IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 25 November 2023, accepted 19 December 2023, date of publication 22 December 2023,
date of current version 5 January 2024.

Digital Object Identifier 10.1109/ACCESS.2023.3346327

RESEARCH ARTICLE

Improving Smart Grids Security: An Active Learning Approach for Smart Grid-Based Energy Theft Detection

SIDRA ABBAS[1], (Graduate Student Member, IEEE), IMEN BOUAZZI[2], STEPHEN OJO[3], GABRIEL AVELINO SAMPEDRO[4,5], (Senior Member, IEEE), AHMAD S. ALMADHOR[6], (Senior Member, IEEE), ABDULLAH AL HEJAILI[7], AND ZUZANA STOLICNA[8]

[1]Department of Computer Science, COMSATS University Islamabad, Islamabad 45550, Pakistan
[2]Department of Industrial Engineering, College of Engineering, King Khalid University, Abha 61421, Saudi Arabia
[3]Department of Electrical and Computer Engineering, College of Engineering, Anderson University, Anderson, SC 29621, USA
[4]Faculty of Information and Communication Studies, University of the Philippines Open University, Los Baños 4031, Philippines
[5]Center for Computational Imaging and Visual Innovations, De La Salle University, Manila 1004, Philippines
[6]Department of Computer Engineering and Networks, College of Computer and Information Sciences, Jouf University, Sakaka 72388, Saudi Arabia
[7]Faculty of Computers and Information Technology, Department of Computer Science, University of Tabuk, Tabuk 71491, Saudi Arabia
[8]Faculty of Management, Comenius University in Bratislava, 820 05 Bratislava, Slovakia

Corresponding authors: Zuzana Stolicna (zuzana.stolicna@fm.uniba.sk) and Sidra Abbas (sidraabbas@ieee.org)

This work was supported by the Deanship of Scientific Research at King Khalid University through the Large Group Research Project under Grant RGP2/470/44.

ABSTRACT Energy providers and the power grid are severely harmed by electricity theft, which also causes economic and non-technical losses. Energy theft causes a decline in power quality and overall profitability. Smart grids may address the problem of power theft by merging data and energy flow. The analysis of smart grid data helps to find power theft. The prior methods, however, could have done a better job of identifying energy theft. In this research, we presented an active learning-based machine learning model for energy theft identification and classification of a smart grid. The suggested approach is based on the following steps. We use a dataset from the Open Energy Data Initiative (OEDI), an energy research database that gets information from numerous OEDI offices and labs. Next, we pre-process the data and employ machine learning methods like Active Learning (AL) based Random Forests (RFAL), eXtreme Gradient Boosting (XGboostAL), Decision Tree (DTAL), Gradient Boosting (GBAL), K-Nearest Neighbors (KNNAL), Categorical Boosting (CatboostAL) and Light Gradient Boosting Machine (LGBMAL) classifier. Using the smart grid-based energy theft detection dataset, the proposed RFAL model outperforms competing models and obtains an accuracy of 70.61%. The principles of smart grid tasks streamline decisions and enhance interaction between humans and machines by combining AL with machine learning. The application of this technology in this area has the potential to enhance the accuracy of energy theft detection and electricity-related problems and consequences.

INDEX TERMS Active learning, energy theft detection, machine learning, privacy, smart grid, security.

I. INTRODUCTION

Electricity is needed for our daily existence. Energy losses frequently happen throughout energy generation, transmission, and distribution. The two categories of electrical losses are Technical Losses (TLs) and Non-Technical Losses (NTLs) [1]. The NTLs are defined as distinguishing between total losses and TLs, which mainly occur by electricity theft. The TL is intrinsic to electricity transport and is triggered by inner activities in the power plan's parts, such as the gearbox liner and transformers [2]. This unlawful behavior frequently involves manipulating, tampering, or

The associate editor coordinating the review of this manuscript and approving it for publication was Yiqi Liu.

circumventing electrical meters. These fraudulent electricity practices could result in power companies losing money. Consider the projected $4.5 billion in annual losses from electricity theft in the United States (US) [3]. Electricity theft is assessed to cost utility companies more than 20 billion yearly. Energy points, heavy burdens on electrical systems, major remuneration losses for the power supplier, and risks to general security can all result from electricity theft.

The three energy theft detection tactics categories are the network-oriented approach, the data-oriented approach, and a hybrid technique that merges the two methodologies [4]. The network design must be changed frequently when employing network- and hybrid-oriented resolutions and adding new devices [1]. It is problematic to use these concepts widely because the grid architecture cannot be accessed due to safety considerations, and installing new devices is expensive. Data-oriented strategies increase the efficiency of doubted energy theft detection and evaluation by concentrating solely on the data generated by smart meters and neglecting network models or further devices. Thus, data-driven methods of predicting power theft have become increasingly popular recently [5].

A ''smart grid'' is a system that combines traditional electricity networks with automated communication tools. The smart grid may confirm that electrical energy is utilized effectively, according to previous research [6], [7], [8]. Effectual energy management is essential to accomplishing this goal, which has attracted global attention to the acceptance and commencement of smart cities (SCs) [9]. SGs utilize the Internet of Things (IOTs) by incorporating sensing instruments, actuators, transmission and network devices for sound smart cities (SCs) planning, essentially employing the Internet of Things (IOTs) to deliver advanced, secure, economically inclined, economically driven, effective energy supplies and services [10]. The authors of [11] proposed an architecture to lessen the effects of peak demands while permitting the trade of more power for a lower price. The unpredictable conduct of bearable power is reduced using a technique based on an information-gap decision approach. A smart meter transmits data to and from particular energy consumers and the power grid.

Advanced metering infrastructure (AMI) data is useful for identifying electricity theft [12]. The main methods of detecting power theft are inspecting problematic hardware or equipment, examining unauthorized line diversions, and contrasting malevolent and legitimate meter records. However, these techniques are quite expensive and time-consuming when verifying all the meters in a system. Additionally, these manual methods are unable to prevent cyberattacks. Machine learning and deep learning techniques are examples of artificial intelligence-based methods that are trending nowadays. Classification and clustering models are further subcategories of existing machine learning techniques.

To our best knowledge, this research is the first to apply an active learning-based strategy to theft detection successfully. Customer energy use data is subjected to machine learning and deep learning techniques to discover and spot odd patterns. This work used an artificial intelligence-based methodology and developed an active learning model based on machine learning that uses various classifiers. We focus on utilizing active learning to achieve the following benefits. The first benefit of active learning lies in transforming the traditional, passive data labeling process into a dynamic and intelligent decision-making system. Unlike conventional supervised learning, active learning takes an inquisitive approach by autonomously selecting the most informative data points for labeling, optimizing data acquisition efficiency. This means fewer labeled examples are needed, reducing the cost and time associated with manual labeling. Active learning adapts and evolves with the model's growing knowledge, effectively improving its performance while leveraging domain expertise when required. It ensures diversity in sample selection, reducing overfitting risks, and excels in scenarios where data is scarce, expensive or where the data distribution evolves.

A. RESEARCH CONTRIBUTION

This study makes several important contributions, including:

- Proposed an active learning-based machine learning model for energy theft identification and classification using a smart grid-based dataset.
- Pre-processing is carried out to improve the accuracy of analyzing and categorizing energy theft data. Eliminating noise and converting the categorical data into numerical using a well-known, one-hot encoding in energy theft classes.
- The accuracy of the suggested method can be significantly improved by incorporating pooled-based active learning approaches with machine learning algorithms. The active learning framework enhances prediction skills by selecting insightful samples for manual annotation and iteratively improving the models. Together, they guarantee a more focused and efficient method, increasing precision for the task.
- ALRF classifier performed well, outperforming the outcomes of the conventional method and obtaining an accuracy of 70.61%. This significant improvement highlights the random forest algorithm's effectiveness and potential to improve classification accuracy in various applications.

B. RESEARCH ORGANIZATION

To make understanding easier, the paper is divided into separate sections. Section II presents an extensive overview of earlier studies. A summary of the dataset selection procedure and the proposed approach is presented in Section III. Section IV thoroughly analyses the experiments and accompanying results. Section V concludes the paper and suggests future research directions.

II. LITERATURE REVIEW

This section looks further into the exhaustive details of prior scholarly research on energy theft detection using

smart grids. It has been split into three sections exploring a distinct strategy. These strategies include machine learning, deep learning, and ensemble methods based on smart grids detecting energy theft.

A. THEFT DETECTION USING MACHINE LEARNING ALGORITHMS

Authors in [13] proposed a machine learning-based approach to automated theft identification in a smart grid setting. The OEDI platform owns the publicly available data, which is utilized to model and create a multi-class theft detection dataset. The authors used multiple classifiers with several assessment alternatives (mechanisms) to evaluate the suggested dataset. According to empirical findings, the theft detection-based RF model outperforms other developed models in performance metrics by 10% or more. The author in [14] proposed an ensemble machine-learning model for identifying energy theft in smart grids using consumer usage trends. Several ML algorithms are examined to determine their false positive and detection rates. The performance of detection is enhanced by using a data pre-processing technique.

To combat over-fitting, the statistical technique of minority over-sampling is also used. Bagging models beat other algorithms, according to a thorough investigation of a real-world dataset of 5,000 clients. The models with more trees and random forests have the highest AUC scores (0.90). The precision study reveals that the suggested bagging techniques outperform existing ones. A detection method that uses statistical and machine learning is presented in this study to gauge the confidence in a theft [15]. An energy theft prediction unit established on a fine tree regression architecture uses an anomaly detection approach to identify questionable data. The suggested approach's training phase uses chronological data on moderate load consumption per unit location, smart meter readings, and temperature. A possibility density process determines doubtful data and calculates the conviction in theft to fit the difference between the true and estimated data. The findings collected show how well the designed detection method works. Author in [16] presented a methodology using randomized tree algorithms to predict electricity theft in smart grids. SMOTE technique is applied to the used dataset to balance the classes, and the hyperparameter of the presented methodology is optimized by applying a grid search optimization approach. The proposed methodology uses different evaluation measurements and achieves 98% accuracy in model evaluation.

B. THEFT DETECTION USING DEEP LEARNING ALGORITHMS

Authors in [17] proposed a technique for predicting electricity theft using smart meter data that tracks energy consumption. This method would help energy supply companies overcome energy scarcity, unforeseen power consumption, and poor power management. They devised the Convolutional Neural Network (CNN). This methodology-based DL algorithm first distinguishes between the periodic energy, and that is not while maintaining the fundamental characteristics of data on power usage. According to the findings, the deep CNN model surpasses earlier models and has the highest level of accuracy for identifying energy theft. The results show that it is possible to recognize atypical unmoving behavior, and an adjustable premises system can do so with high accuracy and long-term usage. This [18] article presents a cost-effective data-driven ETD method that greatly lowers the expenses of data labeling without compromising the accuracy of ETD. The method is implemented systematically using an intellectual deep active learning (DAL) system. The DAL technique effectively chooses the best cases for the ETD model. The effectiveness of the suggested approach is demonstrated by experimental test results using a real ETD dataset provided by the State Grid Corporation of China.

The authors in [19] examined electricity thefts in the distributed generation (DG) space. In this incursion, unscrupulous consumers misuse smart meters, scrutinizing their renewable-based DG units to falsely declare that they have contributed more electricity to the grid and overcharged the utility company. The detection of such harmful behavior is being researched using deep machine learning. To address the issue of electricity theft, the article [20] suggests a deep reinforcement learning (DRL) method that uses samples from real datasets. Multiple alternative scenarios that use the suggested approach are given. A deep Q network (DQN) and a double deep Q network (DDQN) with various deep neural network designs are first used to build a global detection model. Second, using the global detector alters the consumption habits of current customers and increases the difficulty of fending off recently launched cyberattacks. Results show that the proposed DRL approach can effectively learn new consumption patterns. Author in [21] research uses Long and Short Term Memory (LSTM) and Convolutional Neural Network (CNN) to extract abstract features from electricity usage data. The prototype per class, which is utilized to predict the labels of unidentified data, is generated after computing the parameters of the abstract feature. The prototype is symbolic because it trained the network using several balanced portions of the training data. When anomalous data only make up 2.5% and 1.25% of normal data, respectively, the suggested method has been shown to detect electricity theft more successfully than certain standard techniques like CNN, RF, etc. The suggested strategy performs better than other cutting-edge techniques.

C. THEFT DETECTION USING HYBRID MODEL

Authors in [22] investigated the issue of a change and transmitting (CAT) AMI system in which the system operator needs to supply readings of the power consumption regularly. First, process a dataset of actual power usage values to assemble a benign dataset for the CAT AMI. After that, they suggest a fresh series of attacks designed specifically for the CAT AMI to produce a harmful dataset. Then,

to detect fraudulent clients, offer a broad and hybrid DL electricity theft sensor. The presented sensor is trained on benign and harmful data from all customers utilizing the documented CAT measurements. The results of their simulations show that their models have a heightened detection rate and a lower percentage of false alarms when identifying malevolent clients. In [23], the author offers deep learning-based detectors that can successfully block cyber-attacks on smart grid AMI networks that steal electricity. First, they introduce a customer-specific detector based on recurrent neural networks (RNNs) and deep feed-forward networks (DFFN). Then, instead of creating customer-specific detectors, they create generalized electricity theft detectors that are more resistant to contamination attempts. Hyperparameter optimization is researched for all detectors to increase the effectiveness of the developed detectors. Grid search techniques based on sequential, random, and genetic optimization are used in particular to optimize the hyperparameters of the detectors.

Authors in [24] present FedDetect, an innovative federated learning system for energy theft prediction that protects privacy. A system comprises a data center, a control center, and numerous prediction stations. Every detection station in this system can only see data from nearby clients, who can process their data using a local differential privacy approach to maintain privacy. They create a safe protocol so detecting stations can send encrypted training parameters to the CC and the DC, facilitating the model's training. The aggregated parameters are then computed using homomorphic encryption, and the modified model's parameters are sent back to the detection stations. On the cutting-edge temporal convolutional network (TCN), they created a deep learning model to identify energy theft. The results of the experiments show that the suggested federated learning architecture may accomplish high detection accuracy with less computational overhead.

Unlike current approaches, the suggested strategy contains a novel active learning mechanism that uses uncertainty sampling to specify the query strategy. This method demonstrates excellent effectiveness in learning from unlabeled data. A promising new path for improvements in this area is introduced by incorporating active Learning in theft detection using smart grids.

III. PROPOSED METHODOLOGY

The main purpose of the presented methodology is to use active learning to create an accurate predictive model for theft detection. This method uses machine learning algorithms with active learning for implementation and includes several stages, such as dataset preprocessing and model selection. The widely utilized smart grid-based theft detection is used in this proposed method to increase the model's efficacy. Figure 1 shows a graphical illustration of the approach.

The starting point of the proposed method consists of the two processes of dataset preparation and model selection. The dataset with 10 classes for smart grid-based theft detection is used in this investigation. A label encoder is used in dataset preparation to apply one hot coding technique to classes. An AL-based ML model is trained to determine theft in the smart grid. The proposed Algorithm 1 is an active learning pipeline using machine learning methods to classify text input according to its intended use. Data preprocessing, which includes label encoding using one hot encoding technique, is the first step in the procedure. The pipeline introduces four classifiers: the decision tree, gradient boosting, catboost, K-nearest neighbor, and LGBM. These classifiers implement the active learning methodology by creating an ActiveLearner object. With this method, the theft detection class is trained and classified iteratively by examining the most ambiguous data examples. In a three-iteration iteration count, the samples are labeled according to the classes to which they belong. The accuracy, macro-averaged F1 score, and weighted average F1 score of each classifier are then assessed on the test data. Each classifier receives a final classification report and evaluation data for all classifiers, allowing for performance comparison. This process is useful when there is a lack of labeled data, and iterative labeling of more data samples is required to improve classification performance.

A. DATASET DESCRIPTION

OEDI is the main source of the data, an energy research database compiling information from numerous OEDI offices and labs [25]. The original data (12 months) includes measurements of energy use from diverse users throughout the year. Readings are taken at regular intervals. The dataset includes sixteen consumer categories, each with a different energy usage pattern. Dataset shapes for the training dataset are 22,330 with 10 classes, and for testing, they are 22,330.

B. DATA PREPROCESSING

Dataset preprocessing is an essential phase in data analysis and ML that involves cleaning, converting, and systematizing natural data into a structure appropriate for further analysis or training ML classifiers. Appropriate preprocessing enhances models' quality and effectiveness by addressing missing values, outliers, inconsistent formats, and noise in the data. One-hot encoding creates binary columns (0 or 1) for each category, effectively creating a ''dummy'' variable for each category. StandardScaler is a common data preprocessing technique in machine learning and data analysis. It is a part of the scikit-learn library (sklearn) in Python and is used to standardize or normalize the features of a dataset. Standardization is particularly useful when the features in your dataset have different scales, and you want to bring them to a common scale with a mean of 0 and a standard deviation of 1. In this research, we use these two techniques for data preprocessing.

C. ACTIVE LEARNING

Despite potential fluctuations, it aims to maximize architecture interpretation while undervaluing labeling expenses.

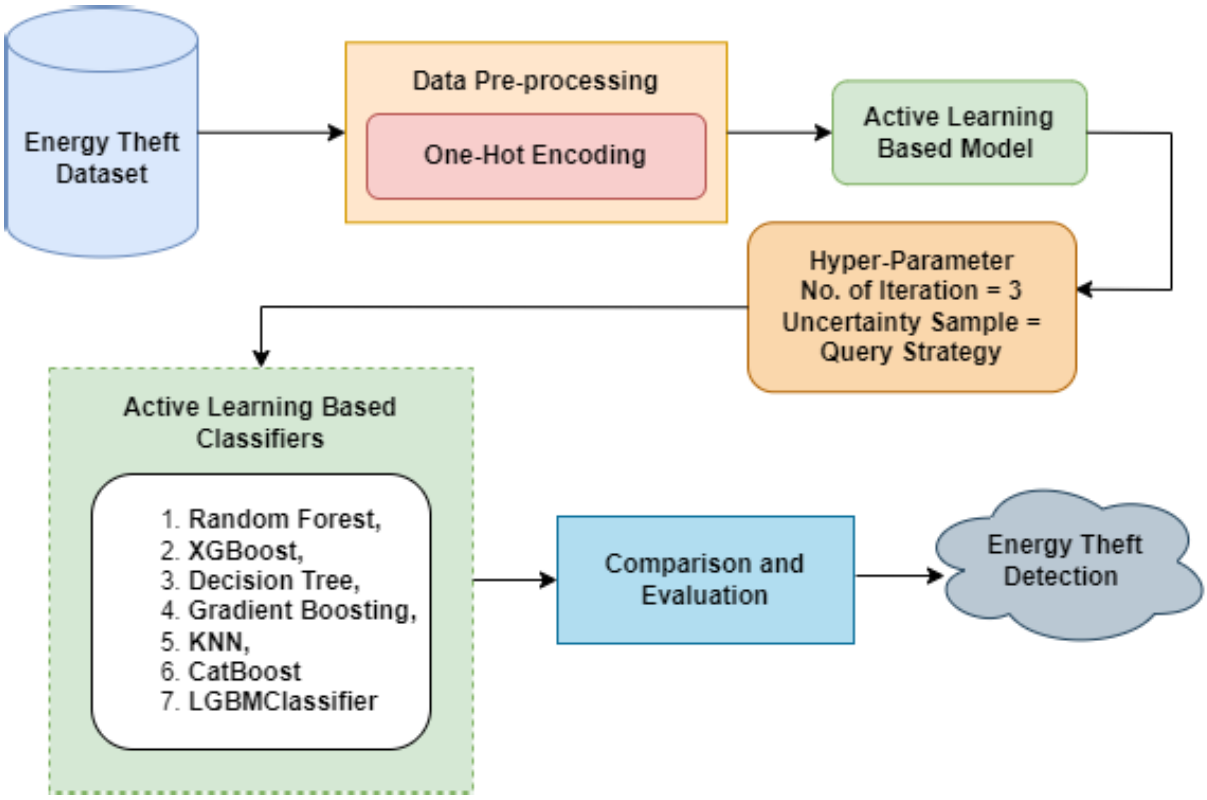

FIGURE 1. Proposed model for the detection of energy theft utilizing active learning.

AL incorporates several sampling techniques, such as uncertainty sampling, query by committee, and data density sampling, to choose samples with distinctive characteristics into a single framework. In pooled-based Active Learning, the semantic annotation is started with a pool of unlabeled instances from the dataset. The most useful data points from this pool are chosen and annotated by a human professional, which causes the model to be retrained.

This iterative procedure is continued until the necessary degree of accuracy is attained. The research uses hyperparameters, particularly ''n_queries'' and ''uncertainty_sample,'' to regulate the active learning process. The ''n_queries'' option controls how often the model iterates, which involves three training processes. The algorithm chooses the most instructive samples for labeling at each iteration. The ''uncertainty_sample'' parameter, configured to a query strategy value, is also thought to be the most instructive in each iteration. This decision is supported by uncertainty sampling, a well-known and successful method for picking data points with high electability. The active learning algorithm improves the model's performance by iteratively choosing and labeling the most relevant samples from the pool of unlabeled data by combining these parameters and methods. The equations 1,2,3,4,5,6 of the pooled-based active learning algorithm [26] reads as follows:

$$g\_it = train\ model\ with(x\_it, y\_it) \tag{1}$$

$$w_{1:it} = g_{it}(x_{1:it}) \tag{2}$$

$$V_{it} = i \epsilon V : argmaxw \epsilon\ yp(w|x_i, w_{1:it}) \tag{3}$$

$$j_{it} = j_{it} \epsilon\ argmax(j) \epsilon\ U_{it} H(p(w|xj_{it}, y_{1:it-1})) \tag{4}$$

$$x_{it+1} = x_{it}\ \upsilon\ xj_{it} \tag{5}$$

$$y_{it+1} = y_{it}\ \upsilon\ yj_{it} \tag{6}$$

where j_it is the index comparing to the considerable illuminating instance from the set, g_it represents the architecture trained on the labeled dataset at iteration it, w_1:it is a vector of the architecture outcome for every labeled sample, V_it illustrates the group of unlabeled samples at iteration it, and H(.) and signifies the entropy. The suggested research employs data preprocessing approaches and an active and unpredictable learning methodology. Three iterations and seven classification algorithms, such as RF, XGBoost, DT, GB, CB, KNN and LGBM, are used to calculate the active learning parameters. RF is an ensemble algorithm based on decision trees and combines the predictions of multiple decision trees to improve accuracy and reduce overfitting. Several parameters we used in this research, such as n_estimators, max_depth,min_samples_split and min_samples_leaf. A DT recursively splits the dataset into subsets based on the most significant attribute at each node, ultimately creating a tree-like structure of decisions. Criterion, splitter and max_depth are used as parameters in this research. KNN is a non-parametric algorithm that makes predictions based on the majority class or mean value of the k-nearest data points to a given query point. This

Algorithm 1 Pseudo Code of Energy Theft Detection Using Active Learning

Input Energy consumption data as input feature, Energy Theft as labels
N_{it} = 3 {Number of Iteration}
$Q_{strategy} = U_{Sample}$ {Query Strategy = Uncertain sample}

Output Energy Theft Detection
$E_{Measure}$: Accuracy, Precision, Recall, F1-Score
Initialization 'a'
List of Algorithms: Random Forest, XGBoost, Decision Tree, Gradient Boosting, KNN, CatBoost and LGBM
'Alg_names' represents their names, and 'algorithms' provide a list
Split training and testing sets.
for Algorithm in algorithms **do**
    Set the Active Learner's initialization parameters as follows:
    Create an Active Learner object on training data.
    Create a custom_query_strategy method called X_pool that accepts a pool of unlabeled samples as input.
    The tasks give back the function uncertainty_sampling with the active learner, X_pool
    Set 'n_iterations' to 3 and give it the value ''active learning iterations.''
    for Each Iteration(n) **do**
        Increment iteration
        Command ''custom_query_strategy'' with ''X_test'' as the parameter and allocate the outcome to ''query_idx''
        choose ''X_test'' and ''y_test'' established on the indicator in ''query_idx''.
        X_pool and y_pool should be given the resulting values accordingly
        Take as an input the predictions for the cases in the ''X_pool'' that are the most uncertain
        Transfer the values to the ''y_pred_pool'' variable.
        With the inputs ''X_pool'' and ''y_pred_pool,'' add the ambiguous samples and their anticipated labels to the ''learner'' object.
    end for
    Provide the overall classifier assessment metrics
    For every classifier, generate their final classification results.
end for

research used n_neighbors and weights parameters in the KNN algorithm.[1] Gradient Boosting builds an ensemble of decision trees to make predictions.[2] It is known for its high predictive accuracy and ability to handle complex relationships in data. XGBoost combines the predictions of multiple decision trees to create a more accurate and robust model, and the parameters used in this research are learning_rate, max_leaf_nodes, min_samples_leaf and l2_regularization. CatBoost is particularly designed for categorical feature support and is known for its efficiency and ease of use. The parameters used in this research are num_of_iterations, learning_rate and depth. LightGBM is a gradient-boosting framework designed for efficiency and speed. boosting_type, num_leaves and learning_rate are used parameters in this research. The method gradually expands the defined dataset while removing unknown samples by adding annotated examples from the test set. The purpose of pragmatic performance metrics is to efficiently identify and rank critical aspects of classifier performance.

[1]https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html

[2]https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.GradientBoostingClassifier.html

IV. EXPERIMENTAL RESULT AND DISCUSSION

This section thoroughly examines the application of an active learning strategy that blends machine learning techniques to the smart-grid-theft-detection dataset. The data set is split into 20% for model testing and 80% for model training. This model uses the strength of machine learning classifiers to learn from the provided dataset. Additionally, several metrics, including accuracy, precision, recall, F1-score, and confusion matrix, are used to assess this model's performance. These metrics all demonstrate the model's efficacy. This section describes the test results, evaluates them, and offers a thorough and insightful interpretation.

A. EVALUATION MEASUREMENTS

This study assesses the model's efficacy using many evaluation criteria, including accuracy, precision, recall, F1 score, and confusion matrix, which are crucial and offer insightful information about the model's performance. The first statistic is accuracy, which is frequently considered the foundation of performance evaluation. With the total number of samples, it calculates the percentage of correctly classified samples. The following equation 7 makes this simple to understand. Despite being calculated relatively uncomplicated, the measure has much significance.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{7}$$

Since accuracy represents the proportion of correctly positive predictions to all positive predictions made by the model, it is a crucial evaluation parameter used in performance evaluation. Equation 8 illustrates this value proportionally, enabling a precise understanding of the metric conceptual equation.

$$Precision = \frac{TP}{TP + FP} \tag{8}$$

Recall, often called sensitivity, offers an evaluation metric that concentrates on the proportion of accurate positive predictions to all positive cases. This balanced viewpoint offers a special benefit throughout the estimation process,

as seen by the calculation in Equation 9. This equation is an excellent example of the value of recall as an intuitive indicator for assessing model performance.

$$Recall = \frac{TP}{TP + FN} \tag{9}$$

The appropriately titled F1 score serves as a harmonic mean of memory and precision since it can accurately convey the underlying meaning of balanced performance. The F1-score, which combines these two measurements, is a frequently used estimation of model performance and is particularly helpful during evaluation. Equation 10, which appears complex but offers much insight, accurately describes this fundamental estimation computation.

$$F1 - score = 2 \times \frac{Precision + Recall}{Precision + Recall} \tag{10}$$

The confusion matrix, which is meticulously built to give precise insight into the effectiveness of the classification model, is a notable and distinctive metric utilized in the evaluation process. This crucial tool expertly contrasts the actual and anticipated numbers to demonstrate the model's effectiveness. The confusion matrix is a special type of display that shows four values: true positive (TP), true negative (TN), false positive (FP), and false negative (FN). The columns of the matrix correlate to the actual class labels, whereas the rows indicate the actual class labels. The correctly identified examples are situated along the diagonal, whereas the incorrectly classified cases are centered on the diagonal elements, which is an intriguing characteristic. The values in the confusion matrix serve as a useful evaluation tool for identifying model strengths and shortcomings, resulting in insights that considerably improve the model and yield positive results.

B. ANALYSIS AND RESULTS

The study assessed the accuracy and F1-score of seven active learning-based classifiers for detecting energy theft using smart grids: RF-AL, XGBoost-AL, DT-AL, GB-AL, KNN-AL, CB-AL, and LGBM-AL. Three iterations of each classifier are examined to ascertain its accuracy rate. Compare the outcomes of active learning with straightforward models as well. Table 1 gives a detailed presentation of the experimental macro-averaged metrics outcomes for AL.

The findings reveal that at iteration 2, the RFAL model has the greatest accuracy (0.7061) and f1-score (0.6952). XGBoostAL obtained the greatest accuracy of 0.6913 at iteration 2 with a f1-score of 0.6734. DTAL attained the greatest accuracy of 0.6490 at iteration 2 with a f1-score of 0.6483. GBAL model attained the maximum accuracy of 0.6151 with a f1-score of 0.5897 at iteration 2. KNNAL model attained the greatest accuracy of 0.6242 with an f1-score of 0.6123 at iteration 3. At iteration 3, CBAL attained the best accuracy of 0.6762 with a f1-score of 0.6540.

Table 2 displays RFAL results on weighted average metrics. The table includes the evaluation measurement values. The model column represents the proposed model result on classes from "RFAL #0" to "RFAL #6." RFAL #2 performs exceptionally well, achieving good results in precision, recall, and an F1-score with a value of 1.00. RFAL #6 performance is very low as it achieves precision with a value of 0.43, recall of 0.33, and an F1-score of 0.37, while RFAL #0, #1, #3, #4, and #5 achieve precision, recall, and F1-score between the range of 0.45 to 0.97. RFAL#5 achieved good precision, recall and f1-score.

TABLE 1. Result based on micro average metrics.

Model	Accuracy	F1-Score
RFAL # 1	0.7044	0.6930
RFAL #2	0.7061	0.6952
RFAL # 3	0.7029	0.6917
XGBoostAL # 1	0.6905	0.6722
XGBoostAL #2	0.6913	0.6734
XGBoostAL # 3	0.6903	0.6720
DTAL # 1	0.6475	0.6470
DTAL #2	0.6490	0.6483
DTAL # 3	0.6437	0.6427
GBAL # 1	0.6125	0.5867
GBAL #2	0.6151	0.5897
GBAL # 3	0.6137	0.5884
KNNAL # 1	0.6240	0.6119
KNNAL #2	0.6239	0.6117
KNNAL # 3	0.6242	0.6123
CBAL # 1	0.6815	0.6602
CBAL #2	0.6727	0.6505
CBAL # 3	0.6762	0.6540
LGBMAL # 1	0.6865	0.6653
LGBMAL #2	0.6904	0.6700
LGBMAL # 3	0.6879	0.6673

TABLE 2. Random forest result on weighted average metrics.

Model	Precision	Recall	F1-Score
RFAL # 0	0.67	0.71	0.69
RFAL # 1	0.53	0.49	0.512
RFAL # 2	1.00	1.00	1.00
RFAL # 3	0.51	0.45	0.48
RFAL # 4	0.74	0.90	0.81
RFAL # 5	0.88	0.97	0.92
RFAL # 6	0.43	0.33	0.37

TABLE 3. XGBoost result on weighted average metrics.

Model	Precision	Recall	F1-Score
XGBoostAL # 0	0.68	0.82	0.75
XGBoostAL # 1	0.48	0.45	0.47
XGBoostAL # 2	1.00	1.00	1.00
XGBoostAL # 3	0.48	0.39	0.43
XGBoostAL # 4	0.67	0.87	0.76
XGBoostAL # 5	0.88	0.97	0.92
XGBoostAL # 6	0.45	0.25	0.32

Table 3 provides XGBoosstAL results on weighted average metrics. The table includes the proposed model's precision, recall, and F1-score values. The model column represents the proposed model result on classes from "XGBoostAL #0" to "XGBoostAL #6." XGBoostAL #2 performs exceptionally well, achieving good results in precision, recall, and an

F1-score with a value of 1.00. XGBoostAL #6 performance is very low as it achieves precision with a value of 0.45, recall of 0.25, and an F1-score of 0.32, while XGBoostAL #0, #1, #3, #4, and #5 achieve precision, recall, and F1-score between the range of 0.39 to 0.97. XGBoostAL#5 has good precision, recall and f1-score.

TABLE 4. Decision tree result on weighted average metrics.

Model	Precision	Recall	F1-Score
DTAL # 0	0.63	0.64	0.64
DTAL # 1	0.44	0.42	0.43
DTAL # 2	1.00	1.00	1.00
DTAL # 3	0.45	0.41	0.43
DTAL # 4	0.73	0.74	0.73
DTAL # 5	0.88	0.90	0.89
DTAL # 6	0.32	0.34	0.33

Table 4 displays DTAL results on weighted average metrics. The table includes the proposed model's precision, recall, and F1-score values. The model column represents the proposed model result on classes from "DTAL #0" to "DTAL #6." DTAL #2 performs exceptionally well, achieving good results in the precision, recall, and an F1-score with a value of 1.00. DTAL #6 performance is very low as it achieves precision with a value of 0.32, recall of 0.34, and an F1-score of 0.33, while DTAL #0, #1, #3, #4, and #5 achieve precision, recall, and F1-score between the range of 0.41 to 0.90. DTAL#5 has good precision, recall and f1-score.

TABLE 5. Gradient boosting result on weighted average metrics.

Model	Precision	Recall	F1-Score
GBAL # 0	0.62	0.76	0.68
GBAL # 1	0.40	0.33	0.37
GBAL # 2	1.00	1.00	1.00
GBAL # 3	0.38	0.25	0.30
GBAL # 4	0.54	0.77	0.64
GBAL # 5	0.74	0.90	0.81
GBAL # 6	0.36	0.20	0.26

Table 5 displays GBAL results on weighted average metrics. The table includes the proposed approach's precision, recall, and F1-score values. The model column represents the proposed approach results on classes from "GBAL #0" to "GBAL #6." GBAL #2 performs exceptionally well, achieving good results in precision, recall, and an F1-score with a value of 1.00. GBAL #6 performance is very low as it achieves precision with a value of 0.36, recall of 0.20, and an F1-score of 0.26, while GBAL #0, #1, #3, #4, and #5 achieve precision, recall, and F1-score between the range of 0.25 to 0.90. GBAL#5 has good precision, recall and f1-score.4 balance recall and f1-score well. GBAL#5 has good recall and f1-score.

Table 6 displays KNNAL results on weighted average metrics. The table includes the proposed approach's precision, recall, and F1-score values. The model column represents the proposed approach results on classes from "KNNAL #0" to "KNNAL #6." KNNAL #2 performs exceptionally well,

TABLE 6. K-nearest neighbour result on weighted average metrics.

Model	Precision	Recall	F1-Score
KNNAL # 0	0.58	0.66	0.61
KNNAL # 1	0.39	0.43	0.41
KNNAL # 2	1.00	1.00	1.00
KNNAL # 3	0.41	0.34	0.37
KNNAL # 4	0.69	0.73	0.71
KNNAL # 5	0.80	0.91	0.85
KNNAL # 6	0.34	0.22	0.27

achieving good results in precision, recall, and an F1-score with a value of 1.00. KNNAL #6 performance is very low as it achieves precision with a value of 0.34, recall of 0.22, and an F1-score of 0.27, while KNNAL #0, #1, #3, #4, and #5 achieve precision, recall, and F1-score between the range of 0.34 to 0.91. KNNAL#5 has good precision, recall and f1-score.

TABLE 7. CatBoost result on weighted average metrics.

Model	Precision	Recall	F1-Score
CBAL # 0	0.67	0.79	0.72
CBAL # 1	0.44	0.41	0.42
CBAL # 2	1.00	1.00	1.00
CBAL # 3	0.44	0.35	0.39
CBAL # 4	0.67	0.90	0.77
CBAL # 5	0.83	0.97	0.90
CBAL # 6	0.46	0.24	0.32

Table 7 displays CBAL results on weighted average metrics. The table includes the proposed approach's precision, recall, and F1-score values. The model column represents the proposed approach result on classes from "CBAL #0" to "CBAL #6." CBAL #2 performs exceptionally well, achieving good results in precision, recall, and an F1-score with a value of 1.00. CBAL #6 performance is very low as it achieves precision with a value of 0.46, recall of 0.23, and an F1-score of 0.32, while CBAL #0, #1, #3, #4, and #5 achieve precision, recall, and F1-score between the range of 0.39 to 0.97. CBAL#5 has good precision, recall and f1-score.

TABLE 8. LGBM result on weighted average metrics.

Model	Precision	Recall	F1-Score
LGBMAL # 0	0.67	0.82	0.74
LGBMAL # 1	0.48	0.46	0.47
LGBMAL # 2	1.00	1.00	1.00
LGBMAL # 3	0.47	0.37	0.41
LGBMAL # 4	0.67	0.88	0.76
LGBMAL # 5	0.88	0.98	0.93
LGBMAL # 6	0.43	0.23	0.30

Table 8 provided the LGBMAL results on weighted average metrics. The table includes the proposed approach's precision, recall, and F1-score values. The model column represents the proposed approach results on classes from "LGBMAL #0" to "LGBMAL #6." LGBMAL #2 performs exceptionally well, achieving good results in precision, recall, and an F1-score with a value of 1.00. LGBML #6

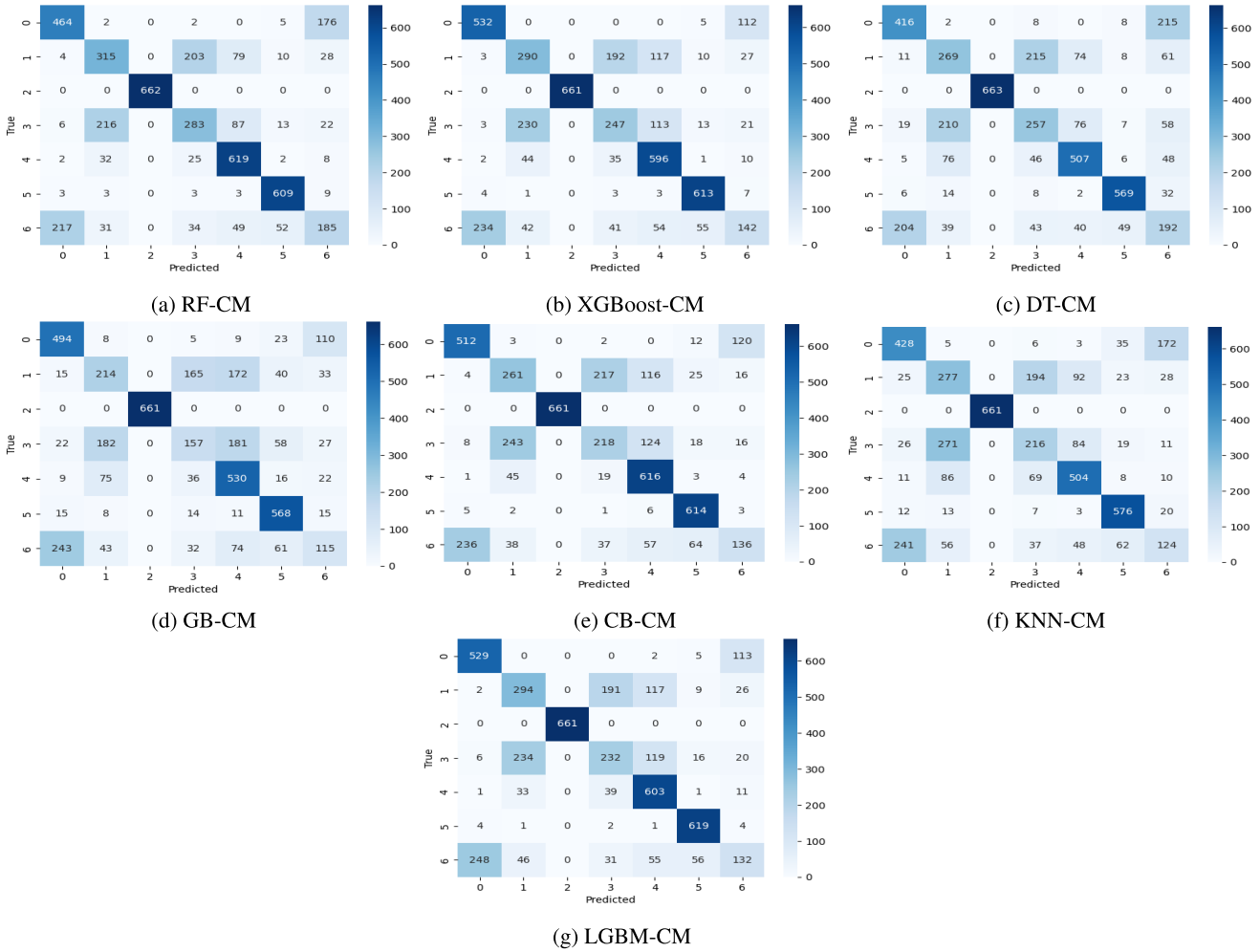

FIGURE 2. (a) CM for RFAL (b) CM for XGBoostAL(c) CM for DTAL (d) CM for GBAL (e) CM for CBAL (f) CM for KNNAL (g) CM for LGBMAL.

performance is very low as it achieves precision with a value of 0.43, recall of 0.23, and an F1-score of 0.30, while LGBMAL #0, #1, #3, #4, and #5 achieve precision, recall, and F1-score between the range of 0.37 to 0.93. LGBMAL#5 outperforms precision, recall and f1-score.

Figure 2 displays the confusion matrix (CM) for active learning-based algorithms. Figure 2a shows the CM for an RFAL, and for XGBoostAL is displayed in Figure 2b. Figure 2c shows the CM for a DTAL, and for GBAL is shown in Figure 3d. The CM for the CBAL is shown in Figure 2e and for KNNAL is shown in Figure 2f and for the LGBMAL is shown in Figure 2g. The Y-axis displays the actual labels, and the X-axis displays the expected labels.

RF accurately predicted 464 occurrences of label 0, 325 occurrences of label 1, 662 occurrences of label 2, 283 occurrences of label 3, 619 occurrences of label 4, 609 occurrences of label 5, and 185 occurrences of label 6. The cases that XGBoost correctly predicted instances included 532 for label 0, 290 for label 1, 661 for label 2, 247 for label 3, 596 for label 4, 613 for label 5, and 142 for label 6. The DT scored admirably, making 416 accurate predictions for label 0, 269 for label 1, 663 for label 2, 257 for label 3, 502502 for label 4, 569 for label 5, and 192 for label 6, among other labels. GB produced 469 accurate predictions for label 0, 214 for label 1, 661 for label 2, 157 for label 3, 530 for label 4, 568 for label 5, and 115 for label 6. The KNN demonstrated case prediction with 428 for label 0, 277 for label 1, 661 for label 2, 216 for label 3, 504 for label 4, 576 for label 5, and 124 for label 6. For labels 0, 1, 2, 3, 6, 14, 5, and 6, CB obtained 512 precise predictions, 261 for label 1, 661 for label 2, 218 for label 3, and 136 for label 6. In the end, the LGBM classifier attained 132 for label 6 and 529 correct predictions for label 0, 294 for label 1, 661 for label 2, 232 for label 3, 603 for label 4, and 619 for label 5.

Essential evaluation measures that set new standards in learning-based classification tasks include the Receiver Operating Characteristic (ROC) curve. Due to their unchanging capacity to comprehend classifying model performance, these measures are widely acknowledged. To this purpose, the ROC curve reveals, for various threshold levels, an asso-

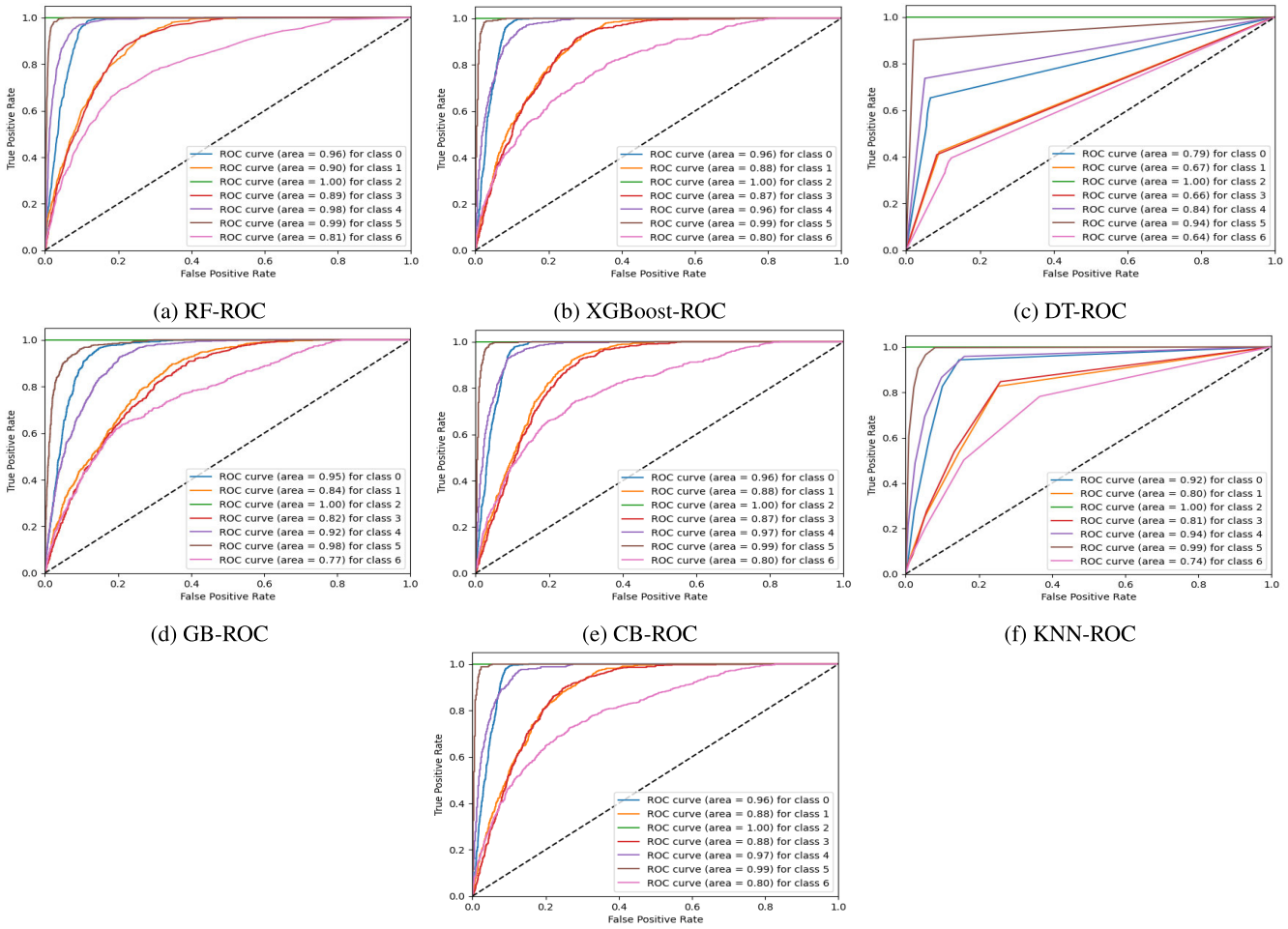

FIGURE 3. (a) ROC curve for RFAL (b) ROC curve for XGBoostAL (c) ROC curve for DTAL (d) ROC curve for GBAL (e) ROC curve for CBAL (f) ROC curve for KNNAL (g) ROC curve for LGBMAL.

ciation between the TPR and the FPR. The AUC completely embraces the genuine value in its entirety by performing on an aggregate of the two-dimensional domain discovered under the ROC curve. The best algorithm has a pronounced ROC curve that crosses the graph's upper-left intersection.

Contrary to this, an AUC of 0 would be required for an algorithm with an ROC curve meeting the bottom-right intersection. The ROC Curves for RFAL are displayed in Figure 3a and for XGBoostAL are displayed in Figure 3b. The ROC Curves for DTAL are displayed in Figure 3c and for GBAL are displayed in Figure 3d. The ROC Curves for CBAL are displayed in Figure 3e and for KNNAL are displayed in Figure 3f. The ROC Curves for the LGBMAL are displayed in Figure 3g. The TPR is shown on the Y-axis, and the FPR is on the X-axis. Effective AUC values for RF varied from 0.96 to 0.99. AUC values of 1.00 are displayed by RF, XGBoost, DT, GB, CB, KNN and LGBM, all performing well in making accurate and trustworthy predictions. These encouraging outcomes show how this methodology can manage complex, high-dimensional databases.

V. CONCLUSION AND FUTURE DIRECTION

Dataset-based energy theft detection has gained several noteworthy advantages due to integrating smart grids with active learning-based algorithms. In addition to addressing label encoding datasets, applying machine learning models and data preprocessing approaches improves the generalization capacities of the algorithms. The models become more resilient and flexible by training on various preprocessed data. The classification models can also be improved iteratively by adopting active learning methodologies. To use these tactics, informative samples must be chosen for manual labeling. Human knowledge must also be incorporated into training, and the algorithm's performance must be updated and improved constantly. By using a dynamic approach, the models are kept current and accurate. This paradigm's good accuracy (70.06%) suggests that it can minimize false positives and negatives in energy theft identification.

The suggested active learning-based classification system offers intriguing future research and application directions when integrated with electricity smart grids. These involve

expanding the model to classify additional smart grid conditions, continuously enhancing the model with feedback and collaboration, strengthening the explainability of the classification model, and combining additional data sources for a thorough understanding of energy theft. They also include integrating the system into electricity platforms for monitoring and detecting, real-time monitoring using various devices, and expanding the model to classify other smart grid conditions. The only limitation of this research is that it has low accuracy compared with other studies that could be addressed by combining deep learning algorithms with active learning-based methodologies; these future possibilities can advance the detection of energy theft and enhance results. Improved energy theft detection in smart grids can lead to various economic benefits, including increased revenue, reduced operational costs, and enhanced grid reliability and efficiency. However, economic consequences, such as initial investments, privacy concerns, and regulatory challenges, must be carefully managed. Balancing these factors is essential to maximize the economic advantages of smart grid technology while minimizing potential drawbacks.

REFERENCES

[1] R. Jiang, R. Lu, Y. Wang, J. Luo, C. Shen, and X. Shen, "Energy-theft detection issues for advanced metering infrastructure in smart grid," *Tsinghua Sci. Technol.*, vol. 19, no. 2, pp. 105–120, Apr. 2014.

[2] S. McLaughlin, B. Holbert, A. Fawaz, R. Berthier, and S. Zonouz, "A multi-sensor energy theft detection framework for advanced metering infrastructures," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 7, pp. 1319–1330, Jul. 2013.

[3] P. McDaniel and S. McLaughlin, "Security and privacy challenges in the smart grid," *IEEE Secur. Privacy*, vol. 7, no. 3, pp. 75–77, May 2009.

[4] G. M. Messinis and N. D. Hatziargyriou, "Review of non-technical loss detection methods," *Electr. Power Syst. Res.*, vol. 158, pp. 250–266, May 2018.

[5] P. Glauner, N. Dahringer, O. Puhachov, J. A. Meira, P. Valtchev, R. State, and D. Duarte, "Identifying irregular power usage by turning predictions into holographic spatial visualizations," in *Proc. IEEE Int. Conf. Data Mining Workshops (ICDMW)*, Nov. 2017, pp. 258–265.

[6] H. Gul, N. Javaid, I. Ullah, A. M. Qamar, M. K. Afzal, and G. P. Joshi, "Detection of non-technical losses using SOSTLink and bidirectional gated recurrent unit to secure smart meters," *Appl. Sci.*, vol. 10, no. 9, p. 3151, Apr. 2020.

[7] M. Adil, N. Javaid, U. Qasim, I. Ullah, M. Shafiq, and J.-G. Choi, "LSTM and bat-based RUSBoost approach for electricity theft detection," *Appl. Sci.*, vol. 10, no. 12, p. 4378, Jun. 2020.

[8] M. Nazari-Heris, M. A. Mirzaei, B. Mohammadi-Ivatloo, M. Marzband, and S. Asadi, "Economic-environmental effect of power to gas technology in coupled electricity and gas systems with price-responsive shiftable loads," *J. Cleaner Prod.*, vol. 244, Jan. 2020, Art. no. 118769.

[9] A. Chojecki, M. Rodak, A. Ambroziak, and P. Borkowski, "Energy management system for residential buildings based on fuzzy logic: Design and implementation in smart-meter," *IET Smart Grid*, vol. 3, no. 2, pp. 254–266, Apr. 2020.

[10] A. O. Otuoze, M. W. Mustafa, O. O. Mohammed, M. S. Saeed, N. T. Surajudeen-Bakinde, and S. Salisu, "Electricity theft detection by sources of threats for smart city planning," *IET Smart Cities*, vol. 1, no. 2, pp. 52–60, Dec. 2019.

[11] H. R. Gholinejad, A. Loni, J. Adabi, and M. Marzband, "A hierarchical energy management system for multiple home energy hubs in neighborhood grids," *J. Building Eng.*, vol. 28, Mar. 2020, Art. no. 101028.

[12] 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," *Knowl.-Based Syst.*, vol. 71, pp. 376–388, Nov. 2014.

[13] S. Zidi, A. Mihoub, S. M. Qaisar, M. Krichen, and Q. A. Al-Haija, "Theft detection dataset for benchmarking and machine learning based classification in a smart grid environment," *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 35, no. 1, pp. 13–25, Jan. 2023.

[14] S. K. Gunturi and D. Sarkar, "Ensemble machine learning models for the detection of energy theft," *Electr. Power Syst. Res.*, vol. 192, Mar. 2021, Art. no. 106904.

[15] A. Ali, M. Mokhtar, and M. F. Shaaban, "Theft cyberattacks detection in smart grids based on machine learning," in *Proc. 5th Int. Conf. Commun., Signal Process., their Appl. (ICCSPA)*, Dec. 2022, pp. 1–4.

[16] S. Y. Appiah, E. K. Akowuah, V. C. Ikpo, and A. Dede, "Extremely randomised trees machine learning model for electricity theft detection," *Mach. Learn. with Appl.*, vol. 12, Jun. 2023, Art. no. 100458.

[17] E. U. Haq, C. Pei, R. Zhang, H. Jianjun, and F. Ahmad, "Electricity-theft detection for smart grid security using smart meter data: A deep-CNN based approach," *Energy Rep.*, vol. 9, pp. 634–643, Mar. 2023.

[18] L. Zhu, W. Wen, J. Li, C. Zhang, B. Zhou, and Z. Shuai, "Deep active learning-enabled cost-effective electricity theft detection in smart grids," *IEEE Trans. Ind. Informat.*, 2023.

[19] M. Ismail, M. F. Shaaban, M. Naidu, and E. Serpedin, "Deep learning detection of electricity theft cyber-attacks in renewable distributed generation," *IEEE Trans. Smart Grid*, vol. 11, no. 4, pp. 3428–3437, Jul. 2020.

[20] A. T. El-Toukhy, M. M. Badr, M. M. E. A. Mahmoud, G. Srivastava, M. M. Fouda, and M. Alsabaan, "Electricity theft detection using deep reinforcement learning in smart power grids," *IEEE Access*, vol. 11, pp. 59558–59574, 2023.

[21] X. Sun, J. Hu, Z. Zhang, D. Cao, Q. Huang, Z. Chen, and W. Hu, "Electricity theft detection method based on ensemble learning and prototype learning," *J. Mod. Power Syst. Clean Energy*, vol. 2023, pp. 1–12, Jan. 2023.

[22] M. I. Ibrahem, S. Abdelfattah, M. Mahmoud, and W. Alasmary, "Detecting electricity theft cyber-attacks in CAT AMI system using machine learning," in *Proc. Int. Symp. Netw., Comput. Commun. (ISNCC)*, Oct. 2021, pp. 1–6.

[23] M. Nabil, M. Ismail, M. Mahmoud, M. Shahin, K. Qaraqe, and E. Serpedin, "Deep learning-based detection of electricity theft cyber-attacks in smart grid AMI networks," in *Deep Learning Applications for Cyber Security*, 2019, pp. 73–102.

[24] M. Wen, R. Xie, K. Lu, L. Wang, and K. Zhang, "FedDetect: A novel privacy-preserving federated learning framework for energy theft detection in smart grid," *IEEE Internet Things J.*, vol. 9, no. 8, pp. 6069–6080, Apr. 2022.

[25] J. B. Leite and J. R. S. Mantovani, "Detecting and locating non-technical losses in modern distribution networks," *IEEE Trans. Smart Grid*, vol. 9, no. 2, pp. 1023–1032, Mar. 2018.

[26] S. Tong, *Active Learning: Theory and Applications*. Stanford, CA, USA: Stanford Univ., 2001.

SIDRA ABBAS (Graduate Student Member, IEEE) received the B.S. degree from the Department of Computer Science, COMSATS University Islamabad, Islamabad, Pakistan. Her research interests include computer forensics, machine learning, criminal profiling, software watermarking, intelligent systems, and data privacy protection.

IMEN BOUAZZI was born in Kasserine, Tunisia, in 1988. She received the engineering degree in applied science in technology (specialty-electronic and microelectronics) from the Higher Institute of Computer Science and Mathematics of Monastir, Tunisia, in 2013, and the Ph.D. degree in science and technology from the University of Monastir, Tunisia, in 2018. She is currently with the Department of Industrial Engineering, King Khalid University, Saudi Arabia. Her research interest includes wireless technology management.

STEPHEN OJO received the B.Sc. degree (Hons.) in electrical and electronics engineering from The Federal University of Technology Akure, Nigeria, in 2014, and the M.Sc. and Ph.D. degrees in information systems degree from Girne American University, Cyprus, in 2017 and 2021, respectively. He is currently an Assistant Professor with the Department of Electrical and Computer Engineering, College of Engineering, Anderson University, Anderson, SC, USA. Before Joining Anderson University, he was a Lecturer with Girne American University, Cyprus, where he taught courses in distributed computing, advanced programming, and electric circuits. He was also a Research Scholar with Vodafone Telecommunication Company, Cyprus, where he developed a multiplicative-based model for signal propagation in wireless networks. He is a full-time Faculty Member with Anderson University, where he teaches computer programming, electric circuits, machine learning, artificial intelligence in wireless mobile networks, and biomedical applications. He has authored and coauthored several peer-reviewed journals. His research interests include wireless networks, machine learning for wireless mobile networks, machine learning, and AI in biomedical devices. He was awarded the Mobil and Full Ph.D. Scholarships throughout the undergraduate program.

GABRIEL AVELINO SAMPEDRO (Senior Member, IEEE) is currently with the Faculty of Information and Communication Studies, University of the Philippines Open University, Philippines. His research interests include the Internet of Things and artificial intelligence.

AHMAD S. ALMADHOR (Senior Member, IEEE) received the B.S.E. degree in computer science from Jouf University (formerly Al-Jouf College), Al-Jouf, Saudi Arabia, in 2005, the M.E. degree in computer science and engineering from the University of South Carolina, Columbia, SC, USA, in 2010, and the Ph.D. degree in electrical and computer engineering from the University of Denver, Denver, CO, USA, in 2019. From 2006 to 2008, he was a Teaching Assistant and the College of Sciences Manager, and then a Lecturer with Jouf University, from 2011 to 2012. Then, he became a Senior Graduate Assistant and a Tutor Advisor with the University of Denver, in 2013 and 2019. He is currently an Assistant Professor of CEN and the VD of the College of Computer and Information Science, Jouf University. His research interests include AI, blockchain, networks, smart and microgrid cyber security, integration, image processing, video surveillance systems, PV, EV, machines, and deep learning. His awards and honors include the Jouf University Scholarship (Royal Embassy of Saudi Arabia in D.C.) and the Al-Jouf Governor Award for Excellence.

ABDULLAH AL HEJAILI received the bachelor's degree in computer science from the Tabuk Teachers College, Saudi Arabia, in 2007, and the master's degree in computer science from CLU, USA, in 2011. He is currently pursuing the Ph.D. degree with the Informatics School, University of Sussex. He is a Lecturer in computer science with the University of Tabuk. His research interests include technology-enhanced learning, image processing, virtual, augmented reality, motion capture, and education applications.

ZUZANA STOLICNA received the Ph.D. degree from Comenius University in Bratislava, in 2013. The subject of the habilitation thesis was the development of the economic policy of the Slovak Republic from the transformation period to the present. She is currently a Docent with the Department of Economics and Finance, Faculty of Management, Comenius University in Bratislava. In addition to pedagogy, she is also engaged in scientific and publishing activities. She has more than 20 years of experience in teaching economics with the University. She has written several university textbooks, contributions at domestic and foreign conferences, and magazine articles. As for research projects, she was a member of the research team of the VEGA projects.