

SecureGrid: A Robust Voting Ensemble Model for Non-Technical Loss (NTL) Detection in Smart Meter Data

Md. Sabbir Hossain Molla

Dept. of EECE

Military Institute of Science and Technology
Dhaka, Bangladesh

Email: sabbir.eece@gmail.com

Md. Tanjil Islam Aronno

Dept. of EECE

Military Institute of Science and Technology
Dhaka, Bangladesh

Email: aronnondc@gmail.com

Md Tanvir Rahman Sahed

Dept. of EECE

Military Institute of Science and Technology
Dhaka, Bangladesh

Email: tanvir.sahed00@gmail.com

Md Nasif Mahmud Zishan

Dept. of EECE

Military Institute of Science and Technology
Dhaka, Bangladesh

Email: nasif.jisan@gmail.com

A K M Nazrul Islam

Dept. of EEE

University of Asia Pacific
Dhaka, Bangladesh

Email: nazrul@eece.mist.ac.bd

Abstract—Power theft is a serious concern for electricity providers all over the world. It makes the power supply less reliable, raises the load consumption, and costs more money. As the global energy sector moves towards using smart grid (SG) technologies, the use of advanced data analysis tools has created new ways to deal with these problems, but electricity theft detection (ETD) is still a major issue. Traditional methods often do not work well because of issues like imbalanced data distributions, overfitting, high-dimensional data, and too much dependency on individual detection techniques. This study's goal is to address these problems by developing a more effective and robust ETD model using modern prediction algorithms. This study addresses one of the voting ensemble models that employs some of the best-known machine learning boosting classifiers, like extreme gradient boosting (XGBoost), light gradient boosting (LightGBM), and categorical boosting (CatBoost) to enhance the prediction accuracy. We use a gated recurrent unit (GRU) to extract features, which makes it easier to work with time series data resulting in the improvement of model performance. There are three major tasks performed: synthetic minority oversampling technique with tomes link (SMOTETomek) was used for data pre-processing, a recurrent neural network (RNN) was used for the temporal feature extraction, and SecureGrid was proposed as a new voting-based ensemble model. The proposed model shows that it can achieve a test accuracy of 93%, an F1 score of 78%, and 97% ROC-AUC in the data set of the State Grid Corporation of China (SGCC). The proposed model outperforms traditional boosting models that have been used on similar datasets. We found that our proposed SecureGrid model has the ability to significantly enhance ETD in SGs. It offers a robust and near to accurate solution which is much better than current approaches.

Index Terms—Boosting Classifiers, CalibratedClassifierCV, ETD, GridSearchCV, Machine Learning, RNN-GRU, SGCC, SMOTETomek, Voting Ensemble Model

I. INTRODUCTION

The theft of electrical power is a global problem that undermines the economic stability of electricity providers and imposes additional costs on consumers. Electrical energy

losses in the energy sector are classified into two main categories: technical losses (TLs) and non-technical losses (NTLs). TLs occur in power distribution systems naturally as electricity flows through transmission and distribution lines. However, NTLs are not caused by electricity transmission but by human factors like tampering meters, bypassing meters, direct tapping from the lines, and billing irregularities. These losses are very concerning because they are preventable and directly impact the financial health of electricity providers, leading to higher costs for consumers and reduced funds for infrastructure improvements. Energy theft, a significant contributor to NTLs, leads to both revenue loss and disruption in energy demand forecasting. One of the most significant challenges regarding both traditional power grids and SGs nowadays is electric power loss. The incorporation of advanced metering infrastructure (AMI) in power sectors has enabled power companies to acquire large-scale energy usage data at a high frequency via smart meters. This data becomes valuable in identifying incidents of electricity theft. Meter tampering, network penetration, and measurement interruption were the three main categories into which the security vulnerabilities of AMI resulting in NTL fraud were classified [1]. To address the issue of NTLs, several classification models based on ML and DL have recently been proposed [2]. Some normal and fraudulent samples are classified from the energy consumption database using the classification-based support vector machine (SVM) approach [3]. In addition, data analysis techniques for detecting energy theft include linear regression (LR) [4] and state estimation [5]. However, these methods are not applicable to situations involving large volumes of data. A deep convolutional neural network model was presented by Zheng et al [6] to examine the energy theft patterns of specific users. The concept of SGs has emerged recently, initiating a new era in the pursuit of energy theft. The authors provide a deep model based on clustering in [7] to identify energy

thieves in the power sector. Using a density-based clustering strategy, the suggested method creates several groups with comparable EC data inside them. Energy thieves are defined as EC samples that are not part of any cluster. Gradient boosting models are also used in [8] to create an ETD model. A larger improvement in the accuracy of ETD is achieved using the boosting model. Moreover, by incorporating synthetic stealing assaults, the issue of class imbalance is handled.

However, some of these methods have low detection accuracy. There are many reasons why these poor performances occur. To obtain increased performance in ETD, no voting ensemble model is constructed from various heterogeneous approaches. Instead, ML classifiers are used independently in the aforementioned ways. Additionally, improper handling of the unbalanced class issue of the dataset leads to skewed findings in terms of the majority class samples. DL techniques, in addition to conventional ML techniques, have shown impressive results in fields including computer vision and image classification [9]. DL methods have been used in this work because of their ability to handle large datasets, automatically extract features, and perform complex classification tasks.

Using a voting mechanism, an ensemble of three sophisticated boosting algorithms has been added to improve the model's resilience and prediction accuracy. This ensemble method makes a prediction model that is more accurate and reliable by combining the best features of XGBoost, LightGBM, and CatBoost. The power of cutting edge ML techniques can be effectively applied to tackle practical ETD problems by employing this comprehensive modeling approach.

II. PROPOSED METHODOLOGY

The methods that have been followed during the whole process of the model generation are discussed in this section. Figure 1 illustrates the model's working mechanism.

A. DATA PREPROCESSING

Data preprocessing plays a pivotal role because it uses several approaches to enhance the quality of metadata, which directly improves the effectiveness of ML algorithms. The following steps are included in the data preprocessing phase. These are explained below.

1) *Dataset Information:* The SGCC dataset is used to determine the major quality revealing metrics of the proposed model. This dataset includes consumer IDs, daily electricity consumption, and labels that identify each record as either honest or fraudulent. The dataset contains an irregular distribution of classes, with samples spread throughout a large range. The data is further distinguished by the existence of distorted and missing values. The dataset also contains information on the ground truth, indicating that 7.2% of all customers are involved in stealing electricity. Each customer is labeled as either 0 or 1, where 0 represents an honest customer and 1 indicates a fraud. The SGCC assigns these designations after conducting on-site inspections. The dataset is tabular and contains full records for each customer in its rows, and daily power usage for all users is shown in its columns. Table 1 clarifies the dataset's meta-information for more clarity.

TABLE I: Dataset details

Parameter	Values
Duration	2014–2016
Total User	10152
Honest Consumers	9466
Fraudulent Consumers	686
Ground truth	7.2%

2) *Handling the Missing Values:* In the context of ETD, datasets usually have missing values denoted as Not a Number (NaN). These problems can occur for many reasons, such as smart meter malfunctions, unscheduled system maintenance, broken distribution lines, or data storage concerns. But performance may be severely hampered when training ML models on datasets with missing values. While deleting customer records with missing values is a typical solution to this problem, doing so might result in the loss of important data that is encoded in the records. A more sophisticated method is used in this study to deal with missing data. To recover missing data, a linear imputation approach is used via the following function [6].

$$f(x_i) = \begin{cases} \frac{x_{i-1} + x_{i+1}}{2} & \text{if } x_i \text{ is missing, neighbors exist} \\ 0 & \text{if } x_i \text{ and any neighbor are missing} \\ x_i & \text{otherwise} \end{cases} \quad (1)$$

Here, x_i represents the load usage data. If both x_{i-1} and x_{i+1} are absurd values in the dataset, they are replaced by the null value zero. If this is not, then absurd or lost values are substituted by taking the mean of the past and post values, respectively, from the dataset. The purpose of this imputation method is to recover missing values while maintaining the temporal continuity of the EC data.

3) *Normalization:* The Min-Max method of normalization is applied to the dataset once the missing values have been fixed. Since DL models are naturally sensitive to a variety of inputs, this normalization step is essential. Experimental findings show that training DL models on normalized data results in excellent performance. The following mathematical formula is used for the min-max technique [10].

$$x_{i,j} = \frac{x_{i,j} - \min(x_i)}{\max(x_i) - \min(x_i)} \quad (2)$$

Here, $\min(x_i)$ denotes the minimum and $\max(x_i)$ denotes the maximum electricity consumption values of consumer i . Scaling the data within a certain range improves the performance of models by employing more stable training and improved convergence.

4) *Handling the Class Imbalance Problem:* We used SMOTE with totem link in this paper since the dataset had class imbalances [11]. The phenomenon of class imbalance arises when there is a notable under-representation of one class (FLAG = 0) in comparison to the other class (FLAG = 1). Because of this imbalance, ML models may perform poorly

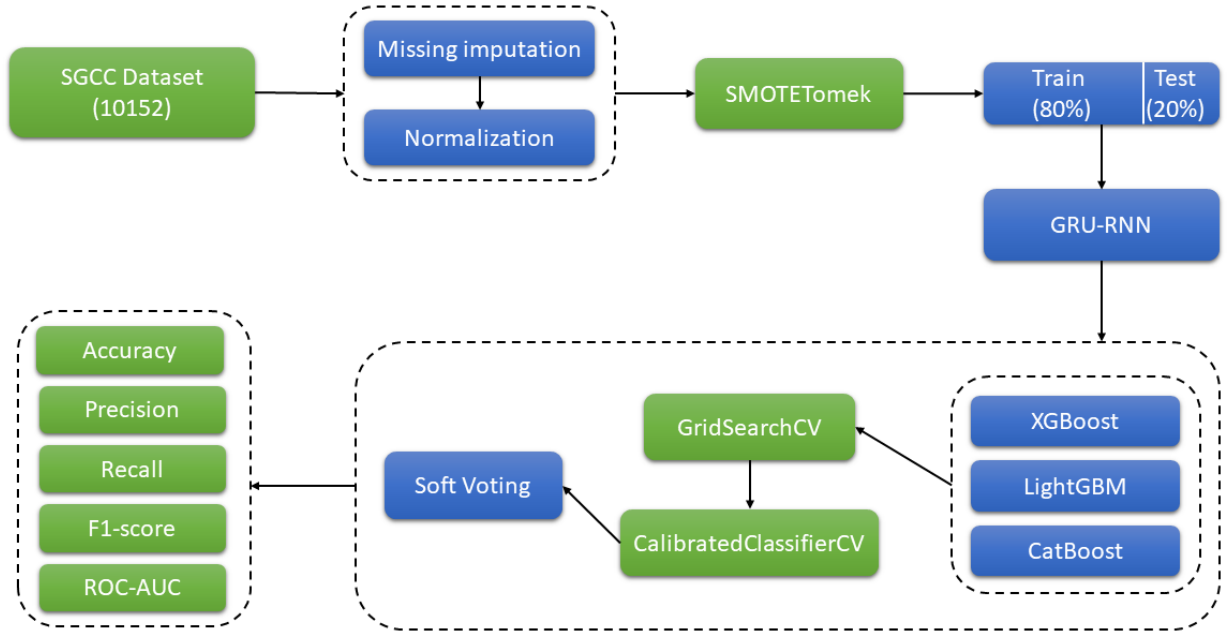


Fig. 1: The proposed SecureGrid model

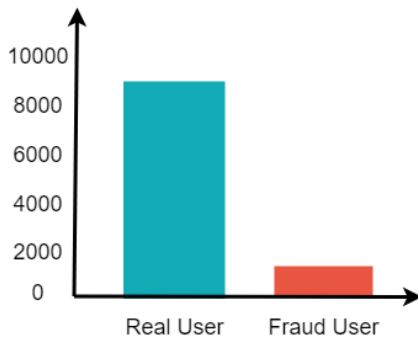


Fig. 2: Distribution before applying SMOTETomek

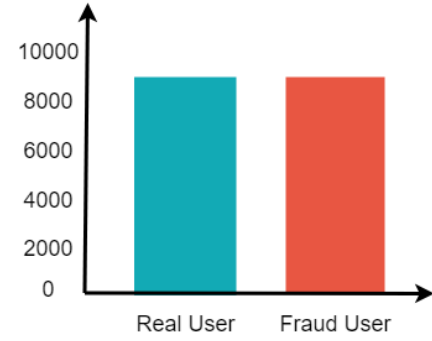


Fig. 3: Distribution after applying SMOTETomek

when attempting to identify samples of a class that is minor, assigning more weights to the class that is major instead of the ideal distribution. This problem is addressed by SMOTE, which ensures the model gets enough data to learn from both groups by creating synthetic examples for minority classes. Then applies Tomek Links to clean the boundary between classes by removing overlapping samples from both classes that are likely noise or borderline cases. Figure 2 and Figure 3 give a visual representation of the state of real and fraud user class before and after performing SMOTE respectively. This approach produces synthetic data so that during the training

process, the models do not get biased and overfit to a specific class and improve generalization.

B. PROPOSED MODEL'S ALGORITHM

1) *Temporal Feature Extraction using RNN*: From available RNNs, the Gated Recurrent Unit (GRU) [12] was used to learn important temporal features from input data. To obtain the temporal dynamics of consumption, a '*GRUFeatureEngineer*' was created using a GRU neural network architecture. A sequence-to-sequence reconstruction approach was used, where the GRU was trained to predict the final time step in energy consumption sequences based on normal (non-

theft) data. The latent vector from the final GRU layer was extracted as the learned features, and the reconstruction error was used as a measure of anomaly. These features, derived from the GRU, were then combined with static input features for classification.

2) *Development of the Boosting Models*: Three gradient boosting classifiers, which are XGBoost, LightGBM, and CatBoost, were utilized. Model training was conducted using *5-fold stratified cross-validation*, with early stopping guided by validation performance to reduce the risk of overfitting. The primary optimization metric in this work was the F1 score, and the model that performed the best for each fold was selected according to its F1 validation score.

Subsequently, each model had undergone isotonic calibration using *CalibratedClassifierCV* to improve the probabilistic outputs, which were then used for threshold tuning and ensemble. A further description of the boosting models depicts there key features.

- **Extreme Gradient Boosting**: This is a very sophisticated gradient boosting approach. It works better in terms of performance and efficiency on classification datasets. The main theme of XGBoost is to develop a group of decision trees successively, each tree fixing the mistakes made by the ones before it.
- **Light Gradient Boosting**: It is another fast and efficient gradient boosting approach that is well-known. It works especially well with large datasets and activities where we need to be careful with our computing power.
- **Categorical Boosting**: This boosting algorithm works well with categorical features. It is renowned for being able to automatically handle categorical features, being robust, and doing well in a variety of ML tasks.

3) *Model Ensembling and Calibration*: We have used a voting ensemble approach to put together the forecasts from three different ML models, namely XGBoost, LightGBM, and CatBoost. A soft voting ensemble classifier was created from the calibrated versions of the above mentioned boosting schemes. The soft voting scheme enables each model to contribute its predicted class probabilities to the final decision, improving robustness and generalization.

$$\hat{y}_{\text{ensemble}} = \text{mode}(\hat{y}_{\text{XGBoost}}, \hat{y}_{\text{CatBoost}}, \hat{y}_{\text{LGBost}}) \quad (3)$$

where $\hat{y}$ represents the predicted class label by each model. In this method, each model has an equal say in the final prediction, and the output of the ensemble is based on the majority class across the three models.

4) *Hyper-parameter tuning with Grid Search CV*: To optimize the performance of the ML models, we have used *Grid Search* to identify the best hyperparameters. CatBoost, XGBoost, and LightGBM so that the models would work better. Grid search looks over a whole hyperparameter grid and evaluates the model's performance for each combination.

The best hyperparameters θ^* are the ones that make the model work best on the validation set:

$$\theta^* = \underset{\theta \in S_1 \times S_2 \times \dots \times S_n}{\operatorname{argmax}} A(\theta) \quad (4)$$

where:

- θ is the vector of hyperparameters.
- $S_1 \times S_2 \times \dots \times S_n$ is the Cartesian product of the individual hyperparameter spaces.
- $A(\theta)$ is the model performance (e.g., accuracy or loss) for hyperparameter configuration θ .

5) *Voting Ensemble*: We propose a soft voting ensemble combining XGBoost, LightGBM, and CatBoost classifiers. Each model outputs a probability distribution $P_i(y = c | x)$ for class $c \in \{0, 1\}$. The final decision is made by averaging the predicted probabilities from all models and assigning the class with the highest overall score:

$$\hat{y} = \arg \max_{c \in \{0, 1\}} \sum_{i=1}^3 w_i P_i(y = c | x) \quad (5)$$

where w_i are model weights (set equally in this work). This ensemble exploits the complementary strengths of boosting models to enhance fraud detection performance.

III. RESULTS AND DISCUSSION

A. PERFORMANCE METRICS

The overall performance of the individual machine learning boosting algorithms, along with the proposed voting ensemble-based model, is calculated using important metrics. The confusion matrix [13] is the foundation for the computation of all the chosen metrics. The performance metrics are explained in more detail below.

1) *Accuracy*: Accuracy of a model is the calculation of the ratio of properly identified incidents to the sum of incidents, as shown in equation 6 [13].

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \quad (6)$$

where TP , FP , TN , and FN refer to true positives, false positives, true negatives, and false negatives, respectively.

2) *Precision*: It is the ratio of accurately predicted positive samples to all the predicted positive samples. It is evaluated by equation 7 [13].

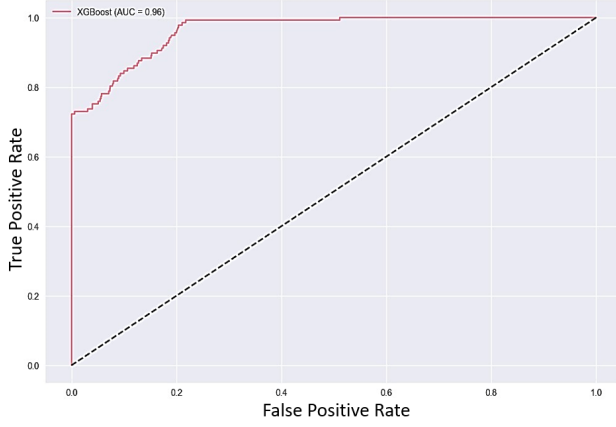
$$\text{Precision} = \frac{TP}{TP + FP} \quad (7)$$

3) *Recall*: It calculates the ratio of all the samples in the actual class that were correctly predicted to be positive. It is evaluated using equation 8 [13].

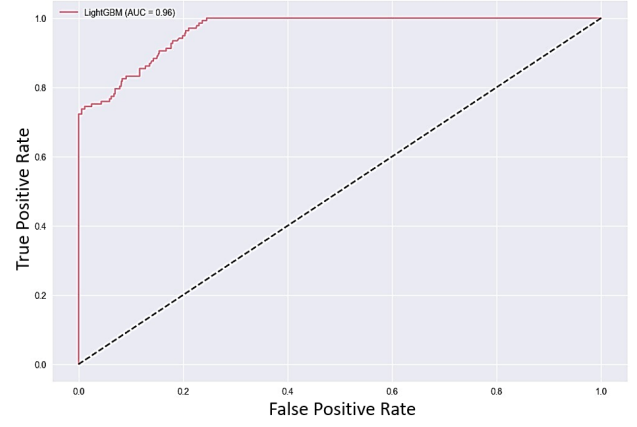
$$\text{Recall} = \frac{TP}{TP + FN} \quad (8)$$

4) *F1-score*: The F1-score is a balanced measurement of both precision and recall, providing the harmonic mean of the two as shown in equation 9 [13].

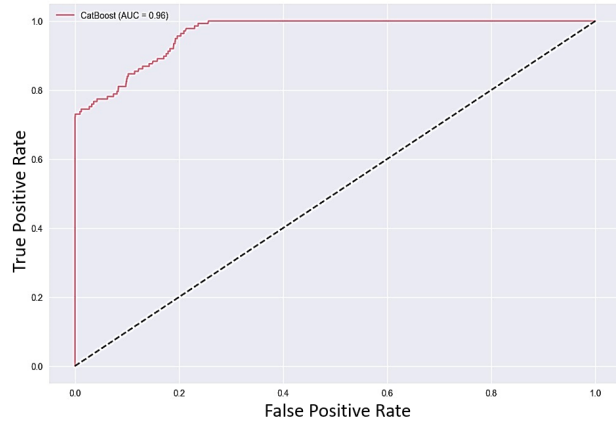
$$\text{F1 score} = \frac{P \times R}{P + R} \times 2 \quad (9)$$



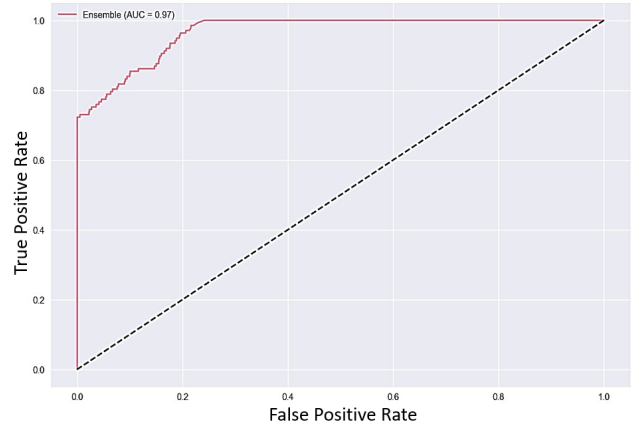
XGBoost (AUC = 0.96)



LightGBM (AUC = 0.96)



CatBoost (AUC = 0.96)



Ensemble (AUC = 0.97)

Fig. 4: Combined ROC curves for all models

5) *ROC-AUC*: It is a sophisticated parameter for the evaluation of classification task-performing models. It represents how well the model can tell the difference between the two classes. The following equation is how it is calculated [13].

$$AUC-ROC = \frac{FP}{FP + TN} \quad (10)$$

B. PROPOSED MODEL'S PERFORMANCE EVALUATION

The best hyperparameters for the baseline classifiers had been chosen using Grid Search method of machine learning. Tuning to the best hyperparameters, the baseline as well as the ensemble model had been trained. Then the classifiers are concatenated in our proposed voting ensemble model. The training data was taken as 80% and the test data was taken as 20%. For each trained model along with the ensemble model, probabilistic predictions were made on the test set. Thresholds between 0.1 and 0.5 were scanned to identify the optimal operating point with respect to F1-score.

Analyzing Figure 5 it can be evaluated that, the proposed SecureGrid model has achieved the highest accuracy of **93%**, surpassing the individual models ranging from 88%

to 90%. Pamir et al [14] has proposed a stacking based ensemble model, namely MLBCSR, along with the integration of general boosting algorithms. Our model has outperformed their stacking ensemble algorithm in terms of ROC-AUC and accuracy. As we see, the basic ML-based classification models, namely LR and SVM, have poor performance as seen in the metrics, whereas boosting methods perform better compared to the basic algorithms. In terms of precision, our proposed SecureGrid model outperforms the baselines with a score of **74%**, which demonstrates its improved ability to minimize false positives compared to XGBoost (69%), LGBost, and CatBoost (67%). The recall score remains the same for the CatBoost and our model because it has a default overfitting detection process. Moreover, the F1-Score further highlights the improved performance with the achievement of **78%** of F1-Score compared to 74% achieved by XGBoost. As seen from Figure 4, among the three ROC-AUC curves, the ensemble model achieves the highest AUC of of 97%, indicating a robust performance of classification compared to the AUC of 96% for the individual boosting models. In fact, outperforming the baseline boosting models validate the better

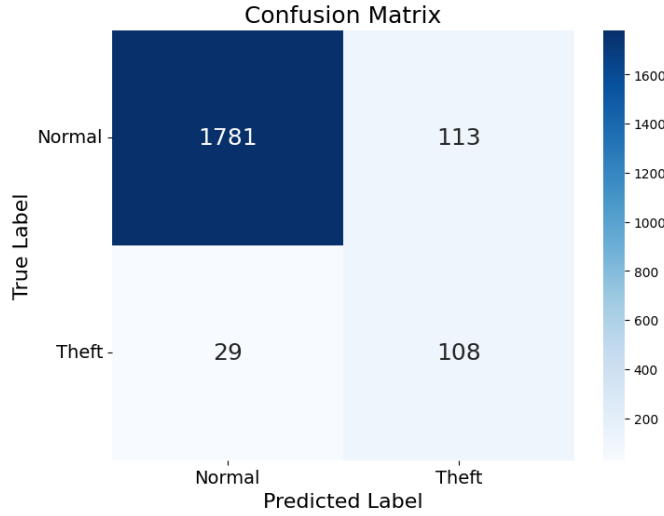


Fig. 5: Confusion matrix of the SecureGrid model

TABLE II: Key performance metrics of SecureGrid compared to other models

Models	Accuracy	Precision	Recall	F1-Score	ROC-AUC
LR	93%	54%	64%	56%	76%
SVM	88%	57%	56%	57%	72%
XGBoost	90%	69%	87%	74%	96%
LGBBoost	88%	67%	86%	71%	96%
CatBoost	88%	67%	86%	71%	96%
MLBCSM	92%	91%	93%	92%	92%
SecureGrid	93%	74%	86%	78%	97%

performance of our proposed SecureGrid model. The feature extraction using recurrent neural network, like GRU, and implementing the Grid Search algorithm for best hyperparameter finding has contributed to the overall better performance of our proposed SecureGrid model. The SecureGrid model's overall performance comparison with baseline models (XGBoost, LightGBM and CatBoost) has been depicted in Table II.

IV. CONCLUSION

This study is a new addition to advanced research in addressing the critical issue of ETD by introducing a voting ensemble model integrating advanced machine learning boosting classifiers such as XGBoost, CatBoost, and LightGBM with GRU-RNN for feature extraction. The key findings demonstrate that our proposed SecureGrid model achieves a test accuracy of 93%, an F1-score of 78%, and a 97% AUC, outperforming baseline methods. Although the study provides promising results, the purpose of the research will be fulfilled when this is integrated with smart metering systems of smart grids to identify theft cases. Future research may focus on implementing this model with smart meters, which may be a new addition to the existing metering technology, and also minimizing the expenditure of the government due to theft.

REFERENCES

- [1] W. Han and Y. Xiao, "Non-technical loss fraud in advanced metering infrastructure in smart grid," in *Cloud Computing and Security: Second International Conference, ICCCS 2016, Nanjing, China, July 29-31, 2016, Revised Selected Papers, Part II 2*. Springer, 2016, pp. 163–172.
- [2] S. O. Tehrani, A. Shahrestani, and M. H. Yaghmaee, "Online electricity theft detection framework for large-scale smart grid data," *Electric Power Systems Research*, vol. 208, p. 107895, 2022.
- [3] S. S. S. R. Depuru, L. Wang, and V. Devabhaktuni, "Support vector machine based data classification for detection of electricity theft," in *2011 IEEE/PES Power Systems Conference and Exposition*. IEEE, 2011, pp. 1–8.
- [4] S.-C. Yip, C.-K. Tan, W.-N. Tan, M.-T. Gan, and A.-H. A. Bakar, "Energy theft and defective meters detection in ami using linear regression," in *2017 IEEE International Conference on Environment and Electrical Engineering and 2017 IEEE Industrial and Commercial Power Systems Europe (EEEIC/I&CPS Europe)*. IEEE, 2017, pp. 1–6.
- [5] S. Salinas, C. Luo, W. Liao, and P. Li, "State estimation for energy theft detection in microgrids," in *9th International Conference on Communications and Networking in China*. IEEE, 2014, pp. 96–101.
- [6] Z. Zheng, Y. Yang, X. Niu, H.-N. Dai, and Y. Zhou, "Wide and deep convolutional neural networks for electricity-theft detection to secure smart grids," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 4, pp. 1606–1615, 2017.
- [7] J. L. Viegas, P. R. Esteves, and S. M. Vieira, "Clustering-based novelty detection for identification of non-technical losses," *International Journal of Electrical Power & Energy Systems*, vol. 101, pp. 301–310, 2018.
- [8] R. Punmiya and S. Choe, "Energy theft detection using gradient boosting theft detector with feature engineering-based preprocessing," *IEEE Transactions on Smart Grid*, vol. 10, no. 2, pp. 2326–2329, 2019.
- [9] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, 2012.
- [10] W. Liao, Z. Yang, K. Liu, B. Zhang, X. Chen, and R. Song, "Electricity theft detection using euclidean and graph convolutional neural networks," *IEEE Transactions on Power Systems*, 2022.
- [11] R. Yang, C. Zhang, R. Gao, and L. Zhang, "A novel feature extraction method with feature selection to identify golgi-resident protein types from imbalanced data," *International journal of molecular sciences*, vol. 17, no. 2, p. 218, 2016.
- [12] R. Dey and F. M. Salem, "Gate-variants of gated recurrent unit (gru) neural networks," in *2017 IEEE 60th international midwest symposium on circuits and systems (MWSCAS)*. IEEE, 2017, pp. 1597–1600.
- [13] I. U. Khan, N. Javeid, C. J. Taylor, K. A. Gamage, and X. Ma, "A stacked machine and deep learning-based approach for analysing electricity theft in smart grids," *IEEE Transactions on Smart Grid*, vol. 13, no. 2, pp. 1633–1644, 2021.
- [14] Pamir, N. Javaid, M. Akbar, A. Aldegheishem, N. Alrajeh, and E. A. Mohammed, "Employing a machine learning boosting classifiers based stacking ensemble model for detecting non technical losses in smart grids," *IEEE Access*, vol. 10, pp. 121 886–121 899, 2022.