Fırat Üniversitesi Deneysel ve Hesaplamalı Mühendislik Dergisi

FUJECE
Fırat University Journal of Experimental and Computational Engineering

Dağıtım Şebekelerinde Teknik Olmayan Kayıpların Makine Öğrenme Yöntemleriyle Tespiti

Mahmut TÜRK [1], Cem HAYDAROĞLU [2*], Heybet KILIÇ [3]

[1]Diyarbakır Organize Sanayi Bölgesi Müdürlüğü, Diyarbakır, Türkiye.
[2]Elektrik-Elektronik Mühendisliği Bölümü, Mühendislik Fakültesi, Dicle Üniversitesi, Diyarbakır, Türkiye.
[3] Elektrik ve Enerji Bölümü, Teknik Bilimler Meslek Yüksek Okulu, Dicle Üniversitesi, Diyarbakır, Türkiye.
[1]mahmut.turk@diyarbakirosb.org.tr, [2]cem.haydaroglu@dicle.edu.tr, [3]heybet.kilic@dicle.edu.tr

Geliş Tarihi: 17.09.2024
Kabul Tarihi: 07.12.2024

Düzeltme Tarihi: 28.10.2024

doi: https://doi.org/10.62520/fujece.1551601
Araştırma Makalesi

Alıntı: M. Türk, C. Haydaroğlu ve H. Kılıç, “Dağıtım şebekelerinde teknik olmayan kayıpların makine öğrenme yöntemleriyle analizi”, Fırat Üni. Deny. ve Hes. Müh. Derg., vol. 4, no 1, pp. 192-205, Şubat 2025.

Öz

Bu çalışma, elektrik şebeke sistemlerinde enerji hırsızlığından kaynaklanan teknik olmayan kayıpların (NTL) oluşturduğu ciddi sürdürülebilirlik ve güvenilirlik sorununa odaklanmaktadır. Bu kayıpları azaltmak amacıyla, farklı kaçak türlerinin (gerilim kaçağı, akım kaçağı ve gerilim-akım kaçağı) tespitinde derin öğrenme mimarilerinden yararlanan yapay zeka tabanlı bir yaklaşım öneriyoruz. Literatürdeki çalışmalardan farklı olarak veri seti iki boyutlu matrislere dönüştürülerek, günümüzün popüler yaklaşımları olan Convolutional Neural Network (CNN) ve Long Short-Term Memory (LSTM) modelleri ile analiz edilmiş; CNN, %97,50 doğruluk oranı ile LSTM'nin %64,17 doğruluk oranını geride bırakmıştır. Ayrıca, klasik yöntemlerden, k-En Yakın Komşu (k-NN) yöntemi ile 67,5 doğruluk oranı ve Destek Vektör Makineleri (SVM) yöntemi ile 62,25 doğruluk oranı elde edilmiştir. Bu gibi geleneksel yöntemlerle yapılan karşılaştırmalar, CNN'in karmaşık kaçak desenlerini belirlemedeki üstünlüğünü ortaya koymuştur. Bulgular, CNN'in akıllı şebeke sistemlerine entegre edilerek gerçek zamanlı hırsızlık tespiti için güvenilir bir araç olarak kullanılma potansiyelini vurgulamaktadır. Gelecekteki araştırmalar, gerçek zamanlı verilerin entegrasyonunu ve hibrit model yaklaşımlarını inceleyerek bu çözümün ölçeklenebilirliğini ve etkinliğini daha da artırmayı hedefleyecektir.

Anahtar kelimeler: Teknik olmayan kayıplar (TOK), Kaçak elektrik tespiti, Derin öğrenme, Konvolüsyonel sinir ağı (CNN), Akıllı şebeke

[*]Yazışılan Yazar

İntihal Kontrol: Evet – Turnitin
Şikayet: fujece@firat.edu.tr

Firat University Journal of Experimental and Computational Engineering

Machine Learning-Based Detection of Non-Technical Losses in Power Distribution Networks

Mahmut TÜRK [1], Cem HAYDAROĞLU [2*], Heybet KILIÇ [3]

[1]Diyarbakir Organized Industrial Zone Directorate, Diyarbakir, Türkiye.
[2]Department of Electrical and Electronics Engineering, Faculty of Engineering, Dicle University, Diyarbakir, Türkiye.
[3]Department of Electrical and Energy, Technical Sciences Vocational School, Dicle University, Diyarbakir, Türkiye
[1]mahmut.turk@diyarbakirosb.org.tr, [2]cem.haydaroglu@dicle.edu.tr, [3]heybet.kilic@dicle.edu.tr

Received: 17.09.2024
Accepted: 07.12.2024

Revision: 28.10.2024

doi: https://doi.org/10.62520/fujece.1551601
Research Article

Citation: M. Türk, C. Haydaroğlu, and H. Kılıç, “Analysis of non-technical losses in distribution networks using machine learning methods”, Firat Univ. Jour.of Exper. and Comp. Eng.., vol. 4, no 1, pp. 192-205, February 2025.

Abstract

This study focuses on the serious sustainability and reliability problem caused by non-technical losses (NTL) due to energy theft in electrical grid systems. In order to reduce these losses, we propose an artificial intelligence-based approach that utilizes deep learning architectures in the detection of different types of leakage (voltage leakage, current leakage and voltage-current leakage). Unlike the studies in the literature, the data set is converted into two-dimensional matrices and analyzed with today's popular approaches, Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) models; CNN surpassed LSTM's 64.17% accuracy rate with 97.50% accuracy rate. In addition, from the classical methods, 67.5 accuracy rate was obtained with the k-Nearest Neighbor (k-NN) method and 62.25 accuracy rate was obtained with the Support Vector Machines (SVM) method. Comparisons with such traditional methods have revealed the superiority of CNN in determining complex leakage patterns. The findings highlight the potential of CNN to be used as a reliable tool for real-time theft detection by integrating it into smart grid systems. Future research will aim to further increase the scalability and effectiveness of this solution by examining the integration of real-time data and hybrid model approaches.

Keywords: Non-technical losses (NTL), Electricity theft detection, Deep learning, Convolutional neural network(CNN),SmartGrid

[*] Corresponging author

Plagiarism Checks: Yes – Turnitin
Complaints: fujece@firat.edu.tr

1. Introduction

According to worldwide studies, the cost of illegal energy use causes approximately $96 billion in lost revenues annually [1]. For example, in the USA and the UK, the loss of revenue due to electricity theft is approximately $6 billion each year [2-3]. Since high loss rates are reflected in the cost of electricity, honest customers are also affected by the increasing unit prices of electricity [4].

Non-technical losses in distribution networks are considered as non-technical losses that occur when the energy transmitted to the user via the energy transmission line by the supplier is invoiced incompletely or not invoiced at all by third parties or subscribers intervening in the distribution network or measurement system, or when the supplied energy cannot be measured or is measured incompletely due to the supplier company's personnel or incorrect adjustment of the measurement system [5]. Non-technical losses can be listed as illegal electricity use, wiring errors, meter errors, billing errors [6]. Wiring errors, errors made by supplier personnel in facility installations, use of poor quality cables and use of inadequate, low quality measuring devices are among the factors that cause these losses. Failure or incorrect connection of meters, electricity meters are electronic devices used to measure the energy consumed or produced in a circuit. Failure or incorrect connection of these devices causes the consumed energy to be invoiced incompletely or not invoiced at all. Billing errors are equivalent to invoicing the consumed energy in monetary terms, everything included. In this case, it is important to bill the consumed energy in a healthy way. Otherwise, the energy consumed in the wrong case will be reflected as a loss to the supplier. Separating losses and leakages from each other, minimizing the leakages detected by measurements and observations to be made in the distribution networks will increase the energy quality. Reducing leakages will also ensure that the electricity network, which acts as an electricity transmission system, operates healthily and correctly under suitable conditions. There are many studies in the literature as a solution to the problems mentioned above. These studies can be evaluated in two categories. These are: i) Classifications made with traditional methods, ii) Algorithmic and regression-based approaches. Classifications made with traditional methods are summarized below. Ghori et al. used three classifiers to predict electricity theft from a real dataset obtained from a distribution company containing 80,000 monthly consumption records. Fourteen performance evaluation metrics were calculated among the three classifiers. Based on these data, the performance of the classifiers in scenarios to detect electricity theft yielded 99.12% for Random Forest, 99.27% for K-Nearest Neighbors (KNN), and 98.37% for Linear Support Vector Machine (SVM) [7]. Çelikpençe attempted to detect electricity theft using six machine learning-based classification algorithms. For this purpose, five large-scale and real-time system data, including the Customer Information System (CIS), Geographic Information System (GIS), Field Management System (FMS), and two different Automatic Meter Reading Systems (AMRS), were utilized. During the data preprocessing stage, data cleaning was performed using SQL techniques, and data normalization was conducted using min-max scaling methods. In the next step, feature extraction was carried out, and the extracted features were fed into the related classification methods. In the final step, the validation process was performed. Using the GridSearch method, an accuracy rate of 68.6% was achieved. The accuracy, sensitivity, and precision values were obtained as 67.11%, 81.70%, and 69.16%, respectively. Using the same parameters with SVM, the values were found as 66.70%, 83.07%, and 68.97% [8].

Ghori et al. used 15 different machine learning algorithms, including recently developed classifiers such as CatBoost, LGBoost, and XGBoost, for theft detection. The study showed that the Artificial Neural Network (ANN) outperformed other classifiers in detecting theft, achieving a 77% success rate in detecting electricity theft. Additionally, the study identified the best 14 features out of the 71 generated features [9]. Gerardo Figueroa et al. presented a framework to address the data imbalance problem in supervised classification techniques for detecting non-technical losses in power grids through resampling techniques. Support Vector Machine (SVM) and Artificial Neural Network (ANN) were used as classifiers. Classification accuracies of 89% and 90% were achieved with the SVM and ANN classifiers, respectively [10].

Algorithmic and regression-based approaches are summarized below. Capeletti et al. proposed an end-to-end solution for the detection of non-technical losses (NTL) using supervised classification learning, specifically through a multilayer perceptron (MLP) model of artificial neural networks (ANN). In this model, data mining concepts were applied to external data (especially ambient temperature) and internal data from energy companies. By correlating these data, improvements were made in the input data of the model, which allows for the identification of consumer units with NTL. The test results indicate a success with an increase of 0.0213 in ROC-AUC and a recall rate of 6.26% [11]. Žarković et al. utilized various artificial intelligence (AI) algorithms, including artificial neural networks (ANN), adaptive neuro-fuzzy inference systems (ANFIS), autoencoder neural networks, and K-means clustering, to detect losses in distribution networks. The artificial neural network (ANN) demonstrated a classification error rate of 7.62% by accurately classifying different consumer types. In contrast, the K-means algorithm had a slightly higher error rate of 9.26%, while ANFIS reached an error rate of 11.11% by failing to detect the initial type of anomaly [12]. Pengwah et al. developed a weighted least squares approach for predicting non-technical losses by utilizing voltage sensitivity coefficients and residual values obtained from the ordinary least squares method. In this study, voltage measurements from smart meters and the predicted sensitivity coefficients were used to estimate the actual consumption of customers. The differences between the measured and predicted consumption values were evaluated against a threshold; if the difference exceeded this threshold, the customer was labeled as a "fraudulent customer [13]. Jené-Vinuesa et al. demonstrated that a Random Forest classifier trained for the detection of non-technical losses effectively identifies types of fraud with a weighted F1 score of 0.859. Additionally, an unsupervised detection model integrating clustering and correlation methods has facilitated the accurate identification of tampered meters. The study presented two adjustable parameters that enable utilities to finely tune their meter tampering detection strategies based on economic factors. High-accuracy fraud detection was achieved using the Fuzzy C-Means algorithm, which obtained an F1 score of 0.9 [14]. Souza et al. used real electricity consumption data to detect non-technical losses and analyzed additional statistical and temporal features based on this data. In the study, these features were utilized to improve the detection rates of various types of NTL. Furthermore, a model that combines the electricity consumption data with these additional features was developed, achieving higher detection rates for all types of fraud considered [15]. Kara et al. proposed gradient boosting decision tree-based models, LightGBM, XGBoost, and CatBoost, to detect illegal electricity usage from smart meter data of consumers. In the study, the CatBoost model achieved accuracy and AUC scores of 78.10% and 74.17%, respectively, while the LightGBM model achieved a sensitivity performance of 72.48% [16]. Saeed et al. used a dataset containing records of both honest customers and those who engaged in illegal energy use. They employed Pearson's chi-square algorithm to select the most relevant features from the data. Subsequently, the Boosted C5.0 Decision Tree (DT) algorithm was applied to classify honest and fraudulent consumers based on these features. To assess the effectiveness of their proposed electricity theft detection method, its performance was compared with several state-of-the-art machine learning algorithms, including Random Forest (RF), Support Vector Machine (SVM), Artificial Neural Network (ANN), and Extreme Gradient Boosting (XGBoost). The proposed Non-Technical Loss (NTL) detection method outperformed these algorithms, achieving an accuracy of 94.6%, a precision of 78.1%, a specificity of 98.2%, an F1 score of 84.9%, and an overall performance of 93.2% [17].

Chatterjee et al. tried to detect electricity theft using Recurrent Neural Networks (RNN), a state-of-the-art artificial intelligence algorithm. The algorithm utilized Long Short-Term Memory (LSTM) units to process sequential electricity consumption data. They proposed a model that can be used to shortlist potential electricity theft consumers in real-time [18]. Khan et al. proposed two schemes for detecting electricity theft by predicting illegal usage. Both schemes combined Autoregressive Integrated Moving Average (ARIMA) and machine learning techniques to effectively model consumer behavior related to electricity theft. Extensive simulations on real-world electricity consumption datasets demonstrated that the proposed schemes outperformed the latest solutions, achieving 98% accuracy, 98.6% precision, 98.2% recall, 97.9% AUC, and a 98.4% F1 score [19]. Xing et al. employed the Maximal Information Coefficient algorithm to capture the global and periodic patterns in consumption data to prevent electricity theft. Extensive experiments conducted under various attack scenarios demonstrated the effectiveness of the proposed method [20].

Barros et al. used the Rotation Forest algorithm to detect electricity theft and the eXtreme Gradient Boosting (XGBoost) algorithm to predict energy recovery. The study analyzed a dataset of 261,489 consumers from a public utility in Brazil. The results indicated that the proposed approach increased the financial return of field inspections by up to 11.5 times [21]. Saeed et al. proposed a new approach for theft detection in Electricity Distribution Companies (EDC) using the Ensemble Bagged Tree (EBT) algorithm. Their model generates a list of suspicious consumers based on irregularities in consumption data for further investigation. The EBT algorithm demonstrated an accuracy of 93.1% in detecting electricity theft [22].

Shih-Che Huang et al. proposed a state estimation-based approach for load forecasting at distribution transformers to detect theft. They used the overall fit of estimated values to feeder bar measurements, based on customer meter readings collected at distribution transformers, to identify irregularities in electricity usage. The results from state estimation were then analyzed using variance analysis (ANOVA) [23]. Yurtseven proposed multiple linear regression models to estimate non-technical losses energy losses caused by theft and fraud by third parties in electricity distribution systems. These models aim to quantify the impact of such losses on the operations of electricity organizations [24]. Sook-Chin Yip et al. proposed two regression-based algorithms to assess anomaly coefficients for analyzing consumers' energy usage behaviors and detecting energy theft caused by meter tampering, as well as identifying faulty smart meters. The model incorporated categorical variables and detection coefficients to pinpoint periods and locations of energy theft and faulty meters. Simulations demonstrated that the proposed algorithms successfully identified all fraudulent consumers [25]. Esther Villar-Rodriguez et al. presented an algorithmic approach that combines probabilistic data mining and time series analysis concepts to identify consumption outliers in smart grids [26].

In order to detect leakage in electricity meters, it is quite difficult to detect leakage in manual/physical inspections. In order to overcome this difficulty, it is inevitable to use artificial intelligence tools developed in a situation where there are millions of meters. In this study, three different leakage electricity meter data and non-leakage electricity meter data were classified using both deep learning architectures and traditional methods (SVM, k-NN). The data set used for Convolutional Neural Networks (CNN), one of the deep learning architectures, was transformed from one dimension to two dimensions. The obtained matrices were given to the proposed CNN architecture and Long-Short Term Memory (LSTM) architecture. In traditional methods, the feature matrix was created by taking the mean, standard deviation, maximum and minimum values from the data set in the pre-processing phase. These obtained features were evaluated in SVM and k-NN classifiers. The original aspects of this study are given below.

- This study involved the development of a unique dataset tailored to address multiple types of illegal electricity usage, specifically designed to capture voltage leakage, current leakage, and combined voltage-current leakage scenarios. Unlike datasets used in prior studies, this dataset incorporates diverse leakage types, providing a robust foundation for machine learning models to identify and classify distinct patterns of non-technical losses (NTLs) in electrical grids.

- For the first time in the literature, this study presents an artificial intelligence-based method capable of detecting and classifying multiple leakage types within electrical distribution systems. This approach leverages advanced deep learning techniques to process and analyze complex consumption patterns, allowing for improved accuracy and adaptability in identifying non-technical losses, a significant step beyond traditional detection methods focused on single leakage types.

- The study introduces a novel approach by transforming the dataset into a two-dimensional matrix format, enabling more effective analysis through deep learning architectures like Convolutional Neural Networks (CNN). This matrix-based data processing method is particularly suited to extracting spatial patterns inherent in the data, enhancing the model's ability to differentiate between leakage types and improving overall classification performance. This methodology represents a pioneering effort in leveraging data structure transformation to optimize AI model efficacy in NTL detection.

2. Method

In this study, three different leakage data and non-leakage electricity meter data were classified using both deep learning architectures and traditional methods (SVM, k-NN). The data set was transformed from one dimension to two dimensions using deep learning architectures such as Convolutional Neural Network (CNN) and Long-Short Term Memory (LSTM). The flow diagram suggested in this study is given in Figure 1.

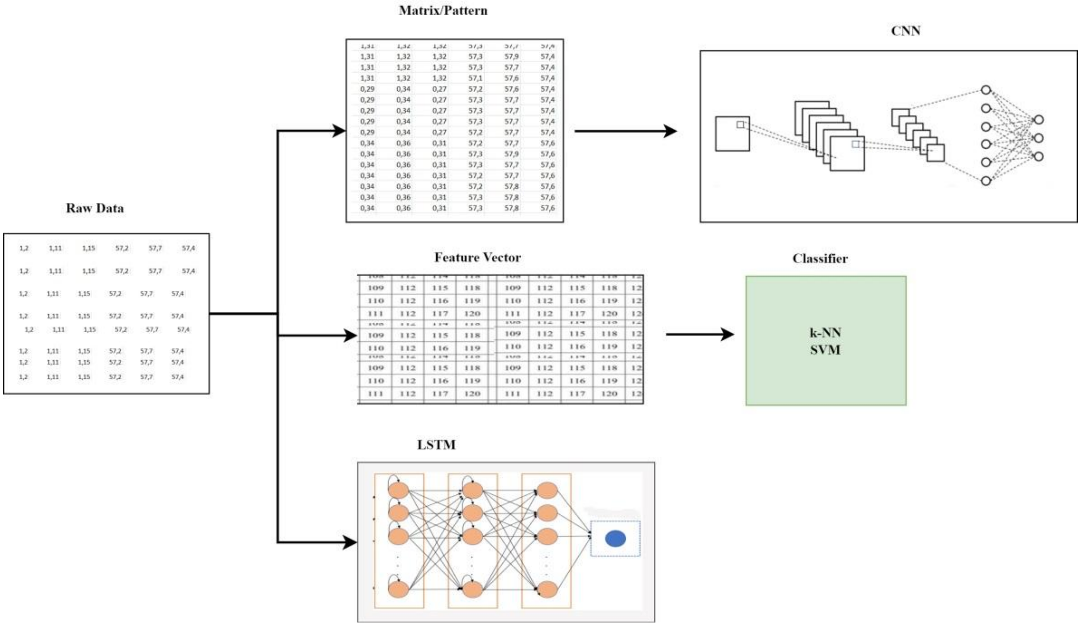


Figure 1. Classification methods proposed in this study

The obtained matrices were classified in the proposed CNN architecture. In traditional methods, the feature matrix was created by obtaining the mean, standard deviation, maximum and minimum values from the data set in the pre-processing stage. These obtained features were evaluated in SVM and k-NN classifiers.

2.1. Dataset

In the study, current and voltage information of 400 electricity meters were taken hourly and daily. Each of these data contains 100 electricity meter data and there are 3 different types of leakage. These are: 1-Voltage Leakage, 2-Current Leakage, 3-Current-Voltage Leakage. The remaining 100 electricity meter data consists of data without leakage. Each meter data from these data is 300x6 in size. Columns (6 piece) in the data set consist of current and voltage data drawn by the meters. Information on the types of leakage in the data set is given below. The consumption values of an electricity meter belonging to a customer consuming electricity are obtained by multiplying the Current X Voltage X Hourly data. The leakage methods examined are; 1-Voltage Leakage, 2-Current Leakage, 3-Current-Voltage Leakage, respectively.

Voltage Leakage is a leakage method that occurs when a customer consuming electricity intervenes in the measurement system and the voltage peaks fed to the electricity meter and included in the consumption calculation of the used electrical energy fall to zero value or to values lower than the required values.

Current Leakage, This leakage method, which has similar features to Voltage Leakage, is a leakage method that occurs when the current values corresponding to the electricity consumed by the customer on the electricity meter are intervened in the measurement system and the current values are reduced to zero or to lower values than they should be.

Current-Voltage Leakage, this leakage method, unlike the other two leakage methods, is a leakage method that occurs when both current and voltage values are intervened in the measurement system from outside, and the current and voltage values that should be zero or lower than the values that should be passed through the electricity meter in order to bill the consumer less or not billed at all for the electricity used. The types of leaked data and non-leaked data are given in Table 1.

Table 1. Sample class data

Class	Current I1	Current I2	Current I3	Voltage L1	Voltage L2	Voltage L3
Voltage Leakage	0.53	0.59	0.53	0	59	58.5
	0.54	0.63	0.53	0	59	58.5
	0.54	0.62	0.55	0	59.1	588
	0.57	0.67	0.61	0	59.1	58.8
Current Leakage	0.75	0.73	0.001	57.4	57.8	57.7
	0.9	0.88	0	57.6	58	57.7
	0.69	0.69	0	57.4	57.8	57.8
	0.84	0.86	0	57.4	57.8	57.7
Voltage-Current Leakage	0	0	1.52	0	57.7	57.3
	0	0	1.57	0	57.7	57.5
	1.55	0	1.67	0	57.9	57.6
	1.63	0	1.65	58.2	57.9	0.01
Non-Illegal	1.77	1.57	1.52	58.7	58.1	57.8
	1.77	1.57	1.52	58.7	58.1	57.8
	1.77	1.57	1.52	58.7	58.1	57.8
	1.77	1.57	1.52	58.7	58.1	57.8

2.2. Convolutional neural network architecture

In the traditional methods used in previous studies of deep learning networks, data losses are high in the feature acquisition phase [27]. Especially the positive developments in computer hardware have paved the way for the use of deep learning architectures. Deep learning architectures consist of layers [27]. Each layer has different functions. In general, an CNN architecture consists of convolution, activation, pooling and a full layer, which is a classical neural network [27]. In the convolution layer, the data is subjected to data convolution with a specified filter. After the convolution process, the data is subjected to the activation process. The pooling layer is a subsampling process. Here, the largest value or average of the data in the frame obtained as a result of stepping and stepping is taken. Feature maps are created as a result of these processes. A CNN architecture can consist of many layers depending on the data size. Finally, the data is classified in the fully connected layer, which has a classical artificial neural network structure. In CNN architectures, the lower layers contain general features, while the upper layers obtain more specific features. The CNN structure proposed in this study is shown in Figure 2.

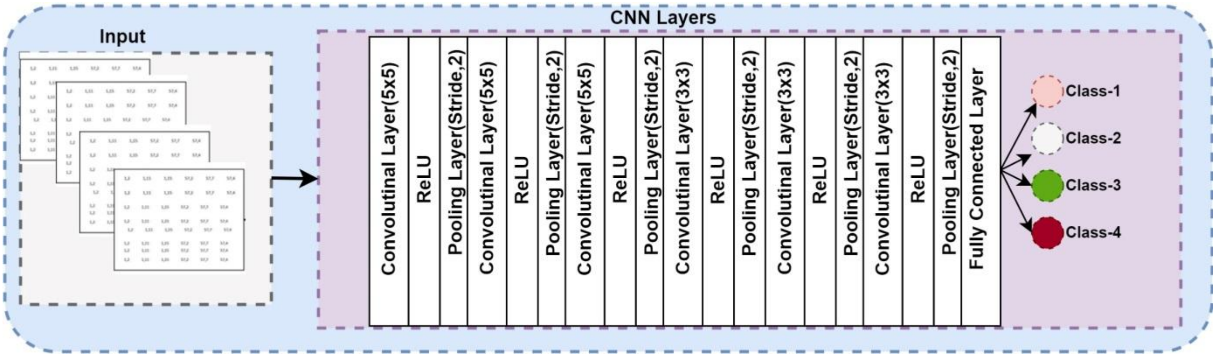


Figure 2. CNN structure proposed in this study

In the study, 6 convolution layers were used in the CNN architecture. 5x5 filters were used in the first three convolution layers, and 3x3 filters were used in the last three layers. A ReLu activation function and a pooling layer were added after each convolution layer. 250 neuron inputs were used in the fully connected layer. From the CNN architecture parameters, learning rate was set as 0.001, batch size as 2 and number of epochs as 100. This study was performed on a computer with 12th Gen Intel(R) Core(TM) i5-1235U 1.30 GHz and 16 GB RAM capacity.

2.3. Long-short term memory architecture

The UKSB architecture is an CNN architecture. It is mostly used in processing one-dimensional data [28]. Its features have come to the fore in the field of language processing. It has been proposed to overcome the shortcomings of continuous recurrent (RNN) networks. In the UKSB architecture, a previous state and input information are kept in memory cells [29]. This memory combines the current and previous inputs. In this way, it ensures the continuity of the data series. In the study, 4 3x3 filtered 2D LSTM convolution layers were used. 40 feature maps were obtained in each layer. LSTM memory blocks are given Figure 3.

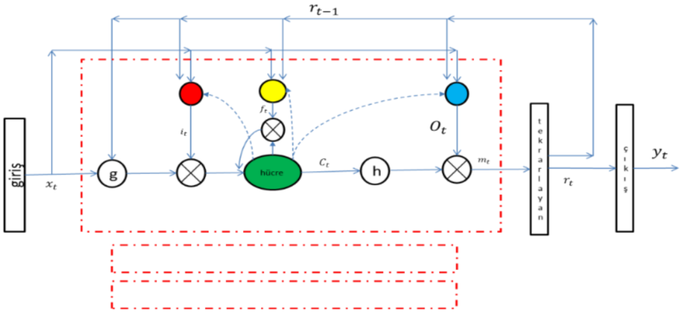
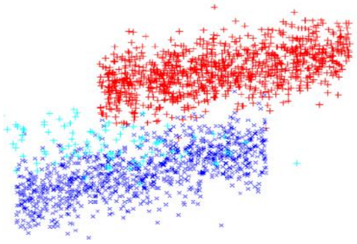


Figure 3. LSTM memory blocks [30]

2.4. k-EN nearest neighbor algorithm

The K-NN classifier algorithm is widely used in classification problems[31]. It is an algorithm that can give very good results even in noisy data. It is a supervised machine learning method. The K-NN classifier algorithm compares a test data and calculates the distance metric k between the features of the test data and the features of the closest classes. In this study, the k value is determined as 3. The K-nearest neighbor (KNN) classification for sample data is given in Figure 4.



. **Figure 4.** The K-nearest neighbor (KNN) classification for sample data [32]

2.5. Support vector machine

Support Vector Machine (SVM) is a supervised machine learning method [33]. It is a vector space-based approach that identifies the decision boundary between two classes, aiming to maximize the margin between the boundary and the nearest points in the training data [34]. Hyper plane structure of support vector machines (SVM) is given Figure 5.

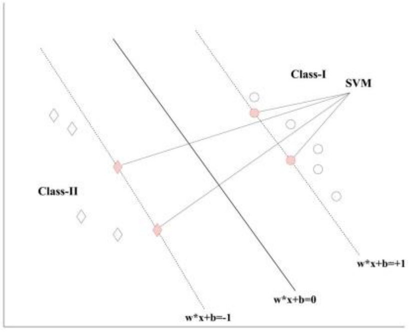


Figure 5. Hyper-plane structure of support vector machines (SVM) [35]

3. Results

In the study, 80% of the dataset was used for training and 20% for testing. Additionally, 10% of the training data was set aside for validation. Three different classifiers were employed in the study. The training accuracy and loss graphs for the convolutional neural network (CNN) architecture are shown in Figure 6.

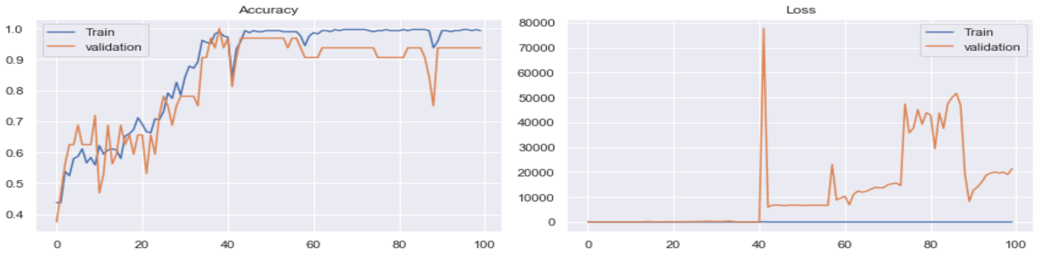


Figure 6. The training accuracy and loss graph of the convolutional neural network architecture

The classification accuracy obtained in the convolutional neural network is 97.50%. The classification performance metrics obtained on a class basis are given in Table 2.

Table 2. Performance measurements obtained from CNN architecture

Class	Precision	Recall	F1- Score
Voltage Leakage	1.00	0.92	0.96
Current Leakage	1.00	1.00	1.00
Voltage-Current Leakage	1.00	1.00	1.00
Non-Illegal	0.88	1.00	0.94
Average	0.97	0.98	0.975

Among the metrics obtained in the table 3, the precision metric of Voltage Leakage, Current Leakage and Voltage-Current Leakage was found to be 1.00. The recall metric was obtained as 1.00 in the Current Leakage, Voltage-Current Leakage and Non-Leakage classes. The F-1 score was found as 1.00 in the Current Leakage and Voltage-Current Leakage classes. The study revealed an average brightness of 0.97% .The confusion matrix obtained as a result of the classification is given in Table 3.

Table 3. Performance measurements obtained from CNN architecture

Actual Class	Predicted Class			
Voltage Leakage	24	0	0	0
Current Leakage	0	21	0	10
Voltage-Current Leakage	0	0	18	0
Non Leakage	2	0	0	15

In the study, only 2 of the test data given to the CNN classifier were classified as Voltage Free Leakage class. The test data of all other classes were classified as error-free.

Table 4. Performance metrics obtained from the LSTM architecture.

Class	**Precision**	**Recall**	**F1- Score**
Voltage Leakage	0.54	0.62	0.58
Current Leakage	0.63	0.61	0.62
Voltage-Current Leakage	0.77	0.69	0.73
Not Illegal	0.60	0.62	0.61
Average	0.635	0.635	0.635

As can be seen in the table 4, in the Voltage Leakage, Current Leakage, Voltage-Current Leakage and No Leakage sensitivity metrics, the lowest value was obtained as 0.54 for Voltage Leakage, while the highest value was obtained as 0.77 for Voltage-Current Leakage. In the recall metric, in the Voltage Leakage, Current Leakage, Voltage-Current Leakage and No Leakage classes, the lowest value was seen as 0.61 for Current Leakage, while the highest value was seen as 0.69 for Voltage-Current Leakage. When we look at the F-1 score values, in the Voltage Leakage, Current Leakage, Voltage-Current Leakage and No Leakage classes, Voltage Leakage showed the lowest value with 0.58, while the highest value was obtained as 0.73 for Voltage-Current Leakage. The study revealed an average brightness of 0.635%

Table 5. Confusion matrix obtained from LSTM architecture

Actual Class	Predicted Class			
Voltage Leakage	15	0	15	1
Current Leakage	0	17	0	10
Voltage-Current Leakage	8	0	27	0
Non Leakage	1	11	0	18

In the confusion matrix obtained in the LSTM classifier, 8 out of 24 Voltage Leakage test data were classified as Voltage-Current Leakage, while 1 was classified as Not Leakage. 11 out of 28 Current Leakage data were classified as Not Leakage, 12 out of 29 Voltage-Current Leakage test data were classified as Voltage Leakage, 10 out of 29 No Leakage test data were classified as Current Leakage, and 1 was classified as Voltage Leakage.

In the study, in the classification made with k-NN classifier, the data was divided into 5 parts and one part at a time was used as test data. In this way, 5 cross-validation (k-Fold) was performed. The classification accuracy obtained from each cross-validation in the study is given in Table 5.

Table 6. K-Fold performance values of K-NN classifier

K-Fold-1	K-Fold-2	K-Fold-3	K-Fold-4	K-Fold-5
71.25	65	58.75	75	67.5
Average	67.5			

When Table 6, is examined, it is seen that the highest success is Fold-4 and the lowest success is Fold-3. An average of 67.5% classification accuracy was obtained in the study.

Table 7. K-Fold performance values of SVM classifier

K-Fold-1	K-Fold-2	K-Fold-3	K-Fold-4	K-Fold-5
67.5	55	62.5	0.6	66.25
Average	62.25			

When the results given in Table 7, are examined, it is seen that the highest value is Fold-1 with 67.5% and the lowest value is Fold-4 with 0.6%. As a result of the study, an average accuracy of 62.25% is observed.

Table 8. Comparison of the classification accuracy of the classifiers used in the study.

Classifier	Accuracy
CNN	97.50
K-NN	67.50
LSTM	64.17
SVM	62.25

Table 8, compares the classification accuracy performance of the classifiers used in the study. It was observed that the highest classification accuracy was obtained from the CNN architecture, whereas the lowest performance was obtained from the SVM classifier.

In deep learning architectures, no feature method was used. On the other hand, in classical methods, the feature method was used. Therefore, in deep learning architectures, all of the data was given to the classifier without any data loss. In classical methods, data loss occurs due to the features obtained. As a result, it can be said that deep learning processes more data, whereas classical methods process less data. At the same time, this is the reason for this performance difference.

In Table 9, the accuracy rates and other performance metrics of this study are compared with previous studies. This table aims to reveal the performance of this study by comparing the results obtained with the results of various studies in the literature.

Table 9. Comparison with studies in the literature

Study	Method	Success(%)	Disadvantages
Çelikpençe.[8]	Classifications made with traditional methods	68.6	The success of the study depends on the accuracy and completeness of the data used. Data in electricity distribution networks may be incomplete or incorrect, which will negatively affect the accuracy of the model.
Jené-Vinuesa et al.[14]	Algorithmic and regression-based approaches	90	How the model works and why it makes certain predictions may not be easily understood by users. This will create transparency and reliability issues.
Kara et al.[16]	Algorithmic and regression-based approaches	78.10-74.17	Although using gradient boosting decision trees to detect illegal electricity usage with smart meter data, the data quality, risk of overfitting, and lack of transparency of the model
Saeed et al.[17]	Algorithmic and regression-based approaches	94.6	Boosted C5.0 algorithm requires more computational power. This can increase processing time when working with large datasets
Khan et al.[19]	Algorithmic and regression-based approaches	98	Hybrid approaches are often more complex and can create additional challenges in data processing, model training, and integration
Saeed et al.[22]	Algorithmic and regression-based approaches	93.1	can cause problems such as ensemble methods sometimes becoming too complex and overfitting the training dataset.
This study	Deep learning based approach	97.50	

The results obtained in this study achieved a higher accuracy rate compared to the studies conducted by Çelikpençe [8], Jené-Vinuesa et al. [14], Kara et al. [16], and Saeed et al. [17, 22]. On the other hand, a 0.50% lower accuracy rate was obtained compared to the study by Khan et al. [19].

4. Conclusion

Electrical energy has an important place all over the world. Electrical energy, which has become indispensable for modern life, is used in many areas such as health, transportation, industry, shelter (home), etc. It is important that electrical energy, which has an important position in economic development, technology and production, is transmitted safely and healthily from its production to the final consumption point. Many negative situations occur in this cycle from the production point to the consumption point. Non-technical losses in distribution networks come to the forefront among these negativities. These losses have a negative effect both in terms of cost and the operation of the systems that transfer electrical energy. Many studies have been carried out in order to prevent these negativities. The cost of the studies carried out is also important. Artificial intelligence, which has become indispensable in the field of technology recently, provides serious gains in terms of both cost and labor in the face of such negative situations.

In this study, an artificial intelligence-based method based on the illegal electricity meter is proposed. A dataset consisting of different types of illegal electricity was created. The data obtained from this dataset was classified in different classifiers. The results obtained in this study show that different types of illegal electricity can be detected using machine learning methods. Thanks to the use of such artificial intelligence-based tools, electrical illegal electricity losses can be minimized. With this method, it is possible to obtain healthy, uninterrupted and efficient energy. The advantage of this study over other studies is that it automatically detects leaks from raw data. It has been observed that it gives good results in deep learning architecture without any preprocessing on the data. If this AI-based method is integrated into real-world systems, smart meters or real-time grid monitoring systems, it will make significant contributions to reducing illegal electricity use and increasing grid reliability. The next study aims to use Siamese networks to detect electrical leaks.

6. Acknowledgment

I would like to thank Diyarbakır Organized Industrial Zone Directorate for providing support during the conduct of this study.

7. Author Contribution Declaration

The authors contributed equally to the article.

8. Ethics Committee Approval and Conflict of Interest Declaration

There is no need to obtain ethics committee permission for the article prepared. "There is no conflict of interest with any person/institution in the article prepared.

9. References

[1] L. J. Lepolesa, S. Achari, and L. Cheng, "Electricity theft detection in smart grids based on deep neural network," IEEE Access, vol. 10, pp. 39638–39655, 2022.

[2] T. Sharma, K. K. Pandey, D. K. Punia, and J. Rao, "Of pilferers and poachers: Combating electricity theft in India," Energy Res. Soc. Sci., vol. 11, pp. 40–52, 2016.

[3] S. Sahoo, D. Nikovski, T. Muso, and K. Tsuru, "Electricity theft detection using smart meter data," in Proceedings of the 2015 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT), Washington, DC, USA, 18–20 February 2015, pp. 1–5.

[4] E. Villar-Rodriguez, J. Del Ser, I. Oregi, M. N. Bilbao, and S. Gil-Lopez, "Detection of non-technical losses in smart meter data based on load curve profiling and time series analysis," Energy, vol. 137, pp. 118–128, 2017.

[5] H. O. Henriques, R. L. S. Corrêa, M. Z. Fortes, B. S. M. C. Borba, and V. H. Ferreira, "Monitoring technical losses to improve non-technical losses estimation and detection in LV distribution systems," Measurement, vol. 161, p. 107840, 2020.

[6] E. S. Ibrahim, "Management of loss reduction projects for power distribution systems," Elect. Power Syst. Res., vol. 55, pp. 49–56, 2000.

[7] K. M. Ghori, M. Imran, A. Nawaz, R. A. Abbasi, A. Ullah, and L. Szathmary, "Performance analysis of machine learning classifiers for non-technical loss detection," J. Ambient. Intell. Humaniz. Comput., pp. 1–16, 2023.

[8] M. Çelikpençe, "Elektrik dağıtım şebekelerinde teknik olmayan kayıp kaçakların makine öğrenmesi ile tespiti," 2023.

[9] K. M. Ghori, R. A. 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.

[10] 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, pp. 1–5, 2017.

[11] M. B. Capeletti, B. K. Hammerschmitt, R. G. Negri, F. G. K. Guarda, L. R. Prade, N. Knak Neto, and A. D. R. Abaide, "Identification of nontechnical losses in distribution systems adding exogenous data and artificial intelligence," Energies, vol. 15, no. 23, p. 8794, 2022.

[12] M. Žarković and G. Dobrić, "Artificial Intelligence for Energy Theft Detection in Distribution Networks," Energies, vol. 17, no. 7, p. 1580, 2024.

[13] A. B. Pengwah, R. Razzaghi, and L. L. Andrew, "Model-less non-technical loss detection using smart meter data," IEEE Trans. Power Deliv., vol. 38, no. 5, pp. 3469–3479, 2023.

[14] M. Jené-Vinuesa, M. Aragüés-Peñalba, and A. Sumper, "Comprehensive Data-Driven Framework for Detecting and Classifying Non-Technical Distribution Losses," IEEE Access, 2024.

[15] M. A. Souza, H. T. Gouveia, A. A. Ferreira, R. M. de Lima Neta, O. Nóbrega Neto, M. M. da Silva Lira, and R. R. de Aquino, "Detection of Non-Technical Losses on a Smart Distribution Grid Based on Artificial Intelligence Models," Energies, vol. 17, no. 7, p. 1729, 2024.

[16] Y. Kara and A. Aksu, "Tüketicilerde Kaçak Elektrik Kullanımının Akıllı Sayaç Verisi Üzerinden Gradyan Artırmalı Karar Ağacı Tabanlı Makine Öğrenmesi Yöntemleriyle Tespiti," J. Investig. Eng. Technol., vol. 6, no. 1, pp. 1–12, 2023.

[17] M. Salman Saeed, M. W. Mustafa, U. U. Sheikh, T. A. Jumani, I. Khan, S. Atawneh, and N. N. Hamadneh, "An efficient boosted C5.0 decision-tree-based classification approach for detecting non-technical losses in power utilities," Energies, vol. 13, no. 12, p. 3242, 2020.

[18] S. Chatterjee, V. Archana, K. Suresh, R. Saha, R. Gupta, and F. Doshi, "Detection of non-technical losses using advanced metering infrastructure and deep recurrent neural networks," in 2017 IEEE International Conference on Environment and Electrical Engineering and 2017 IEEE Industrial and Commercial Power Systems Europe (EEEIC/I&CPS Europe), pp. 1–6, 2017.

[19] H. M. Khan, F. Jabeen, A. Khan, S. A. Badawi, C. Maple, and G. Jeon, "Hybrid non-technical-loss detection in fog-enabled smart grids," Sust. Energy Technol. Assess., vol. 65, p. 103775, 2024.

[20] Y. Xing, L. Guo, Z. Xie, L. Cui, L. Gao, and S. Yu, "Non-technical losses detection in smart grids: An ensemble data-driven approach," in 2020 IEEE 26th International Conference on Parallel and Distributed Systems (ICPADS), pp. 563–568, Dec. 2020.

[21] R. M. R. Barros, E. G. da Costa, and J. F. Araujo, "Maximizing the financial return of non-technical loss management in power distribution systems," IEEE Trans. Power Syst., vol. 37, no. 2, pp. 1634–1641, 2021.

[22] M. S. Saeed, M. W. Mustafa, U. U. Sheikh, T. A. Jumani, and N. H. Mirjat, "Ensemble bagged tree based classification for reducing non-technical losses in Multan Electric Power Company of Pakistan," Electronics, vol. 8, no. 8, p. 860, 2019.

[23] S. C. Huang, Y. L. Lo, and C. N. Lu, "Non-technical loss detection using state estimation and analysis of variance," IEEE Trans. Power Syst., vol. 28, no. 3, pp. 2959–2966, 2013.

[24] Ç. Yurtseven, "The causes of electricity theft: An econometric analysis of the case of Turkey," Utilities Policy, vol. 37, pp. 70–78, 2015.

[25] S. C. Yip, K. Wong, W. P. Hew, M. T. Gan, R. C. W. Phan, and S. W. Tan, "Detection of energy theft and defective smart meters in smart grids using linear regression," Int. J. Electr. Power Energy Syst., vol. 91, pp. 230–240, 2017.

[26] E. Villar-Rodriguez, J. Del Ser, I. Oregi, M. N. Bilbao, and S. Gil-Lopez, "Detection of non-technical losses in smart meter data based on load curve profiling and time series analysis," Energy, vol. 137, pp. 118–128, 2017.

[27] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436–444, 2015.

[28] A. H. Mirza, "Low complexity efficient online learning algorithms using LSTM networks," M.S. thesis, Bilkent Univ., Ankara, Turkey, 2018.

[29] Y. Yu, X. Si, C. Hu, and J. Zhang, "A review of recurrent neural networks: LSTM cells and network architectures," Neural Comput., vol. 31, no. 7, pp. 1235–1270, 2019.

[30] H. Sak, A. W. Senior, and F. Beaufays, "Long short-term memory recurrent neural network architectures for large scale acoustic modeling," in Proc. ICASSP, 2014.

[31] P. K. Syriopoulos, N. G. Kalampalikis, S. B. Kotsiantis, and M. N. Vrahatis, "k NN Classification: a review," Ann. Math. Artif. Intell., pp. 1–33, 2023.

[32] Ö. Tomak and O. Ö. Mengi, "K-Nearest Neighbor Classification of Harmonics Using Akaike Information Criterion," Karadeniz Fen Bilim. Derg., vol. 7, no. 1, pp. 1–8, 2017.

[33] J. Ringelberg and E. Van Gool, "On the combined analysis of proximate and ultimate aspects in diel vertical migration (DVM) research," Hydrobiologia, vol. 491, pp. 85–90, 2003.

[34] K. Blachowiak-Samolyk, S. Kwasniewski, K. Richardson, K. Dmoch, E. Hansen, H. Hop, et al., "Arctic zooplankton do not perform diel vertical migration (DVM) during periods of midnight sun," Mar. Ecol. Prog. Ser., vol. 308, pp. 101–116, 2006.

[35] Ö. Türk, "Classification of electroencephalogram records related to cursor movements with a hybrid method based on deep learning," Int. J. Imaging Syst. Technol., vol. 31, no. 4, pp. 2322–2333, 2021.