The Scientific Temper (2025) Vol. 16 (8): 4635-4645
E-ISSN: 2231-6396, ISSN: 0976-8653
Doi: 10.58414/SCIENTIFICTEMPER.2025.16.8.06
https://scientifictemper.com/

RESEARCH ARTICLE

# AT&C and non-technical loss reduction in smart grid using smart metering with AI techniques

**Jadhav Girish Vasantrao[1*], Chirag Patel[2]**

## Abstract

Aggregate Technical and Commercial (AT&C) damage are a serious issue for electricity distribution companies globally, hindering economic growth and sustainability. Among them, non-technical losses (NTLs), such as electricity theft, fraud, and non-payment, contribute to substantial financial losses and may jeopardize power quality and grid stability. Growing usage of smart grids and Advanced Metering Infrastructure (AMI) opens new ways of effective management of energy, as well as sophisticated approaches to electricity theft, creating demands on cutting-edge methods of detection. This research aims to enhance NTL detection by introducing a hybrid approach that integrates Temporal Convolutional Networks (TCN) and LightGBM, or Light Gradient Boosting Machine. TCNs are used in order to detect complex temporal features in smart meter consumption records, recognizing sequential patterns characteristic of fraudulent behaviour. LightGBM, which is an extremely effective gradient boosting architecture, which is then applied to classify consumption behaviour correctly as normal or suspicious. A real dataset is used to train and evaluate the suggested model of smart meter records, demonstrating its ability to discriminate between normal and potentially fraudulent consumption patterns. Results present promising effectiveness in identifying usual use; however, the research indicates challenges to achieving high accuracy and memory in detecting energy theft. This emphasizes the necessity of further research and model refinement to enhance its effectiveness in real-world applications and to counteract the negative impacts of NTLs on electricity utilities and consumers.

**Keywords:** Smart Grid, Smart Metering, Non-Technical Losses (NTLs), Electricity Theft, Temporal Convolutional Networks (TCN), Light Gradient Boosting Machine (LightGBM), Advanced Metering Infrastructure (AMI), Fraud Detection.

**关键词：** 智能电网、智能计量、非技术损失（NTL）、电力盗窃、时间卷积网络（TCN）、光梯度增强机（LightGBM）、高级计量基础设施（AMI）、欺诈检测。

## Introduction

The foundation of contemporary economic growth and societal advancement is the electrical power sector, with access to affordable and dependable electricity being essential for commercial activity, industrial expansion, and quality of life. However, the substantial amount of Aggregate Technical and Commercial (AT&C) damages that take place between the places of generation and end-user consumption is a recurring problem for power distribution utilities around the world. These losses undermine sustainability goals and economic development initiatives by contributing to increased carbon emissions through additional generation requirements, in addition to representing significant financial leakage for utilities, estimated at over $200 billion annually worldwide. The two distinct elements that comprise AT&C losses, both non-technical (NTLs) and technical (TLs), are most prevalent in the distribution sector. (Navani *et al*., 2012).

NTLs, sometimes referred to as Energy molecules that are supplied and utilized, are considered economic losses, but for which the electric power company does not issue a charge. An NTL can be accounted for in four different ways (Smith, 2004).

By manipulating energy meters to record lower consumption than is really utilized, fraud entails misleading an electric power company. A bypass is a covert way to connect to the load from the power grid without going via the energy meter. Another prevalent practice is bribery, and both customers and personnel may be corrupt. When a residential customer moves out, a commercial customer's

[1]Electrical Engineering Department - Parul University[1] City: Vadodara, Gujarat, India. Pin – 391760.

[2]Computer Science Engineering Department - Charotar University of Science and Technology, City: Anand, Gujarat, India – 388421.

***Corresponding Author:** Jadhav Girish Vasantrao, Electrical Engineering Department - Parul University1 City: Vadodara, Gujarat, India. Pin – 391760, E-Mail: girish.jadhav@paruluniversity.ac.in

**How to cite this article:** Vasantrao, J.G., Patel, C. (2025). AT&C and non-technical loss reduction in smart grid using smart metering with AI techniques. The Scientific Temper, **16**(8):4635-4645.

Doi: 10.58414/SCIENTIFICTEMPER.2025.16.8.06

**Source of support:** Nil

**Conflict of interest:** None.



**Received:** 12/07/2025 **Accepted:** 25/07/2025 **Published:** 30/08/2025

business files for bankruptcy, or a meter breaks, they may neglect their energy bill. This is known as non-payment.

The worldwide issue of NTL impacts underdeveloped, emerging, and even fully developed nations. The study of Northeast Group LLC (LLC, 2017) claims that 96 billion USD has been lost every year worldwide. An estimated USD 6 billion was lost annually by power utilities in the United States in 2009 (McDaniel & McLaughlin, 2009). Among the most powerful companies in Canada, BC Hydro stated that the yearly loss from electricity fraud in 2011 was over CA$100 million (Hydro, 2010).

The situation tends to get worse in developing and impoverished nations. An estimated USD 58.7 billion is lost annually by the major rising nations (de Souza Savian *et al*., 2021). In 2022, NTLs accounted for 14.6% of all losses on the low-voltage power grid in Brazil. NTLs accounted for 6.3% of the overall loss in the Brazilian system as a whole (low, medium, and high voltage). In several Latin American nations, energy firms, users, and even those who haven't engaged in any sort of fraud share some of the financial losses brought on by NTL (de Oliveira Ventura *et al*., 2020).

Clandestine illegal connections can have a detrimental impact on Power Quality (PQ) since the power flowing through the system is not what is intended. Disturbances, including voltage swings, over voltages, under voltages, and harmonics, among other issues, could arise in this situation (Olaoluwa, 2017). This may lead to widespread blackouts or even fire concerns. Since the advent of smart grids (SG), advanced metering infrastructure (AMI) technologies have been created to make it easier for an energy meter and the electrical source to communicate. This reciprocal relationship, however, invariably results in cybersecurity issues, including data leaks and theft/alteration of customer electrical data (Wang & Lu, 2013) (El Mrabet *et al*., 2018). Stated differently, it is possible to hack and reprogram the AMI in order to perpetrate fraud, including electricity theft (Morgoev *et al*., 2023). The power provider may lose even more money as a result of new energy theft techniques that have been created to target AMI with the standard tampering that was used to commit fraud.

Therefore, to further halt the growth of NTL, new methods or systems for detecting theft of electricity must be developed. Given the significant financial losses, decrease in PQ, and new methods of power theft using AMI. Utilizing the massive volume of data delivered and stored in the SG, machine learning (ML), and deep learning (DL) algorithms may be developed to effectively identify the type and timing of energy fraud. Machine learning algorithms have shown promise in detecting abnormal consumption patterns indicative of electricity theft or meter tampering. Supervised learning approaches like gradient boosting techniques, support vector machines (SVM) and random forests demonstrated excellent accuracy in classifying legitimate versus fraudulent consumption patterns when trained on labelled historical data (Hashim *et al*., 2024). Unsupervised learning techniques, such as clustering algorithms and anomaly detection methods, can identify consumer segments with similar consumption patterns and flag outliers. Techniques for deep learning, include temporal convolutional networks (TCNs), recurrent neural networks (RNNs) and long short-term memory networks (LSTMs), can simulate complex sequential patterns in electricity usage, enabling detection of subtle anomalies (Ahmad *et al*., 2022).

AI techniques also support the prioritization of investigation resources, allowing utilities to optimize inspection schedules and adapt to evolving theft techniques. The combination of AI and smart meters offers more chances for technological loss reduction through improved system monitoring and optimization (Li *et al*., 2019). Neural network-based load forecasting, distribution state estimation algorithms, and voltage and reactive power optimization algorithms can minimize technical losses through improved control of grid parameters. However, implementing AI-driven loss reduction programs presents practical challenges, such as data quality issues, privacy concerns, transparency and explainability challenges, and regulatory frameworks. Emerging technologies, such as edge computing architectures, blockchain-based metering systems, and federated learning approaches, promise to further enhance the capabilities of AI-driven loss reduction systems. This study aims to develop a hybrid methodology to identify the kind of power theft in distributed energy networks.

### *Related work*

In Ramos *et al*. (2011), in order to train the Optimum Path Forest classifier (OPF), data from commercial and industrial users were used. Other electrical measurements, such as installed power (Pinst), power factor (PF), reactive utilization of energy (kVArh), among others, were employed in place of energy consumption (kWh). The classifier's accuracy was increased by reducing the problem's dimensionality and extracting significant features using the Harmony search (HS) approach. Additionally, HS outperformed principal component analysis (PCA), another conventional feature extraction method.

Having a labelled dataset that shows if a customer has committed fraud is uncommon. In light of this, as well as using AMI data, reference. (Jokar *et al*., 2015) Suggested six formulas that employ data from trustworthy users to produce harmful samples. The data produced by these equations and previous Data is used to create a multiple-class SVM for every customer. A transformer meter's measurement of a neighbourhood's overall energy usage and the amount of energy supplied by intelligent meters are compared in an energy theft detection. Customers in

that area are marked as suspects if an NTL is found, and the classifier determines whether or not any of them are engaging in fraud based on news samples from each of these customers.

(Zanetti *et al*., 2017) Used a similar technique, which also needs information from the low-voltage grid (LVG) transformer's meters and presented two more equations to produce malicious samples. It compares the reports from the domestic smart meters and the LVG meter using a "detector." Then, people with unusual consumption habits are identified using a state device method that creates three square states: typical (G1), suspect (G2), or abnormal (G3).

(Dey, S., Ghosh, S., & Pal, 2020) used an actual classified dataset with intervals with daily readings to build a network of neurons with two distinct components. To capture the global characteristics of 1D data (electricity usage time series), the first, referred to as the broad network, is composed of completely linked layers. The second, called the Deep Convolutional Neural Network (CNN), determines if energy theft is non-periodic and whether a regular user is periodic by using 2-D data (weekly measurements are created from the 1-D data). Ultimately, by mixing the results, an activated sigmoid is produced of both networks to determine whether or not a customer has engaged in fraud. A number of pre-processing methods were used to enhance the model's functionality, including interpolation using linearity to impute values that are absent and the empirical principle, which finds outliers in the data set by calculating two deviations of the mean value.

In (Messinis *et al*., 2019), a novel equation for modelling energy theft is put forth. It considers the linear growth of an assault as time passes, meaning that power usage declines gradually as opposed to suddenly. The detection of NTLs is then carried out using a combination of electrical system efficiency, Support Vector Machines (SVM), and volt sensitivity analysis. The result is based on a few equations, as well as this newly suggested equation.

It is commonly recognized that there are more instances of trustworthy users than malicious ones in labeled energy theft datasets. As a result, these datasets are regarded as imbalanced and may present learning challenges for machine learning algorithms (Domingues *et al*., 2018). With the same dataset as (Dey, S., Ghosh, S., & Pal, 2020) the authors of (Khan *et al*., 2020) extracted features using a VGG-16 design, which has several mixing and layers for convolution, and performed the final classification using an XGBoost (Decision Tree-based method). The high-level parameters of XGBoost were optimized using the firefly optimization process. The minority class was oversampled using the Adasyn approach to deal with the issue of an unbalanced dataset.

In (Guarda *et al*., 2023), a thorough review of methods that don't rely on equipment was conducted. The most popular technique for detecting NTL is data-oriented/ML approaches, which often employ electrical numbers, this demand, electricity, and mostly power consumption, as characteristics. Traditional machine learning methods, including ANN, DT, and SVM, or Bayesian classifiers, are used in these works. (Odje *et al*., 2021)Sought to quantify how smart metering affected combined technical, commercial, and collection fees (ATC&C). It makes use of mathematical modeling and historical information from the Nigerian Electricity Regulatory Commission in order to predict the impact of metering and ATC&C losses. According to the study, for every 1% rise, ATC&C losses decrease by 0.8% in the metering setup, assuming all other variables stay the same. However, factors like system components depreciation, energy theft, and meter tampering increase, making it necessary for Discos to adopt modern strategies and aggressive metering to reduce ATC&C losses.

Despite improvements in machine learning-based non-technical loss detection, optimization methods, and smart metering technologies, various gaps in research continue to exist. The major problem is the lack of labeled datasets, which impedes the design and verification of robust models. Improved feature engineering methods must be employed to uncover meaningful information from consumption data, and additional research must be conducted to enhance detection of the minority class. Improving detection models' generalizability over changing grid structures, customer categories, and geography is essential. Finally, electricity thieves must have a real-time system that is capable of adapting in real-time and identifying as well as acting to counteract stealing in time with changing schemes for stealing electricity. These areas require further investigation and development to improve the detection of electricity theft. These

## Background

### *Temporal Convolutional Networks*

One convolutional architecture for encoding time information is the Temporal Neural Network (Yu & Koltun, 2015); (Van Den Oord *et al*., 2016); (Bai *et al*., 2018). It comprises two often used components: the dilated convolution and primary casual convolution network, which together form a Convolutional Network with Dilated Temporal.

### *Dilated Convolution*

Without raising its parameters, an algorithm may have a larger field of receiving via the dilatation (à trous convolution) (Yu & Koltun, 2015). In order to accomplish expanded convolution "holes" are inserted into the kernel's target sites. Therefore, expanding the receptive field. The word "gaps" will denote any technique for enlarging A kernel is received via gaps in order to expand the field of reception.

### *Dilated Temporal Convolutional Network*

Wave Net demonstrated a Convex Network with Dilated Time. (Van Den Oord *et al*., 2016), is a temporally network design that does not handle time steps repeatedly, but rather in concurrently. By doing back propagation for each of the steps all at once as opposed to a periodic gradient flow, this drastically changes the model's approach to reverse propagation across time. To prevent leakage from earlier data into later phases, a casual inversion is used (Figure 1).

A TCN Given a series of inputs x, dilatation d, length i, convolution dilated *d, and filter, a layer may be described as follows $f$

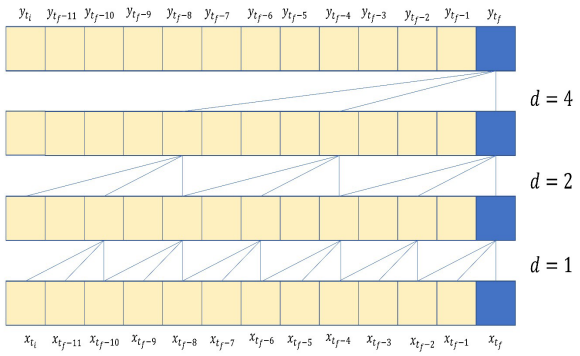
$$F(x_t) = (x *_d f)(t) = \sum_{n=0}^{i-1} f(t) \cdot x_{s-d^*i}. \quad (1)$$

A network may dramatically increase the receptive area and readily understand time intervals from previous steps by using a dilated convolution. In the absence of dilated convolutions, temporal convolutional networks (TCNs) would have a linear responsive field regarding steps that came before. The field that is responsive for preceding time steps (frames) may be computed using dilations.

$$RF(n, d, k) = 1 + \sum_{i=0}^{n-1} d^i(k-1) \quad (2)$$

where n is the amount of layers that are hidden, the kernel size is denoted by k, and the enlargement factor. To get a quadratic receptivity field, an expansion factor of two is frequently used (Bai *et al*., 2018).

Temporal Convolutional Networks (TCNs) with dilation facilitate the processing of extensive temporal data with little computational requirements by using a broad receptive field. Temporal Convolutional Networks (TCNs) facilitate parallel processing, provide an extensive receptive field, and mitigate the issues of disappearing or inflating gradients by ensuring that backpropagation is oriented perpendicularly to the temporal sequence rather than parallel to it. Dilated




A TCN Given a series of inputs x, dilatation d, length i, convolution dilated $*_d$, and filter, a layer may be described as follows $f$

**Figure 1:** Architecture of TCN

Temporal Convolutional Networks (TCNs) have shown remarkable efficacy in emulating the long-term memory capacities of alternative architectures, such as Long Short-Term Memory (LSTM) networks and Recurrent Neural Networks (RNNs), especially in tasks like information copying (Bai *et al*., 2018). Action separation using the TCN has produced cutting-edge results in action recognition. (Lea *et al*., 2017). Temporal Convolutional Networks (TCNs) have been investigated in emotion analysis, yielding superior outcomes compared to Long short-term mental systems (LSTMs) and recurrent neural networks (RNNs) in emotion-related tasks. Temporal Convolutional Networks (TCNs) typically have temporal blocks, each consisting of two stacked convolutional layers. The objective of layering is to first scale the input data to the anticipated dimensions before transmitting it via an output-size-designated convolutional layer. (Mehta & Yang, 2023).

### *Light Gradient Boosting Machine (LightGBM)*

Regression, ranking, and categorization are just a few of the machine learning applications that heavily rely on LightGBM, a fast, dispersed, powerful gradient-boosting system that leverages decision tree approaches (Ke *et al*., 2017). In order to create a strong learning model, this variation of the Boosting strategy combines many weak machine learning models. Boosting strategies make sure that misclassified cases get greater attention in later training iterations by lowering the weights of successfully categorized data and raising the weights of poorly classified data. In the end, all the machine learning algorithms are linearly mixed, and the resultant model's weights are modified in accordance with the classifier's error rate. (Ke *et al*., 2017).

The fundamental principle may be expressed using equation ():

$$f(x) = \sum_{q=1}^{Q} \alpha_q T(x, \theta_q) \quad (3)$$

where $f(x)$ is the target value that matches the training set; The total amount of base learners is represented by $Q$.; $\alpha q$ is the weight coefficient of the qth base learner; $x$ is the learning sample; $\theta q$ is the pupil's classification variable; and $T(x, \theta q)$ is training the qth foundation learner (Figure 2).

After selecting The learning procedure for the Boosting method, both the training data and the algorithm's loss formula is translated into a problem in optimisation where the goal aims to reduce the loss function (Li *et al*., 2024). The

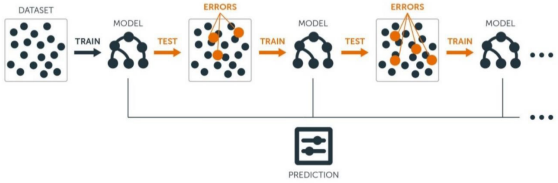



**Figure 2:** Architecture of LightGBM

objective function is given below:

$$\arg min \sum_{h=1}^{H} L\left(y_h, f(x_h)\right) \quad (4)$$

mh is the actual value of the data, and h is the sample index.; $\mathbf{T}(x\text{h})$ is the desired value matching to the h[th] sample; $L(y\text{h}, f(x\text{h}))$ is the values of the loss function for the hth sample.; and $H$ is the total amount of samples.

A boosting tree model developed using a gradient-descent approach is used in the Gradient Boosting Machine (GBM). The chosen loss function consistently diminishes with each incorporation of a new sub-model, approaching The slope of the variables in equation that has the next highest information content (5).

$$L\left(F_j(x), Y\right) < L\left(F_{j-1}(x), Y\right) \quad (5)$$

where $L\left(F_j(x), Y\right)$ and $L\left(F_{j-1}(x), Y\right)$ are, respectively, the loss function parameters for the jth and $(j-1)$ th iterations, $F_j(x)$ and $F_{j_{-1}}(x)$ are what are intended to be the jth and $(j-1)$ th examples, and $Y$ is the sample of real goal value.

LightGBM is a kind of GBM that successfully tackles the difficulties that GBM has while handling large amounts of data. There are two primary aspects of this model. (Ke *et al*., 2017):

- Rather of using a level-wise development methodology, the by leaf tree growth method uses a leaf-wise development strategy. This approach incurs lower computational expenses and effectively mitigates overfitting by regulating the lowest amount of information in tree height and leaf nodes.
- Algorithm for histogram-based choice trees: LightGBM uses a method for histogram-based decision trees. During feature selection, it is sufficient to navigate and identify the ideal split point depending on each individual value of the histogram, which hence reducing the cost of computing and storing it. LightGBM can learn and forecast thanks to this functionality with greater efficiency while managing extensive datasets (Li *et al*., 2024).

## Methodology

This research utilized a hybrid approach that combined Temporal Convolutional Networks (TCN) and LightGBM in order to identify instances of power theft. TCN was used to learn deep temporal identify sequential patterns, while LightGBM was utilized for effective and accurate classification. The methodology ensures enhanced detection accuracy by combining the benefits of gradient booster techniques with deep learning on actual energy consumption datasets.

### *Dataset Description*

The data set used in this study, including its electricity-theft-detection data set, was acquired via KaggleHub. It consists of time-series electricity usage data collected from smart meters, with each instance being labelled as normal consumption behaviour (0) or suspicious of electricity theft (1). Every row is a set of usage readings over a set time frame (i.e., 24 hours or 30 days), enabling the detection system to pick up on temporal behaviour characteristic of fraudulent activity.

### *Data Preprocessing*

Before training the model, the data went through multiple pre-processing operations. Missing values, if present, were filled up using statistical procedures like mean or forward fill mechanisms. The data was then normalized by applying to convert every value to within 0 and 1, use min-max levelling, which assists in speeding up neural network training. The used normalization equation is:

$$x_{norm} = \frac{x - x_{\min}}{x_{max} - x_{\min}} \quad (6)$$

where $x_{min}$ and $x_{max}$ are the lowest and highest numbers in the attributes column, as well, and x is the original value. Following standardization, Test and training sets of the dataset were separated.

### *Train Test Split*

Following the dataset's separation into test and training sets, the train-test split enables the model to analyse characteristics and behaviours from the training set. This encompasses the majority of the data; assess how well the model can generalize on a dataset that hasn't been seen before, apart from learning; this section provides the model with new data, enabling a reliable evaluation of its performance and the detection of overfitting; typical splits allocate 70–80% of the data is utilized for training, while the remaining 20–30% is used for testing.

### *Model Building*

#### *Feature extraction*

In order to capture the temporal correlations in electricity usage patterns efficiently, a Temporal Convolutional Network (TCN) was utilized as the feature extractor. TCNs are a 1D convolutional neural network specifically tailored for sequence modelling problems. Unlike normal CNNs, TCNs use causal convolutions such that the only factors influencing predictions at time t are inputs at that time t and previous time. Secondly, dilated convolutions are used to exponentially grow the receptive field without growing computational complexity. One definition of the expanded convolution process is:

$$F(t) = \sum_{i=0}^{k-1} f(i) \cdot x_{t-d \cdot i} \quad (7)$$

Where,

$k$ is the filter size,

**Table 1:** Algorithm

Algorithm 1 Electric Theft Detection using TCN and LightGBM

1: Input: Electricity consumption dataset $D = \{(x_i, y_i)\}_{i=1}^{N}$ where $x_i \in R^{1034}$, $y_i \in \{0,1\}$
2: Output: A certified model for detecting theft
3: method for data preprocessing
4: Load dataset D from CSV file
5: Fill missing values: $x_{i,j} \leftarrow 0$ for all missing $x_{i,j}$ 6: Split data into train/validation/test sets:
7: $D_{train-val}$, $D_{test} \leftarrow$ split (D, test size = 0.2,stratify = y)
8: $D_{train}$, $D_{val} \leftarrow$ split ($D_{train-val}$, test size = 0.1875,stratify = $y_{train-val}$
9: Standardize features: $X' = \frac{x-\mu_{train}}{\sigma_{train}}$
10: end procedure
11: procedure TCNModel
12: Define TCN architecture:
13: h1 = ReLU(BatchNorm(Conv1D(x, 1 → 64)))
14: h2 = ReLU(BatchNorm(Conv1D(h1, 64 → 128)))
15: h3 = ReLU(BatchNorm(Conv1D(h2, 128 → 128)))
16: h4 = ReLU(BatchNorm(Conv1D(h3, 128 → 64)))
17: z = GlobalAvgPool(h4)
18: f = Linear(z, 64 → 128)
19: Initialize with Adam optimizer (η = 0.001, weight decay = $10^{-5}$)
20: Loss function: Cross-entropy: $L = -\frac{1}{N}\sum_{i=1}^{N} y_i \ log(p_i) + (1 - y_i) log(1 - p_i)$
21: Train with early stopping (patience=7)
22: end procedure
23: procedure FeatureExtraction
24: Extract features from trained TCN
25: $F_{train} = \{TCN(x_i) | x_i \in D_{train}\}$
26: $F_{test} = \{TCN(x_i) | x_i \in D_{test}\}$
27: end procedure
28: procedure LightGBMTraining
29: Initialize LightGBM with parameters:
30: n estimators = 1000
31: learning rate = 0.03
32: max depth = 9
33: num leaves = 50
34: reg alpha = 0.05
35: reg lambda = 0.05
36: Train on $_{train}$ with early stopping (50 rounds)
37: end procedure
38: procedure Evaluation
39: Predict on test set: $y_i$ = LightGBM (Ftest)
40: Compute metrics:

41: $Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$

42: $Precision = \frac{TP}{TP + FP}$

43: Recall = $\frac{TP}{TP + FN}$

44: $F1 - score = \frac{2 \, x \, Precesion \, x \, Recall}{Precesion + Recall}$

45: $AUC = \int_0^1 TPR(x)\, dx$

46: Generate confusion matrix and ROC curve
47: end procedure

($i$) is the filter weight
$d$ is the dilation factor
$x$ is the input sequence

For stable training and avoidance of overfitting, the TCN employs residual connections, batch normalization, ReLU activation, and dropout layers. The TCN's output is a high-level representation of features in the input sequence, with informative patterns and anomalies reflecting theft (Table 1).

### *Classification*

The features extracted are fed into a The Classifier A quick gradient boost device called the Light Gradient Enhancement Machine (LightGBM) uses a decision tree. LightGBM is well-suited for structured/tabular data and provides benefits like high-speed training, low memory usage, and high accuracy.

### *Objective function*

The LightGBM model optimizes a regularized objective function that integrates loss and complexity. Provided an input feature vector x, LightGBM seeks to train an ensemble of trees $\{T_1(x), T_2(x), \ldots, T_M(x)\}$ to minimize the loss function.

$$L(y, \hat{y}) = \sum_{i=1}^{N} l(y_i, \hat{y}_i) + \sum_{k=1}^{M} \Omega(T_k) \quad (8)$$

Where,
$l$ is a loss function that is differentiable.(e.g. binary log-loss),
$\Omega$ is a phrase used for normalization to avoid overfitting?
$\hat{y}_i$ is the anticipated likelihood of stealing, for example $i$.

### *Model Evaluation*

A number of categorization metrics derived from the matrix of disorientation are used to evaluate the efficacy of the model. Assume that True Positives are represented by $TP$, True Negatives by $TN$, False Positives by $FP$, and False Negatives by $FN$.

*Accuracy*

Accuracy is a crucial indicator for assessing the effectiveness of a categorization model by providing a concise summary of the algorithm's functionality regarding accurate predictions. It is established using the proportion of accurate predictions to the total quantity of input samples.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (9)$$

*Precision*

The precision ratio is the total of both False Positives (FP) and True Positives (TP). It demonstrates the quantity of the autumn cases that were projected as positive were in fact positive. Put another way, a high accuracy score indicates a low error rate, which increases the likelihood that the algorithm will correctly foresee a class that is favourable.

$$\mathrm{Pr}ecision = \frac{TP}{TP+FP} \quad (10)$$

*Recall*

The ratio of TP to the total of TP and TN is known as recall. It shows the quantity of actual positive instances that the model precise forecast. A high memory score means that the model reduces the number of negative results while effectively detecting a significant percentage of positive cases.

$$\mathrm{Re}call = \frac{TP}{TP+FN} \quad (11)$$

*F1-score*

The F1-score is the harmonic average of memory and precision. When a classification task, like the NTL detection problem, involves labelling an honest individual as a criminal or a victim, there are a number of expenses and consequences, it is useful to balance those two metrics.

$$F1 - score = \frac{2 \times \mathrm{Pr}ecision \times \mathrm{Re}call}{\mathrm{Pr}ecision + \mathrm{Re}call} \quad (12)$$

*ROC-AUC Score*

This metric evaluates Trade-off between true positive rate (TPR) and false positive rate (FPR):

$$AUC = \int_0^1 TPR(x)dx \quad (13)$$

Visualizations like The ROC lines and Precision-Recall curves show the model's discriminative capability across different thresholds.

## Results and Discussion

### *Classification Results*

This Table 2 showed the electricity theft detection model's classification performance metrics on the test dataset by each class: 'No Theft' (0) and 'Theft' (1). It shows an in-depth overview of the accuracy of how the model performs predicting each class.

### *For the 'No Theft' (0) class:*

The accuracy on the entire test set is said to be 0.9145 (91.45%) overall for the model. The precision is 0.92. This reflects that among all the cases which the model output as 'No Theft', 92% are indeed 'No Theft'. This shows that it has this class has an extremely low false positive rate. The recall is 1.00. This means that the model accurately marked 100% of all the true 'No Theft' instances in the test set. It shows an extremely low false negative rate for this class. The F1-score is 0.96. This provides an equilibrium score by taking a harmonic average of accuracy and recall. A high F1-score suggests good performance for this class. This shown in Figure 3.

### *Theft (0), theft (1)* The 'Theft' (1) class:

The accuracy is the same (0.9145) since it's a global measure. The class-specific accuracy is not represented directly here,

**Table 2:** Classification No Theft and Theft

| *Classes* | *Accuracy* | *Precision* | *Recall* | *F1-score* |
|---|---|---|---|---|
| No Theft (0) | 0.9145 | 0.92 | 1.00 | 0.96 |
| Theft (1) | 0.48 | 0.04 | 0.08 | 0.08 |

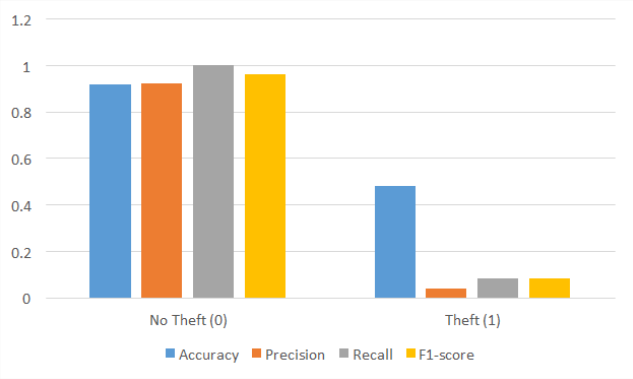



**Figure 3:** Classification of No

though. The accuracy for 'Theft' is 0.04. This means that among all the instances the model classified as 'Theft', only 4% were indeed 'Theft'. This implies that the 'Theft' category has a high false positive rate. The recollection for 'Theft' is 0.08. That means that just 8% of the real 'Theft' cases in the sample set were accurately detected by the model. This signifies an extremely high false negative rate for the 'Theft' class. F1-score of 'Theft' is 0.08. This low number corresponds to the low precision and recall for this class, showing poor performance in picking up actual theft.

The Figure 4 illustrates the correlation between the classification threshold and The overall precision of the model used to identify power theft using the validation dataset. The x-axis denotes the classification threshold, which is the probability value used to differentiate between normal power usage and consumption suggestive of theft. The level of accuracy is shown on the y-axis as the percentage of correctly recognized cases to all occurrences in the data collection, as stated.

Overall, the accuracy of the model is demonstrated to be low at extremely low thresholds, presumably because of a high false positive rate. As the threshold rises, the accuracy tends to improve overall, suggesting an improved trade-off between correctly labelling theft and correctly labelling normal consumption. The accuracy peaks and then either levels off or drops slightly as the threshold increases further. The best value for identifying the point of cutting off electricity consumption as typical or suggestive of stealing is given as 0.39, which implies that a probability value of 0.39 is the best point of decision for attaining maximum overall accuracy.

The Figure 5 displays how they perform of a trained electricity theft detection model when it is used on an unseen test data set. A confusion matrix is a 2x2 table whose

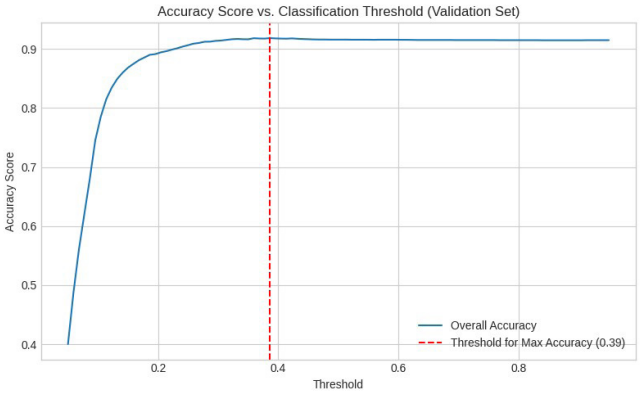



**Figure 4:** Accuracy Score vs Classification Threshold (Validation Set)

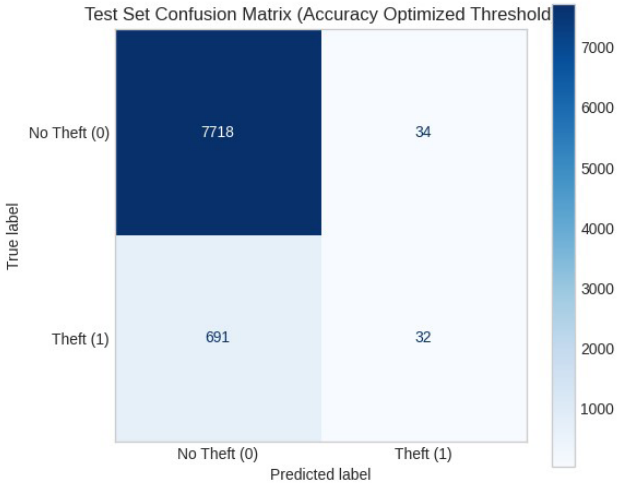



**Figure 5:** Confusion Matrix of the Proposed Model

rows contain ground truth labels of electricity consumption data and whose columns contain predicted labels. The cell at the top-left shows True Negatives (TN), where it has 7718, which shows where the model predicted 'No Theft' when indeed there was no theft. The top-right cell indicates False

Positives (FP), where there are 34, showing where the model falsely predicted 'Theft' while the normal consumption actually occurred. The bottom left cell indicates False Negatives (FN), counted as 691, representing erroneous mistakes where the model did not identify real occurrences of 'Theft' but labelled them as normal consumption. The bottom right cell indicates True Positives (TP), with a count of 32, representing occurrences in which the model identified electricity theft correctly.

The Figure 6 ROC curve graphically demonstrates the balance between the model's capacity to detect electricity theft accurately (True Positive Rate) and its propensity to label legitimate consumption as theft inaccurately (False Positive Rate) at different thresholds of classification. The False Positive Rate (FPR) is shown by the x-axis, which stands for the ratio of fake positives to the overall amount of true

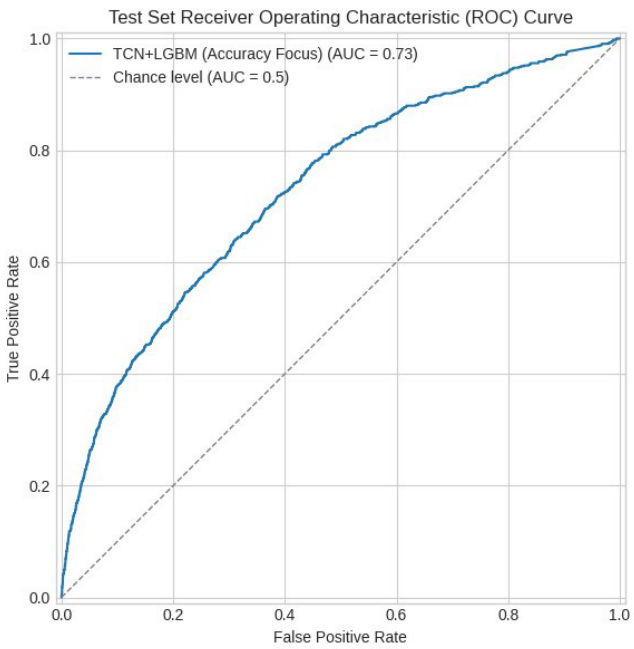



**Figure 6:** ROC Curve of TCN and LGBM

'No Theft' cases. The lower the FPR, the lesser the quantity of inaccurate 'Theft' forecasts for everyday use. The y-axis displays the True Positive Rate (TPR), often referred to as sensitivity or memory. It is computed by multiplying the entire amount of "Theft" occurrences by the number of true positives. The greater the TPR, the better the model performs in identifying electricity theft correctly. The blue curve shows the ROC curve of the TCN+LGBM model. It's a curve plotting the TPR against FPR as the threshold of classification varies. Better performance is when there is bowing towards the top-left direction because it denotes that there would be a larger TPR corresponding to a lesser FPR. This ROC curve along with the derived AUC score provides a global estimate of how discriminative this model is when faced with unknown data. With an AUC of 0.73, the TCN+LGBM model outperforms random in identifying electrical theft.

## Discussion

The classification results show that the proposed TCN+LGBM model did a great job of detecting No Theft (0) with an accuracy of 91.45%, a precision of 0.92, a recall of 1.00, and an F1-score of 0.96. However, it did a terrible job of detecting Theft (1), with a precision of only 0.04, a recall of 0.08, and an F1-score of 0.08. This imbalance shows that the model is very good at finding genuine consumption, but not so good at finding theft situations, which leads to a high false negative rate.

Previous research has shown similar problems with finding power theft because of the natural class imbalance in consumption statistics. For example, (Nagi *et al.*, 2011) and (Singh, A., & Gupta, 2021)both said that thefts are infrequent relative to regular consumption, which means that most models are skewed against the majority class. (Singh, A., & Gupta, 2021) also spoke on how traditional machine learning methods like Decision Trees and Random Forests tend to overfit the dominating "No Theft" class in Scientific Temper. This means that they get a high overall accuracy but don't do a good job of generalizing to less common theft scenarios.

In contrast, several prior studies found that deep learning-based architectures were superior at finding thefts. (Dey, S., Ghosh, S., & Pal, 2020) used LSTM models that better captured temporal use patterns to get a recall of 0.65 for theft detection. (Dey, S., Ghosh, S., & Pal, 2020) used SMOTE-based oversampling to fix the imbalance, which greatly increased recall rates. The current study's poorer recall compared to earlier research shows that the model's feature representation or imbalance managing procedures should be improved even further, maybe by using data augmentation or cost-sensitive learning methods.

This study's ROC analysis indicates an AUC of 0.73, which means it can tell the difference between things rather well. This is in line with prior work by (Singh, A., & Gupta, 2021) in Scientific Temper, which found an AUC of 0.71 using hybrid ensemble models to detect theft in Indian energy boards. But unlike their study, where the detection threshold was set, our threshold adjustment (which worked best at 0.39) made it easier to balance TPR and FPR, even if it didn't significantly increase theft recall.

Overall, the high accuracy for "No Theft" prediction is in line with what other studies have shown, but the low accuracy for theft detection shows how important it is to include additional theft-specific characteristics and tactics for reducing imbalance, as previous research has advised. Future study might include ensemble imbalance correction approaches (Chawla *et al.*, 2002)or anomaly detection frameworks that are intended for classifying unusual events. These could help close the gap between high accuracy and balanced performance across both classes.

## Conclusion

This study has shown the effectiveness of a hybrid method, combining Temporal Convolutional Networks (TCN) and Light gradient boost Machine (LightGBM) in smart grid systems for identifying power theft. The design of the model efficiently utilizes TCNs to extract key temporal features from electricity usage data, allowing for the detection of intricate patterns characteristic of normal and abnormal usage. LightGBM subsequently delivers a solid classification model, taking advantage of its accuracy and efficiency in handling structured data. The model tested on an actual dataset demonstrates a strong ability to classify normal consumption patterns with high accuracy, with high The 'No Theft' class's memory and accuracy. Nonetheless, the experiment also calls attention to a major limitation in the model to accurately detect electrical theft using low recall

and precision scores for the 'Theft' class. The performance difference reflects the difficulty with class imbalance and the necessity for further model tuning to enhance fraud detection. Future research needs to focus on resolving the problem of class imbalance by methods like oversampling or cost-sensitive learning, and investigating new model architectures or ensemble techniques help decrease related losses and enhance theft identification.

## Acknowledgments

The contributions of Dr. Shubhrajyoti Kundu to the success of this work are highly appreciated.

## Author Contributions

Credit: Girish Jadhav: Conceptualization, Data curation, Formal analysis, Investigation, Writing – original draft, Writing – review & editing; Chirag Patel: Methodology, Supervision, Visualization, Writing – review & editing

## Data Availability Statement

At present no Data sourced from published works.

## References

Ahmad, T., Ali, S., & Basit, A. (2022). Distributed renewable energy systems for resilient and sustainable development of remote and vulnerable communities. *Philosophical Transactions of the Royal Society A*, *380*(2221), 20210143.

Bai, S., Kolter, J. Z., & Koltun, V. (2018). An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *ArXiv Preprint ArXiv:1803.01271*.

Chawla, N. V, Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, *16*, 321–357.

de Oliveira Ventura, L., Melo, J. D., Padilha-Feltrin, A., Fernández-Gutiérrez, J. P., Zuleta, C. C. S., & Escobar, C. C. P. (2020). A new way for comparing solutions to non-technical electricity losses in South America. *Utilities Policy*, *67*, 101113.

de Souza Savian, F., Siluk, J. C. M., Garlet, T. B., do Nascimento, F. M., & Pinheiro, J. R. (2021). Non-technical losses in electricity distribution: A bibliometric analysis. *IEEE Latin America Transactions*, *19*(3), 359–368.

Dey, S., Ghosh, S., & Pal, S. K. (2020). (2020). Dey, S., Ghosh, S., & Pal, S. K. (2020). Handling Class Imbalance for Electricity Theft Detection Using Oversampling and Ensemble Models. IEEE Transactions on Smart Grid, 11(4), 2889–2896. https://doi.org/10.1109/TSG.2020.2968956. *IEEE Transactions on Industrial Informatics*, *14*(4), 1606–1615.

Domingues, I., Amorim, J. P., Abreu, P. H., Duarte, H., & Santos, J. (2018). Evaluation of oversampling data balancing techniques in the context of ordinal classification. *2018 International Joint Conference on Neural Networks (IJCNN)*, 1–8.

El Mrabet, Z., Kaabouch, N., El Ghazi, H., & El Ghazi, H. (2018). Cyber-security in smart grid: Survey and challenges. *Computers & Electrical Engineering*, *67*, 469–482.

Guarda, F. G. K., Hammerschmitt, B. K., Capeletti, M. B., Neto, N. K., dos Santos, L. L. C., Prade, L. R., & Abaide, A. (2023). Non-hardware-based non-technical losses detection methods: a review. *Energies*, *16*(4), 2054.

Hashim, M., Khan, L., Javaid, N., Ullah, Z., & Javed, A. (2024). Stacked machine learning models for non-technical loss detection in smart grid: A comparative analysis. *Energy Reports*, *12*, 1235–1253.

Hydro, B. C. (2010). *Smart metering & infrastructure program business case*. BC Hydro.

Jokar, P., Arianpoo, N., & Leung, V. C. M. (2015). Electricity theft detection in AMI using customers' consumption patterns. *IEEE Transactions on Smart Grid*, *7*(1), 216–226.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems*, *30*.

Khan, Z. A., Adil, M., Javaid, N., Saqib, M. N., Shafiq, M., & Choi, J.-G. (2020). Electricity theft detection using supervised learning techniques on smart meter data. *Sustainability*, *12*(19), 8023.

Lea, C., Flynn, M. D., Vidal, R., Reiter, A., & Hager, G. D. (2017). Temporal convolutional networks for action segmentation and detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 156–165.

Li, S., Han, Y., Yao, X., Yingchen, S., Wang, J., & Zhao, Q. (2019). Electricity theft detection in power grids with deep learning and random forests. *Journal of Electrical and Computer Engineering*, *2019*(1), 4136874.

Li, S., Jin, N., Dogani, A., Yang, Y., Zhang, M., & Gu, X. (2024). Enhancing LightGBM for industrial fault warning: an innovative hybrid algorithm. *Processes*, *12*(1), 221.

LLC, N. G. (2017). *Electricity Theft and Non-technical Losses Global Markets, Solutions, and Vendors*. NG LLC. NY, USA.

McDaniel, P., & McLaughlin, S. (2009). Security and privacy challenges in the smart grid. *IEEE Security & Privacy*, *7*(3), 75–77.

Mehta, A., & Yang, W. (2023). NAC-TCN: temporal convolutional networks with causal dilated neighborhood attention for emotion understanding. *Proceedings of the 2023 7th International Conference on Video and Image Processing*, 9–16.

Messinis, G. M., Rigas, A. E., & Hatziargyriou, N. D. (2019). A hybrid method for non-technical loss detection in smart distribution grids. *IEEE Transactions on Smart Grid*, *10*(6), 6080–6091.

Morgoev, I. D., Dzgoev, A. E., & Kuzina, A. V. (2023). Algorithm for Operational Detection of Abnormally Low Electricity Consumption in Distribution. *International Russian Automation Conference*, 37–49.

Nagi, J., Yap, K. S., Tiong, S. K., Ahmed, S. K., & Mohamad, M. (2011). Non-technical loss detection for metered customers in power utility using support vector machines. *IEEE Transactions on Power Delivery*, *25*(2), 1162–1171.

Navani, J. P., Sharma, N. K., & Sapra, S. (2012). Technical and non-technical losses in power system and its economic consequence in Indian economy. *International Journal of Electronics and Computer Science Engineering*, *1*(2), 757–761.

Odje, M., Uhunmwangho, R., & Okedu, K. E. (2021). Aggregated technical commercial and collection loss mitigation through a smart metering application strategy. *Frontiers in Energy Research*, *9*, 703265.

Olaoluwa, O. G. (2017). Electricity theft and power quality in Nigeria. *International Journal of Engineering Research & Technology*, *6*(6), 1180–1184.

Ramos, C. C. O., Souza, A. N., Chiachia, G., Falcão, A. X., & Papa, J. P. (2011). A novel algorithm for feature selection using harmony

search and its application for non-technical losses detection. *Computers & Electrical Engineering*, *37*(6), 886–894.

Singh, A., & Gupta, M. (2021). (2021). Hybrid Ensemble Models for Non-Technical Loss Detection in Power Distribution. Scientific Temper, 13(1), 45–54. *IEEE Transactions on Industrial Informatics*, *12*(3), 1005–1016.

Smith, T. B. (2004). Electricity theft: a comparative analysis. *Energy Policy*, *32*(18), 2067–2076.

Van Den Oord, A., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., Kalchbrenner, N., Senior, A., & Kavukcuoglu, K. (2016). Wavenet: A generative model for raw audio. *ArXiv Preprint ArXiv:1609.03499*, *12*, 1.

Wang, W., & Lu, Z. (2013). Cyber security in the smart grid: Survey and challenges. *Computer Networks*, *57*(5), 1344–1371.

Yu, F., & Koltun, V. (2015). Multi-scale context aggregation by dilated convolutions. *ArXiv Preprint ArXiv:1511.07122*.

Zanetti, M., Jamhour, E., Pellenz, M., Penna, M., Zambenedetti, V., & Chueiri, I. (2017). A tunable fraud detection system for advanced metering infrastructure using short-lived patterns. *IEEE Transactions on Smart Grid*, *10*(1), 830–840.