ELECTRICITY THEFT DECTECTION VIA DEEP LEARNING

By

Sehrish Farid

00000277686

Supervisor

Assoc Prof Dr. Naima Iltaf

A thesis submitted to the faculty of Computer Software Engineering, Military College of Signals, National University of Sciences and Technology, Islamabad, Pakistan, in partial fulfilment of the requirements for the degree of MS in Software Engineering

September 2022

DECLARATION

I, *Sehrish Farid* declare that this thesis titled ”Electricity Theft Detection via Deep Learning” has not been already submitted for a degree or some other qualification at NUST or some other institution.

Sehrish Farid
00000277686/MSCS25

Copyright

DEDICATION

This thesis is dedicated to

MY HUSBAND and MY FAMILY for their unconditional support, TEACHERS for their respective guidance

AND FRIENDS for their help at difficult times

ABSTRACT

Electricity theft is a major problem facing many countries around the world. Its adverse effects include loss in revenue for power distribution companies and government economy, the distribution quality of electricity, increased generation load and high electricity cost which affects honest consumers as well. With the advancement in smart meters in electricity infrastructure, massive data is generated that can be analyzed for electricity consumers consumption patterns. By the help of these consumption patterns of consumers several machine learning and deep learning models and techniques are built to detect fraudulent consumers. By the help of the detection of the theft location and fraud consumers energy distribution companies can impose fines on these consumers that help them to reduce revenue losses. In this research a combination of CNN and STN based model is proposed to detect electricity theft. STN is used in many image classification problems along with CNN to enhance the performance of the models. STN along with CNN is used for the first time in electricity theft detection. STN is used to translate, rotate and scale the the original input. The dataset used in this research is real customers electricity usage data publicly provided by the State Grid Corporation of China (SGCC). SGCC dataset has missing values, and also has class imbalance problems. The number of theft users is significantly lower than the honest consumers, which is addressed by using Synthetic Minority Oversampling Technique (SMOTE). Proposed model is compared with state of the art machine learning and deep learning models and techniques and results show that the proposed model can identify theft and normal users with greater accuracy.

ACKNOWLEDGEMENTS

All praise to ALLAH, The Magnificent who's blessing has bestowed me with strength required to complete this thesis and helped me to cross every hurdle.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

INTRODUCTION

The well-being of living in society is in association with the basic necessities of life. Availability of electricity is one of the basic necessities of life. The continual electricity supply depends on effective and efficient transmission and distribution infrastructure. Electrical losses are broadly categorized as Technical and Non-technical losses[1].Figure 1.1 shows the electricity distribution network. The Technical losses

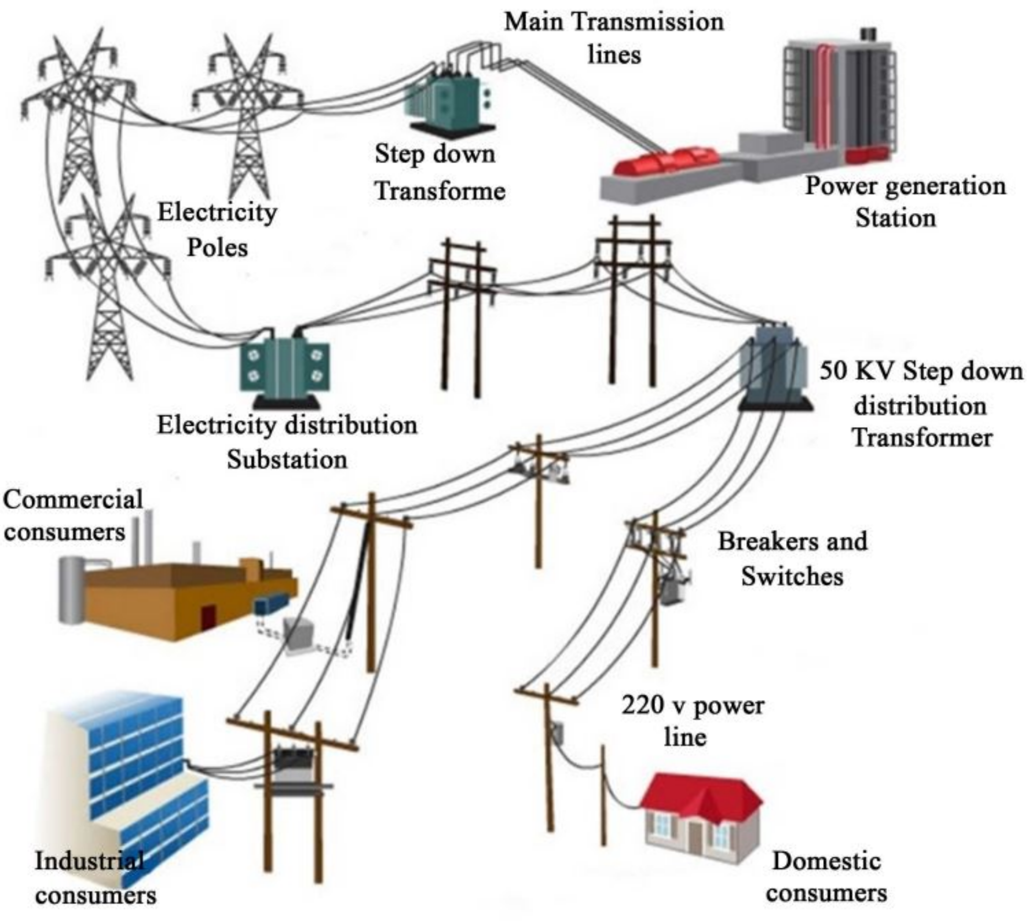

Figure 1.1: Electricity Distribution Networks

are due to various factors including energy heat debauched in the magnetic losses in transformers, equipment used for the energy distribution, transmission lines, conductors, and other equipment. Technical losses are typically 22.5%[2] and directly depend on the transmission network characteristics. Estimation and calculation of technical loss are exceedingly complex, making it hard to pinpoint the loss area. The technical loss cannot completely eliminated but can be reduced by applying advanced techniques throughout the system. . Figure 1.2 shows the categories of Technical and Non Technical Loses.

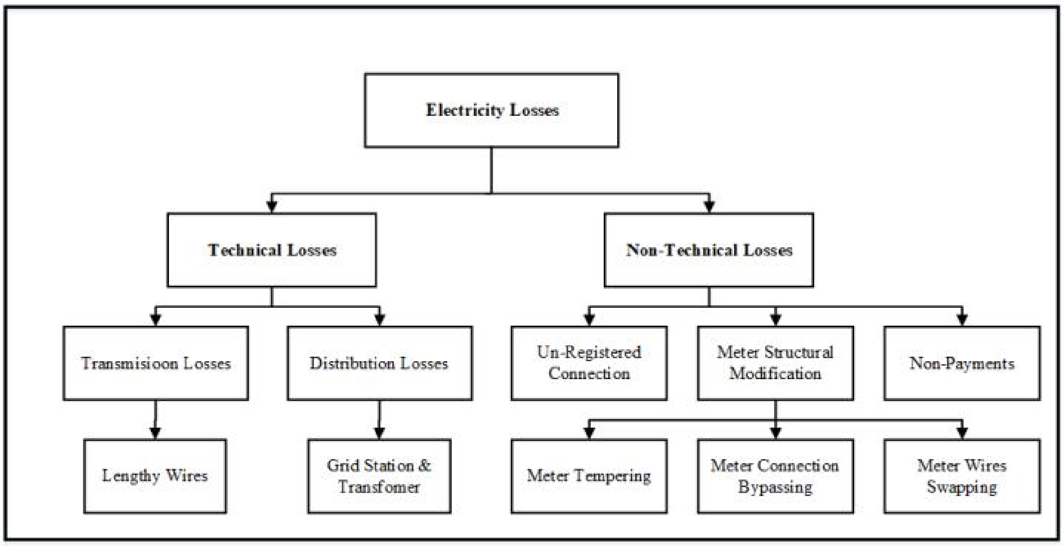

Figure 1.2: Technical and Non Technical Loses

Non-technical losses are related to administrative losses that happen because of non-billed electricity, errors in billing, faults of the equipment, low-quality framework, and deceptive consumption of electricity. Energy theft by consumer deception includes illegal connection points, and bypassing meters. Stop the rotating disk of electricity meters, physical destruction of meters, and tampering, regularized corruption, and organized crime [3]. All the mentioned reasons for Non-Technical losses, electricity theft is the most advent issue faced by countries around the world. Electricity theft or NTLs highly affects developed and developing countries, and it can affect developed economies [4].

According to research conducted by North East Group in 2015, the world economy loses US$ 89.3 billion per year amid Non-Technical losses [5]

Some Repercussion of economic and power companies' losses includes the decrease

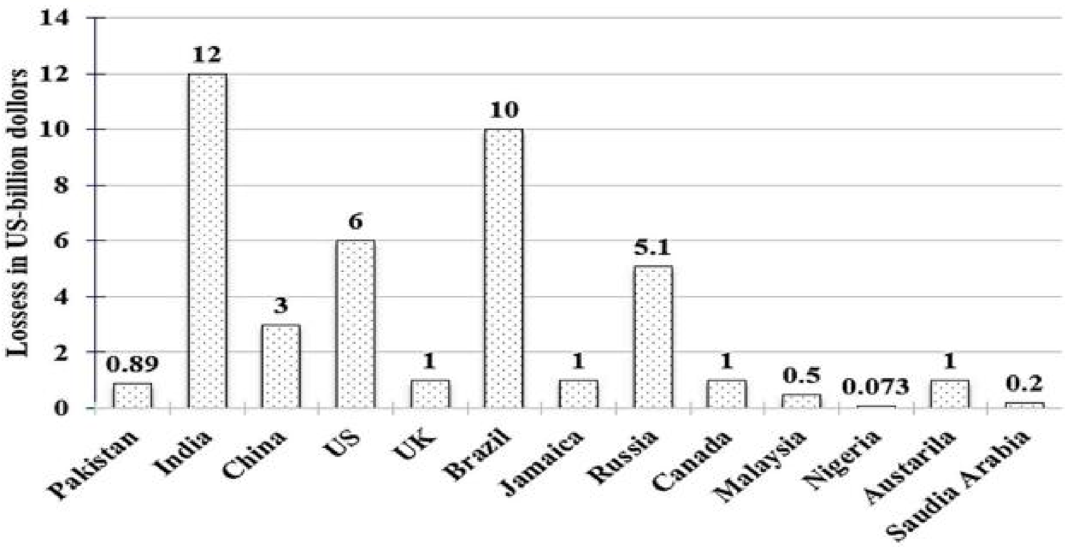

Figure 1.3: Electricity Losses of Countries

in the capital for investment in power utilities in the future, an increase in the prices of electricity on the consumer side, adverse damage in the infrastructure of electricity transportation, the necessity of investments from the government when it could invest in other areas and more. The common ways of ETD are to conduct manual inspections of selected consumers and audit their previous electricity bills. However, these approaches are inefficient because they are both time and labor-intensive. The advancement of traditional grids to smart grids allows a two-way flow of information and energy that enables real-time energy management, billing, and load surveillance [6].

Due to the high cost of electricity, along with the limited amount of power resources, effective and efficient utilization of resources is a very crucial part of the economic and social well-being of any county. To achieve this, a smart grid is an essential solution for efficient monitoring of the usage of power resources. Figure 1.4 shows a smart grid infrastructure. Smart grid systems can be explained as the whole electricity network comprising the energy infrastructure and system to supervise and control the electricity usage as well as an intelligent system that monitors the consumer electricity usage patterns and mode of action of all the customers in the system. [7]. Figure 1.5 shows

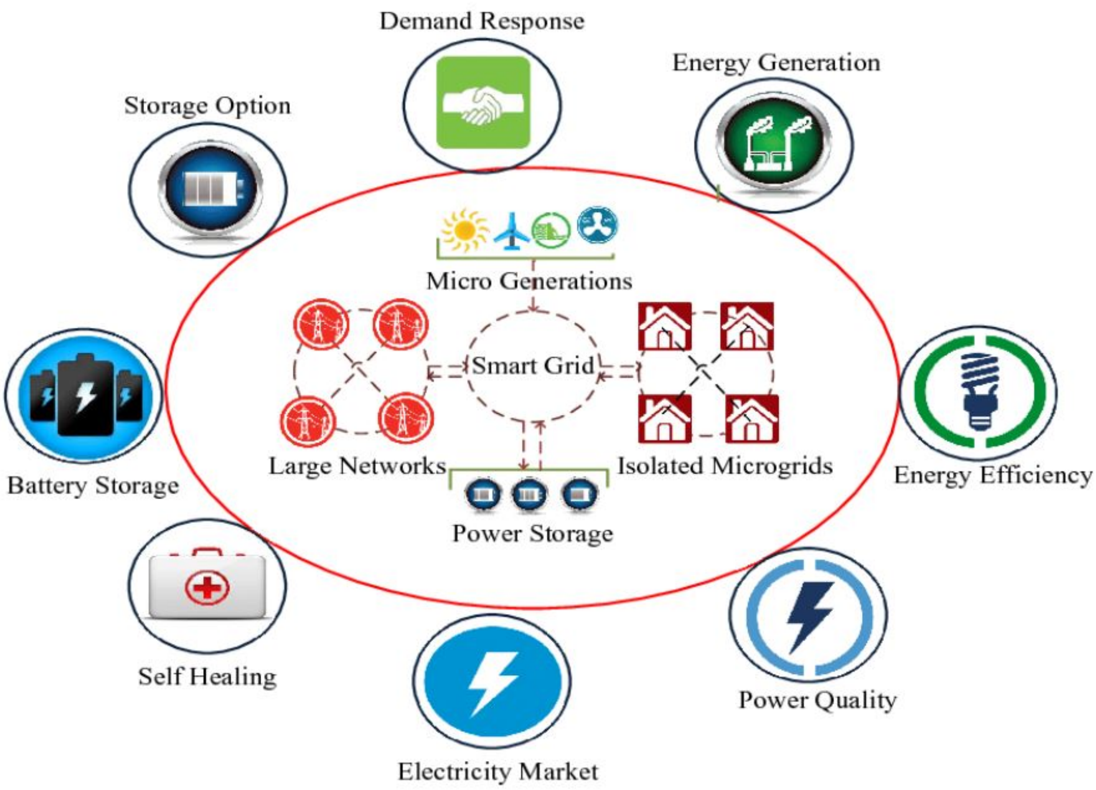

Figure 1.4: Smart Grid Infrastructure

the smart meter.

With the help of an intelligent grid system, energy usage readings are sent to the operational centers using the network, and electricity transmission companies do the billing for each consumer based on the readings of their smart meters[8].

AI based models and techniques comprise Machine Learning and Deep Learning. Machine Learning approaches are based on the training of a model with an algorithm to process huge datasets effectively by predictive modeling. While deep learning [48] consists of artificial neural networks, which employ the functioning of the human brain model and helps the models assess and learn irrational functions. Due to the ability to process and manage massive datasets, and automate feature extraction, and its classification process, deep learning techniques are used to create techniques and models to work with smart meter data originating from the smart grids.

Deep Learning has achieved remarkable performance in several fields such as

Figure 1.5: Smart Meter

speech[9], Natural language processing[10], and computer vision which motivates us to apply it for the detection of fraud in electricity. Deep learning techniques are being applied to several other areas of astronomy[11], and genomics[12]. In this research CNN and STN-based method is proposed for theft detection in electricity, STN is proven to be higher performance in combination with CNN. Experiments are conducted on the SGCC data set of real electricity consumers data to evaluate the performance of the model

1.1 Categorization of NTL Detection Methods

Different methodologies and research are explored by researchers to identify fraudster customers efficiently and effectively. Existing NTL detection methods are broadly categorized as hardware and non-hardware-based methods. Installation of meters with special devices that are able to PDCs in detecting malicious activity by consumers [13,14].

Non Hardware-based solutions prove efficient in NTL detection with advancements in communication and data processing of the behavior of energy consumers. Furthermore, the association between socio-economic problems and demographic aids PDCs to investigate the process of NTLs is classified into an affront theoretical method.
Non-hardware NTL detection methods are classified into three major categories:

1. Data-based methods.

2. Network-based methods

3. Hybrid methods

A general classification of NTL detection is shown in Figure 1.6

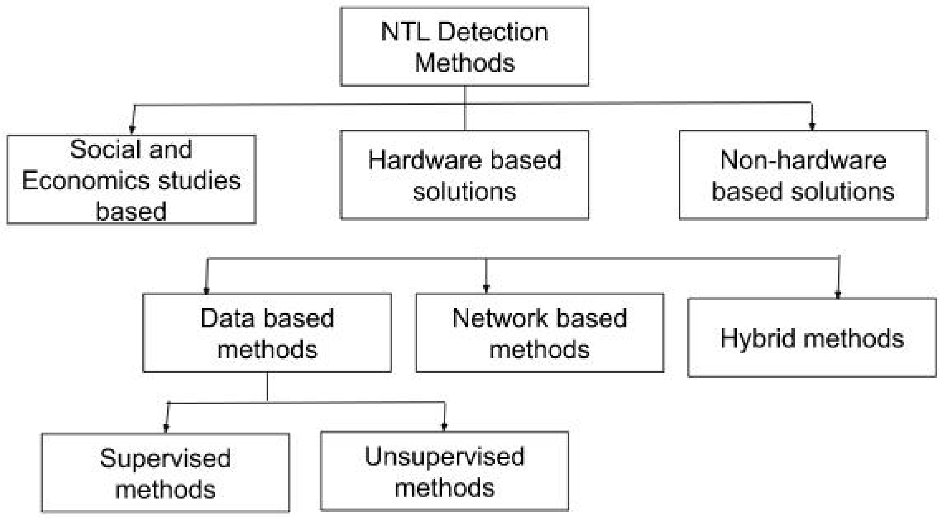

Figure 1.6: Classification of NTL Detection Method

1.1.1 STATE AND ECONOMIC STUDIES BASED

In this category of research the existence of non-technical losses over the specific geographical area or population and social implications associated with the fraud are explored. The authors in [15,16] utilized the statistical methods to make relevance among economic, market variables, and social demographics to the nature of fraud. The real benefit of these kinds of research is that they are helpful in formulating principles and have a huge influence on decreasing NTL. The major limitation of these methods is that their application scope is specific to the region or country.

1.1.2 HARDWARE-BASED SOLUTIONS

Hardware-based solutions based on the design and characterization of physical equipment that enables the detection and estimation of any malicious activity [15]. The author in [17] proposed a method that detects the tampering of power meters. This helps the power distribution companies to detect the bypassing main line and neutral line and reverse problems by sending messages. The drawback of these methods is the huge amount needed for physical installation of the systems[47], which is not economical for most power distribution companies, specifically for underdeveloped countries.

1.1.3 Non-Hardware Based Solutions

Non-Hardware based techniques depend on the research, whose main focus is to detect electricity theft from users' electricity usage data. These techniques are categorized into three methods:

1. Data-based methods.
2. Network-based methods
3. Hybrid based methods

1.1.4 Data Based Methods

Data-based methods are focused on machine learning algorithms and data analytic. Data-based methods consist of supervised and unsupervised methods. Each has its own advantages and limitations.

1.1.5 Network Based Methods

Network based methods are based on the information obtained from smart grid systems and the estimation of different physical features of the electrical network for effectively detecting non-technical losses [18,19]. Methods and techniques that fall in this category are more accurate but these methods' applicability is not easy. To calculate the number of dedicated sensors and their specific placement is determined by the AI algorithms.

1.1.6 Hybrid Model

Hybrid models are a combination of techniques or algorithms from data-driven and network implementation methods from network-based models. Hybrid models are based on models. Hybrid models are applied due to the higher accuracy of the detection of theft.

1.2 Problem Statement

Electricity theft detection is very crucial for every country around the globe. As it affects the country's economy and electricity system infrastructure. If electricity theft is not detected accurately and effectively its consequences become adverse. Consequences are faced not only by power distribution companies but also by honest consumers too. It decreases the profits or revenue returns of the power companies, increases the cost of electricity per unit at the consumer side, load shedding, and inflation. As power distribution companies and the government obtain desired revenue, they cannot invest in the betterment of the electricity distribution system. To reduce these problems electricity theft detection has become a major challenge for developed and underdeveloped countries. NTL losses around the world are US$ 90 billion per year. This includes the USA, Brazil, Malaysia, UK, India, and Pakistan. Pakistan loses US$ 0.89 billion per year which is a huge amount for our underdeveloped country. Therefore an effective and efficient model to detect electricity theft is needed. Deep learning models are of great concern in this regard as they proved remarkable in terms of accuracy and performance.

1.3 Contribution

The main contributions of this study are as follows:

1. Propose an efficient and effective electricity theft detection model based on the combination of STN and CNN.

2. Class imbalance problem of the dataset is studied and a solution is proposed to enhance the performance of the model

3. Experiments are conducted on deep learning models after and before the resolving class imbalance issue

1.4 Relevance to National Needs

Pakistan is an underdeveloped country. Economically weak countries like Pakistan cannot endure economic loss by any kind of electricity theft i.e. fraud, corruption, stealing. Pakistan is facing the issue of electricity theft due to political uncertainty, low-level of government efficiency, appointments of nontechnical staff, poor infrastructure, and lack of efficient theft detection models or techniques. Pakistan is the country that has the benefit of natural resources for the production of electricity, but due to the lack of long-term proper planning and theft detection mechanism, the country bears economic losses. This research focuses on the data-driven model for electricity theft detection by applying deep learning algorithms and techniques.

1.5 Area of Application

The research will help Pakistan electricity companies to detect Non-technical Loses. The detection of fraudulent customers will help to reduce the losses. Power distribution companies can act against them to recover the loss by imposing fines. Honest consumers are no longer need to pay high price of electricity by locating the theft consumers.It will help them to increase revenue return which they will invest in power infrastructure for the betterment of the country. Electricity prices will reduce. Government revenue will increase with the detection of theft.

1.6 Advantages

- Reduces Revenue losses

- Reduces Electricity prices

- Detect the actual fraudulent consumers

- Power companies can invest for the betterment of energy distribution infrastructure.

1.7 Thesis Outline

The outline for chapters can be given as follows:

- Chapter 1: Introduction and objectives of electricity theft detection included in this chapter.
- Chapter 2: This chapter comprises the literature and background along with a brief description of existing techniques and approaches to electricity theft detection.
- Chapter 3: Proposed deep learning model to detect electricity theft detection, which includes dataset description, preprocessing, and proposed model details.
- Chapter 4: This chapter contains a comparison of machine learning and deep learning techniques and the results of the proposed technique.
- Chapter 5: This chapter discusses the conclusion and possible future work directions.

Chapter 2

LITERATURE REVIEW

This research explores an effective and efficient technique that detects electricity theft, which helps power companies to locate theft areas and consumers and take action accordingly. To develop a model, it is worth analyzing existing techniques and approaches developed by different researchers in this regard. This section presents a brief description of existing techniques in terms of data availability, data preprocessing, feature extraction and engineering, and model training and testing mechanisms.

In [20] Applied wide and deep convolution Neural networks to detect fraudulent electricity consumers dataset used in the research is a smart grid provided by State Grid Corporation of China. The major contribution was that they transformed 1D data into 2D data to apply the CNN model so that CNN can detect the periodicity of normal consumers and the non-periodicity of thieves. The author evaluated the model by using multiple classifiers out of RF SVM, LR wide and Deep CNN performs better.

In [21] proposed NG Boost algorithm to classify Normal and thieves consumers. Improve the performance of the model by applying data preprocessing techniques. To deal with missing values utilized the random forest algorithm-based imputation. To deal with the class imbalance issue in data, a weighted minority oversampling technique was used. The NG Boost algorithm was used to train and test the model to accurately detect consumers' profiles as healthy or as theft. Tree SHAP is used to evaluate the performance of a model.

In [22] the author addressed the three challenges of fraud's data set. The challenges addressed were negative impacts on the model's performance due to the imbalanced dataset. I.e. mostly frauds consumers are less as compared to normal consumers, second the higher dimensionality of the time series data in electricity data. Finally selection of the evaluation parameters specifically for the imbalanced dataset. The framework proposed by the authors was Maximal overlap Discrete Wavelet-Packet Transform (MODWPT) for feature extraction and Random Under-Sampling Boosting for Non-technical loss detection. NTL detection quality improves by the combination of MODWPT and RUSBOOST. Experiments were performed using smart meter data readings from Honduras comprising 2271 consumers' records.

In [23] proposed Edge computing enabled non-technical loss-detection Data collected from different sources involving UMass Amherst, the Department of Energy, and the federal open government data repository. The model was performed on the electricity readings of office buildings, homes, and apartments, hence dealing with big data to detect NTL. ENFD performs better than other detectors e.g. NFD, DCI, FNFD, BCGI.

In [24] used fuzzy Machine learning technique that is self-organized neural networks to identify non-technical losses by domestic consumers. Measurement data of the real network was used by the bottom up method. Theft detection was calculated by an anomaly index of every subscriber. The intent of healthy behavior patterns of domestic loads with and without photovoltaic sources are determined as normal patterns. Classes of the anomaly were defined according to the type of fraud. 3 classes vary by passing meters at peak hours, unmetered circuits used as loads at specific hours, and customers selling the grid. Model performance of correctly detected frauds for class A, class B, and class C is 64%, 49%, and 91% respectively.

[25] used Light S.A. company, Brazilian dataset to detect theft. For the identification of theft, the intelligent model comprised a couple of neural network ensembles which were envisaged to increase the accuracy in the detection of irregularities amid low tension consumers. The proposal was divided into three stages: preprocessing and normalization, filtering module and classification module. The Ensemble Neural Network depicts an increase in performance to detect low tension consumers. Performance parameters, Positive Predictive Value (PPV), and class error rate were used for the evaluation of the model. The model depicted PPV of 62.6% and a class error rate of 31.3%.

In [26] conducted data preprocessing on an imbalanced SGCC dataset. 2014, 2015, and 2016 in such a manner that has values of NaN, zeros, and normal data. In the first step of preprocessing 33979 consumer data was obtained out of 42372. Then after the data selection process, missing values NaN and zero values were imputed by arithmetic average. Data were normalized between the range of -5 to 5. According to the cross-fold validation method, the dataset processed is divided into 5 different parts. LSTM was used to compare the accuracy, F1 score Recall, and precision before and after data preprocessing.

In [27] Conducted preprocessing techniques on the SGCC dataset. The interpolation method was used to impute missing values. The empirical decomposition model was exerted for the breakdown of the time series signature. For the maximum with min features for the proposed model i.e. K-Nearest extensive experiments were done on statistical and time domain data. Experiments were concluded as thirteen features were giving the best accuracy. Those features are Peak to Peak (PP), Standard Deviation (STD), Waveform length Ratio (WLR), Mean ()M), Skewness (SK), Log Energy (LF_E), Median Absolute Deviation (MAD), Sparseness (SP), Crest Factor (F), Energy Entropy (EE), Spectral Decrease (SPD) and Total Harmonic Distortion (THD), A feature vector is formed comprised of above mentioned 13 features for the classification. 9 classifiers coarse tree, Linear Discriminant, Medium Tree Logistic Regression, Fine Tree Medium KNN, cosine KNN, Coarse KNN, and Fine KNN perform best in all of these classifiers and got 91% accuracy.

The authors in [28] proposed a semi-supervised framework named RDAE-AG-Triple GAN. The model consisted of relational denoising autoencoder (RDAE) and attention guided (AG) Triple GAN. In the proposed framework RDAE was executed to extract features and their relevance and AG was applied for feature weighting that was extracted and dynamically supervised by the AG Triple GAN. Performance Parameters precision FI, MCC, AUC, and Recall were used for the evaluation of the proposed system. The model depicted a 95% true detection rate.

In [29] Proposed a combination of neural network and random Forest (CNN-RF) to detect electricity usage irregularities. To investigate the features of data among different hours of the day and different days CNN was designed. Features were extracted from huge and varying smart meter data by applying convolution and downs sampling operations. To mitigate the issue of overfitting, a dropout layer was added. In the training phase, the backpropagation algorithm was used to update network parameters. After that Random forest was trained on the bases of extracted features. The proposed hybrid module embraces the grid search algorithm for optimal feature selection. For dealing with the missing values discard the consumer record whose continuous missing value is greater than 10. Which was the limitation of the model. Only selected those consumers having single missing data. The model depicted an 83% F1 score.

The authors in [30] conducted experiments on real-time consumer electricity usage data provided by the state grid of China Corporation. A combination of convolutional neural networks (CNN) and Long Short Term Memory (LSTM) was proposed. In the proposed hybrid architecture CNN was used to automate feature extraction and due to the time series nature of LSTM was used. A hybrid model achieved satisfactory results by applying the synthetic minority oversampling technique (SMOTE) to deal with class imbalance problems. The performance of the model was compared before and after applying SMOTE.

In [31] presented two computational models for theft identification for both low and high-voltage electrical power customers by applying data mining techniques. The electricity usage patterns (i.e. low and high) needed different methodologies and approaches. For high voltage usage consumers, identification of theft was done by using an artificial neural network i.e. self Organizing Maps (SOM) which permits the detection of the consumption profile historically registered, and comparison with current patterns. For the low voltage consumers, hybrid data mining techniques were explored. Research majorly focused on High Voltage consumers.

In [32] authors proposed a Consumption Pattern Based Electricity Theft Detector named CPBETD. In the proposed model, the total course of electricity consumption recorded and reported by smart meters is compared with the calculated usage of every neighborhood by transformer meters. At this level of Non-Technical loss, consumers in the region with abnormal patterns were considered thieves or theft consumers. A multiclass support vector Machine was trained by making use of historic data of the consumer along with a synthetic attack data set. After that classifier was used on over real data of 5000 customers. Performance parameters with high false positive rate (FPR) and detection rate (DR) were used to evaluate the performance of the model. However, the authors did not explain how they collected data from the 5000 customers.

In [33] The issue of class imbalance for power theft identification was explored. Datasets available for theft detection are highly imbalanced. Consumers who perform fraud are much less than normal consumers, which affects the accuracy and performance of the classifiers. To improve the accuracy and performance, different classifiers and techniques, such as optimum path forest, one-class SVM, and C4.5 decision techniques were combined. A combination of the classification techniques depicted a 2% - 10% upgrade over sole classifiers. But the drawback of the proposed method is that it imposed a high computational load and the performance upgradation is not substantial.

In [34] addressed the three main problems in the detection of electricity theft. These were unlabeled data, II) imbalance dataset and III) specific parameters for the validation of the classifier performance. To address these problems, Optimized Classifier Differential Evaluation Random Under Boosting was proposed. The metaheuristic optimization algorithm Differential Equation was used to optimize the proposed model. The dataset used in this research was real consumer data publicly available by State Grid of China corporation. To remove outliers from the dataset Density-Based Spatial Clustering of application with noise was conducted. However, the limitation of the model was the loss of potential information in under-sampling.

In [35] Proposed a committee based on a semi-supervised technique. The authors addressed the problem of less number of fraud records in the utility dataset. For this, they used labeled and unlabeled data. Random forest is first trained by labeled data, after that unlabeled data that the model classifies is confidently added to the labeled data. Labeled sample sets iteratively. The fraud detection performance was improved by using a training by committee semi-supervised learning in comparison to the supervised models that only used labeled data to train the model. The model shows 89% AUC.

In [36] to detect illegal electricity usage patterns of consumers, a mathematical model based on levenberg-Marquardt (LM) and probabilistic neural network (PNN) was applied. They divided consumers' electricity usage patterns into two categories:

- Users consume a portion of required electricity illegally. If the user consumes electricity by deception for a portion of the day
- Experiments were conducted to detect two types of consumers by applying methods individually and in combination - Four experiments were conducted to detect theft for both categories of consumers. The fourth experiment depicts better results than the rest which was the combination of both models.

In [37] the purpose was to propose a solution for the economic return, and the cost estimation required to build electricity theft detection infrastructure, for this both income retrieval and the inspection amount were studied. The proposed model assists the companies in short and long-term planning which includes company budgeting which region is considered first etc. Uruguay's Utility (UTE) a large dataset of real electricity usage consumers was used. Dataset was oversampled before LR and ANN, RF the maximum economic return was achieved with RF for f1-score, precision, and TPR.

In [38] is used ANN, LSTM, decision tree, RF, CNN, and autoencoder to detect theft in power in a labeled dataset that was developed by the authors [43] as a benchmark. The metrics used to compare the performance of the classifiers were AUC, F1 score average, and accuracy. CNN presented a better performance than other methods, and deep learning techniques. Autoencoder and LSTM were better than ANN in accuracy and F1 average. However, AUC is not better.

In [39] Hybrid self attention model was proposed to detect theft in the SGCC dataset. Binary Mask was introduced to deal with the missing values. Experiments were conducted after and before quantile transformation. By applying quantile transformation results were improved. Authors utilized a multihead self attention framework along with dedicated convolutions and unification is performed of kernel size 1. AUC is 92 % but f1 score is 60%.

In [40] spatial transformer networks were proposed. STN consists of 3 components: localization, grid generator, and sampler. It transforms spatially invariance input to translation, scale, rotation, and more generic wrapping. It improves the CNN classifier performance as CNN lacks the ability to perform well on invariant input data. Experiments were done of spatially invariant data, which shows improvement in terms of accuracy.

In [41] authors utilized STN for the first time in CNN-based hyperspectral image (HSI) classification. CNNbased HSI classifiers can take input as a cuboid, the dimensionalities of the cuboid were crucial for the effectiveness of the classifier. STN was used to generate an optimal input feature map for CNN. STN was utilized to rotate, translate and scale the original data to cuboid data effectively. Experiments were conducted which show that with the integration of STN, a significant accuracy.

In [42] authors were inspired by the effectiveness of STN in combination with CNN, and an STN and CNN-based model was proposed for modulation classification for the first time. They apply STN on raw baseband I/Q invariant data to bring effect in channels. Experiments were conducted to investigate the STN and CNN and conventional models in terms of accuracy, STN, and CNN.

Chapter 3

PROPOSED METHODOLOGY

This chapter comprises of the detail description of dataset, data preprocessing, and proposed methodology. Electricity theft is a major issue around the globe. Traditional machine learning models have some limitations in regard to computation or performance. With the remarkable performance of the Deep learning models, they are utilized in many fields such as computer vision, astronomy, image classification, image restoration etc. Deep learning models are utilized in this research to propose classifier for the electricity theft detection. A combination of CNN and STN based model is proposed to detect electricity theft in the real consumers usage data provided publicly by the State Grid Corporation of China (SGCC).

3.1 Dataset

Research is conducted on a dataset provided publicly by the State Grid Corporation of China (SGCC). The nature of data is a time series of data having a series of real consumer electricity readings. Data depicts the daily electricity usage of 42372 user units. The Time window of data is from January 2014 to October 2016, corresponding to o1035 days. Table 3.1 presents the information of metadata of the dataset.

Dataset has two classes i.e. Normal consumers and thieves consumers. Dataset is highly imbalanced as normal users are in 38757 in numbers whereas thieves are in 3615 which consists of 91.45% and 8.55% of the dataset respectively. Dataset is clearly an example of an imbalanced dataset, where one class to another class is biased or skewed. The dataset comprises 42372 rows and 1035 columns. Metadata types include numbers, characters and missing or erroneous values called non-numeric (NaN). Dataset is labeled data, as in the Flag column consists of two values 1 or 0. 1 refers to

Description	Value
Time Window	2014/01/01 - 2016/10/31
Normal consumers	38,757
Fraudulent consumers	3,615
Total consumers	42,372
Missing data cases	Approx. 25%

Table 3.1: Description of SGCC Dataset

the fraud and 0 as a normal user.

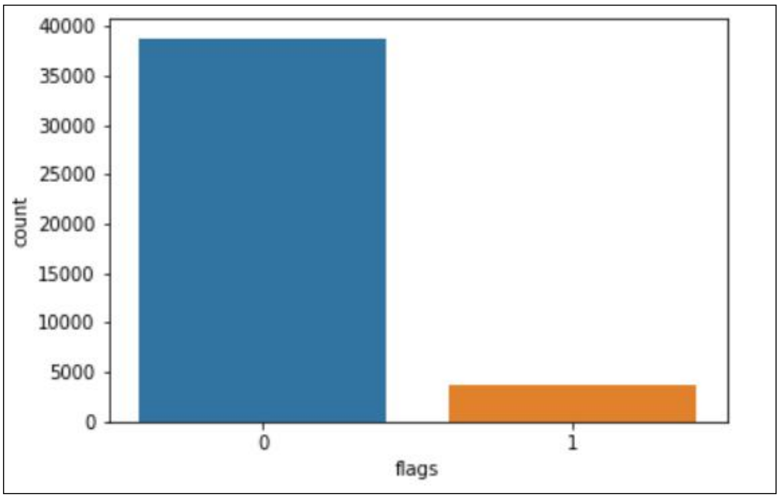

Figure 3.1: Normal and Theft consumers in SGCC dataset

3.1.1 Data Preprocessing

Data preprocessing plays a very important role in determining the success or failure of many deep learning techniques. In our research the real electricity consumers' SGCC dataset has some particular features, a large number of missing entries, periodicity, and non-periodicity of normal and thrive consumers class imbalance problem i.e. theft class is significantly less than normal class, a strong skewness, and kurtosis.

3.1.2 Data Transformation

Data is reshaping from 1D to 2D which is an important contribution of [1], which allows for the analysis of data samples as images and makes it easy to apply it to computer vision and image processing like CNN, spatial domain. Fig 3.2 shows the transformation of 1D data to 2D format.

The motivation behind transforming 1D data to 2D data in [1] is as follows.

- As the electricity consumption data is time series data, it is difficult to analyze the periodicity of normal consumers and the non-periodicity of theft consumers with massive size.
- It is hard to capture key characteristics or features of consumers from 1D data.
- To make it available for use in deep learning techniques.

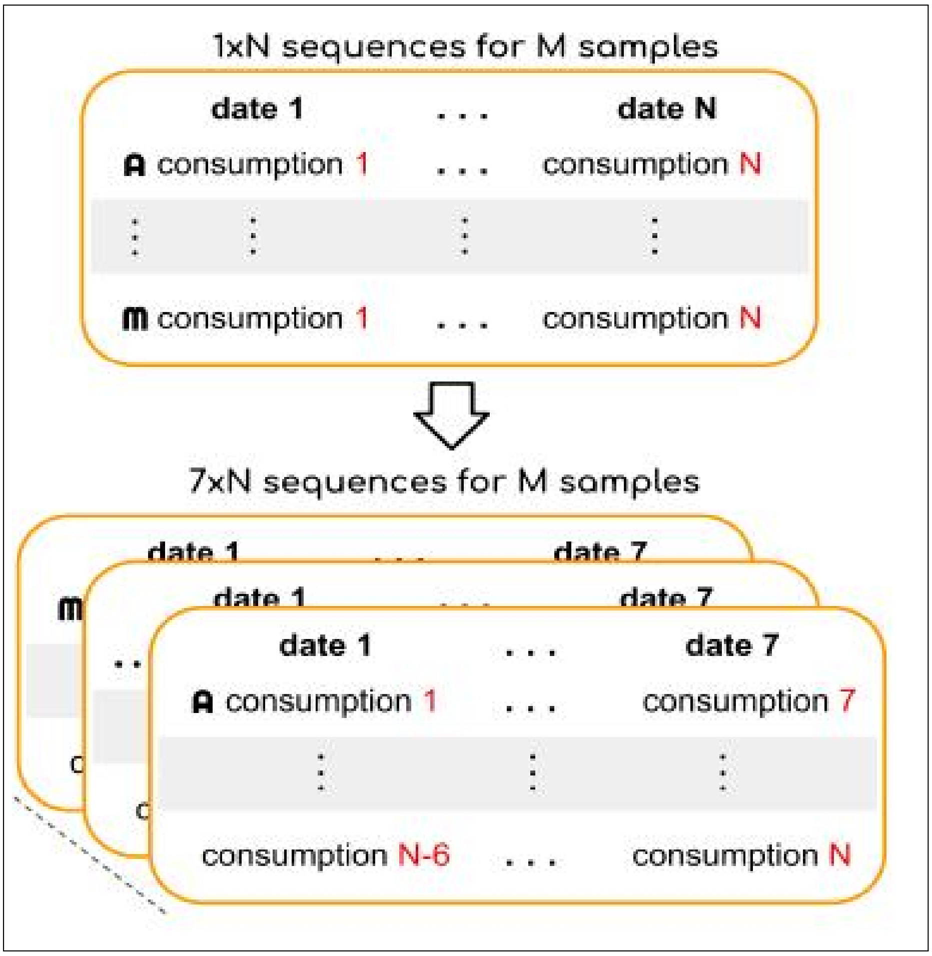

Figure 3.2: Transformation of 1D data to 2D data

3.1.3 Missing Values

After reshaping data, the next challenge for developing an efficient and effective deep learning technique is missing values in the dataset. In [21] presents the detailed description of missing values. Figure 3.3 shows the missing values present in the SGCC dataset.

In the fig, it shows among 60.11% of the total users had less than 200 missing read-

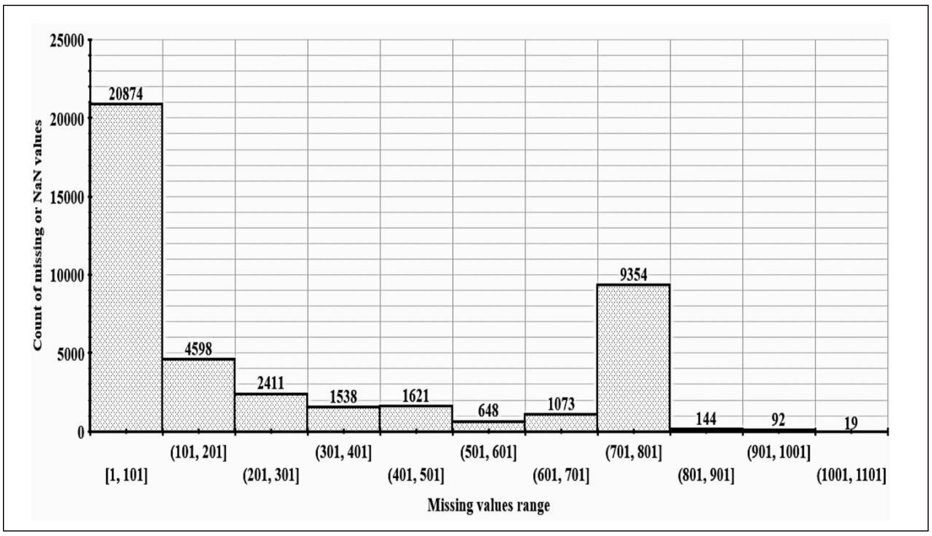

Figure 3.3: Missing Values in SGCC Dataset

ings. There are two approaches to fill in the missing values. One practice to delete the missing entries from the dataset and this may lead to loss of valuable insight of the data. Second, the mostly used approach to estimate the missing readings in the dataset is interpolation or with mean or median of the data [44]. In the same context other techniques such as practical swarm optimization, genetic algorithm and simulated annealing have also been proposed [45]. These techniques proved prohibitively slow when dealing with huge datasets.

To deal with missing values we used the method of Binary mask and contribution from [20]. In this method a binary mask is created as an additional input channel and it is calculated as follows: In the first step the indices of all missing values found, the value 1 is given to these indices and all remaining gets 0. This is the binary mask. The missing value indices at the value channel receive a value of zero. The two inputs that are binary mask and value channel is given to 2D CNN.

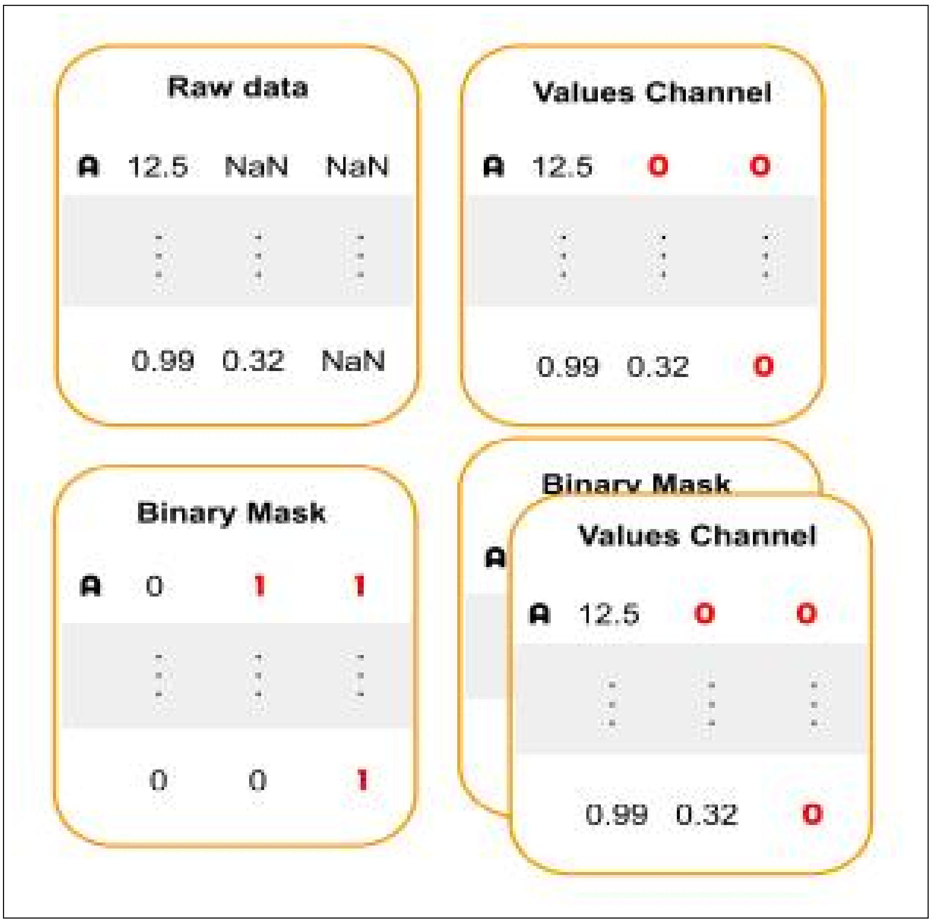

Figure 3.4: Binary Mask of SGCC Dataset

3.1.4 Correlation Matrix

As the nature of the electricity dataset is time series data, due to this only one variable is performed at uniform intervals. To calculate possible correlation and periodicity, two observations are done.

- Correlation matrix is formulated on the electricity usage of 7 days i.e. from Monday to Sunday for honest and fraudulent energy consumers as shown in Figure 3.5
- Autocorrelation function is used to find patterns recognition and periodicity among normal consumer and theft consumer class.

After autocorrelation function analysis indicates that the thieves show a similar pattern because of the higher correlation found among the day of the week for these users which indicates that this feature could be explored more to improve model performances.

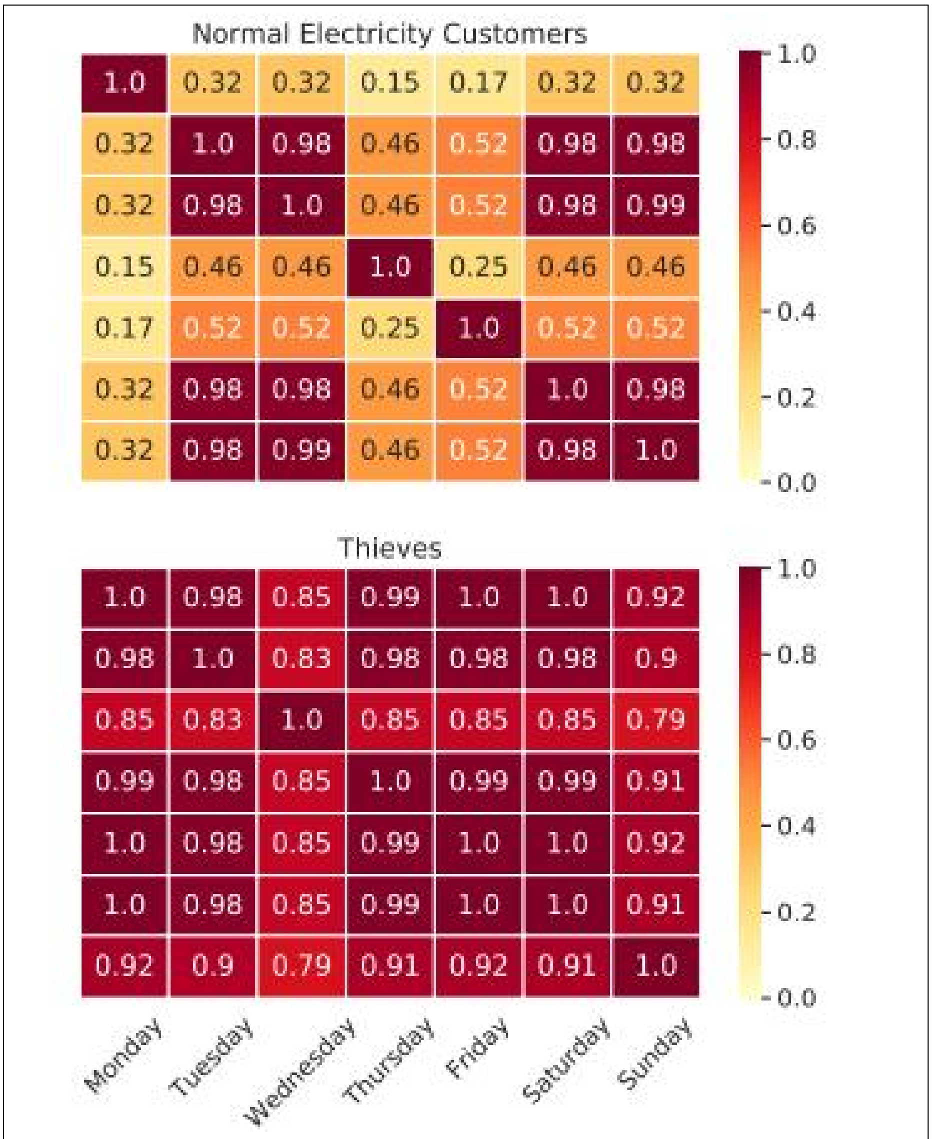

Figure 3.5: Correlation Matrix of SGCC Dataset

3.1.5 SMOTE

The SGCC dataset has class imbalanced problems. The number of theft users is very low than normal users. The ratio is 91% for normal users and 9% for theft users. Such a dataset affects the performance of the classifiers i.e. they may bias toward the majority class.

Class imbalance issues were addressed by different authors who proposed different methods and techniques which can be broadly classified into two categories i.e. the cost function based techniques and the sample based techniques. In this research sampling

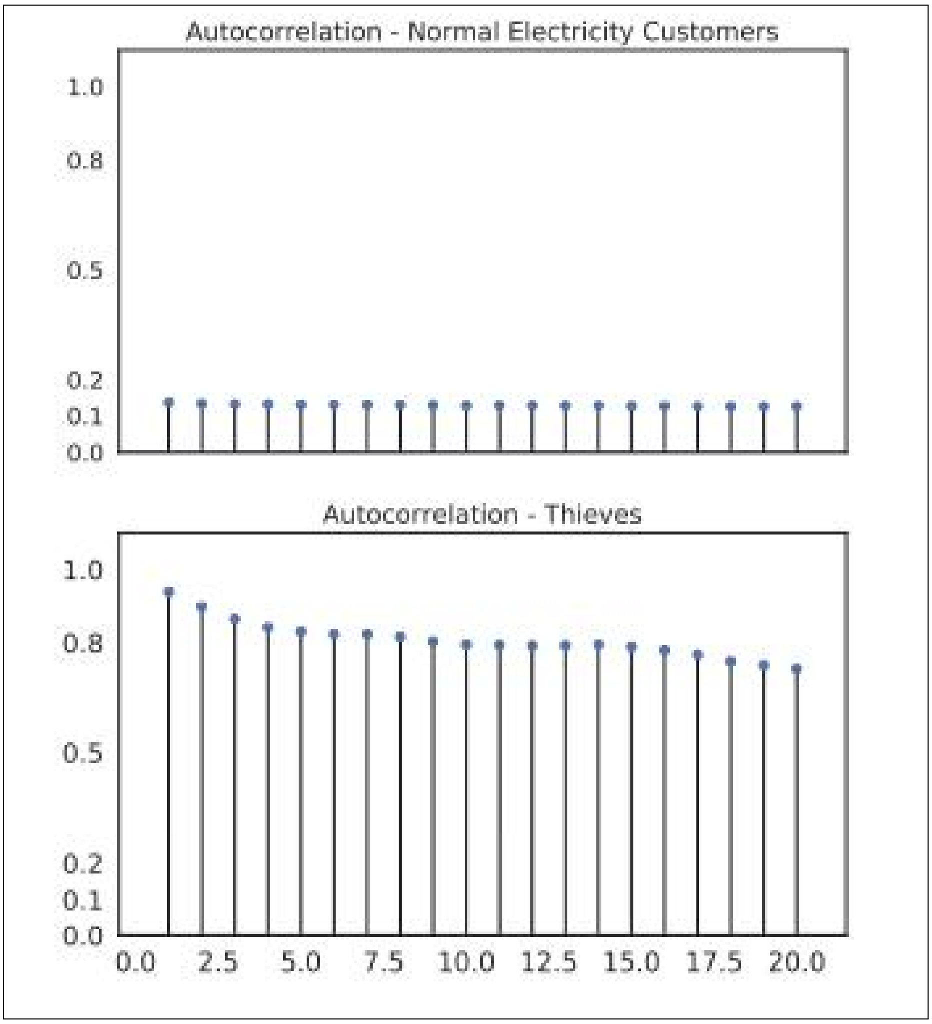

Figure 3.6: Autocorrelation of Normal and Thieve Costumers

based approach is used. Sampling can be done via oversampling and/or under-sampling to mitigate the imbalance among the two classes in the dataset.

Under sampling is the method in which all the instances of minority class are kept while decreasing the size of the majority class which is beneficial in the computation context of the model. But the limitation of this technique is that it may lose potential information, and the model becomes less accurate.

Oversampling replicates the instances of a minority class which increases the size of the class. The advantage is that there is no fear of information loss but the limitation is the overfitting of a classification model.

In this research oversampling approach is used to increase the size of the minority class that is the theft class. To avoid the overfitting problem, the approach of generating synthetic data is utilized. Synthetic Minority Oversampling Technique (SMOTE) is used to produce synthetic instances of the theft class [46]. SMOTE generates synthetic data instances next to the line segment connecting any or all kinds of K nearest neighbors of the theft class in the feature map.

If (X1, X2) is a data point of theft class and its K nearest neighbor is selected as (X'1, X'2)then the synthetic data instance is

$$X_1, X_2 = (X_1, X_2) + random(0,1) \star \triangle \tag{3.1}$$

The Figure 3.7 shows a flow diagram of SMOTE

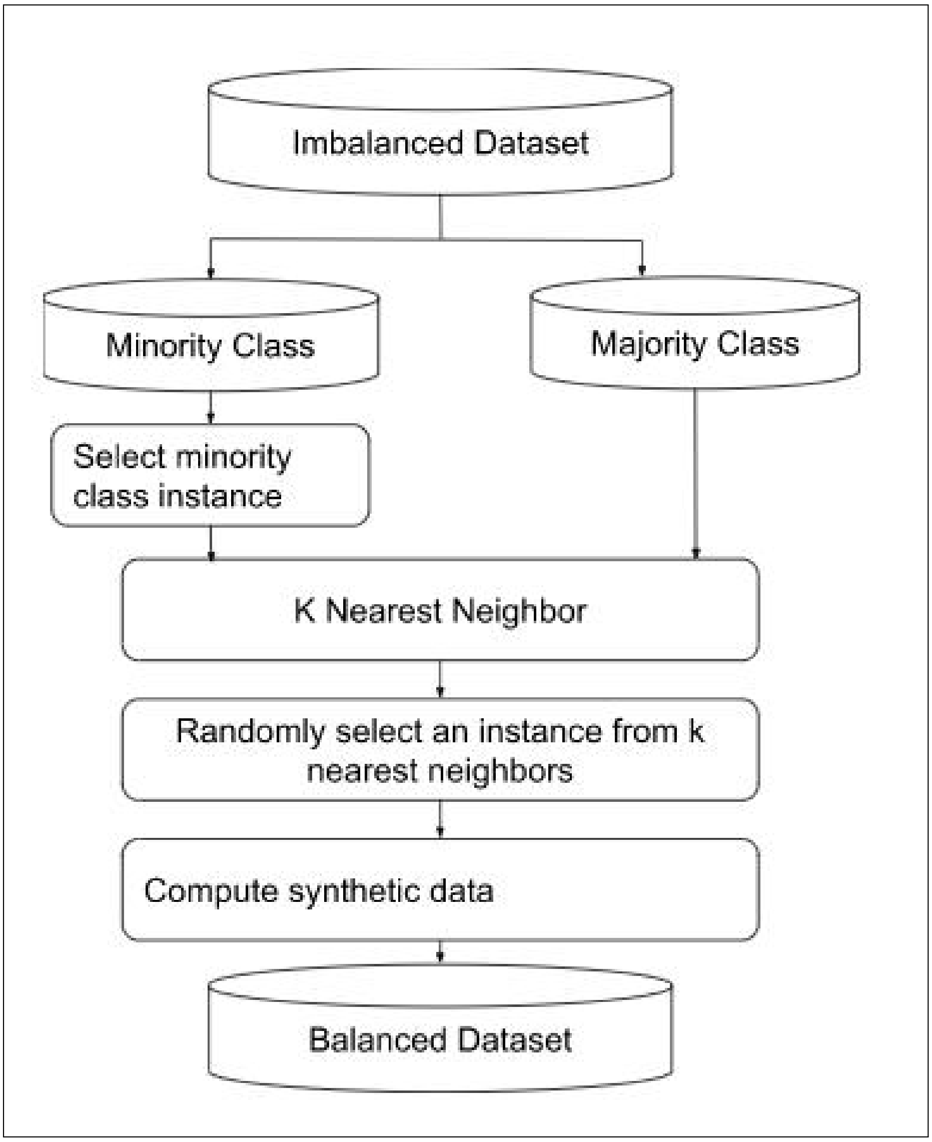

Figure 3.7: Flow Chart of SMOTE

3.2 Proposed Classifier

The proposed classifier in this research is spatial transformer Network and convolution Neural Networks. Both are briefly described below in . The STN is utilized to improve the robustness of CNN classifiers in case of randomness brought by the channel effect. For applying STN its input dimensions are 32 x 32 x 3, while the data array we got preprocessing is 2 x 147 x 7. Therefore for parsing data to STN, the convolution layer is used to change dimensions. Input is reshaped to 34x14x-1, -1 means remaining dimensions, then a convolution layer of kernel 1x36 and padding of 9 is applied. After that input is transferred to STN that rotates, translates, and scales the original input to make it spatially invariant. Affine transformation is applied to input 2D arrays.

In STN two convolution layers each followed by a max pooling layer are applied. Rectified Linear activation function is utilized on each layer. Two fully connected layers are utilized for classification. Softmax activation function is used on fully connected layers.

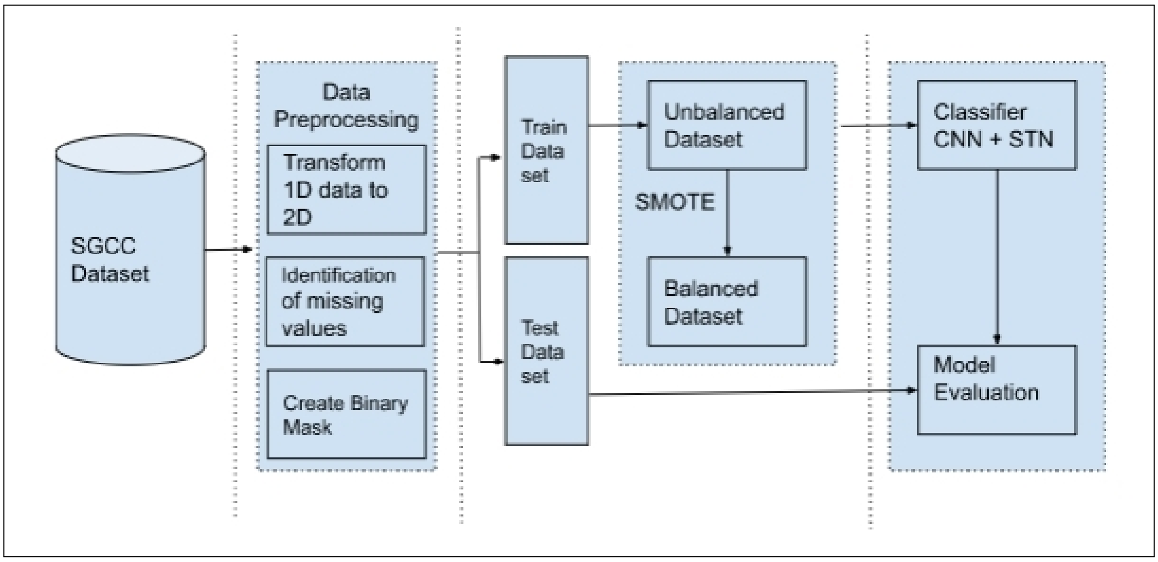

Figure 3.8: Block Diagram of proposed Methodology

3.2.1 Spatial Transformer Networks

Spatial Transformer is a differential module that enables neural networks to investigate how to conduct spatial transformation on the input feature map so that the invariance of the classifier can be enhanced. STN consists of three main modules.

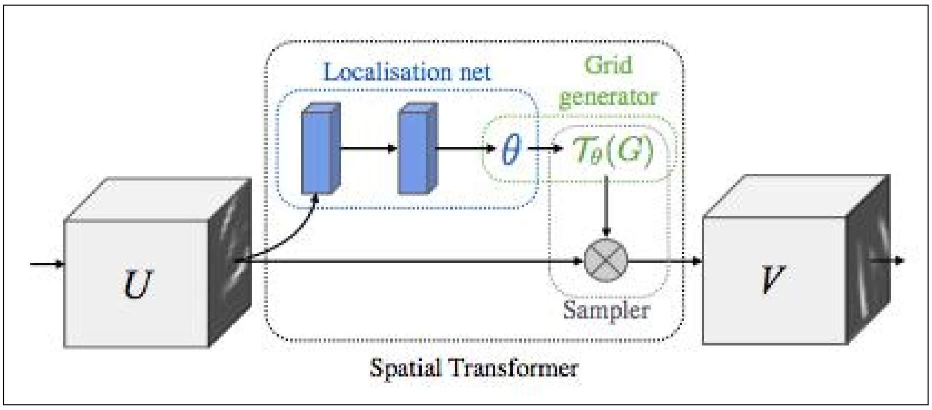

Figure 3.9: Flow Diagram of STN

3.2.2 Localization Network

It is the first part of the Spatial Transformer Network, which takes the input feature map (U) and generates the parameters or features for the spatial transformer

$$\theta = f_{loc}(U) \tag{3.2}$$

The function floc () can be of any shape e.g. it can be a fully connected layer or CNN but should have a regression layer that generates the transformation parameter.

3.3 Grid Generator

Grid generator is responsible to generate an input sampling grid by using the transformation parameters produced in the localization network along with the regular input feature

$$[p, q] = \theta[x, y] \tag{3.3}$$

[x, y] is the coordinate in input feature (U) and [p, q] is the coordinate in the sampling grid.

3.4 Sampler

The third module is the sampler. It utilizes U and the sampling grinds points [p, q] as an input to generate the required input U the sampling grid generated by sampler is not accurately aligned with discrete grid values of U, the bilinear sample is utilized to improve the input.

$$U = S(J, U) \tag{3.4}$$

Where J is the sampling grid and S is the bilinear sampler. The epitome of using a spatial transformer network along with CNN is to improve the performance of the classifier.
The purpose of applying affine transform input is to produce a rotated, translated and scaled version of the original input U, and then optimal input U can be generated for CNN.

3.5 Convolution Neural Network

Convolution Neural Networks is proposed by [49], and it is a subclass of neural networks. CNN framework consists of multiple convolution layers and pooling layers. One or more fully connected layers are also after the convolution and pooling layer is used. The convolution layer is the main component of CNN. It works on two sets of data or inputs i.e. input data points and a convolution filter or kernel.

The activation function is utilized after performing a convolution operation Recti-

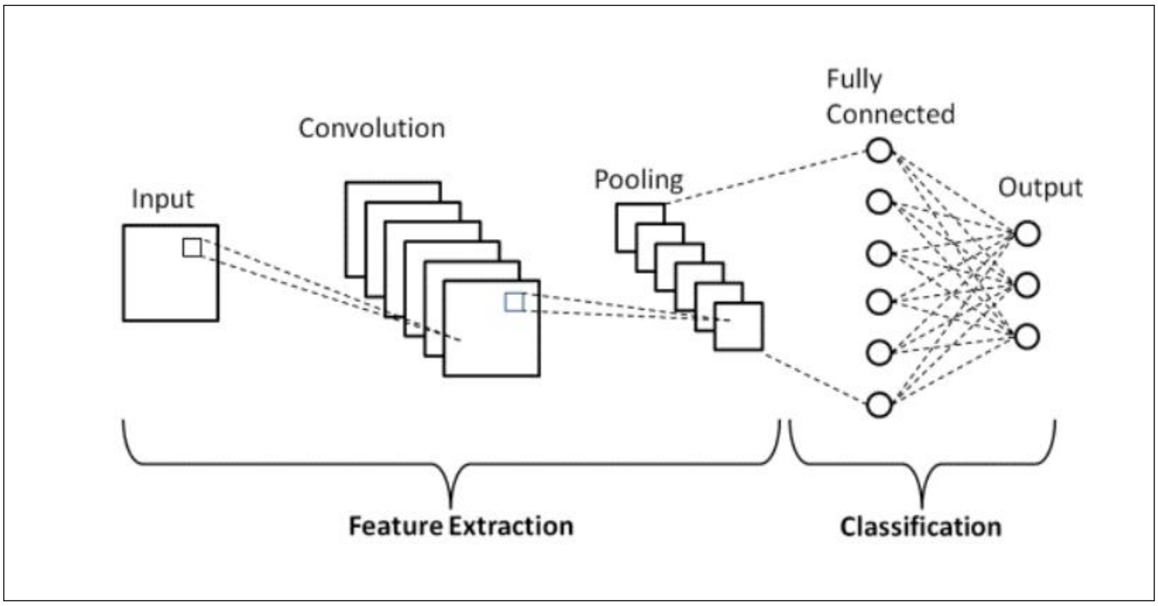

Figure 3.10: Convolution Neural Networks

fied Linear United (ReLU) is used in this analysis. After convolution operation the pooling layer comes. It optimizes every input feature map to decrease the dimensions, which benefits training time and avoids overfitting problems. Max pooling and average pooling layers are mostly used. A fully connected layer is an artificial neural network that conducts the actual classification.

Chapter 4

RESULT AND ANALYSIS

4.1 Evaluation Metrics

The electricity theft detection model is based on time series data and it is a binary classification problem. To evaluate the quality of model parameters the confusion matrix is mostly taken into account. Confusion matrix labels the output of the classifier into four groups: True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN). F1-score, Precision, Accuracy, and Recall are the metrics of the confusion matrix.

4.1.1 F1-Score

F1-Score is calculated by taking the harmonic mean of the classifier Recall and Precision. It is used to assess binary classification problems which classify samples into positive or negative. It is used to estimate the accuracy of the model. The formula for the F1 score is:

$$F1Score = 2\frac{\times(Precision \times Recall)}{(Precision + Recall)} \tag{4.1}$$

4.1.2 Precision

The ratio of correctly predicted theft consumers to the total theft consumers in the dataset specifies the precision of the model. In other words, precision is the performance indicator of the model, that is, how much the model correctly predicts. A high precision value depicts that the model has higher accuracy.

$$Precision = \frac{TruePositive}{(Precision + Recall)} \tag{4.2}$$

4.1.3 Recall

The recall is the ratio of the number of theft consumers correctly detected as theft to the results that should have been returned.

$$Recall = \frac{TruePositives}{(TruePositives + FallNegatives)} \tag{4.3}$$

4.1.4 AUC

Area under Curve (AUC) is used to measure the ability of a model to classify among classes in this research as theft or normal. Higher AUC[50] depicts the better performance of the model.

$$AUC = \frac{\sum_{i} \epsilon positiveclassRank_i - \frac{M(1+M)}{2}}{M \times N} \tag{4.4}$$

Where Ranki is the number of samples i, M is the number of positive samples and N is the number of negative samples.

4.1.5 MAP

MAP stands for Mean Average Precision. It is used to measure the quality of the model's information retrieval[51].

$$MAP@N = \frac{\sum_{i=1}^{r} P@k_i}{r} \tag{4.5}$$

Where r = number of fraudulent consumers in the top N sample.

4.2 Confusion Matrix

Confusion matrix is a measure of the performance of binary class classification problem as well as multiple class classification problem. It depicts the number of predicted and actual class values.It has four values True Positive(TP), False Positive (FP), True Negative (TN) and False Negative (TN). Figure 4.1 shows the proposed model confusion matrix.

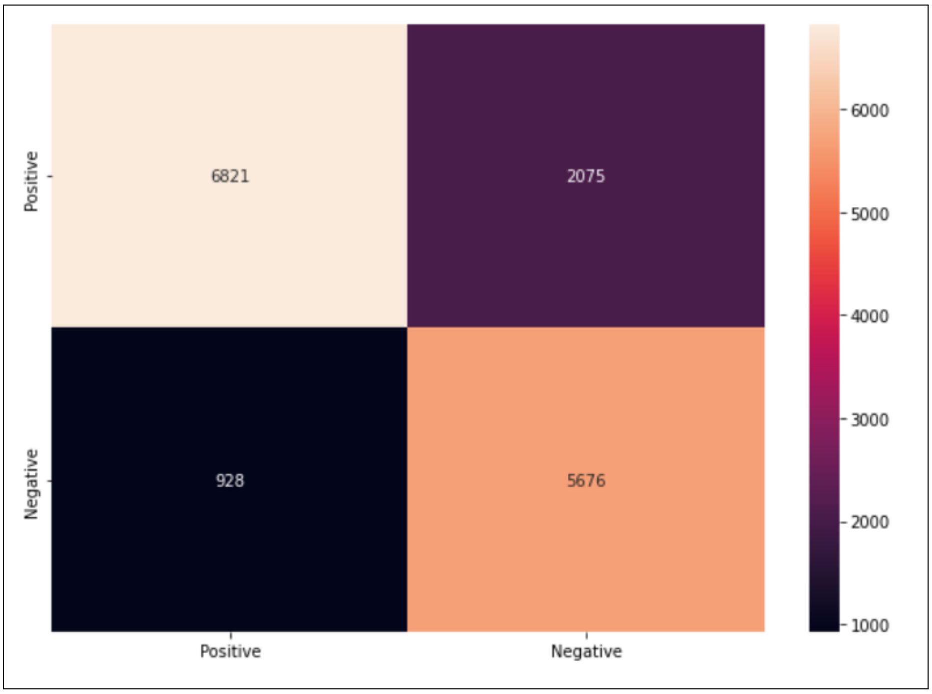

Figure 4.1: Confusion Matrix

4.3 Parametric Results of proposed model

Table 4.1 shows the F1score, Precision, Recall, value of MAP@100, value of MAP@200, and value AUC of proposed model.The proposed model shows good results in terms of F1 score, Recall, AUC and MAP, while shows moderate result in precision.

4.4 Comparison

Different traditional machine learning and deep learning models are used to compare with the proposed model in this research. This section will present the results of chosen techniques for comparison over the given dataset. Performance metrics used for comparison of techniques are f1 score, AUC, MAP@100, and MAP@200. Brief description of the techniques chosen for comparison is as follows:

Technique	CNN+STN (Proposed model)
F1 score	0.87
Precision	0.77
Recall	0.88
AUC	0.87
MAP@100	0.97
MAP@200	0.98

Table 4.1: Results

4.4.1 Logistic Regression

Logistic Regression (LR) is a basic model and is used mostly in binary classification problems. It consists of one neural network layer. It utilizes the Sigmoid activation function.

4.4.2 Random Forest

Random Forest (FR) is important and essential integration of multiple decision trees. The decision tree was used in [52] and comparing the performance of single and multiple decision trees, RF presented better performance.

4.4.3 Support Vector Machine

Support Vector Machine (SVM) is utilized in many studies to detect the presence of electricity theft. It is a supervised learning technique used for many classification and regression problems[53].

4.4.4 Convolution Neural Networks

Convolution Neural Networks are a subclass of neural networks and depict remarkable performance in classification problems, and can be integrated with other models to enhance the performance.

4.4.5 Wide and Deep CNN

Wide and Deep CNN model is proposed [20] to detect theft detection using 2D data for Deep CNN and 1d data in Wide component

4.4.6 Hybrid Multi. Attention Dilated Convolution Network

It implements self-attention model [39] for theft detection by using converted data from 1d to 2D to apply 2D CNN

4.4.7 Main Results

Table 4.2 shows the comparison of the models, same dataset i.e. SGCC dataset is used to compare the models. LR performs least in terms of AUC and the proposed model outperforms other techniques in terms of AUC. CCN+STN model shows highest f1 score among all the techniques chosen for comparison

Technique	AUC	F1 score	MAP@100	MAP@200
LR	0.68	0.69	0.64	0.57
RF [52]	0.73	0.72	0.62	0.60
SVM [53]	0.71	0.78	0.69	0.60
CNN	0.86	0.53	0.90	0.88
Wide and Deep [20]	0.7	0.56	0.94	0.89
Self Attention [39]	0.92	0.60	0.92	0.97
CNN+STN (proposed)	**0.92**	**0.87**	**0.97**	**0.98**

Table 4.2: Comparison with existing techniques

4.4.8 F1 Score

Figure 4.2 shows the comparison of F1 Score of the techniques.

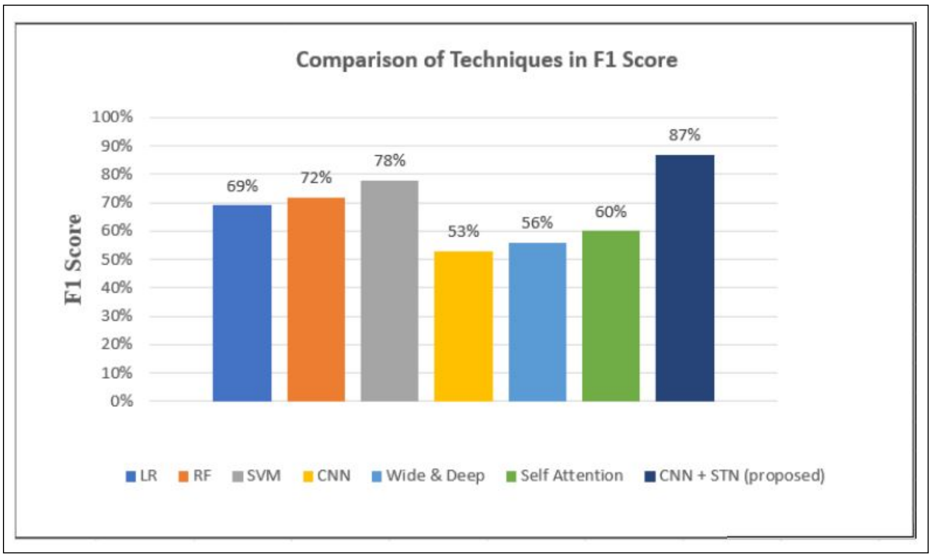

Figure 4.2: Comparison of F1 Score of the techniques

4.4.9 MAP@100

Figure 4.3 shows the comparison of MAP@100 value of the techniques.

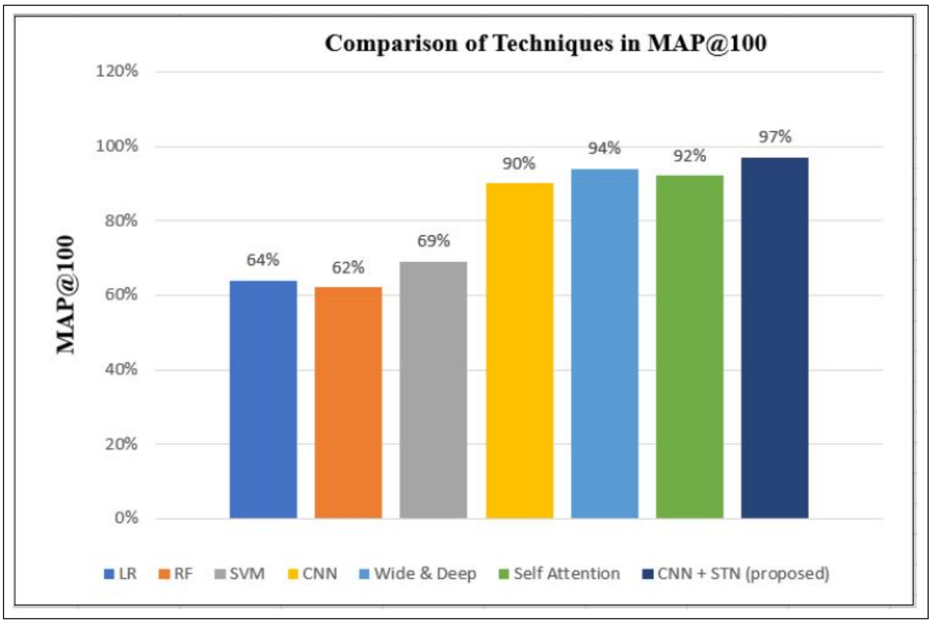

Figure 4.3: Comparison of MAP@100 value of the techniques

4.4.10 MAP@200

Figure 4.4 shows the comparison of MAP@200 value of the techniques.

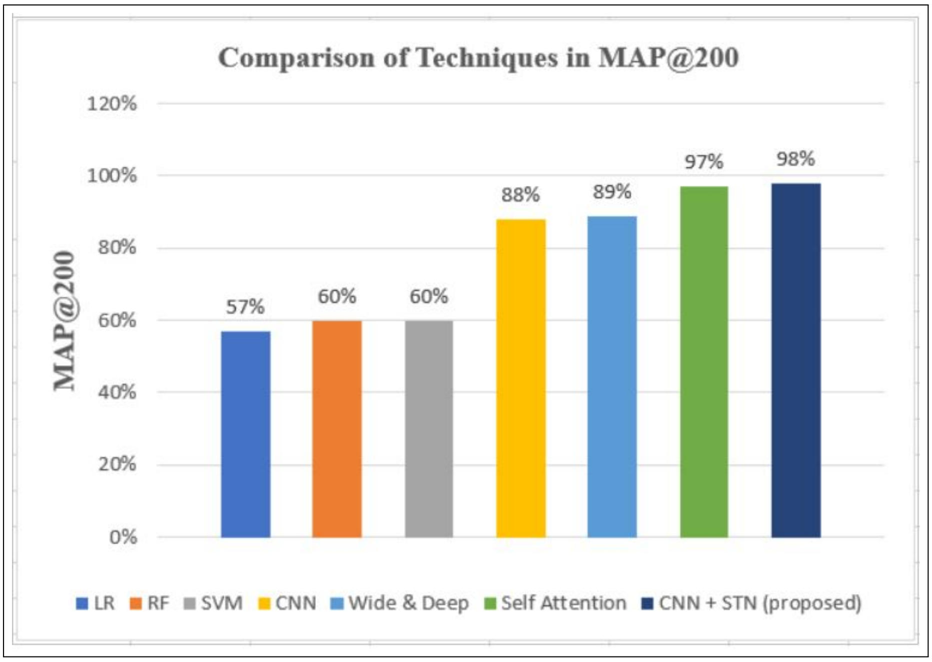

Figure 4.4: Comparison of MAP@200 value of the techniques

Chapter 5

CONCLUSION AND FUTURE WORK

In this research the adverse effects of the electricity theft detection are studied. Machine learning and deep learning models and techniques are analyzed. SGCC dataset is used to detect theft consumers i.e. a real electricity usage values of the consumers. Missing values are deal with binary mask and for imbalance dataset SMOTE is used which is effective in terms of performance. Inspired by the effectiveness of the STN in image and modulation classification, a combination of CNN +STN model is proposed to detect electricity theft in the SGCC dataset. STN consists of three components localization, grid generator and sampler. CNN automates the features extraction . The proposed model is compared with traditional machine learning and deep learning models and techniques. Parameters used for comparison of the models are f1 score, AUC, MAP@100 and MAP@200. The proposed model performs better than other models. The proposed model is also evaluated in terms of recall and precision, it shows highest F1 score i.e. 87% and moderate precision. The proposed will test on other datasets that may contain region information, temperature and weather insights of that specific regions in future. Model still have area of improvement in terms of precision.

BIBLIOGRAPHY

[1] Depuru, S.S.S.R.; Wang, L.; Devabhaktuni, V.; Nelapati, P. A, "A hybrid neural network model and encoding technique for enhanced classification of energy consumption data.", *In Proceedings of the 2011 IEEE Power and Energy Society General Meeting, Detroit, MI, USA*, pp. 1–8, 2011.

[2] Parmar, J., "Total Losses in Power Distribution and Transmission Lines. Electrical Engineering.", *Journal of biomedical informatics*

[3] Dos Angelos, E.W.S.; Saavedra, O.R.; Cortés, O.A.C.; De Souza, A.N., " Detection and identification of abnormalities in customer consumptions in power distribution systems. ", *IEEE Trans. Power Deliv.*, PP. 2436–2442, 2015.

[4] Nagi, J.; Mohammad, A.M.; Yap, K.S.; Tiong, S.K.; Ahmed, S.K.; Mohamad, M.; Mohammad, A.M.; Yap, K.S.; Tiong, S.K.; Ahmed, S.K, "Nontechnical Loss Detection for Metered Consumers in Power Utility Using Support Vector Machines. ", *IEEE Trans. Power Deliv.*, pp. 1162–1171, 2010.

[5] Worldwide Annual NTLs Revenue Loss Report. Accessed: Jan. 6, 2021. [Online]. Available: https://www.prnewswire.com/news-releases/96-billion-is-lost-every-year-to-electricity-theft-300453411.htm

[6] Patrick Glauner, Jorge Augusto Meira, Petko Valtchev, Radu State, and Franck Bettinger, "The challenge of non-technical loss detection using artificial intelligence: A survey. ", *International Journal of Computational Intelligence Systems*, 10(1):760, 2017communities. Concurrency and Computation: Practice and Experience, 33(17), e6316..

[7] Yu, X., Cecati, C., Dillon, T., and Simoes, M. G. "The new frontier of smart grids.”, *IEEE Industrial Electronics Magazine* 5(3), PP. 49-63., 2011.

[8] Mavridou, A., and Papa, M. , "A situational awareness architecture for the smart grid.", *In Global Security, Safety and Sustainability and e-Democracy*,pp. 229-236, 2011.

[9] G. Hinton, L. Deng, D. Yu, G. E. Dahl, A.-r. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. N. Sainath et al. "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups", *IEEE Signal Process. Mag.*, 29(6), pp. 82–97, 2012.

[10] T. Mikolov, M. Karafiat, L. Burget, J. Cernock ´ y, and S. Khudanpur, ."Recurrent neural network based language model",*in Interspeech*, vol. 2, pp. 3, 2010,

[11] R. Gravet, G. Cabrera-Vives, P. G. Perez-Gonz ´ alez, J. Kartaltepe, ´ G. Barro, M. Bernardi, S. Mei, F. Shankar, P. Dimauro, E. Bell et al., "A catalog of visual-like morphologies in the 5 candels fields using deep learning", *The Astrophysical J. Supplement Series*, vol. 221, no. 1, p. 8, 2015

[12] J. Zhou and O. G. Troyanskaya, " Predicting effects of noncoding variants with deep learning-based sequence model ", *Nature Methods vol. 12*, no. 10, pp. 931–934, 2015

[13] Viegas, J.L.; Esteves, P.R.; Melício, R.; Mendes, V.M.F.; Vieira, S.M "A Solutions for detection of non-technical losses in the electricity grid: A review. ", *Renew. Sustain. Energy Rev.* , 2017, 80, 1256–1268.

[14] Xia, X.; Xiao, Y.; Liang, W. SAI,"A Suspicion Assessment-Based Inspection Algorithm to Detect Malicious Users in Smart Grid ", *IEEE Trans. Inf. Forensics Secure*, 15, 361–374 2020.

[15] Mimmi, L.M.; Ecer, S , "An econometric study of illegal electricity connections in the urban favelas of Belo Horizonte, Brazil. ", *Energy Policy.*38, 5081–5097. 2010.

[16] Jamil, F. , "On the electricity shortage, price and electricity theft nexus.", *Energy Policy*,54, 267–272, 2013 .

[17] Lydia, M.; Kumar, G.E.P.; Levron, Y. ,"Detection of Electricity Theft based on Compressed Sensing.", Proceedings of the 5th International Conference on Advanced Computing Communication Systems ICACCS 2019, Coimbatore, India pp. 995–1000.,15–16 March 2019.

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

[19] Otuoze, A.O.; Mustafa, M.W.; Mohammed, O.O.; Saeed, M.S.; Surajudeen-Bakinde, N.T.; Salisu, S., "Electricity theft detection by sources of threats for smart city planning. ", *IET Smart Cities*, 1, 52–60, 2019.

[20] Zibin Zheng, Yatao Yang, Xiangdong Niu, Hong-Ning Dai, and Yuren Zhou. , "Wide deep convolutional neural networks for electricity-theft detection to secure smart grids.", *IEEE Transactions on Industrial Informatics,*,14:1606–1615, 04 2018.

[21] Hussain, S., Mustafa, M. W., Al-Shqeerat, K., Saeed, F., Al-Rimy, B., "A Novel Feature-Engineered-NGBoost Machine-Learning Framework for Fraud Detection in Electric Power Consumption Data.", *Sensors (Basel, Switzerland),https://doi.org/10.3390/s21248423*,21(24), 8423. , 2021.

[22] N. F. Avila, G. Figueroa and C. -C. Chu, "NTL Detection in Electric Distribution Systems Using the Maximal Overlap Discrete Wavelet-Packet Transform and Random Undersampling Boosting", *in IEEE Transactions on Power Systems*, vol. 33, no. 6, pp. 7171-7180, Nov. 2018.

[23] Han, W., Xiao, Y., " Edge computing enabled non-technical loss fraud detection for big data security analytic in Smart Grid.", *J Ambient Intell Human Comput*, 11, 1697–1708 , 2020.

[24] Vahabzadeh, Alireza, et al., "A fuzzy-SOM method for fraud detection in power distribution networks with high penetration of roof-top grid-connected PV" *Energies 13.5*,1287, 2020

[25] Muniz, C., Figueiredo, K., Vellasco, M., Chavez, G., Pacheco, M., "Irregularity detection on low tension electric installations by neural network ensembles.", *In IEEE 2009 International Joint Conference on Neural Networks*,(pp. 2176-2182), 2009.

[26] Rei, Marek, Gamal KO Crichton, and Sampo Pyysalo., "Attending to characters in neural sequence labeling models.", *J Ambient Intell Human Comput*,, 2016

[27] Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin., "Attention is all you need.", *Advances in neural information processing systems*, 2017.

[28] Xiao, J. and Zhou, Z. , "Research progress of RNN language model", *In 2020 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA)*, pp. 1285-1288, 2020.

[29] Hori, T., Cho, J., and Watanabe, S.,"End-to-end speech recognition with word-based RNN language models.", *In 2018 IEEE Spoken Language Technology Workshop (SLT)*, pp. 389-396, 2018.

[30] Kim, S., Shangguan, Y., Mahadeokar, J., Bruguier, A., Fuegen, C., Seltzer, M. L., and Le, D., "Improved neural language model fusion for streaming recurrent neural network transducer.", *In ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 7333-7337, 2021.

[31] Baliyan, A., Batra, A., and Singh, S. P., "Multilingual sentiment analysis using RNN-LSTM and neural machine translation", *In IEEE 2021 8th International Conference on Computing for Sustainable Global Development (INDIACom)*, pp.710-713, 2021.

[32] S. Hochreiter and J. Schmidhuber, "Long short-term memory", *Neural computation*, vol. 9(8) pp. 1735-1780, 1997.

[33] L. Zhang, M. Yu, T. Chen, Z. Shi, C. Bao and K. Ma, "Auxiliary Training: Towards Accurate and Robust Models", *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 369-378, 2020.

[34] C.C. Yang, H. Yang, L. Jiang, , "Postmarketing drug safety surveillance using publicly available health-consumer-contributed content in social media", *ACM Trans. Manag. Inf. Syst.* pp. 1-21, 2014.

[35] S. Ramamoorthy and S. Murugan., "An attentive sequence model for adverse drug event extraction from biomedical text.", 2018.

[36] Zhang, T., Lin, H., Ren, Y., Yang, L., Xu, B., Yang, Z., and Zhang, Y."Adverse drug reaction detection via a multihop self-attention mechanism.", *BMC bioinformatics*,20(1), pp. 1-11, 2019.

[37] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfifitting.", *The journal of machine learning research*, 15(1), pp. 1929-1958, 2014.

[38] Gurulingappa, "Development of a benchmark corpus to support the automatic extraction of drug-related adverse effects from medical case reports", *Journal of biomedical informatics*, pp. 885-892, 2012.

[39] S. Gupta, S. Pawar, N. Ramrakhiyani, G. K. Palshikar, and V. Varma, "Semi-supervised recurrent neural network for adverse drug reaction mention extraction", *BMC bioinformatics*, 19(8) pp.1-7, 2018.

[40] Sahu, Sunil Kumar, and Ashish Anand, "Drug-drug interaction extraction from biomedical texts using long short-term memory network", *Journal of biomedical informatics*, pp. 15-24, 2018.

[41] El-allaly, Ed-drissiya, Mourad Sarrouti, Noureddine En-Nahnahi, and Said Ouatik El Alaoui, "MTTLADE: A multi-task transfer learning-based method for adverse drug events extraction.", *Information Processing and Management*, 58(3), pp. 1-7, 2021.

[42] Mirmohammadsadeghi, M., Hanna, S. S., Cabric, D., "Modulation classification using convolutional neural networks and spatial transformer networks.", *In 2017 51st Asilomar Conference on Signals, Systems, and Computers*,pp. 936-939, 2017.

[43] Trevizan, R. D., "Detection and identification of commercial losses in distribution systems: methodology based on optimum-path forest.", *M. Sc. dissertation, Dept. Electr. Eng., Federal Univ. of Rio Grande do Sul.*, 2014.

[44] Mohd Mustafa Al Bakri Abdullah., "Fitting distribution.", *Key Engineering Materials*, 594-595:889–895, 2014.

[45] Collins Leke, Bhekisipho Twala, and Tshilidzi Marwala., "Missing data prediction and classification: The use of auto-associative neural networks and optimization algorithms.", *CoRR*, abs/1403.5488, 2014.

[46] Yang, R., Zhang, C., Gao, R., Zhang, L., "A novel feature extraction method with feature selection to identify Golgi-resident protein types from imbalanced data.", *International journal of molecular sciences*, 17(2), 218.,2016.

[47] Liao, C.; Ten, C.-W.; Hu, S. , "Strategic FRTU deployment considering cybersecurity in secondary distribution network.", *IEEE Trans. Smart Grid*, 1264–1274., 2013.

[48] R. M. Mohammad, F. Thabtah, and L. McCluskey, "An assessment of features related to phishing websites using an automated technique.", *in Proceedings of the 2012 International Conference for Internet Technology and Secured Transactions*, 2012.

[49] LeCun, Y., Bengio, Y., Hinton, G., "Deep learning. nature", *Advances in neural information processing systems*, 521(7553), 436-444.,2015.

[50] J. Davis and G. Mark, "“The relationship between precision-recall and roc curves.", Proceedings of The 23rd International Conference on Machine Learning. ACM, pp. 233–240., 2006.

[51] M. Turpin and F. Scholer, "User performance versus precision measures for simple search tasks", *In Proceedings of the 29th annual international ACM SIGIR Conference on Research and Development in Information Retrieval. ACM*, pp. 11–18, 2006.

[52] M. Biswal and P. K. Dash, "SMeasurement and classification of simultaneous power signal patterns with an s-transform variant and fuzzy decision tree", *IEEE Transactions on Industrial Informatics*, vol. 9, no. 4, pp. 1819–1827, 2013.

[53] J. Nagi, K. S. Yap, S. K. Tiong, S. K. Ahmed, and M. Mohamad, "Nontechnical loss detection for metered customers in power utility using support vector machines", *IEEE Transactions on Power Delivery*,vol. 25, no. 2, pp. 1162–1171, 2010.

[54] Visa, S., Ramsay, B., Ralescu, A. L., Van Der Knaap, E.,"Confusion matrix-based feature selection.", *MAICS.*, 710(1), 120-127, 2011.

[55] Xia, L.,"Historical profile will tell? A deep learning-based multi-level embedding framework for adverse drug event detection and extraction.", *Decision Support Systems.*, 160, 113832., 2022.

[56] Joos Korstanje.,"https://towardsdatascience.com/the-f1-score-bec2bbc38aa6", ,2021.

[57] Saeed MS, Mustafa MW, Hamadneh NN, Alshammari NA, Sheikh UU, Jumani TA, Khalid SB, Khan I.,"Detection of non-technical losses in power utilities—A comprehensive systematic review.", *Energies.*,13(18):4727.,2020.

[58] Ximing Zhang, Mingang Wang, Jinkang Wei, Can Cui.,"Robust Visual Tracking based on Deep Spatial Transformer Features.", *Chinese Control And Decision Conference (CCDC).* 2018.