzy logic for improving the classification performance of the model [44, 51–53].

Generally, electricity theft datasets contain an imbalanced class, whereby anomalous (thieves) are smaller than normal consumers [35]. The problem with an imbalanced dataset is known in the literature. The effects of the imbalanced dataset are that the ML model will study the ways to categorize the most regular classes and will not learn the less common classes [36]. Therefore, small classes will not be detected in the confusion matrix caused by the learning of the machine learning model as it performs well in most of the common classes [36]. As a result, the ML model will later become worthless in solving the problem.

Recently, in [54], the study compared several machine learning methods with and without the imbalanced data handling technique on the SGCC dataset. The results showed that SVM and ANN yielded high accuracies even though the imbalanced data handling technique was not applied. In addition, there are many classifiers used without applying the imbalanced data handling technique for solving electricity theft problem, including ANN, Naive Bayes, logistic regression (LR), linear discriminant analysis, quadratic discriminant analysis, random forest (RF), SVM, DT, K-nearest neighbor (KNN), stochastic gradient descent, AdaBoost, CatBoost, LightGBM, and XGBoost [38].

It can be clearly seen that some machine learning models can be applied without an imbalanced data handling technique for solving the classification task. Moreover, no research focused on comparing machine learning classifiers with DT, ANN, DANN, and AdaBoost. With these motivations, this paper aims to contribute to the literature by comparing a few supervised learning algorithms to analyze the best performance of the comparative methods without applying any imbalanced data handling techniques.

### 3. Supervised Learning Algorithms

This section describes and presents the concept and equation of each supervised learning algorithm for electricity theft detection. Figure 1 shows the standard steps in supervised learning algorithms, in which the ML algorithm uses training data, features vectors, and label data as input to produce a predictive model before utilizing new data to provide the expected label as output.

**3.1. Decision Tree (DT).** Decision Tree (DT) is defined as a supervised learning method that is able to solve problems pertaining to regression and classification [55]. DT [31] categorizes instances according to attribute values. In the DT algorithm, a tree consists of node and branch, whereby every node symbolizes a feature of an instance to be categorized. The tree also presumes that the node of every branch represents a value. A simple representation of DT is as shown in Figure 2.

**3.2. Artificial Neural Network (ANN).** The architecture of ANN consists of the input layer (one layer), hidden layers (one or more layers), and output layer (one layer) [48] as shown in Figure 3. ANN is also known as multilayer perceptron or multilayer feed-forward neural network. This algorithm is inspired by interconnections constructed in the human brain. In ANN, the inputs are represented as dendrites (located in the human brain) that receive electrochemical signals generated by neurons and then send them to the cell body.

Each input has a weight and carries signals to a specific hidden layer. Usually, a neuron is driven via an activation function called the sigmoid function. Among the various activation functions such as step function, Gaussian function ramp function, and linear function, it is noted that hyperbolic tangent function can be applied as well [57]. The last layer in ANN refers to the axon that extends to the synapse and connects two distinct neurons. Generally, the simple construction of ANN contains a hidden layer, two inputs, and a single output. An epoch of neural network (NN) is the movement of neurons between the input and output that occurs back and forth. The best epoch depends on the tolerable error in the training of NN. The equation of the ANN output is as indicated in the following equation:

$$y_i = \psi \left( \sum_{i=1}^n \omega_{ji} x_i + \theta_i \right). \quad (1)$$

**3.3. Deep Artificial Neural Network (DANN).** ANN that has two or more hidden layers is considered as a deep neural network (DNN) [58, 59]. Deep learning (DL), or known as DNN, uses an NN algorithm that involves vast computing power and humongous data to capture a high degree of information depending on the raw input data of other layers. Another name of DL and DNN is deep artificial neural network (DANN). Each layer of DANN is able to classify attributes with various forms that exist. All these layers are

achieved by understanding the various forms in which information from the preceding layer is put together in order to create distinguishing features. The architecture of DANN is shown in Figure 4.

**3.4. AdaBoost.** AdaBoost is defined as an ensemble learning approach that has been introduced by Freund and Schapire [60]. It has obtained tremendous success in classification. AdaBoost is noted to be less inclined to overfitting of learning methods in a majority of prediction concerns. The approach develops weak learners through a group of weights that are kept in the training dataset. Then, it will adaptively modify the learners after every weak learning cycle. In the training dataset, weights that are incorrectly classified by the current weak learner will tend to increase, while weights that are correctly classified will tend to decrease [61]. An example of AdaBoost is shown in Figure 5.

### 4. Dataset Description

The dataset was made available by the State Grid Corporation of China (SGCC) [29]. It contained the electricity consumption data of 42,372 consumers, whereby 91.47% (38,757) were normal consumers and 8.53% (3615) had abnormal consumption patterns that could be suspected of electricity theft. The dataset was collected over a time interval from 1<sup>st</sup> January 2014 until 31<sup>st</sup> October 2016 (1,035 days). To achieve an even value of 148 weeks, one more day of data was added. Table 1 displays the description of the data.

### 5. Dataset Preprocessing

The dataset contained some missing data that were represented as not a number (NaN). The missing values were handled by imputing the NaN cell with the mean of row, which signified the mean of consumption for each consumer. The dataset was then normalized by using MinMax Scaler since there were different types of consumers that increased the diversity of the data. The scaling is defined as follows:

$$f(x_i) = \frac{x_i - \min(x)}{\max(x) - \min(x)}. \quad (2)$$

Scaling the data is important since some algorithms are sensitive to diverse data, especially the neural network algorithm. The dataset was then divided into training and testing sets with several different ratios of splitting percentage to observe which ratio gave the best result in predicting anomaly users. The ratios used were 90/10, 80/20, 70/30, and 60/40.

### 6. Evaluation Metrics

The dataset used in this study was an imbalanced dataset, in which the number of true consumers varied significantly as compared to the false consumers. The classifier in the imbalanced dataset was found to be biased when it regarded the real electricity thieves as true customers. For further

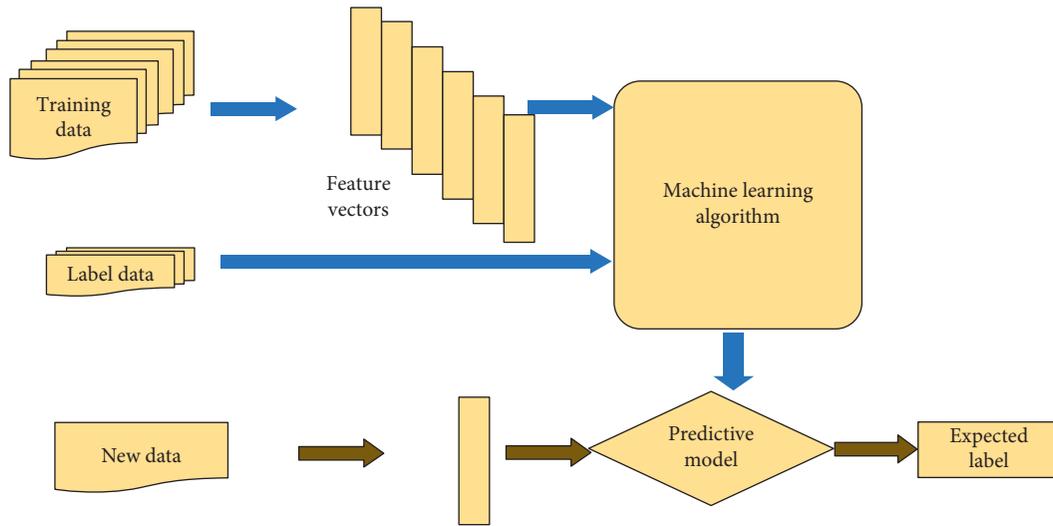


FIGURE 1: Standard steps in supervised learning for classification.

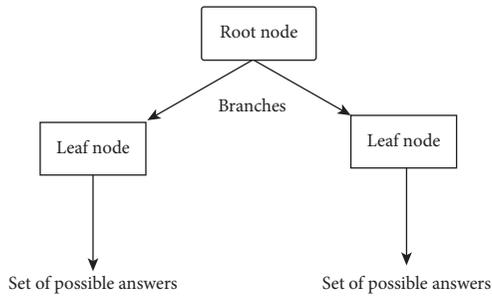


FIGURE 2: Representation of Decision Tree.

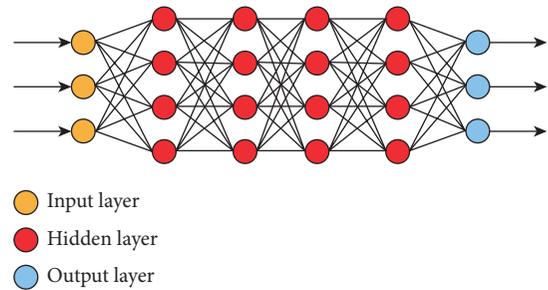


FIGURE 4: Deep artificial neural network (DANN) (adopted from [56]).

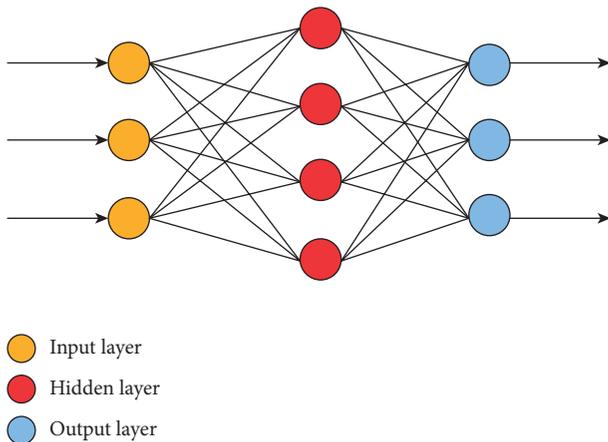


FIGURE 3: Architecture of artificial neural network (ANN) (adopted from [56]).

calculation, a simple accuracy metric was deemed as unreliable. For that reason, numerous evaluation measures were taken into consideration in the current research. The values of each performance indicator were verified based on the confusion matrix. An example of the confusion matrix is as shown in Table 2.

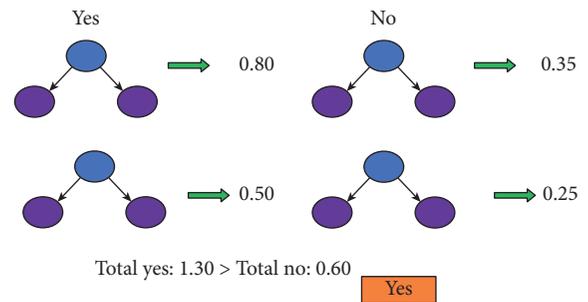


FIGURE 5: Example of AdaBoost.

Normal consumer is represented in the negative class, while anomaly consumer is represented in the positive class. Information driven from the confusion matrix is as follows:

- (i) TP: anomaly consumer accurately predicted as anomaly
- (ii) TN: normal consumer accurately predicted as normal
- (iii) FP: normal consumer predicted as anomaly
- (iv) FN: anomaly consumer predicted as normal

TABLE 1: Data description.

Explanation	Value
Time window for electricity consumption	1 <sup>st</sup> January 2014–31st October 2016
No. of the customers (total)	42,372
No. of normal electricity customers	38,757
No. of electricity thieves	3,615

TABLE 2: Confusion matrix.

		Predicted	
		Positive	Negative
Actual	Positive	True positive (TP)	False positive (FP)
	Negative	False negative (FN)	True negative (TN)

The results were evaluated using accuracy, precision, recall, F1-score, and AUC (area under ROC curve). Accuracy refers to the number of instances that are correctly categorized by classifiers and divided by all the instances. The calculation is presented as follows:

$$\text{accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FN} + \text{FP}} \times 100\%, \quad (3)$$

where TP, FP, TN, and FN indicate True Positives, False Positives, True Negatives, and False Negatives, respectively [48].

The value of TN refers to normal consumers that composed a very high number of this dataset. Therefore, the value for accuracy would also be high. The classifiers learned from the dataset and did not ignore that the dataset was correctly classified because it did not impact the accuracy metric. Due to the fact that SGCC was an imbalanced dataset, depending on accuracy alone was not suitable for this kind of issue.

Owing to that, the evaluation metrics suitable for imbalanced datasets were F1-score and AUC. It was noted that AUC was performed by the exchange between TP rates and FP rates. This kind of measurement determined how well classifiers are correctly classifying the classes. AUC presented the exchange between TP rates and FP rates with a range of 0 and 1. A good classifier was considered if the classifier had a value of ROC-AUC close to 1. For instance, if AUC was equal to 1, it meant that the classes were correctly classified by the classifier, whereas if AUC was equal to 0.5, it indicated that the classifier performed a random prediction [36].

AUC denotes the ability of the model to separate between normal and anomaly classes. The value range was from 0 to 1. A value near 1 indicated a high measure of separability, whereas a value lower than 0.5 signified that the classifier was unable to distinguish between the two classes; thus, it was considered as performing random guessing.

Precision measures the proportion of actual positive out of the total predicted positives. In this case, it denotes the proportion of correctly identified anomaly consumers out of all predicted anomaly consumers. High precision indicates a low FP rate. Recall measures the proportion of actual

positives from the total actual positives. For this study, it signifies the percentage of correctly identified anomaly consumers out of all actual anomaly consumers. High recall indicates a low false negative rate.

F1-score is favored when the data are unfairly distributed, whereby the balance is determined by considering both the precision and recall values. As the classifier in high skewness data tends to be biased toward the majority class, evaluating the F1-score is more reliable than individually using recall or precision.

Precision in equation (4) refers to the ratio of correctly categorized positive class (TP) over the total of positive classes (TP + FP). A high precision value indicated a low FP rate. In equation (5), recall, also known as sensitivity or True Positive Rate (TPR), refers to the rate of correctly classified positive class (TP) as compared to all observations in the actual class (TP + FN). The metric analysis helps to identify the number of instances that are correctly categorized:

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100\%. \quad (4)$$

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100\%. \quad (5)$$

F1-score (F-measure) is more suitable for imbalanced class distribution, which includes the weighted average of recall and precision [36] as shown in equation (6). The value is computed from 0 (the worst) to 1 (the best) [62]. If the classes are found to be very imbalanced, it is suggested to observe both measures of recall and precision. On the other hand, F1-score merges the two measures for a more appropriate evaluation metric for a dataset of this type [36]:

$$F = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}. \quad (6)$$

## 7. Experimental Results

This section presents the results for all comparison methods. Two sets were derived from the dataset, which were the training set and testing set as mentioned in Section 5. Different ratios for each algorithm gave different results of recall, accuracy, AUC, precision, and F1-score. AUC and F1-

TABLE 3: Comparison of outcomes of supervised learning.

Technique	T : T	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	AUC
DT	90/10	91.08	47.37	2.39	4.55	0.5106
	80/20	91.10	50.77	<b>4.37</b>	<b>8.05</b>	<b>0.5198</b>
	70/30	<b>91.77</b>	<b>53.97</b>	3.24	6.10	0.5149
	60/40	91.59	48.84	1.47	2.86	0.5067
Average	—	91.39	50.24	2.87	5.39	0.5130
ANN	90/10	92.54	<b>79.03</b>	13.92	23.67	0.5679
	80/20	92.44	62.66	33.69	43.82	0.6588
	70/30	92.42	57.21	<b>43.62</b>	49.50	0.7029
	60/40	<b>92.74</b>	57.31	50.71	<b>53.81</b>	<b>0.7363</b>
Average	—	92.54	64.05	35.49	42.7	0.6665
DANN	90/10	92.47	<b>65.71</b>	19.60	30.20	0.5934
	80/20	92.44	60.91	38.01	46.80	0.6783
	70/30	<b>93.04</b>	62.64	45.10	<b>52.44</b>	0.7130
	60/40	91.29	48.24	<b>61.03</b>	<b>53.89</b>	<b>0.7754</b>
Average	—	92.31	59.38	40.94	45.83	0.6900
AdaBoost	90/10	<b>91.67</b>	<b>63.64</b>	7.67	13.69	0.5363
	80/20	91.54	55.13	5.93	10.71	0.5247
	70/30	92.16	55.49	<b>8.97</b>	<b>15.45</b>	<b>0.5418</b>
	60/40	91.62	55.50	7.70	13.52	0.5356
Average	—	91.75	57.44	7.57	13.34	0.5346

score were used as the dataset contained imbalanced classes. Table 3 demonstrates the outcomes of evaluation methods for all comparison methods.

The best results were highlighted in bold in the table. Based on the results, ANN showed the highest average of accuracy with 92.54%, followed by DANN (92.31%), AdaBoost (91.75%), and DT (91.39%). Generally, DT and DANN achieved the best accuracy in the 70/30 splitting percentage with 91.77% and 93.04%, respectively. Besides, ANN and AdaBoost performed the best accuracy for ratios of 60/40 and 90/10, respectively.

All classifiers at all ratios achieved more than 0.5 in the AUC evaluation. It can be concluded that these classifiers were applicable in performing classification tasks. In the 70/30 splitting percentage, DANN outperformed DT and AdaBoost in terms of AUC when it achieved 0.7310 as compared to 0.5149 and 0.5418, respectively. Another evaluation was an F1-score, which was related to AUC. The best results of AUC could also provide the best results of the F1-score. Based on Table 3 at the 70/30 splitting percentage, it can be clearly seen that while DT, ANN, DANN, and AdaBoost had the highest score in AUC with the values of 0.5149, 0.7029, 0.7130, and 0.5418. F1-score also yielded the highest result with 6.10%, 49.50%, 52.44%, and 15.45%, respectively.

To evaluate how well the class of anomaly was distributed during classification, precision evaluation was used in this study. Based on the experimental result, three classifiers, namely, ANN, DANN, and AdaBoost, performed the best precision at the ratio of 90/10 splitting percentage with 79.03%, 65.71%, and 63.64%, respectively. DT showed the highest precision at the 70/30 splitting percentage with 53.97%. The highest average of precision was achieved by ANN with 64.05%.

As for the average of recall, DANN outperformed other comparison methods with 40.94%, followed by ANN

(35.49%), AdaBoost (7.57), and DT (2.87%). ANN and DANN both produced the highest recall at a 60/40 ratio with 50.71% and 61.03%, respectively. Another classifier, which was DT, achieved the highest recall (4.37%) while the training and testing samples were set into 80 and 20, respectively. Within contrast, AdaBoost achieved the best recall when the ratio was set to 70/30 splitting percentage.

Even though ANN produced the highest average of accuracy and average of precision with 92.54% and 64.05%, respectively, DANN obtained the highest result for three other evaluation metrics, that is, recall, F1-score, and AUC, with values of 40.94%, 45.83%, and 0.69%, respectively. In conclusion, the ratio for training and testing, or well known as splitting percentage, played a significant part in providing the best result.

## 8. Comparative Analysis and Discussion

Figures 6(a)–6(e) show the graph of the performance of DT, ANN, DANN, and AdaBoost for different types of evaluation methods such as precision, accuracy, recall, AUC, and F1-score, respectively. Theoretically, the training dataset was used to fit the model, whereas the testing dataset was utilized to measure the ML method's fitness. Generally speaking, the splitting percentage on the dataset aimed to evaluate the ML model's implementation and execution on new data. Based on Figure 6(a), the results showed very high accuracy in the training and prediction models of DANN at 70/30 splitting percentage, which was more than 93% accuracy as compared to AdaBoost, ANN, and DT. It is also noteworthy to see that AdaBoost and DT's trained model performance interestingly improved when the splitting percentage was 80/20 as compared to the 70/30 splitting percentage. Most classifiers except for AdaBoost significantly reduced the accuracy when the splitting percentage was 90/10.

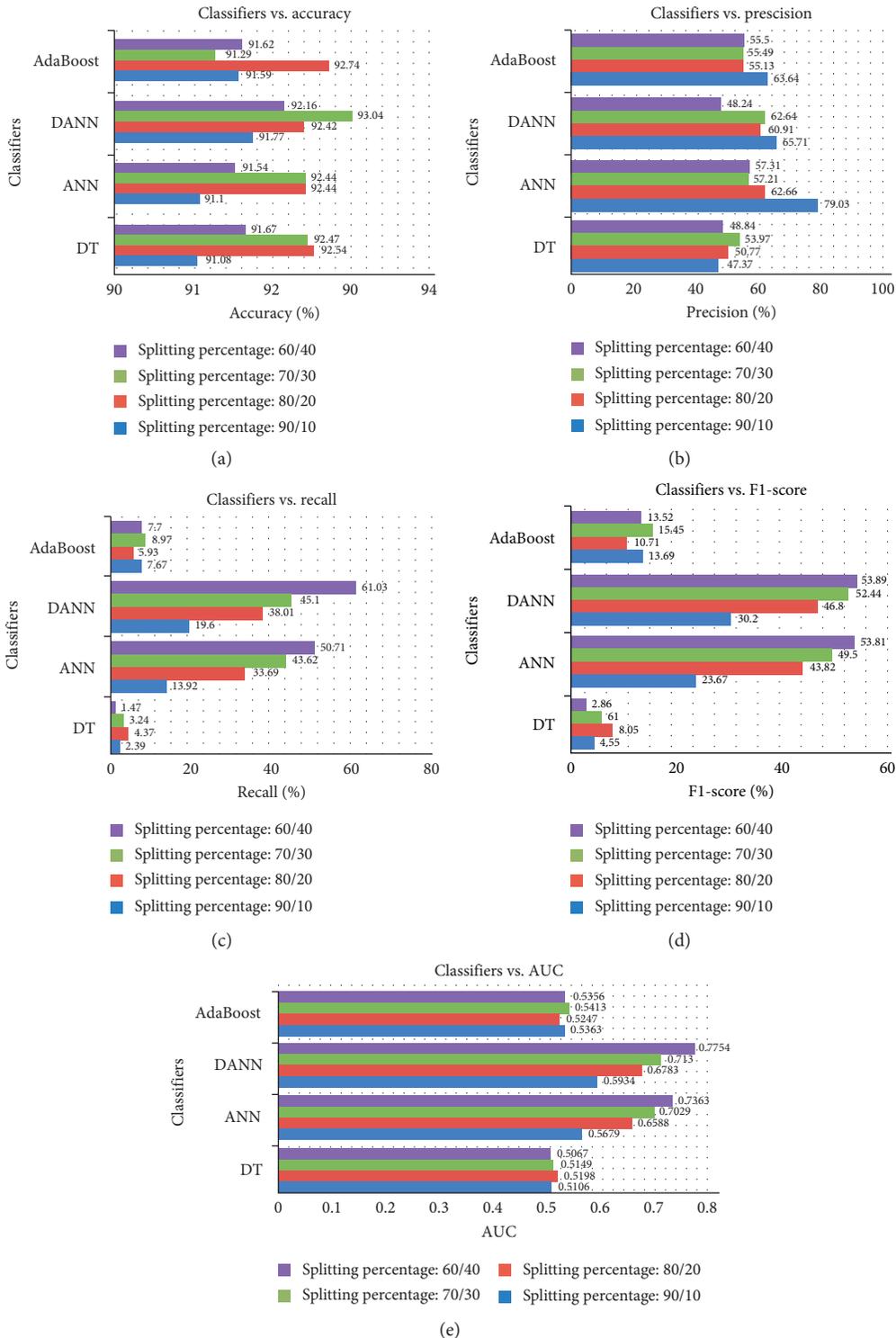


FIGURE 6: Pattern of evaluation methods such as (a) accuracy, (b) precision, (c) recall, (d) F1-score, and (e) AUC for each algorithm.

Based on Figure 6(b), ANN dramatically increased in the precision value at the 90/10 splitting percentage and moved downward at the 70/30 splitting percentage. It can be noticed that AdaBoost achieved the highest precision with 63.64% at the 90/10 splitting percentage. Precision for three other ratios of splitting percentage for the performance of

AdaBoost was almost consistent at about 55%. At the initial step of the precision test, DT seemed to be slightly increased at the 70/30 splitting percentage. However, its precision percentage slowly went downward at the 60/40 splitting percentage. The precision of DANN nearly achieved 66% when the splitting percentage was against 90/10.

Based on Figure 6(c), the recall value of DANN was higher at 60/40 than 70/30 splitting percentage. It can be seen that DANN had a small percentage of recall at the 90/10 splitting percentage. The behavior of ANN was quite similar to DANN when it gradually increased the percentage of recall for each splitting percentage. AdaBoost slightly improved the percentage of recall from 90/10 to 60/40 splitting percentage. It can be clearly seen that the recall value of DT was significantly lower than other comparison methods.

Based on Figure 6(d), DANN achieved a higher F1-score at 60/40 than 70/30, 80/90, and 90/10 splitting percentage. ANN also yielded the highest F1-score at 60/40, which was similar to DANN. Even though AdaBoost poorly achieved a higher F1-Score than DANN and ANN, the score was better than DT.

Based on Figure 6(e), DANN steadily increased over the splitting percentage. It is obvious that the value of AUC for DANN increased when the training set decreased. ANN also had similar behavior to DANN. DT and AdaBoost slightly increased the AUC value at 80/20 and 70/30 splitting percentages, respectively. Both of them provided AUC in the range between 0.50 and 0.53.

Different classifiers will provide different performances of precision, accuracy, recall, AUC, and F1-score in different splitting percentages. For most of them, when the splitting percentage was 90/10, the highest accuracy would be provided.

## 9. Conclusion

This paper analyzed the performance results of supervised learning algorithms with four classifiers for electricity theft detection. Performance evaluated using accuracy, precision, recall, F1-score, and AUC for all classifiers. Compared to other supervised learning classifiers, DANN surpassed the recall, F1-Score, and AUC of other classifiers like ANN, AdaBoost, and DT. For future research, experiments can be performed on other supervised learning algorithms with different types of dataset and suitable preprocessing methods can be applied to produce better performance.

## Data Availability

Previously reported State Grid Corporation of China (SGCC) datasets were used to support this study and are available at <http://www.sgcc.com.cn/>. These prior studies (and datasets) are cited at relevant places within the text as reference [29].

## Conflicts of Interest

The authors declare that there are no conflicts of interest.

## Acknowledgments

This research work was funded by URND TNB Seeding Fund (U-TE-RD-20-08) and BOLD Publication Fund.

## References

- [1] R. R. Bhat, R. D. Trevizan, R. Sengupta, X. Li, and A. Bretas, "Identifying nontechnical power loss via spatial and temporal deep learning," in *Proceedings of the 2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA)*, December 2016.
- [2] R. Rajesh, E. Siva Sankari, and P. Matheswaran, "Detection of non-technical loss in power utilities using data mining techniques," *International Journal for Innovative Research in Science & Technology*, vol. 1, no. 9, pp. 97–101, 2015.
- [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.
- [4] N Group-LLC, "World loses \$89.3 billion to electricity theft annually, \$58.7 billion in emerging markets," 2019, <https://www.prnewswire.com/news-releases/world-loses-893-billion-to-electricity-theft-annually-587-billion-in-emerging-markets-300006515.html>.
- [5] NG LLC, *Electricity Theft and Non-technical Losses Global Markets, Solutions, and Vendors*, NG LLC, NY, USA, 2017.
- [6] P. Glauner, J. A. Meira, P. Valtchev, R. State, and F. Bettinger, "The challenge of non-technical loss detection using artificial intelligence: a survey," arXiv preprint arXiv: 1606.00626, 2016.
- [7] R. M. Mohamad, F. Thabtah, and L. McCluskey, "An assessment of features related to phishing websites using an automated technique," in *Proceedings of the 2012 International Conference for Internet Technology and Secured Transactions*, IEEE, London, UK, December 2012.
- [8] A. Onan and S. Korukoğlu, "A feature selection model based on genetic rank aggregation for text sentiment classification," *Journal of Information Science*, vol. 43, no. 1, pp. 25–38, 2017.
- [9] A. Onan, S. Korukoğlu, and H. Bulut, "A hybrid ensemble pruning approach based on consensus clustering and multi-objective evolutionary algorithm for sentiment classification," *Information Processing & Management*, vol. 53, no. 4, pp. 814–833, 2017.
- [10] A. Onan, S. Korukoğlu, and H. Bulut, "A multiobjective weighted voting ensemble classifier based on differential evolution algorithm for text sentiment classification," *Expert Systems with Applications*, vol. 62, pp. 1–16, 2016.
- [11] A. Onan, "An ensemble scheme based on language function analysis and feature engineering for text genre classification," *Journal of Information Science*, vol. 44, no. 1, pp. 28–47, 2018.
- [12] A. Onan, "Biomedical text categorization based on ensemble pruning and optimized topic modelling," *Computational and Mathematical Methods in Medicine*, vol. 2018, Article ID 2497471, 2018.
- [13] A. Onan, "Ensemble learning based feature selection with an application to text classification," in *Proceedings of the 2018 26th Signal Processing and Communications Applications Conference (SIU)*, May 2018.
- [14] A. Onan, S. Korukoğlu, and H. Bulut, "Ensemble of keyword extraction methods and classifiers in text classification," *Expert Systems with Applications*, vol. 57, pp. 232–247, 2016.
- [15] A. Onan, "Hybrid supervised clustering based ensemble scheme for text classification," *Kybernetes*, vol. 46, pp. 330–348, 2017.
- [16] A. Onan, "Two-stage topic extraction model for bibliometric data analysis based on word embeddings and clustering," *IEEE Access*, vol. 7, pp. 145614–145633, 2019.

- [17] A. Onan, "Classifier and feature set ensembles for web page classification," *Journal of Information Science*, vol. 42, no. 2, pp. 150–165, 2016.
- [18] A. Onan, "On the performance of ensemble learning for automated diagnosis of breast cancer," in *Artificial Intelligence Perspectives and Applications*, vol. 347, pp. 119–129, Springer, Berlin, Germany, 2015.
- [19] S. N. H. Sheikh Abdullah et al., "Round randomized learning vector quantization for brain tumor imaging," *Computational and Mathematical Methods in Medicine*, vol. 2016, Article ID 8603609, 2016.
- [20] U. Sharin, S. N. H. S. Abdullah, K. Omar, A. Adam, and S. Sharis, "Prostate cancer classification technique on pelvis CT images," *International Journal of Engineering and Technology*, vol. 8, no. 1.2, pp. 206–213, 2019.
- [21] N. Bulut, *Breast Cancer and Surgery*, BoD–Books on Demand, Norderstedt, Germany, 2018.
- [22] R. Kurniawan, S. N. H. S. Abdullah, F. Lestari, M. Z. A. Nazri, A. Mujahidin, and N. Adnan, "Clustering and correlation methods for predicting coronavirus COVID-19 risk analysis in pandemic countries," in *Proceedings of the 2020 8th International Conference on Cyber and IT Service Management (CITSM)*, October 2020.
- [23] L. J. Muhammad, E. A. Algehyne, S. S. Usman, A. Ahmad, C. Chakraborty, and I. A. Mohammed, "Supervised machine learning models for prediction of COVID-19 infection using epidemiology dataset," *SN Computer Science*, vol. 2, no. 1, pp. 11–13, 2021.
- [24] S. S. Abdullah, F. A. Bohani, Z. A. Nazri et al., "Amenities surrounding commercial serial crime prediction at greater valley and kuala lumpur using K-means clustering/pengecaman kemudahan awam sekitar lokasi jenayah kormesial bersiri di lembah klang dan kuala lumpur menggunakan kaedah gugusan K-means," *Jurnal Teknologi*, vol. 80, no. 4, 2018.
- [25] M. Arif and M. A. Al Ghamdi, "Seasonal climate prediction using machine learning techniques for the Holy City of Makkah," *Multi-Knowledge Electronic Comprehensive Journal for Education and Science Publications (MECSJ)*, no. 24, 2019.
- [26] A. Ghazvini, S. N. H. S. Abdullah, M. Kamrul Hasan, and D. Z. A. Bin Kasim, "Crime spatiotemporal prediction with fused objective function in time delay neural network," *IEEE Access*, vol. 8, pp. 115167–115183, 2020.
- [27] S. S. S. R. Depuru, L. Wang, and V. Devabhaktuni, "Support vector machine based data classification for detection of electricity theft," in *Proceedings of the 2011 IEEE/PES Power Systems Conference and Exposition*, March 2011.
- [28] 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.
- [29] Z. Zheng, Y. Yang, X. Niu, H. N. Dai, and Y. Zhou, "Wide and deep convolutional neural networks for electricity-theft detection to secure smart grids," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 4, pp. 1606–1615, 2018.
- [30] Z. A. Khan, M. Adil, N. Javaid, M. N. Saqib, M. Shafiq, and J.-G. Choi, "Electricity theft detection using supervised learning techniques on smart meter data," *Sustainability*, vol. 12, no. 19, p. 8023, 2020.
- [31] R. Alizadehsani, J. Habibi, M. J. Hosseini et al., "A data mining approach for diagnosis of coronary artery disease," *Computer Methods and Programs in Biomedicine*, vol. 111, no. 1, pp. 52–61, 2013.
- [32] H. Huang, S. Liu, and K. Davis, "Energy theft detection via artificial neural networks," in *Proceedings of the 2018 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)*, October 2018.
- [33] Z. Chen, D. Meng, Y. Zhang, T. Xin, and D. Xiao, "Electricity theft detection using deep bidirectional recurrent neural network," in *Proceedings of the 2020 22nd International Conference on Advanced Communication Technology (ICACT)*, February 2020.
- [34] P. Finardi, I. Campiotti, G. Plensack et al., "Electricity theft detection with self-attention," arXiv preprint arXiv: 2002.06219, 2020.
- [35] A. Maamar and K. Benahmed, "Machine learning techniques for energy theft detection in AMI," in *Proceedings of the 2018 International Conference on Software Engineering and Information Management*, Gothenburg, Sweden, May 2018.
- [36] P.-N. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining*, Addison, Ed., Wesley Longman, Publishing Co., Inc., Boston MA USA, 2005.
- [37] J. Van Hulse, T. M. Khoshgoftaar, and A. Napolitano, "Experimental perspectives on learning from imbalanced data," in *Proceedings of the 24th International Conference on Machine Learning*, Corvallis, Oregon, June 2007.
- [38] K. M. Ghorri, R. Abbasi, M. Awais, M. Imran, A. Ullah, and L. Szathmary, "Performance analysis of different types of machine learning classifiers for non-technical loss detection," *IEEE Access*, vol. 8, pp. 16033–16048, 2019.
- [39] G. Figueroa, Y.-S. Chen, N. Avila, and C.-C. Chu, "Improved practices in machine learning algorithms for NTL detection with imbalanced data," in *Proceedings of the 2017 IEEE Power & Energy Society General Meeting*, July 2017.
- [40] N. F. Avila, G. Figueroa, and C.-C. Chu, "NTL detection in electric distribution systems using the maximal overlap discrete wavelet-packet transform and random undersampling boosting," *IEEE Transactions on Power Systems*, vol. 33, no. 6, pp. 7171–7180, 2018.
- [41] P. Glauner, J. Meira, L. Dolberg, R. State, F. Bettinger, and Y. Rangoni, "Neighborhood features help detecting non-technical losses in big data sets," in *Proceedings of the 3rd IEEE/ACM International Conference on Big Data Computing, Applications and Technologies*, Shanghai China, December, 2016.
- [42] P. Massaferro, J. M. Di Martino, and A. Fernández, "Fraud detection in electric power distribution: an approach that maximizes the economic return," *IEEE Transactions on Power Systems*, vol. 35, no. 1, pp. 703–710, 2020.
- [43] M. M. Buzau, J. Tejedor-Aguilera, P. Cruz-Romero, and A. Gómez-Expósito, "Detection of non-technical losses using smart meter data and supervised learning," *IEEE Transactions on Smart Grid*, vol. 10, no. 3, pp. 2661–2670, 2019.
- [44] A. Jindal, A. Dua, K. Kaur, M. Singh, N. Kumar, and S. Mishra, "Decision tree and SVM-based data analytics for theft detection in smart grid," *IEEE Transactions on Industrial Informatics*, vol. 12, no. 3, pp. 1005–1016, 2016.
- [45] R. Wu, L. Wang, and T. Hu, "AdaBoost-SVM for electrical theft detection and GRNN for stealing time periods identification," in *Proceedings of the IECON 2018-44th Annual Conference of the IEEE Industrial Electronics Society*, October 2018.
- [46] B. C. Costa, B. L. A. Alberto, A. M. Portela, W. Maduro, and E. O. Eler, "Fraud detection in electric power distribution networks using an ann-based knowledge-discovery process," *International Journal of Artificial Intelligence & Applications*, vol. 4, no. 6, pp. 17–23, 2013.

- [47] 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. 1, pp. 376–388, 2014.
- [48] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, pp. 978–981, Morgan Kaufman Publishers, Waltham, MA, 2012.
- [49] R.-E. Fan, K.-W. Chang, C. J. Hsieh, X.-R. Wang, and C.-J. Lin, "LIBLINEAR: a library for large linear classification," *Journal of Machine Learning Research*, vol. 9, pp. 1871–1874, 2008.
- [50] S. S. R. Depuru, L. Wang, and V. Devabhaktuni, "Support vector machine based data classification for detection of electricity theft," in *Proceedings of the 2011 IEEE/PES Power Systems Conference and Exposition*, pp. 1–8, Phoenix, AZ, USA, March 2011.
- [51] J. Nagi, K. S. Yap, S. K. Tiong, S. K. Ahmed, and F. Nagi, "Improving SVM-based nontechnical loss detection in power utility using the fuzzy inference system," *IEEE Transactions on Power Delivery*, vol. 26, no. 2, pp. 1284–1285, 2011.
- [52] J. Nagi, K. S. Yap, S. K. Tiong, S. K. Ahmed, and A. M. Mohammad, "Detection of abnormalities and electricity theft using genetic support vector machines," in *Proceedings of the TENCON 2008-2008 IEEE Region 10 Conference*, November 2008.
- [53] D. R. Pereira, M. A. Pazoti, L. A. M. Pereira et al., "Social-spider optimization-based support vector machines applied for energy theft detection," *Computers & Electrical Engineering*, vol. 49, pp. 25–38, 2016.
- [54] J. Pereira and F. Saraiva, "A comparative analysis of unbalanced data handling techniques for machine learning algorithms to electricity theft detection," in *Proceedings of the 2020 IEEE Congress on Evolutionary Computation (CEC)*, July 2020.
- [55] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning*, Springer, Berlin, Germany, 2013.
- [56] W. Xing and D. Du, "Dropout prediction in MOOCs: using deep learning for personalized intervention," *Journal of Educational Computing Research*, vol. 57, no. 3, pp. 547–570, 2019.
- [57] H. Kukreja, N. Bharath, C. S. Siddesh, and S. Kuldeep, "An introduction to artificial neural network," *International Journal of Advance Research and Innovative Ideas in Education*, vol. 1, no. 5, pp. 27–30, 2016.
- [58] D. Yao, M. Wen, X. Liang, Z. Fu, K. Zhang, and B. Yang, "Energy theft detection with energy privacy preservation in the smart grid," *IEEE Internet of Things Journal*, vol. 6, no. 5, pp. 7659–7669, 2019.
- [59] S. Shalev-Shwartz and S. Ben-David, *Understanding Machine Learning: From Theory to Algorithms*, Cambridge University Press, Cambridge, UK, 2014.
- [60] Y. Freund and R. E. Schapire, "Experiments with a new boosting algorithm," in *ICML'96: Proceedings of the Thirteenth International Conference on Machine Learning*, ICML, Citeseer, Bari, Italy, July 1996.
- [61] X. Li, L. Wang, and E. Sung, "A study of AdaBoost with SVM based weak learners," in *Proceedings of the 2005 IEEE International Joint Conference on Neural Networks, 2005*, IEEE, Montreal, QC, Canada, July 2005.
- [62] S. Bhattacharyya, H. Bhaumik, S. De, and G. Klepac, *Intelligent Analysis of Multimedia Information*, IGI Global, PA, USA, 2016.