**A Project report on**

#### Electricity Theft Detection in Power Grids with Deep Learning and Random Forests

###### A Dissertation submitted to JNTU Hyderabad in partial fulfillment of the academic requirements for the award of the degree.

**Bachelor of Technology**

### in

**Computer Science and Engineering**

Submitted by

KOKKONDA RAJANI 19H51A05G7 MADALA NAVYA 19H51A05H2 ANUGAM YUVARAJU 19H51A05J5

Under the esteemed guidance of Dr G Ravi Kumar

Associate Professor



**Department of Computer Science and Engineering**

#### CMR COLLEGE OF ENGINEERING & TECHNOLOGY

(An Autonomous Institution under UGC & JNTUH, Approved by AICTE, Permanently Affiliated to JNTUH, Accredited by NBA.)

KANDLAKOYA, MEDCHAL ROAD, HYDERABAD - 501401.

2019- 2023

**CMR COLLEGE OF ENGINEERING & TECHNOLOGY**

KANDLAKOYA, MEDCHAL ROAD, HYDERABAD – 501401

**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**



### CERTIFICATE

This is to certify that the Major Project Phase-1 report entitled **" Electricity Theft Detection in Power Grids with Deep Learning and Random Forests "** being submitted by Kokkonda Rajani (19H51A05G7), Madala Navya (19H51A05H2), Anugam Yuvaraju (19H51A05J5) in partial fulfillment for the award of **Bachelor of Technology in Computer Science and Engineering** is a record of bonafide work carried out his/her under my guidance and supervision.

The results embody in this project report have not been submitted to any other University or Institute for the award of any Degree.

**Dr. G. Ravi Kumar Dr. Siva Skanda Sanagala**

**Associate Professor Associate Professor and HOD**

**Dept. of CSE Dept. of CSE**

**ACKNOWLEDGEMENT**

With great pleasure we want to take this opportunity to express my heartfelt gratitude to all the people who helped in making this project work a grand success.

We are grateful to **Dr. G. Ravi Kumar, Associate Professor**, Department of Computer Science and Engineering for his valuable technical suggestions and guidance during the execution of this project work.

We would like to thank **Dr. Siva Skandha Sanagala,** Head of the Department of Computer Science and Engineering, CMR College of Engineering and Technology, who is the major driving forces to complete my project work successfully.

We are very grateful to **Dr. Vijaya Kumar Koppula**, Dean-Academic, CMR College of Engineering and Technology, for his constant support and motivation in carrying out the project work successfully.

We are highly indebted to **Dr. V A Narayana,** Principal, CMR College of Engineering and Technology, for giving permission to carry out this project in a successful and fruitful way.

We would like to thank the Teaching & Non- teaching staff of Department of Computer Science and Engineering for their co-operation

We express our sincere thanks to **Mr. Ch. Gopal Reddy**, Secretary, CMR Group of Institutions, for his continuous care.

Finally, We extend thanks to our parents who stood behind us at different stages of this Project. We sincerely acknowledge and thank all those who gave support directly and indirectly in completion of this project work.

K. Rajani 19H51A05G7

M. Navya 19H51A05H2

A. YuvaRaju 19H51A05J5

Electricity Theft Detection in Power Grids with Deep Learning and Random Forests

**TABLE OF CONTENTS**

**CHAPTER**

**NO. TITLE PAGE NO.**

[LIST OF FIGURES ii](#_TOC_250015)

[LIST OF TABLES iii](#_TOC_250014)

[ABSTRACT iv](#_TOC_250013)

1. INTRODUCTION 1
   1. Problem Statement 2
   2. [Research Objective 2](#_TOC_250012)
   3. [Project Scope and Limitations 3](#_TOC_250011)
2. BACKGROUND WORK 6
   1. [Machine Learning Algorithm for Efficient Power Theft Detection using Smart Meter Data](#_TOC_250010)
      1. [Introduction 4](#_TOC_250009)
      2. Merits, Demerits and Challenges 6
      3. [Implementation 7](#_TOC_250008)
   2. Detecting and Locating Non-Technical Lossesin Modern
      1. [Introduction 13](#_TOC_250007)
      2. Merits, Demerits and Challenges 15
      3. [Implementation 16](#_TOC_250006)
   3. Using GANCNN and ERNET for Detection of Non Technical Losses
      1. [Introduction 19](#_TOC_250005)
      2. [Merits, Demerits and Challenges 21](#_TOC_250004)
      3. [Implementation 22](#_TOC_250003)
3. RESULTS AND DISCUSSION
   1. Comparison of Existing Solutions 25
   2. [Data Collection and Performance metrics 27](#_TOC_250002)
4. [CONCLUSION](#_TOC_250001)

6.1 Conclusion 29

1. [REFERENCES 30](#_TOC_250000)

**FIGURE**

##### List of Figures

**NO. TITLE PAGE NO.**

* 1. Electricity Theft Identification Model 5
  2. Sample customer data with three types of bogus data 8
  3. Architecture for a smart grid 14
  4. Process diagram 14
  5. Flow chart 16
  6. Graph I vs V 17
  7. UCL graph 18
  8. Algorithm graph 18
  9. Flow chart of data cleaning process 22
  10. Flowchart of ERNET model. 23
  11. Accuracy and loss of the GANCNN and ERNET models 24
  12. AUC of proposed GANCNN and ERNET models 24

3.1 Accuracy 26

##### List of Tables

**TABLE**

**NO. TITLE PAGE NO.**

* 1. Trust worthiness of customers . 7
  2. Voltage magnitudes 17
  3. Network structure 23
  4. Dataset table 27
  5. Accuracy table 28
  6. Success rate table 28
  7. Efficiency table 28

##### ABSTRACT

As one of the major factors of the nontechnical losses (NTLs) in distribution networks, the electricity theft causes significant harm to power grids, which influences power supply quality and reduces operating profits. In order to help utility companies solve the problems of inefficient electricity Inspection and irregular power consumption, a novel hybrid convolutional neural network-random forest (CNN-RF) model for automatic electricity theft detection is presented in this project.

In this model, a convolutional neural network (CNN) firstly is designed to learn the features between different hours of the day and different days from massive and varying smart meter data by the operations of convolution and down sampling. In addition, a dropout layer is added to retard the risk of overfitting, and the back propagation algorithm is applied to update network parameters in the training phase. And then, the random forest (RF) is trained based on the obtained features to detect whether the consumer steals electricity.

# CHAPTER 1

## INTRODUCTION

#### CHAPTER 1 INTRODUCTION

The loss of energy in electricity transmission and distribution is an important problem faced by power companies all over the world. The energy losses are usually classified into technical losses (TLs) and nontechnical losses (NTLs). The TL is inherent to the transportation of electricity, which is caused by internal actions in the power system components such as the transmission liner and transformers the NTL is defined as the difference between total losses and TLs, which is primarily caused by electricity theft. Actually, the electricity theft occurs mostly through physical attacks like line tapping, meter breaking, or meter reading tampering. These electricity fraud behaviours may bring about the revenue loss of power companies. As an example, the losses caused by electricity theft are estimated as about $4.5 billion every year in the United States (US). And it is estimated that utility companies worldwide lose more than 20 billion every year in the form of electricity theft. In addition, electricity theft behaviours can also affect the power system safety. For instance, the heavy load of electrical systems caused by electricity theft may lead to fires, which threaten the public safety. Therefore, accurate electricity theft detection is crucial for power grid safety and stableness

With the implementation of the advanced metering infrastructure (AMI) in smart grids, power utilities obtained massive amounts of electricity consumption data at a high frequency from smart meters, which is helpful for us to detect electricity theft However, every coin has two sides; the AMI network opens the door for some new electricitytheft attacks. These attacks in the AMI can be launched by various means such as digital tools and cyber attacks. The primary means of electricity theft detection include humanly examining unauthorized line diversions, comparing malicious meter records with the benign ones, and checking problematic equipment or hardware. However, these methods are extremely time-consuming and costly during full verification of all meters in a system.Besides, these manual approaches cannot avoid cyber attacks. In order to solve the problems mentioned above, many approaches have been put forward in the past years. These methods are mainly categorized into state- based, game-theory-based, and artificial- intelligence-based models.

#### Problem Definition

* The detection of electricity theft behaviors is a binary classiﬁcation problem which calls for distinguishing of normal and electricity theft users.
* Electricity theft detection in power grids using deep learning algorithms such as the combination of CNN-RF algorithms.
* In order to help utility companies solve the problems of inefficient electricity inspection and irregular power consumption, a novel hybrid convolutional neural network-random forest (CNN- RF) model for automatic electricity theft detection is presented in this project.

#### Research Objective

* The main objective of this project is to detect the non technical losses in the power grids.
* Electricity theft has become a big problem as the losses incurred in the theft is high. As an example, the losses caused by electricity theft are estimated as about $4.5 billion every year in the United States (US) . And it is estimated that utility companies worldwide lose more than 20 billion every year in the form of electricity theft.
* The research on this project has given an idea that the losses incurred can be found using the combination of convolutional neural networks and random forest algorithms.
* Though there were other algorithms the CNN-RF has been efficient and gives better prediction accuracy.
* In power consumption if there is huge consumption in certain period then in dataset we will get value as 1 which indicates energy theft else we will have 0 as class label which means normal usage.

#### Project Scope and Limitations

##### Project scope:

* In this project it is only concerned to the detection of electricity theft but it can be implemented in advance by raising an alarm to the officials.
* And the consumer data can be given with the data base of the records of on-duty officers.
* As this project works with 100% accuracy it can be further implemented to make a device.

##### Limitations:

* There are other implementations of this project which are devices that are easy to use such as iot devices.
* To monitor the electricity theft one should have the knowledge on machine learning and the algorithms unlike other devices.

# CHAPTER 2

## BACKGROUND WORK

#### CHAPTER 2 BACKGROUND WORK

##### Machine Learning Algorithm for Efficient Power Theft Detection using Smart Meter Data

##### Introduction

The electric grid refers to a network of transmission lines, substations, transformers and more that deliver electricity from the power plant to our home or business. Digital technology that allows for two-way communication between the utility and its customers, and the sensing along the transmission lines is what makes the electric grid smart. The smart grid components include Automated Metering Infrastructure (AMI), Phasor Measurement Unit and Communication network. The AMI describe the whole infrastructure from smart meter to two-way communication network to control center equipment and all the applications that enable the gathering and transfer of energy usage information in near real time. The components of AMI include: smart meter, communication network, meter data acquisition system, meter data management system. The AMI improvise the following features: system reliability, energy cost, and electricity theft. The functionality includes service switching, time-based rates, remote programming to control smart devices, power quality measure, and a user interface for real-time monitoring. It is an automated device having the features to collect the consumption data usually in hour basis (may vary).

The Artificial neural network is built to classify the Non-Technical Loss power tampering for intelligently identifying the losses by selecting the most required features from the customer profile. The extreme Learning Machine classification technique elucidates the operation of identifying the customer energy consumption pattern that classifies genuine and illegal profiled customers. The classification models are applied on regular energy consumption data as well as the encoded data to compare corresponding classification accuracies and computational overhead. The previous research works on power theft detection focuses on the customer power usage profile data for theft detection. The specific are where there is dissimilarity in supplied power and billed power. All the customers belonging to that area are considered to be suspects. The drawback of the work discussed is that power theft identification has been carried out based on the assumption that the customers are suspected to be fraud.

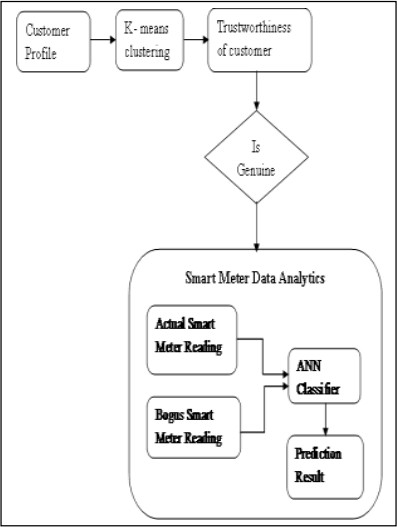
This case could identify the potential customer as the fraudulent customers. This provides the motivation for this research work to include the bogus customer power consumption data into the actual power consumption data. The machine learning algorithms are used to analyze the data that literally cluster and then classify the Customer. The customer’s data are discriminated as genuine and fraud based on their usage pattern.

Fig. 2.1:Electricity Theft Identification Model

##### MERITS, DEMERITS AND CHALLENGES MERITS

* + - * The identification of power theft will also extend its support for load forecasting that permits the utilities to exactly predict the power demand for future specific to individual customer.
      * The information produced through this analytics, increase knowledge of customer usage pattern and the requirement of power for the future.

##### DEMERITS

* + - * Random customers dataset is used as training dataset to the algorithm which was called genuine customers.
      * If the chosen dataset is has fraudulent customers then it will effect the complete result and performance of the detection.

##### CHALLENGES

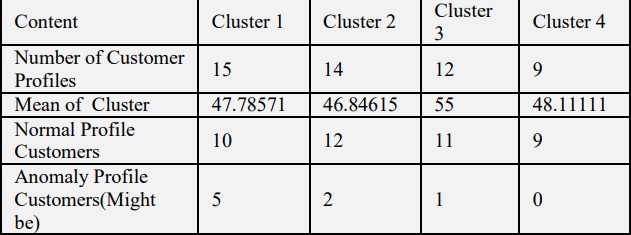
* + - * The customer behavior in power consumption may differ based on variety of feature such as season, special occasion, Temperature, working days.
      * Considered dataset is not enough for the theft detection as it is taken only for 4 weeks of a month which might change the whole results.

##### IMPLEMENTATION

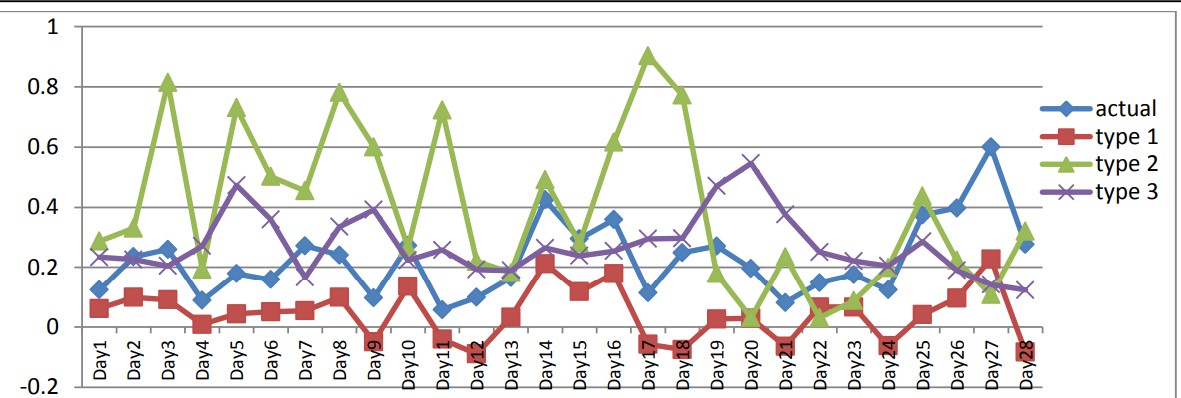
The k-means clustering algorithm allows us to specify the number of expected cluster. It groups the customers based on their 28 days power usage value. ‘K’ cluster are formed to accommodate the customers profile in any one specific cluster. ‘K’ is the number of clusters determined by ourselves. The number of clusters is determined in correspondence to the number of customer profile taken into consideration (E.g. For 10 customer, 3 clusters is acceptable. The clusters are formed with number of customer’s profile, where the cluster with less number of customer profile are not considered. The cluster with the maximum number of customer profile is selected. The customers belonging to the cluster that satisfies these two conditions are considered for the Smart Meter Data Analytics. The prominent customers of the selected cluster are identified by calculating the Euclidean distance of each customer profile to its corresponding cluster head value. The Euclidean distance is used for finding the distance between the cluster member and cluster head. The formula to find the Euclidean distance is as follows: N

D(a,b) = D(b,a) = √∑(ai – bi ) 2 i=1 (1)

Where a, b are the customers profile, n is the total number of customers considered for analytics. After certain number of iteration, the usage pattern of finalized cluster head originated as the source for the anomaly customer detection. The customers profile which are close to the cluster head is identified as genuine profile. The genuine profiles are separated and their energy consumption data are taken for further investigation. The trustworthiness of the customer is verified by using the clustering algorithm



**Table 2.1**: Trustworthiness of Customers

After finding the trustworthiness of the customers, the genuine profiled customers are considered for the classification model by including the bogus data into actual data. The Artificial Neural Network is built to classify the customer’s profile. The three types of bogus data along with the actual data are considered to train the neural network. 60% of the dataset is utilized for training the neural network. After required number of iterations, the neural network is trained to predict any new customer profile to genuine or fraud. The remaining 40% of the dataset is used for testing the dataset. The prediction is made by the ANN classification model. The performance of the proposed system is using two parameters namely accuracy and error rate. The difference in actual class value and the predicted class value is considered for the performance evaluation. The performance of the model depends on the number of dataset taken into consideration.

**Fig 2.2**: Sample customer data with three types of bogus data

import numpy as np import pandas as pd

import matplotlib.pyplot as plt import keras

columns = ['id','daytime','power']

data1 = pd.read\_csv('File1.txt',sep = ' ',names = columns) data1['day']=data1['daytime'][:]

data1['time'] = data1['daytime'][:] data1['day'] = data1['day']/100 data1['day'] = np.int64(data1['day']) data1['time']= data1['time']%100 data1['time'] = np.int64(data1['time']) data1=data1.drop(columns= ['daytime'])

data1 = data1[(data1.day<200) & (data1.day>=195)]

from collections import Counter count = 0

temp = data1['id'].values.tolist() no\_of\_occur = Counter(temp) data1.shape

data1.head() no\_of\_occur

new\_data = data1.drop([1489],axis = 0) new\_data.head() new\_data.reset\_index(drop=True) new\_data.ix[:,2] -= 194

new\_data new\_data.values[:,0].tolist().count(1392) new\_data.ix[:,0]

customers = {} count = 0

for ix in range(new\_data.shape[0]):

if new\_data['id'].iloc[ix] not in customers.keys():

customers[new\_data['id'].iloc[ix]] = new\_data['power'].iloc[ix] else:

temp = {}

temp[new\_data['id'].iloc[ix]] = new\_data['power'].iloc[ix]+ customers[new\_data['id'].iloc[ix]]

CMRCET B. Tech (CSE) Page No 9

customers.update(temp) count +=1

print (count) kk.keys()

kk\_new = pd.DataFrame(kk, columns=['id', 'day\_1.0','day\_2.0','day\_3.0','day\_4.0','day\_5.0'])

kk\_new = pd.DataFrame(kk, columns=['id', 'day\_1.0','day\_2.0','day\_3.0','day\_4.0','day\_5.0']) kk.items()

data\_new = pd.DataFrame(customers.items(), columns=['id', 'totalpower'df2 = pd.DataFrame( [[i,kk[i]['day\_1.0'],kk[i]['day\_2.0'],kk[i]['day\_3.0'],kk[i]['day\_4.0'],kk[i]['day\_5.0'] ] for i in kk.keys()] , columns = ['id','day\_1.0','day\_2.0','day\_3.0','day\_4.0','day\_5.0'])

df2

df3 = df2 import random

multi = random.random()

for ix in range(df3.shape[0]):

for iy in range(df3.shape[1]-1): df3.ix[ix,iy+1] = (df3.ix[ix,iy+1]\*multi)/48

df3

df4 = pd.DataFrame( [[i,kk[i]['day\_1.0'],kk[i]['day\_2.0'],kk[i]['day\_3.0'],kk[i]['day\_4.0'],kk[i]['day\_5.0'] ] for i in kk.keys()] , columns = ['id','day\_1.0','day\_2.0','day\_3.0','day\_4.0','day\_5.0'])

df4

df1\_elements = df4.sample(n=32) type(df1\_elements)

df1\_elements

for ix in range(df1\_elements.shape[0]): gg = df1\_elements['id'].iloc[ix]

for iy in range(df4.shape[0]): if df4['id'].iloc[iy] == gg:

vv = (ix+1)%6 if vv == 0:

vv +=1

col = df4.columns[vv] df4[col].iloc[ix] = 0

df4\

df5 = pd.DataFrame( [[i,kk[i]['day\_1.0'],kk[i]['day\_2.0'],kk[i]['day\_3.0'],kk[i]['day\_4.0'],kk[i]['day\_5.0'] ] for i in kk.keys()] , columns = ['id','day\_1.0','day\_2.0','day\_3.0','day\_4.0','day\_5.0'])

df5

mean = []

CMRCET B. Tech (CSE) Page No10

for ix in range(df5.shape[0]): sum = 0

for iy in range(df5.shape[1]-1):

sum += df5.ix[ix,iy+1] mean.append(sum/5)

for ix in range(df5.shape[0]):

for iy in range(df5.shape[1]-1): df5.ix[ix,iy+1] \*= mean[ix]

df5

orig = pd.DataFrame( [[i,kk[i]['day\_1.0'],kk[i]['day\_2.0'],kk[i]['day\_3.0'],kk[i]['day\_4.0'],kk[i]['day\_5.0'] ] for i in kk.keys()] , columns = ['id','day\_1.0','day\_2.0','day\_3.0','day\_4.0','day\_5.0'])

#orig orig #type1 df3 #type2 df4 #type3 df5 orig['label'] = 1 df3['label']=0 df4['label']=0 df5['label']=0

semifinal = pd.concat([orig,df3,df4,df5],ignore\_index=True) final = semifinal.drop("id", axis=1)

final

divide = int(final.shape[0]\*0.6) #np.random.shuffle(final)

final final.shape final.ix[0,0]

X\_train = final.ix[:divide,:-1] y\_train = final.ix[:divide,-1] X\_test = final.ix[divide:,:-1] y\_test = final.ix[divide:,-1] print (X\_train.shape) X\_train.head()

final2 = final.values np.random.shuffle(final2) X\_train = final2[:divide,:-1] y\_train = final2[:divide,-1] X\_test = final2[divide:,:-1]

y\_test = final2[divide:,-1] y\_test

import random

multi = random.random() print (X\_train.shape) print (X\_test.shape) import keras

from keras.models import Sequential from keras.layers import Dense

from keras.callbacks import ModelCheckpoint print (X\_train.shape)

print (X\_test.shape) classifier2 = Sequential()

classifier2.add(Dense(output\_dim = 16, init = 'uniform', activation = 'relu')) classifier2.add(Dense(output\_dim = 8, init = 'uniform', activation = 'relu')) classifier2.add(Dense(output\_dim = 1, init = 'uniform', activation = 'sigmoid')) classifier2.compile(optimizer = 'adam', loss = 'binary\_crossentropy', metrics = ['accuracy']) checkpointer2 = ModelCheckpoint(filepath='weights.hdf5', verbose=1, save\_best\_only=True) classifier2.summary()

classifier2.fit(X\_train, y\_train, batch\_size=32, epochs=100,validation\_data=(X\_test, y\_test), callbacks=[checkpointer2]) y\_pred = classifier2.predict(X\_test)

y\_pred = (y\_pred > 0.5) classifier2.load\_weights('weights.hdf5') y\_pred = classifier2.predict(X\_test) y\_pred = (y\_pred > 0.5) np.sum(y\_pred==y\_test)/y\_test.shape[0]

y\_test= np.reshape(y\_test,(y\_test.shape[0],1)) y\_pred

from sklearn.metrics import confusion\_matrix cm = confusion\_matrix(y\_test, y\_pred)

cm

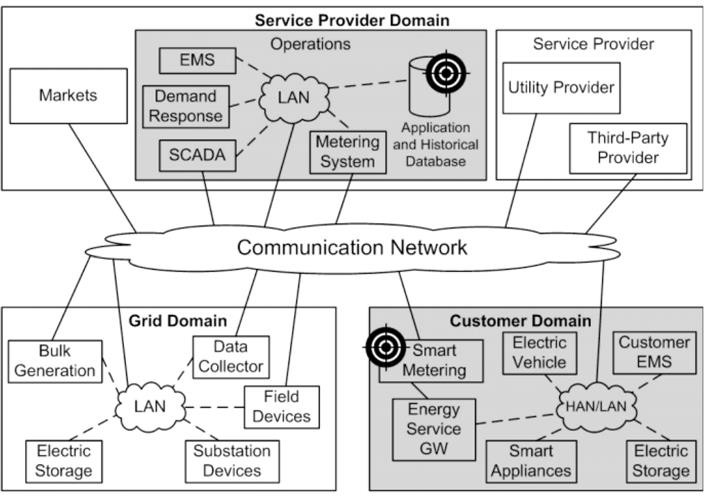
from sklearn.metrics import accuracy\_score print (accuracy\_score(y\_test,y\_pred))

* 1. **Detecting and Locating Non-Technical Losses in Modern Distribution Networks**

##### INTRODUCTION:

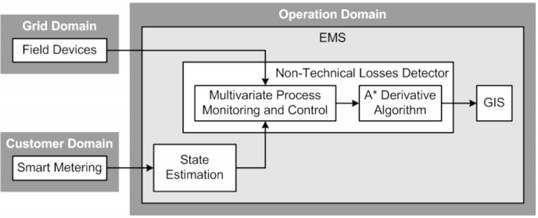
Electrical power loss represents the differencebetween the quantity of energy injected into an electric distribution system and the quantity of energy that is billed.There are two types of electrical power losses: technical and non-technical. Technical losses comprise the power dissipation in the electrical system components (distribution linesand transformers), whereas non-technical losses are causedby unpredicted external actions against the electrical powersystem. Non-technical losses are the major source of commercial loss because of the difficulty of measuring them. Themost probable causes of non-technical losses are related tofrauds, such as the alteration of meter accuracy, consumption of unbilled energy bypassing utility meters, and tapping low-voltage lines.

Smart grid characteristics change the nature of electricity theft. Attacks range from crude physical system manipulation to the remote penetration and control of complex computational systems, new vulnerabilities of the smart grid infrastructure such as different types of cyber-attacksare identified. Cyber-attacks require multiple defense mechanisms that have high cost for protecting all vulnerable loads inlarge power systems. Cost-efficient load protection strategies should minimize the cost and prevent damages in the power grid. In this way, assumes the feasibility of cyber tampering on electronic meters and proposes a framework to perform online data detection of irregularity in the measurements. The distribution network is divided into subsystems limited by feeder remote terminal units (FRTUs). Each subsystem is checked using the distribution power flow module. The non-technical losses are detected when the mismatch ratio is frequently greater than the predefined threshold. The calculation of the mismatch ratio depends on the average three-phase power consumption, power losses and power measurements foreach subsystem. The use of average values requires additional stages to recognize consumption patterns based on historical load profile.



**Fig 2.3 :** Architecture for a smart grid

Figure shows the block diagram of the proposed procedure for detecting and locating non- technical losses in distribution networks. The input data of the non-technical loss detector come from the grid and customer domains. Data from the grid domain are reliable states measured by field devices, such as the phasor measurement unit (PMU) and intelligent electronic device (IED). Field devices measure states at terminals of dis- tribution transformers or automatic switches. Reliable states are compared with states calculated by a state estimator that utilizes data from the smart metering system.



**Fig 2.4 :** Process diagram

##### MERITS, DEMERITS AND CHALLENGES MERITS

* Generates information valuable for policy and decision making
* Can prevent all types of NTLs resulting from the meter and Electrical network
* Low cost and use of available resources
* Precise estimates of performance

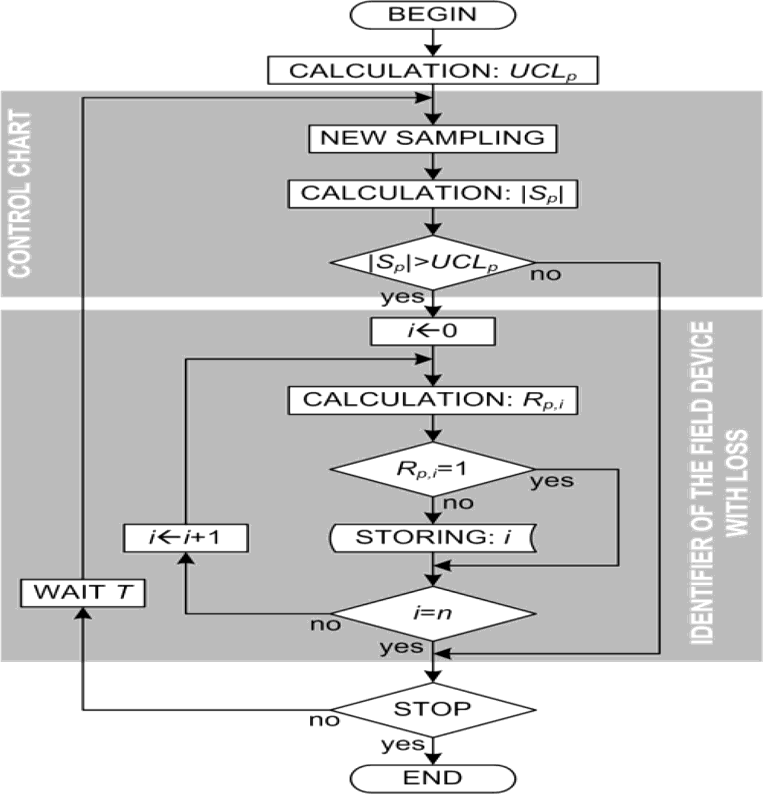
##### DEMERITS

* Unable to detect the specific sources of NTLs
* Costs with equipment’s

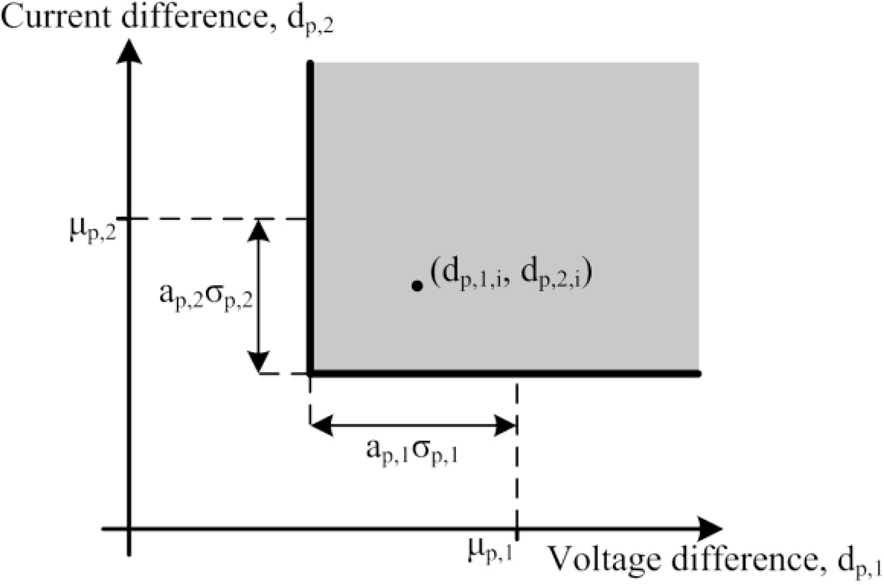
##### CHALLENGES

* Require smart meters which are costly and require high maintainance.
* Detection is not guaranteed and required data may not be available.
  + 1. **IMPLEMENTATION**

The comparison of network states in the power system hasa selectivity problem because large magnitudes are compared to detect small errors. Voltage and current measurements havelarge magnitudes, whereas small errors result from voltage andcurrent differences. The utilization of the multivariate procedure of monitoring and control overcomes the selectivity problem in the power system.

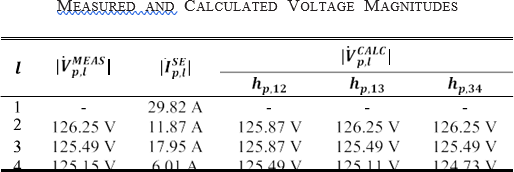


**Fig 2.5 Flow Chart**



**Figure 2.6** Graph I vs V

Confidenceintervals have one lower limit and no upper limit because large dissimilarities produce low values of *dp,j,i*, whereas similarities produce high values, according to (1) and (2). The positionidentification of the point *i (dp,*1*,i, dp,*2*,i)* in the dispersion diagram, i.e., inside or outside of the bounds of the confidenceintervals, is achieved by the calculation of the discrete range, *Rp,i*, according to (12)

Table

**Table 2.2 Voltage Magnitudes**

highlights three points (*N* 10, 20 and 30) thatare used to evaluate the impact of the sample space size on the efficiency and selectivity of the detection algorithm. In addition, one unregistered load is inserted in the LV network for emulating the non-technical loss. The unregistered load is randomly varied from 0 to 10 kVA in a total of 350 power changes for each value of *N*. The random power values are divided into two groups: one group from 0 to 1 VA, and the other group from 1 VA to 10 kVA.

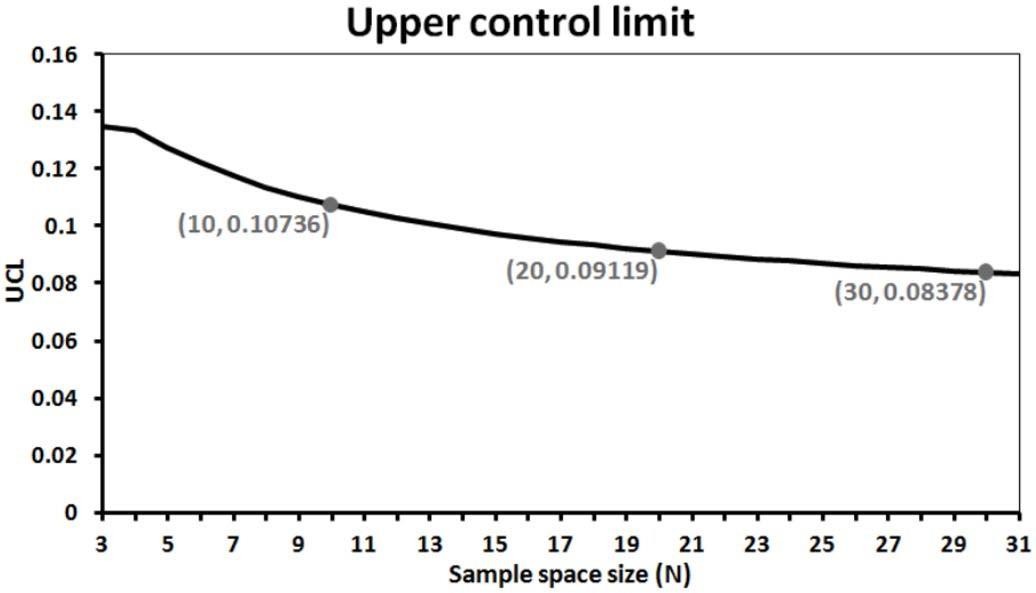
The algorithm behavior is shown through a bar chart ofthe successful rate by the apparent power groups and sample space sizes. The successful rate is obtained by the relation among the amount of detected instances and the total number of simulated instances where each instance is a power change.

Fig 2.7 UCL graph

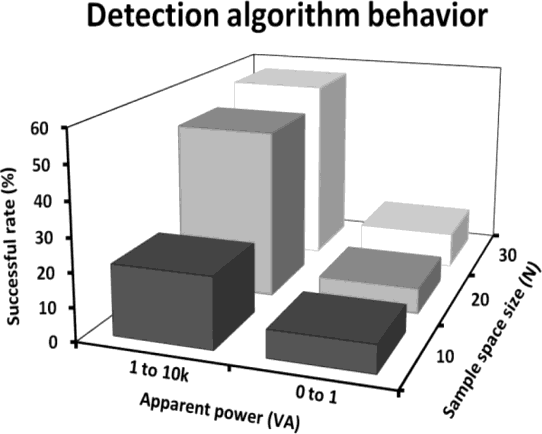


Fig 2.8 Algorithm graph

* 1. **Using GANCNN and ERNET for Detection of Non Technical Losses to Secure Smart Grids**

##### INTRODUCTION:

With the increase in the number of residential homes and industries, the demand of energy increases manifolds. There-fore, power generation companies need to generate more electricity Moreover, there should be a balance betweenelectricity generation and consumption to mitigate the issue of energy shortage. Due to the latest advancements in Advanced Metering Infrastructure (AMI), traditional grids are converted into smart grids where data is collected through smart meters. The balance between demand and supply is also established using bi- directional flow of energy and information. In energy transmission systems, two types of losses occur, which are known as Technical Losses (TLs) and Non-Technical Losses (NTLs). The former losses occurdue to poor infrastructure and energy dissipation. Whereas, the latter losses are defined as the difference between total electricity transmitted through distribution lines and the elec- tricity consumed by the users. Due to the NTLs, power utili-ties face losses worth millions of dollars, which highly affectthe country’s economy. The manual inspection of these losses is both time consuming and expensive.

There are different reasons for the occurrence of NTLs, which are broadly categorized in two categories: human and non-human. The former includes tampering the meter read- ings, hooking with the main lines, etc. Whereas, the lat- ter includes errors in smart meters, fluctuating energy flow, meter inaccuracies, etc., With the NTLs, other losses also occur, such as unbearable load on electrical systems, load shedding, economical loss, etc., With the use of smart meters, flow of both energy and information becomes auto- mated. For the utility companies, the smart meters remotely provide data related to readings of electricity consumption on real time basis. Therefore, it becomes easy to steal the electricity by manipulating the electricity consumption data.

##### METHODOLOGIES

In smart grids, anomaly is defined as the deviation from regular or normal electricity consumption patterns. It occurs due to many factors like arrival of more family membersat home, occurrence of a special occasion, illegal use of electricity, etc. In anomaly detection, data-driven models areused that learn the normal patterns and detect the abnormal patterns to identify the electricity thieves.

The LSTM method may require high memory bandwidth to feed its computational units. Ding *et al.* have proposed a hybrid model, whichis based on Gausian Mixture Model (GMM) and LSTM for the detection of real time anomaly. However, it is difficult todetermine for certain the number of clusters to be created. Also, the class imbalanced problem is not tackled. Authors in have proposed Jaya-LSTM for the forecasting of elec- tricity load. All of the above mentioned methods perform bet- ter in terms of anomaly detection. However, the methods arenot feasible enough to accurately detect electricity fraudsters. Zheng have proposed Wide and Deep Convolutional Neural Network (WDCNN) for ETD. They have used State Grid Corporation of China (SGCC) dataset, which consists of verified electricity thieves. However, the class imbalancedproblem is not addressed.

Moreover, electricity theft is a crucial problem for utility companies, as they have to bear huge losses every year. Many data-driven based solutions are proposed in the literaturefor ETD. However, there exists some limitations in these solutions, which are needed to be addressed. Li *et al.* have proposed a hybrid model, which consists of CNN and Random Forest (RF) for ETD. However, the computational complexity of RF is very high as it takes more time to con- struct decision trees. In addition, FPR is also not calculated.Hasan *et al.* have proposed a hybrid technique by com- bining CNN and LSTM for ETD. The proposed technique efficiently performed in terms of accuracy. However, LSTMrequires a lot of memory for storing long-term sequences. Moreover, LSTM is not hardware friendly because it needs more resources as compared to CNN and Gated Recurrent Unit (GRU). In , authors have used Synthetic Minor- ity Oversampling Technique (SMOTE) to balance the data for training CNN and LSTM models to perform classifica- tion.

##### MERITS, DEMERITS AND CHALLENGES

**MERITS**

* A robust hybrid approach to detect anomaly in the behavior of electricity consumers using the k-means clustering mechanism and DNN algorithm
* Anomaly detection framework based on loss factor and error term to detect NTLs
* A LSTM method for anomaly detection in electricity consumption data
* A WDCNN method is used for ETD
* A simple moving average method is used for ETD

##### DEMERITS

* Class imbalanced problem is not addressed
* The LSTM method may require high memory bandwidth to feed the computational units
* It is difficult to determine for certain the no of clusters to be created

##### CHALLENGES

* The model is not efficient in real life scenarios
* Techniques are not suitable for smaller datasets as the methods may create overfitting problem
  + 1. **IMPLEMENTATION**

##### Using GANCNN and ERNET for Detection of Non Technical Losses to Secure Smart Grids

To overcome the issues identified from the literature, we pro- pose two deep learning models in this work: GANCNN andERNET. The former is the combination of Self-Attention Generative Adversarial Network (SAGAN) and CNN. Whereas, the latter is a hybrid of EfficientNet, Residual Network (ResNet) and GRU. In the GANCNN model, data sampling and feature extraction are done using Adaptive Syn- thetic Edited Nearest Neighbor (ADASYNENN) and Locally Linear Embedding (LLE), respectively. In the ERNET model, GRU is applied for the classification of honest and dishonest consumers.

For the classification of electricity thieves and normal con- sumers, a hybrid GANCNN model is proposed, which is a combination of SAGAN and WDCNN. SAGAN is a deep learning model and is considered as the best training model.It has two modules: generator and discriminator . The former creates synthetic data similar to original data byselecting random input samples from the dataset. The latter discriminates between fake and original data . During GAN’s process, both generator and discriminator modules are trained until discriminator is failed half of the time to distin-guish between fake and original samples, which means that generator is successful in creating fake samples. The randominput samples are selected on the basis of inverse transform technique in which Cumulative Distribution Function (CDF)is used. CDF is given in Equation

*CDF* = *fZ* (*z*) = *P*(*Z <*= *z*)

where *P* is the probability, *Z* is randomly selected input fromdata and *z* is the input sample.

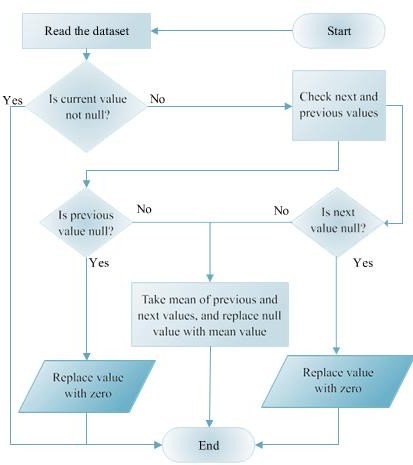


Fig 2.9 Flow chart of data cleaning process.

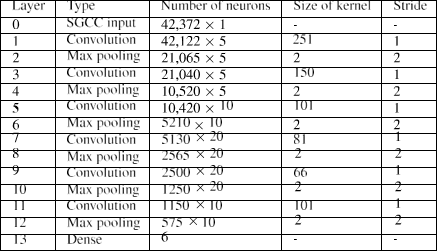


Table 2.3 Network structure

The second model proposed in this work for ETD is com- prised of five stages, as shown in Figure 8. The stagesare same as defined in the GANCNN model. The SGCC dataset is used for this model as well. For dimensionality reduction, Sparse Auto Encoder (SAE) technique is usedas a feature extractor. For data sampling, SMOTEENN is proposed. A hybrid of EfficientNet, ResNet and GRU, named as ERNET, is proposed for classification of theft and normalconsumers. A detailed flowchart of the ERNET model is shown in Figure

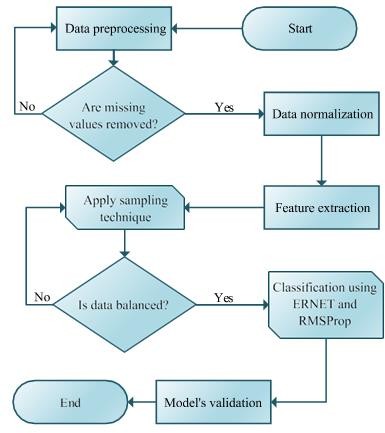
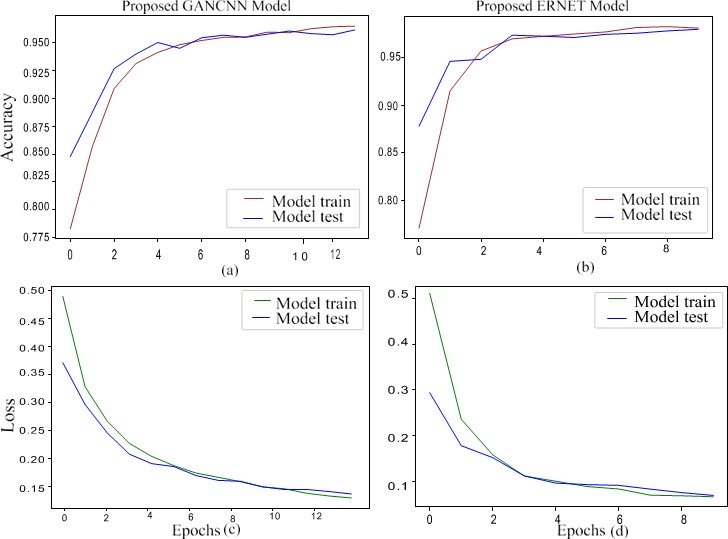


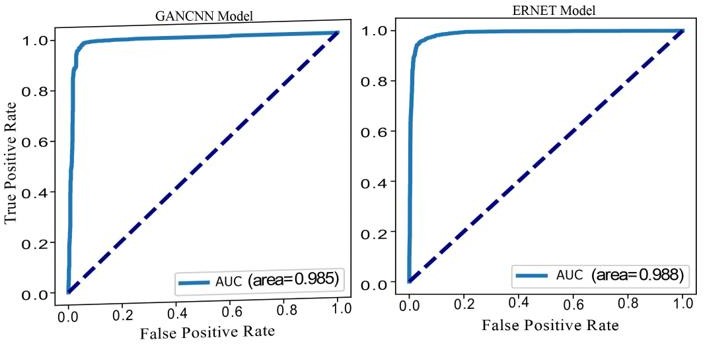
Fig 2.10 Flowchart of ERNET model.

Existing sampling techniques have some limitations. For example, undersampling technique discards important infor-mation by randomly removing samples. Whereas, in over- sampling technique, overfitting problem arises due tothe duplication of samples, which further leads to poor generalization and misclassification. To resolve the afore- mentioned issues, SMOTEENN is used in the proposed work. The working of SMOTEENN is presented in Algorithm 1. It can be seen from the algorithm that SMOTEENN com- prises of two sampling techniques: SMOTE (lines 1-9) and ENN (lines 10-14). The algorithm is presented to give a better understanding to the readers that how SMOTEENN would work. The technique starts with the oversampling of the minority class using SMOTE, which is an enhanced versionof Random Oversampling



* 1. Accuracy and loss of the GANCNN and ERNET models.

The former is used to generate synthetic data using the actual data. The purpose of generating synthetic data is to solve the class imbalanced problem. Whereas, the latter is used for increas- ing the network’s width and depth. The aim of increasing thewidth and depth of the network is to achieve higher predictionaccuracy from a large dataset.



* 1. AUC of proposed GANCNN and ERNET models.

# CHAPTER 3

## RESULTS AND

**DISCUSSION**

Electricity Theft Detection in Power Grids with Deep Learning and Random Forests

##### CHAPTER 3 RESULTS AND DISCUSSION

* 1. **COMPARISION OF EXISTING MODELS Algorithm**
     1. K-means clustering to analyze genuine customers data, Artificial neural networks (ANN)
     2. A-Star algorithm.
     3. GANCNN (Generative Adversarial Network convolutional neural networks) and ERNET (hybrid of Efficient Net, Residual Network (ResNet))

##### Data used for theft detection

1. A sample customer’s smart meter reading for every 30 minutes in (KWH) for 28 days is considered.
2. The testing distribution network provides energy to 834 monitored loads, including 53 loads at the medium-voltage network. There are 47 distribution transformers with a suitable field device for monitoring.
3. The consumption records of 42,372 consumers are present in the dataset.

##### Accuracy

1. The Accuracy (AC) depends on the two factors: Sensitivity (SE) and Specificity (SP). The accuracy of the proposed system is the percentage of correctly classified records. By considering equation, the proposed system provides 97% accuracy.
2. The accuracy is low i.e., 22% for small datasets and it is more i.e., 78% accurate for large datasets.
3. The results of GANCNN for accuracy 95%. On the other hand, the results of ERNET accuracy 98%.

Electricity Theft Detection in Power Grids with Deep Learning and Random Forests

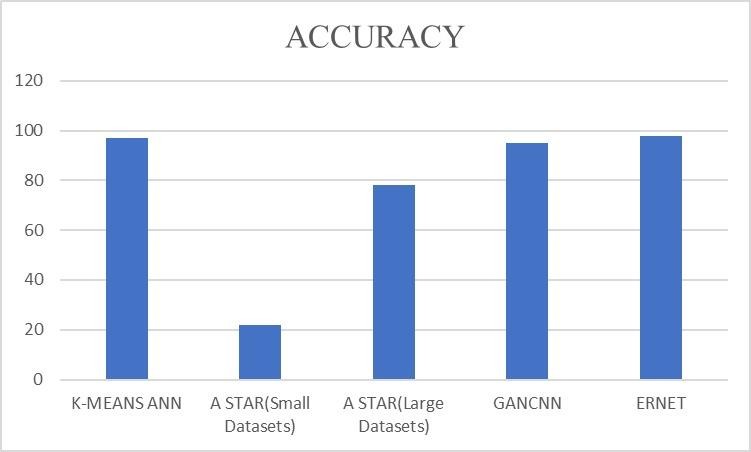


Fig 3.1 : Accuracy

##### Data Collection and Performance Metrics

###### Data Collection

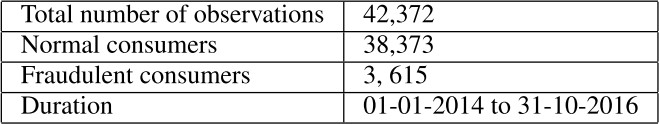
* + 1. The dataset used for the experiment is the subset of smart meter dataset of Ireland in December 2010. The data set considered in this work is residential smart meter data. The data contains the information about the customer id, code for date/time, electricity consumption for every 30 minutes (in KWh). The daily profile for every customer comprises of 48 power consumption readings. This work requires the aggregated data values. The electricity consumption data obtained for every day per customer is 48 readings. These 48 readings of a customer are averaged for per day power consumption data value. For the experimental purpose, a sample dataset comprises of 4 weeks power consumption reading of the customers is considered.
    2. The proposed methodology is evaluated under a real distribution network with the topological diagram. The testing distribution network provides energy to 834 monitored loads, including 53 loads at the medium-voltage (MV) network. There are 47 distribution transformers with a suitable field device for monitoring.
    3. The data used in this work is acquired from SGCC. It is a labeled dataset with a known number of electricity thieves. It consists of customers’ ID, daily consumption and flagged (i.e., target attribute) either as 0 or 1. Daily electricity consumption data from January 2014 to October 2016 is considered. The consumption records of 42,372 consumers are present in the dataset. Out of these, 3,615 are electricity thieves and remaining 38,373 are normal consumers. Below table shows the detail of SGCC dataset.

Table 3.1 : Dataset table

##### Performance Metrics

1. AC = (TP+TN) / (TN + FP + TP + FN)

Sensitivity (SE): Recall is commonly called as Sensitivity. It is the True Positive Rate (TPR).

SE = TP / (TP + FN)

Specificity (SP): Specificity is True Negative Rate (TNR).

SP = TN / (TN +FP) Precision: Precision is Positive Predictive Value (PPV). Precision = TP / (TP+FP)

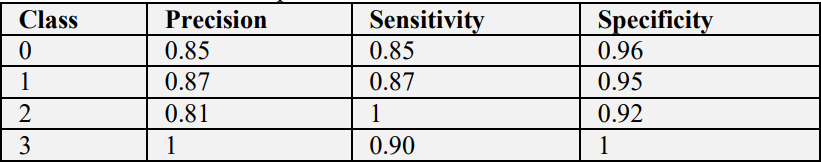


Table 3.2 Accuracy table

2.

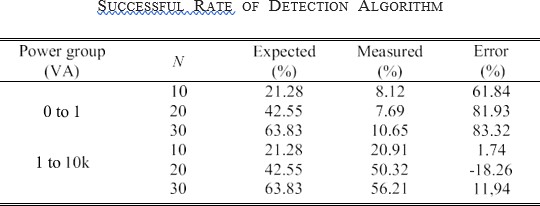


Table 3.3 Success rate table

###### 3.

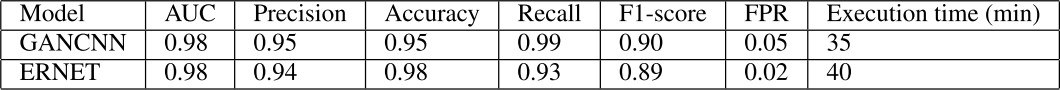


Table 3.4 Efficiency table

CHAPTER 4

## CONCLUSION

#### CONCLUSION

In this paper, a novel CNN-RF model is presented to detect electricity theft. In this model, the CNN is similar to an automatic feature extractor in investigating smart meter data and the RF is the output classifier. Because a large number of parameters must be optimized that increase the risk of overfitting, a fully connected layer with a dropout rate of 0.4 is designed during the training phase. In addition, the SMOT algorithm is adopted to overcome the problem of data imbalance. Some machine learning and deep learning methods such as SVM, RF, GBDT, and LR are applied to the same problem as a benchmark, and all those methods have been conducted on SEAI and LCL datasets. The results indicate that the proposed CNN-RF model is quite a promising classification method in the electricity theft detection field because of two properties: The first is that features can be automatically extracted by the hybrid model, while the success of most other traditional classifiers relies largely on the retrieval of good hand-designed features which is a laborious and time-consuming task. The second lies in that the hybrid model combines the advantages of the RF and CNN, as both are the most popular and successful classifiers in the electricity theft detection field.

Since the detection of electricity theft affects the privacy of consumers, the future work will focus on investigating how the granularity and duration of smart meter data might affect this privacy. Extending the proposed hybrid CNN-RF model to other applications (e.g., load forecasting) is a task worth investigating.

# CHAPTER 5

## REFERENCES

#### REFERENCES

1. S. S. S. R. Depuru, L. Wang, and V. Devabhaktuni, “Electricity theft: overview, issues, prevention and a smart meter based approach to control theft,” *Energy Policy*, vol. 39, no. 2, pp. 1007–1015, 2011.View at: [Publisher Site](https://doi.org/10.1016/j.enpol.2010.11.037) | [Google Scholar](https://scholar.google.com/scholar_lookup?title=Electricity%20theft%3A%20overview%2C%20issues%2C%20prevention%20and%20a%20smart%20meter%20based%20approach%20to%20control%20theft&author=S.%20S.%20S.%20R.%20Depuru&author=L.%20Wang&author&author=V.%20Devabhaktuni&publication_year=2011)
2. J. P. Navani, N. K. Sharma, and S. Sapra, “Technical and non-technical losses in power system and its economic consequence in Indian economy,” *International Journal of Electronics and Computer Science Engineering*, vol. 1, no. 2, pp. 757–761, 2012.View at: [Google Scholar](https://scholar.google.com/scholar_lookup?title=Technical%20and%20non-technical%20losses%20in%20power%20system%20and%20its%20economic%20consequence%20in%20Indian%20economy&author=J.%20P.%20Navani&author=N.%20K.%20Sharma&author&author=S.%20Sapra&publication_year=2012)
3. P. McDaniel and S. McLaughlin, “Security and privacy challenges in the smart grid,” *IEEE Security & Privacy Magazine*, vol. 7, no. 3, pp. 75–77, 2009.View at: [Publisher Site](https://doi.org/10.1109/msp.2009.76) | [Google](https://scholar.google.com/scholar_lookup?title=Security%20and%20privacy%20challenges%20in%20the%20smart%20grid&author=P.%20McDaniel%20&author=S.%20McLaughlin&publication_year=2009) [Scholar](https://scholar.google.com/scholar_lookup?title=Security%20and%20privacy%20challenges%20in%20the%20smart%20grid&author=P.%20McDaniel%20&author=S.%20McLaughlin&publication_year=2009)
4. T. B. Smith, “Electricity theft: a comparative analysis,” *Energy Policy*, vol. 32, no. 1, pp. 2067– 2076, 2004.View at: [Publisher Site](https://doi.org/10.1016/s0301-4215(03)00182-4) | [Google Scholar](https://scholar.google.com/scholar_lookup?title=Electricity%20theft%3A%20a%20comparative%20analysis&author=T.%20B.%20Smith&publication_year=2004)
5. J. I. Guerrero, C. León, I. Monedero, F. Biscarri, and J. Biscarri, “Improving knowledge-based systems with statistical techniques, text mining, and neural networks for non-technical loss detection,” *Knowledge-Based Systems*, vol. 71, no. 4, pp. 376–388, 2014.View at: [Publisher](https://doi.org/10.1016/j.knosys.2014.08.014) [Site](https://doi.org/10.1016/j.knosys.2014.08.014) | [Google Scholar](https://scholar.google.com/scholar_lookup?title=Improving%20knowledge-based%20systems%20with%20statistical%20techniques%2C%20text%20mining%2C%20and%20neural%20networks%20for%20non-technical%20loss%20detection&author=J.%20I.%20Guerrero&author=C.%20Le%C3%B3n&author=I.%20Monedero&author=F.%20Biscarri&author&author=J.%20Biscarri&publication_year=2014)
6. S.-C. Huang, Y.-L. Lo, and C.-N. Lu, “Non-technical loss detection using state estimation and analysis of variance,” *IEEE Transactions on Power Systems*, vol. 28, no. 3, pp. 2959–2966, 2013.View at: [Publisher Site](https://doi.org/10.1109/tpwrs.2012.2224891) | [Google Scholar](https://scholar.google.com/scholar_lookup?title=Non-technical%20loss%20detection%20using%20state%20estimation%20and%20analysis%20of%20variance&author=S.-C.%20Huang&author=Y.-L.%20Lo&author&author=C.-N.%20Lu&publication_year=2013)
7. N. Edison, A. C. Aranha, and J. Coelho, “Probabilistic methodology for technical and non- technical losses estimation in distribution system,” *Electric Power Systems Research*, vol. 97, no. 11, pp. 93–99, 2013.View at: [Publisher Site](https://doi.org/10.1016/j.epsr.2012.12.008) | [Google Scholar](https://scholar.google.com/scholar_lookup?title=Probabilistic%20methodology%20for%20technical%20and%20non-technical%20losses%20estimation%20in%20distribution%20system&author=N.%20Edison&author=A.%20C.%20Aranha&author&author=J.%20Coelho&publication_year=2013)
8. J. B. Leite and J. R. S. Mantovani, “Detecting and locating non-technical losses in modern distribution networks,” *IEEE Transactions on Smart Grid*, vol. 9, no. 2, pp. 1023–1032, 2018.View at: [Publisher Site](https://doi.org/10.1109/tsg.2016.2574714) | [Google Scholar](https://scholar.google.com/scholar_lookup?title=Detecting%20and%20locating%20non-technical%20losses%20in%20modern%20distribution%20networks&author=J.%20B.%20Leite%20&author=J.%20R.%20S.%20Mantovani&publication_year=2018)
9. S. Amin, G. A. Schwartz, A. A. Cardenas, and S. S. Sastry, “Game-theoretic models of electricity theft detection in smart utility networks: providing new capabilities with advanced metering infrastructure,” *IEEE Control Systems Magazine*, vol. 35, no. 1, pp. 66–81, 2015.View at: [Google Scholar](https://scholar.google.com/scholar_lookup?title=Game-theoretic%20models%20of%20electricity%20theft%20detection%20in%20smart%20utility%20networks%3A%20providing%20new%20capabilities%20with%20advanced%20metering%20infrastructure&author=S.%20Amin&author=G.%20A.%20Schwartz&author=A.%20A.%20Cardenas&author&author=S.%20S.%20Sastry&publication_year=2015)
10. 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.View at: [Publisher](https://doi.org/10.1109/tpwrd.2009.2030890) [Site](https://doi.org/10.1109/tpwrd.2009.2030890) | [Google Scholar](https://scholar.google.com/scholar_lookup?title=Nontechnical%20loss%20detection%20for%20metered%20customers%20in%20power%20utility%20using%20support%20vector%20machines&author=J.%20Nagi&author=K.%20S.%20Yap&author=S.%20K.%20Tiong&author=S.%20K.%20Ahmed&author&author=M.%20Mohamad&publication_year=2010)
11. E. W. S. Angelos, O. R. Saavedra, O. A. C. Cortés, and A. N. de Souza, “Detection and identification of abnormalities in customer consumptions in power distribution systems,” *IEEE Transactions on Power Delivery*, vol. 26, no. 4, pp. 2436–2442, 2011.View at: [Publisher Site](https://doi.org/10.1109/tpwrd.2011.2161621) | [Google Scholar](https://scholar.google.com/scholar_lookup?title=Detection%20and%20identification%20of%20abnormalities%20in%20customer%20consumptions%20in%20power%20distribution%20systems&author=E.%20W.%20S.%20Angelos&author=O.%20R.%20Saavedra&author=O.%20A.%20C.%20Cort%C3%A9s&author&author=A.%20N.%20de%20Souza&publication_year=2011)
12. M. Ismail, M. Shahin, M. F. Shaaban, E. Serpedin, and K. Qaraqe, “Efﬁcient detection of electricity theft cyber attacks in ami networks,” in *Proceedings of the IEEE Wireless Communications and Networking Conference*, Barcelona, Spain, April 2018.View at: [Google Scholar](https://scholar.google.com/scholar_lookup?title=Ef%EF%AC%81cient%20detection%20of%20electricity%20theft%20cyber%20attacks%20in%20ami%20networks&author=M.%20Ismail&author=M.%20Shahin&author=M.%20F.%20Shaaban&author=E.%20Serpedin&author&author=K.%20Qaraqe)
13. Z. Zheng, Y. Yang, X. Niu, H.-N. Dai, and Y. Zhou, “Wide and deep convolutional neural networks for electricity-theft detection to secure smart grids,” *IEEE Transactions on Industrial Informatics*, vol. 14, no. 4, pp. 1606–1615, 2018.View at: [Publisher Site](https://doi.org/10.1109/tii.2017.2785963) | [Google](https://scholar.google.com/scholar_lookup?title=Wide%20and%20deep%20convolutional%20neural%20networks%20for%20electricity-theft%20detection%20to%20secure%20smart%20grids&author=Z.%20Zheng&author=Y.%20Yang&author=X.%20Niu&author=H.-N.%20Dai&author&author=Y.%20Zhou&publication_year=2018) [Scholar](https://scholar.google.com/scholar_lookup?title=Wide%20and%20deep%20convolutional%20neural%20networks%20for%20electricity-theft%20detection%20to%20secure%20smart%20grids&author=Z.%20Zheng&author=Y.%20Yang&author=X.%20Niu&author=H.-N.%20Dai&author&author=Y.%20Zhou&publication_year=2018)

P. Jokar, N. Arianpoo, and V. C. M. Leung, “Electricity theft detection in AMI using customers’ consumption patterns,” *IEEE Transactions on Smart Grid*, vol. 7, no. 1, pp. 216–226, 2016.View at: [Publisher Site](https://doi.org/10.1109/tsg.2015.2425222) | [Google Scholar](https://scholar.google.com/scholar_lookup?title=Electricity%20theft%20detection%20in%20AMI%20using%20customers%E2%80%99%20consumption%20patterns&author=P.%20Jokar&author=N.%20Arianpoo&author&author=V.%20C.%20M.%20Leung&publication_year=2016)