##### 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-II 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 CONTENT**

**CHAPTER**

**NO. TITLE PAGE NO.**

LIST OF FIGURES ii

LIST OF TABLES iii

ABSTRACT iv

**1** **INTRODUCTION** 1

1.1 Problem Statement 2

1.2 Research Objective 2

1.3 Project Scope and Limitations 3

**2** **BACKGROUND WORK**

**2.1.** **Machine Learning Algorithm for Efficient Power Theft Detection**

**using Smart Meter Data**

2.1.1. Introduction 4

2.1.2. Merits, Demerits and Challenges 6

2.1.3. Implementation 7

**2.2. Detecting and Locating Non-Technical Losses in**

**Modern Distribution Networks**

2.2.1. Introduction 13

2.2.2. Merits, Demerits and Challenges 15

2.2.3. Implementation 16

**2.3.** **Using GANCNN and ERNET for Detection of Non Technical Losses**

2.3.1. Introduction 19

2.3.2. Merits, Demerits and Challenges 21

2.3.3. Implementation 22

**3 PROPOSED SYSTEM**

3.1. Objective of Proposed Model 25

3.2. Algorithms Used for Proposed Model 25

3.3. Designing 27

3.3.1.UML Diagram 27

3.4. Stepwise Implementation and Code 29

CMRCET B. Tech (CSE) Page No i

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

**4 RESULTS AND DISCUSSION** 38

4.1. Performance Metrics 43

**5 CONCLUSION** 44

5.1. Conclusion and Future Enhancement 45

**6 REFERENCES** 46

**GITHUB LINK** 47

CMRCET B. Tech (CSE) Page No ii

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

**List of Figures**

**FIGURE NO. TITLE PAGE NO.**

2.1 Electricity Theft Identification Model 5

2.2 Sample customer data with three types of bogus data 8

2.3 Architecture for a smart grid 14

2.4 Process diagram 14

2.5 Flow chart 16

2.6 Graph I vs V 17

2.7 UCL graph 18

2.8 Algorithm graph 18

2.9 Flow chart of data cleaning process 22

2.10 Flowchart of ERNET model. 23

2.11 Accuracy and loss of the GANCNN and ERNET models 24

2.12 AUC of proposed GANCNN and ERNET models 24

3.1 CNN Architecture 25

3.2 Random Forest Architecture 26

3.3 USE Case Diagram 27

3.4 Class Diagram 28

3.5 Sequence Diagram 28

3.6 Collaboration Diagram 29

4.1 Output 38

4.2 Uploading Electricity Theft Dataset 38

4.3 Pre Processing Dataset 39

4.4 Generating CNN Model 39

4.5 Result of CNN 40

4.6 Result of CNN with Random Forest 40

4.7 Result of CNN with SVM 41

4.8 Result of Random Forest 41

4.9 Result of SVM 42

4.10 Result of Predict Electricity Theft 42

4.11 Comparison Graph 43

CMRCET B. Tech (CSE) Page No iii

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

**List of Tables**

**TABLE**

**NO. TITLE PAGE NO.**

2.1 Trust worthiness of customers . 7

2.2 Voltage magnitudes 17

2.3 Network structure 23

CMRCET B. Tech (CSE) Page No iv

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

# **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 company 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.

CMRCET B. Tech (CSE) Page No V

# **CHAPTER 1**

**INTRODUCTION**

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

**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 lines 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 electricity theft 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.

CMRCET B. Tech (CSE) Page No 1

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

* 1. **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.

* 1. **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.

CMRCET B. Tech (CSE) Page No 2

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

**1.3 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.

CMRCET B. Tech (CSE) Page No 3

**CHAPTER 2**

**BACKGROUND WORK**

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

**CHAPTER 2**

**BACKGROUND WORK**

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

**2.1.1 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.

CMRCET B. Tech (CSE) Page No 4

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

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.

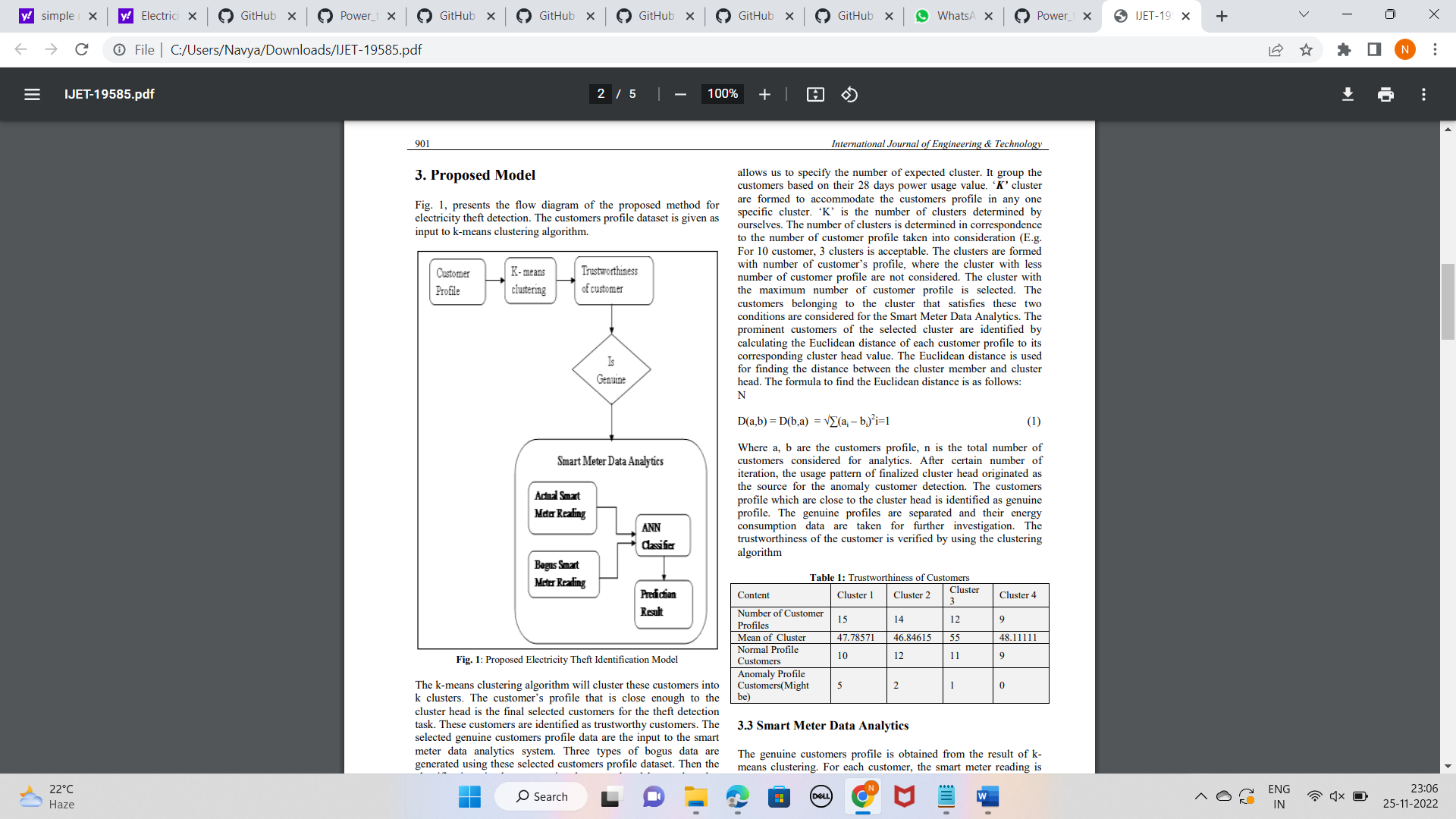
****

Fig. 2.1:Electricity Theft Identification Model

CMRCET B. Tech (CSE) Page No 5

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

**2.1.2 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.

CMRCET B. Tech (CSE) Page No 6

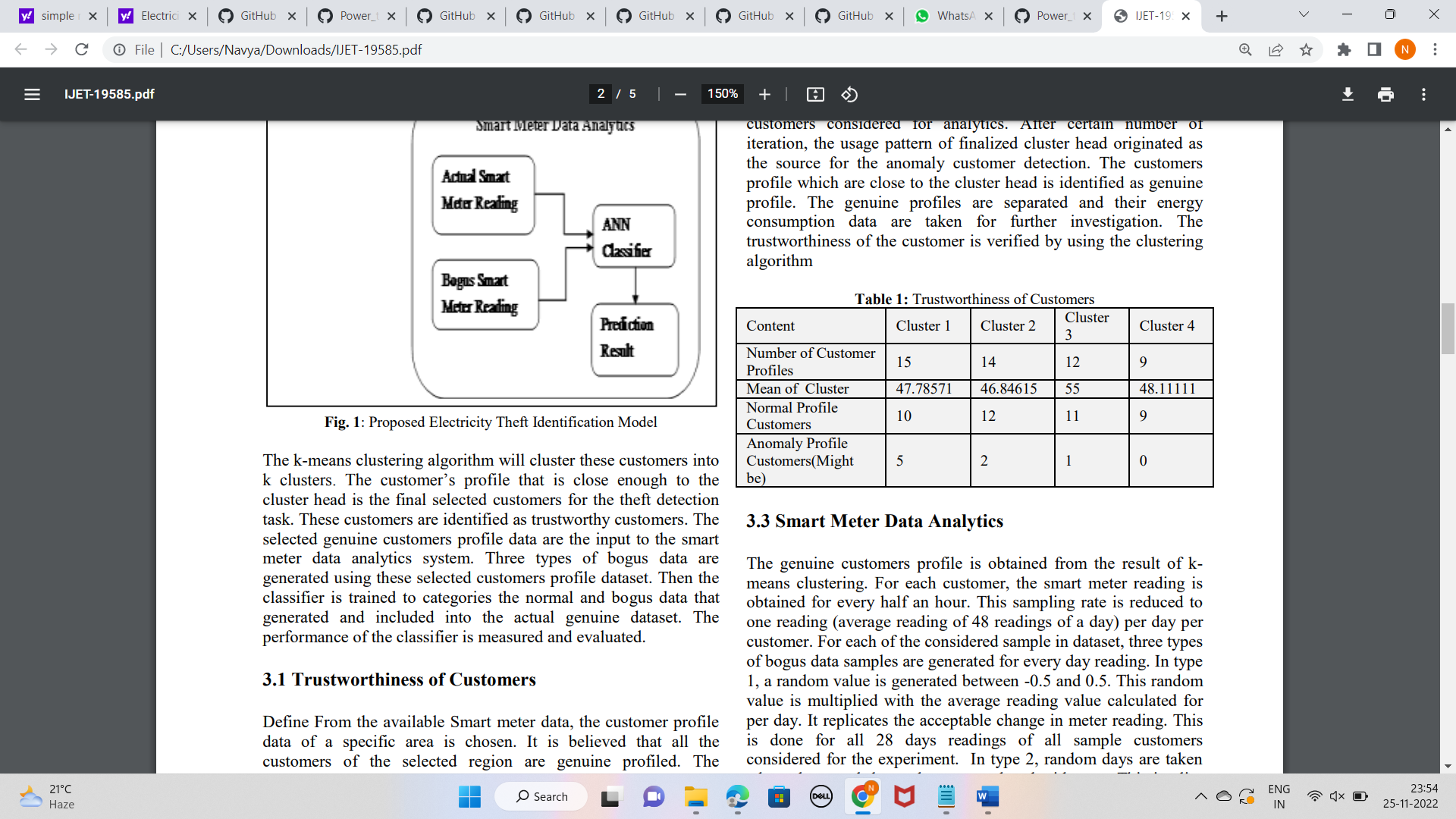
Electricity Theft Detection in Power Grids with Deep Learning and Random Forests

**2.1.3 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

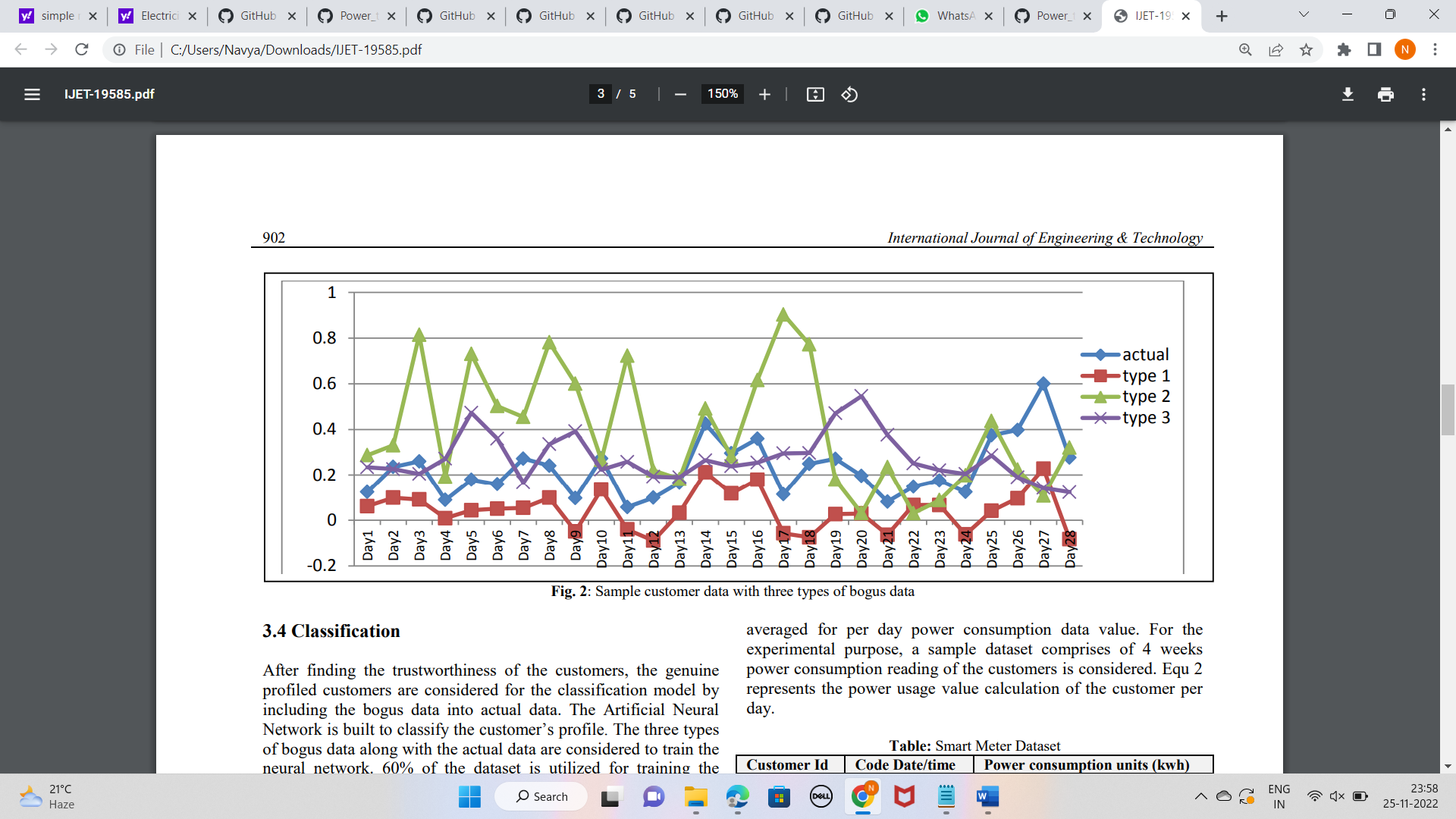
****

**Table 2.1**: Trustworthiness of Customers

CMRCET B. Tech (CSE) Page No 7

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

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

CMRCET B. Tech (CSE) Page No 8

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

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

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

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

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

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]

CMRCET B. Tech (CSE) Page No 11

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

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))

CMRCET B. Tech (CSE) Page No 12

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

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

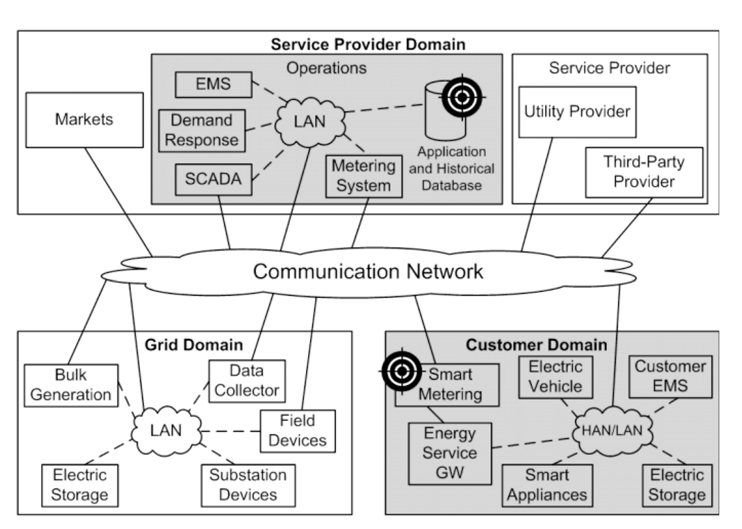
**2.2.1 INTRODUCTION:**

Electrical power loss represents the difference between 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 dissipa tion in the electrical system components (distribution lines and transformers), whereas non-technical losses are caused by unpredicted external actions against the electrical power system. Non-technical losses are the major source of commercial loss because of the difficulty of measuring them. The most probable causes of non-technical losses are related to frauds, such as the alteration of meter accuracy, consump tion 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-attacks are identified. Cyber-attacks require multiple defense mechanisms that have high cost for protecting all vulnerable loads in large 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 for each subsystem. The use of average values requires additional stages to recognize consumption patterns based on historical load profile.

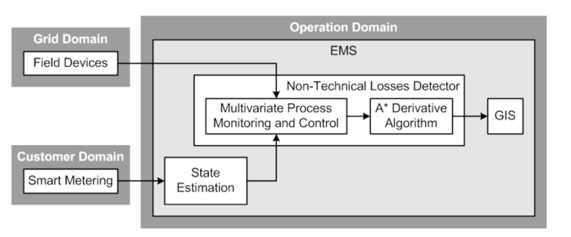
CMRCET B. Tech (CSE) Page No 13

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



**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

CMRCET B. Tech (CSE) Page No 14

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

**2.2.2 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**

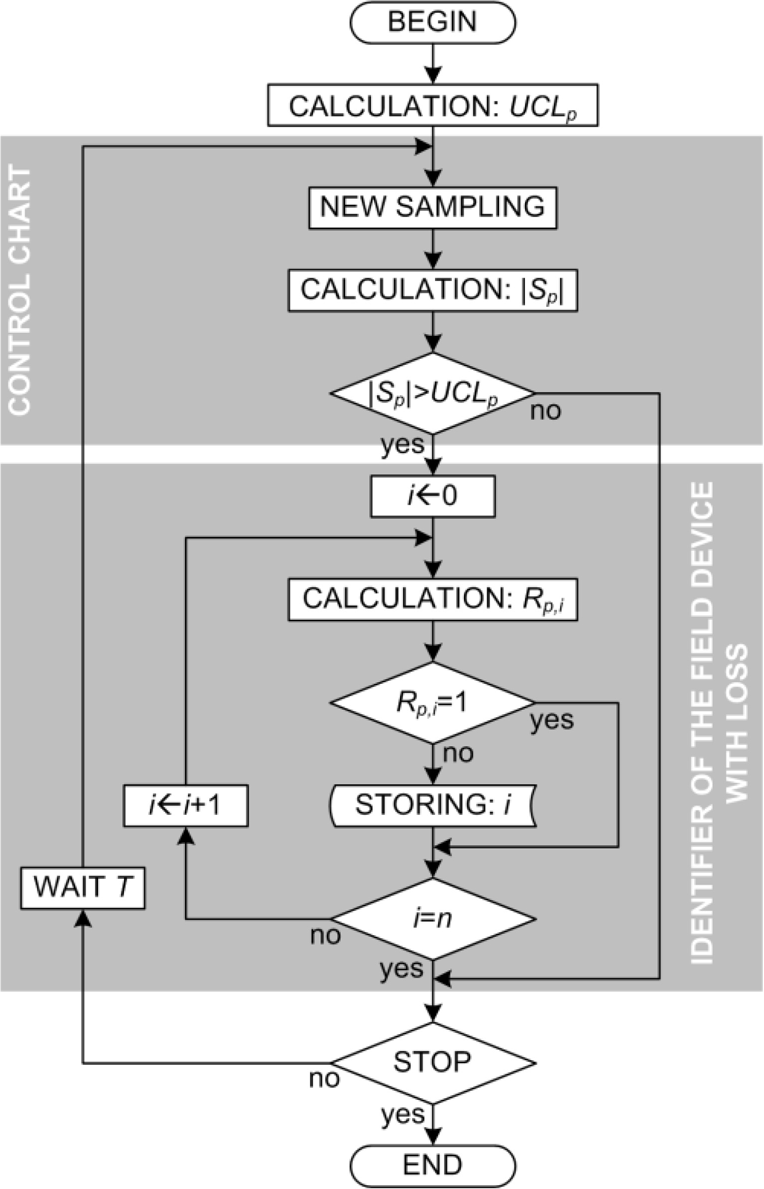
* Require smart meters which are costly and require high maintainance.
* Detection is not guaranteed and required data may not be available.

CMRCET B. Tech (CSE) Page No 15

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

**2.2.3 IMPLEMENTATION**

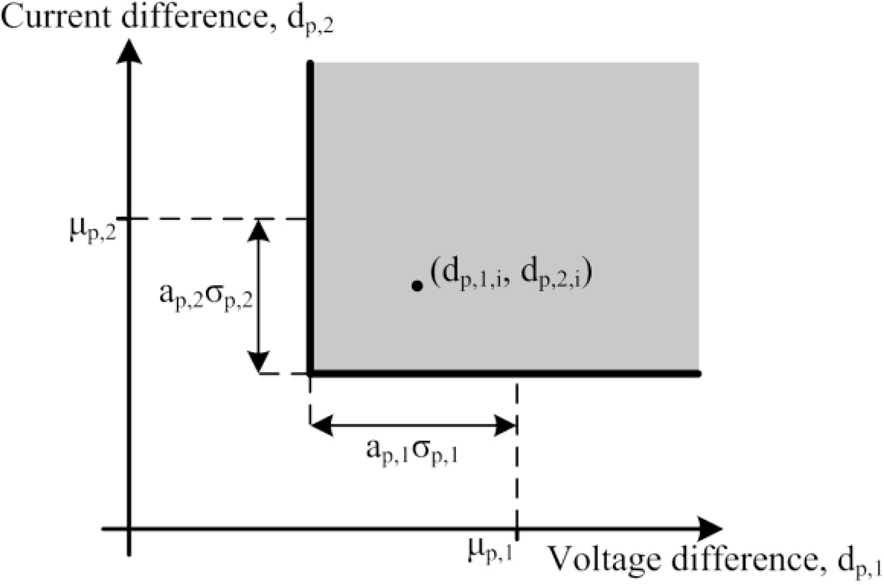
The comparison of network states in the power system has a selectivity problem because large magnitudes are compared to detect small errors. Voltage and current measurements have large magnitudes, whereas small errors result from voltage and current differences. The utilization of the multivariate procedure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



**Fig 2.5 Flow Chart**

CMRCET B. Tech (CSE) Page No 16

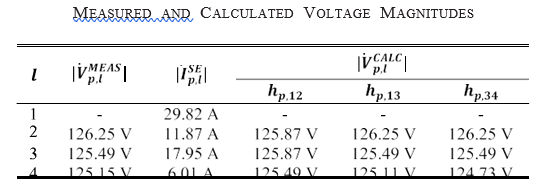
Electricity Theft Detection in Power Grids with Deep Learning and Random Forests



**Figure 2.6** Graph I vs V

Confidence intervals 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](#_bookmark3)) and ([2](#_bookmark3)). The position identification of the point *i (dp,*1*,i, dp,*2*,i)* in the dispersion diagram, i.e., inside or outside of the bounds of the confidence intervals, is achieved by the calculation of the discrete range, *Rp,i*, according to ([12](#_bookmark7))

Table



**Table 2.2 Voltage Magnitudes**

CMRCET B. Tech (CSE) Page No 17

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

highlights three points (*N* 10, 20 and 30) that are 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 of the 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

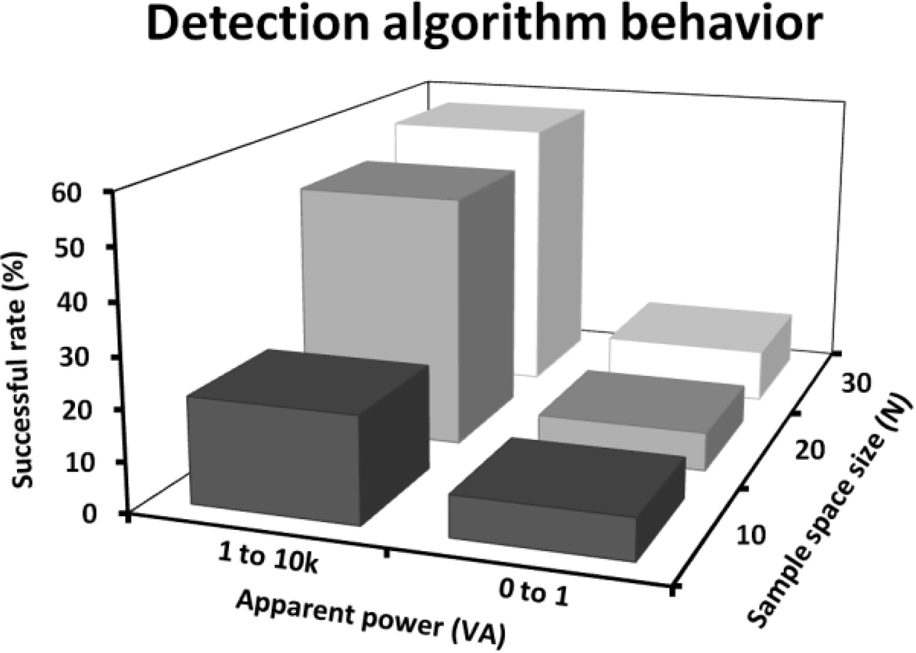


Fig 2.8 Algorithm graph

CMRCET B. Tech (CSE) Page No 18

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

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

**2.3.1 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 between electricity 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 occur due 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 affect the 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 members at home, occurrence of a special occasion, illegal use of electricity, etc. In anomaly detection, data-driven models are used that learn the normal patterns and detect the abnormal patterns to identify the electricity thieves.

CMRCET B. Tech (CSE) Page No 19

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

The LSTM method may require high memory bandwidth to feed its computational units. Ding *et al.* have proposed a hybrid model, which is based on Gausian Mixture Model (GMM) and LSTM for the detection of real time anomaly. However, it is difficult to determine 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 are not 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 imbalanced problem 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 literature for 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, LSTM requires 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.

CMRCET B. Tech (CSE) Page No 20

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

**2.3.2 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
* Anomalydetection 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

CMRCET B. Tech (CSE) Page No 21

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

**2.3.3 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 and ERNET. 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 by selecting 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 random input 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 from data and *z* is the input sample.

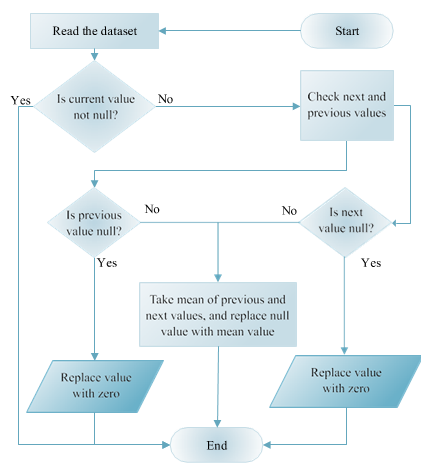


Fig 2.9 Flow chart of data cleaning process.

CMRCET B. Tech (CSE) Page No 22

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

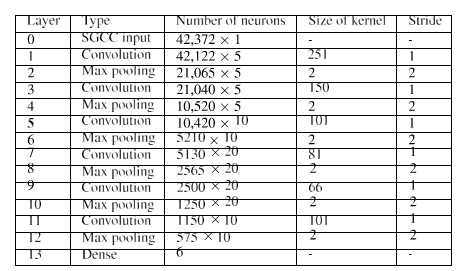


Table 2.3 Network structure

The second model proposed in this work for ETD is com- prised of five stages, as shown in Figure [8](file:///C:\Users\yuvaraj\Downloads\Using_GANCNN_and_ERNET_for_Detection_of_Non_Technical_Losses_to_Secure_Smart_Grids.docx#_bookmark20). The stages are 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 used as 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 normal consumers. A detailed flowchart of the ERNET model is shown in Figure

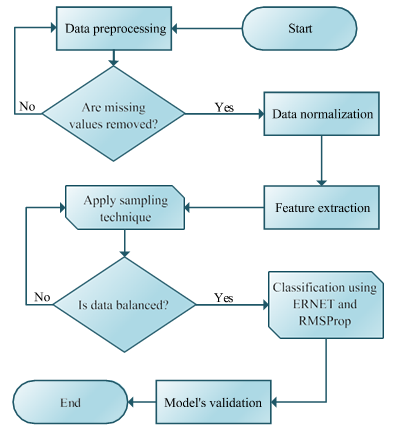
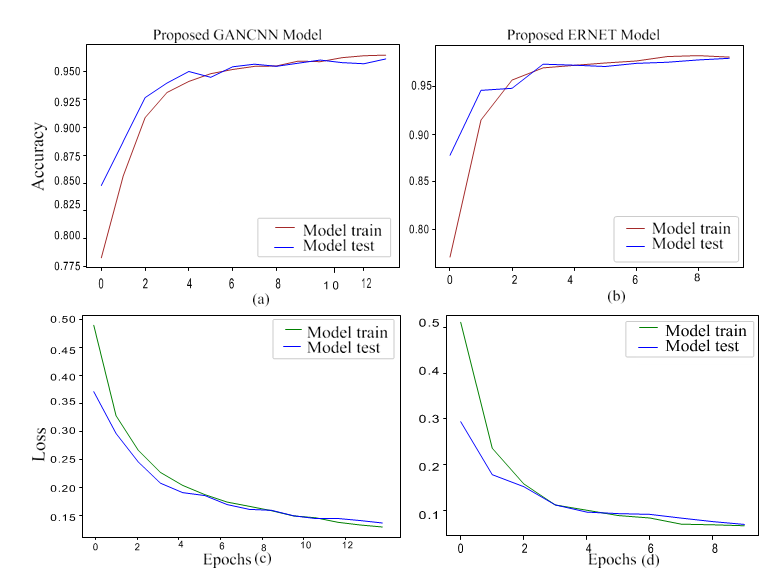


Fig 2.10 Flowchart of ERNET model.

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 to the 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](file:///C:\Users\yuvaraj\Downloads\Using_GANCNN_and_ERNET_for_Detection_of_Non_Technical_Losses_to_Secure_Smart_Grids.docx#_bookmark22). 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 version of Random Oversampling

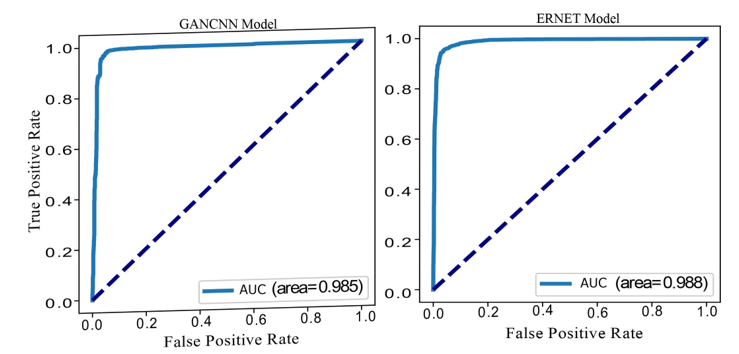
CMRCET B. Tech (CSE) Page No 23

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



2.11 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 the width and depth of the network is to achieve higher prediction accuracy from a large dataset.



2.12 AUC of proposed GANCNN and ERNET models.

CMRCET B. Tech (CSE) Page No 24

**CHAPTER 3**

**PROPOSED SYSTEM**

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

**CHAPTER 3**

**PROPOSED SYSTEM**

**3.1 OBECTIVE OF PROPOSED MODEL**

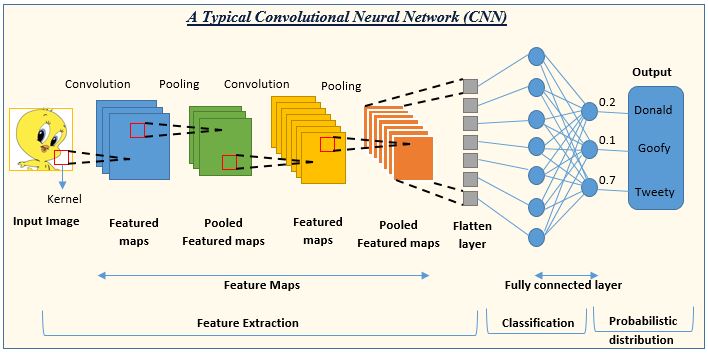
* The detection of electricity theft behaviors is a binary classiﬁcation problem which calls for distinguishing of normal and electricity theft users.
* Electricity theft detection in power grids using deep learning algorithms such as the combination of CNN-RF algorithms.
* In order to help utility companies 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.

**3.2 ALGORITHMS USED FOR PROPOSED MODEL**

**Convolutional Neural Network**

The convolutional layers are the key component of a CNN, where filters are applied to the input image to extract features such as edges, textures, and shapes. The output of the convolutional layers is then passed through pooling layers, which are used to down-sample the feature maps, reducing the spatial dimensions while retaining the most important information. The output of the pooling layers is then passed through one or more fully connected layers, which are used to make a prediction or classify the image.

CNNs are trained using a large dataset of labeled images, where the network learns to recognize patterns and features that are associated with specific objects or classes. Once trained, a CNN can be used to classify new images, or extract features for use in other applications such as object detection or image segmentation.

****

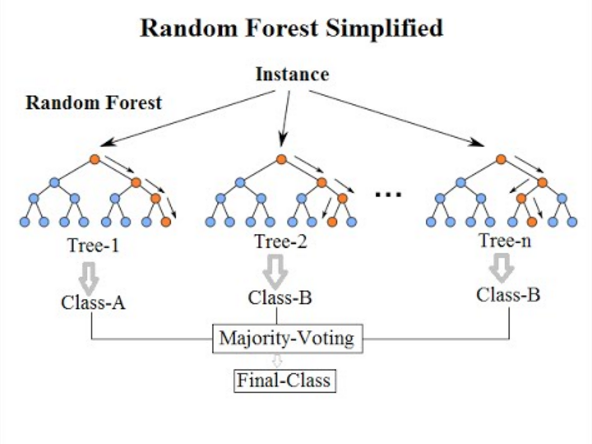
**Figure 3.1** CNN Architecture

CMRCET B. Tech (CSE) Page No 25

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

**Random Forest**

Random Forest is an ensemble technique capable of performing both regression and classification tasks with the use of multiple decision trees and a technique called Bootstrap and Aggregation, commonly known as **bagging**. The basic idea behind this is to combine multiple decision trees in determining the final output rather than relying on individual decision trees. Random Forest has multiple decision trees as base learning models. We randomly perform row sampling and feature sampling from the dataset forming sample datasets for every model.



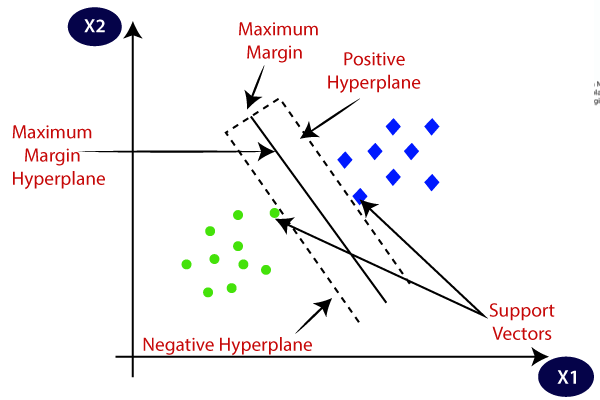
**Figure 3.2** Random Forest Architecture

**Support Vector Machine**

Support Vector Machine (SVM) is a algorithm used for both classification and regression. Though we say regression problems as well it’s best suited for classification. The objective of the SVM algorithm is to find a hyperplane in an N-dimensional space that distinctly classifies the data points. The dimension of the hyperplane depends upon the number of features. If the number of input features is two, then the hyperplane is just a line. If the number of input features is three, then the hyperplane becomes a 2-D plane. It becomes difficult to imagine when the number of features exceeds three.

CMRCET B. Tech (CSE) Page No 26

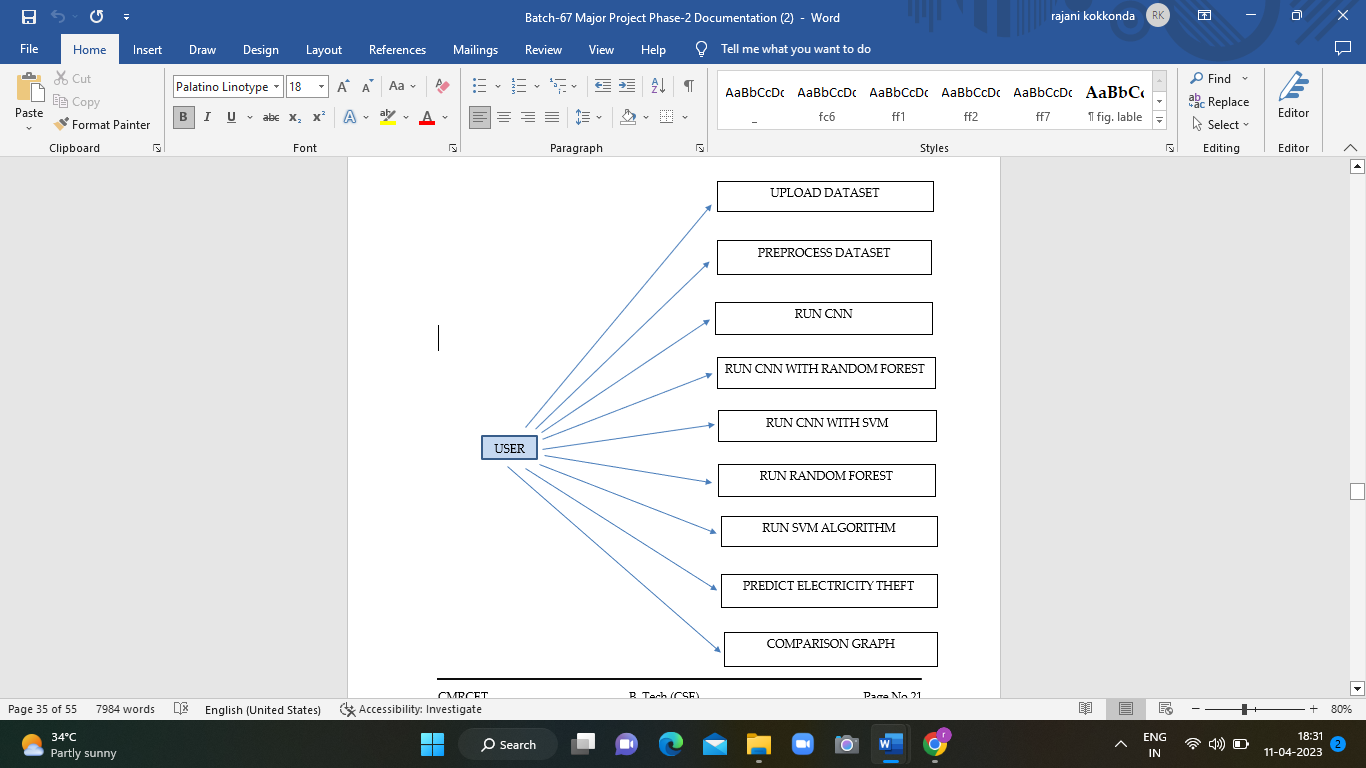
Electricity Theft Detection in Power Grids with Deep Learning and Random Forests



**3.3 DESIGNING**

**3.3.1 UML DIAGRAM**

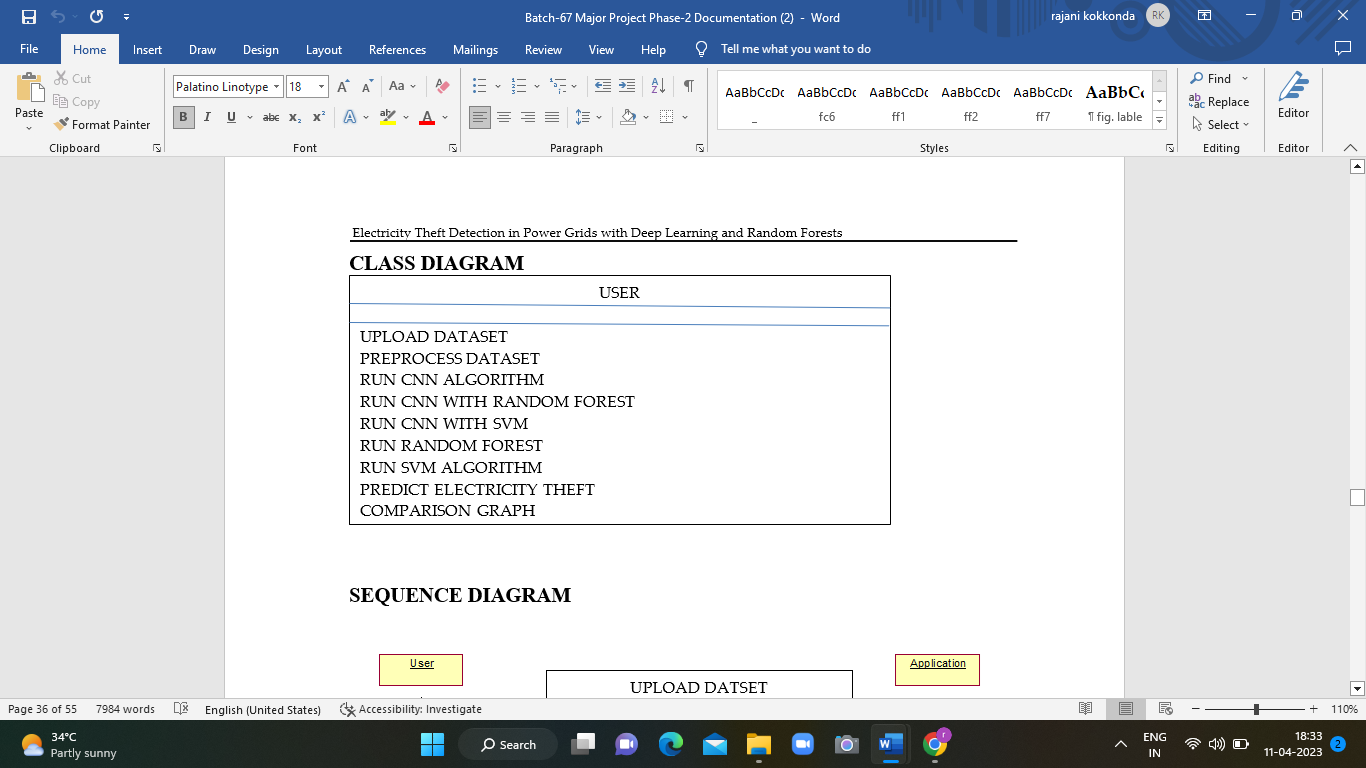
**USE CASE DIAGRAM**

****

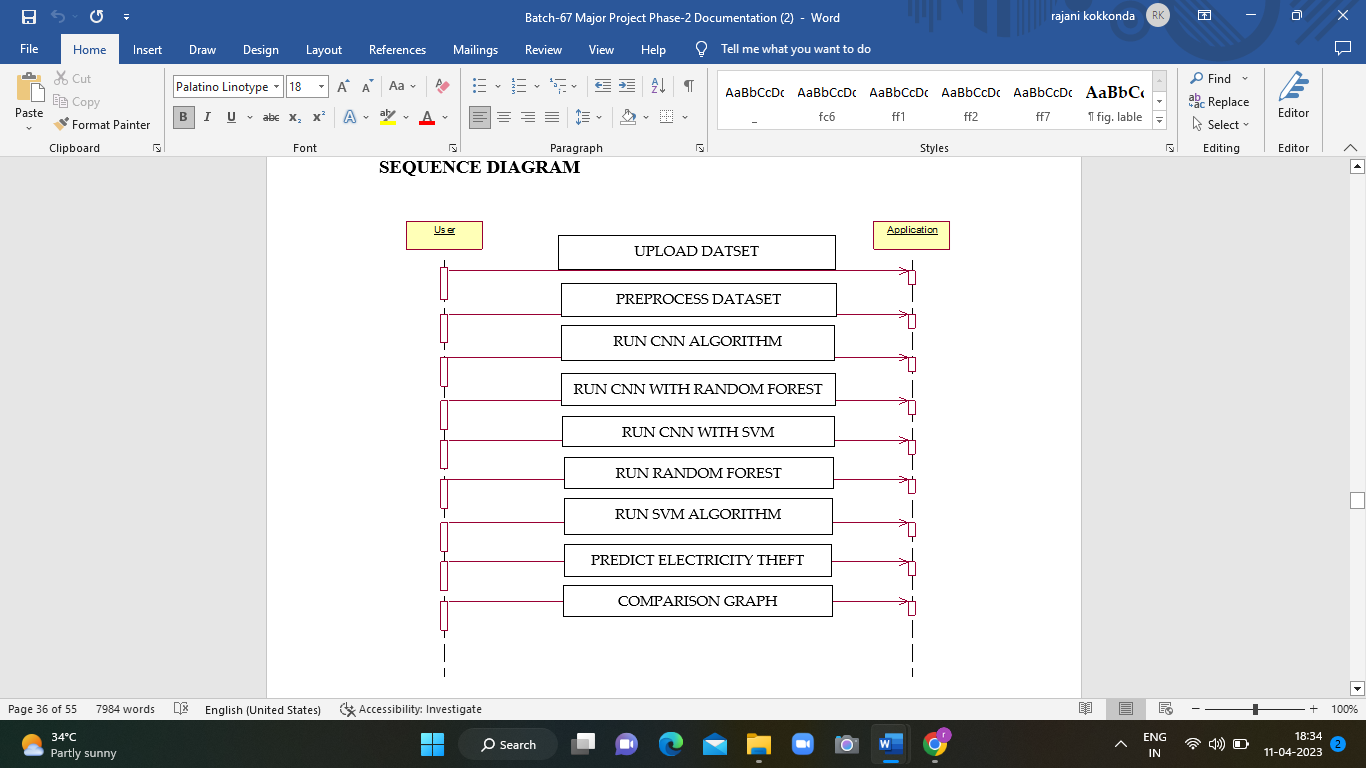
CMRCET B. Tech (CSE) Page No 27

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

**3.2 CLASS DIAGRAM**

****

**3.3 SEQUENCE DIAGRAM**

****

CMRCET B. Tech (CSE) Page No 28

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

**COLLABORATION DIAGRAM**

Upload dataset

Preprocess dataset

Run CNN Algorithm

Run CNN with Random Forest

Run CNN with SVM

Run Random Forest

Run SVM Algorithm

Predict Electricity Theft

Application

User

**Figure 3.3.4** Collaboration Diagram

**3.4 IMPLEMENTATION AND CODE**

from tkinter import \*

import tkinter

from tkinter import filedialog

import numpy as np

from tkinter.filedialog import askopenfilename

import pandas as pd

from tkinter import simpledialog

import pandas as pd

import numpy as np

from sklearn.preprocessing import LabelEncoder

from sklearn.preprocessing import normalize

from keras.models import Sequential, Model

from keras.layers import Dense, Dropout, Activation

from keras.utils.np\_utils import to\_categorical

from keras.models import model\_from\_json

from sklearn.ensemble import RandomForestClassifier

from sklearn.metrics import precision\_score

from sklearn.metrics import recall\_score

from sklearn.metrics import f1\_score

from sklearn.metrics import accuracy\_score

from sklearn.model\_selection import train\_test\_split

from sklearn import svm

import os

import matplotlib.pyplot as plt

CMRCET B. Tech (CSE) Page No 29

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

main = tkinter.Tk()

main.title("Electricity Theft Detection in Power Grids with Deep Learning and Random Forests")

main.geometry("1000x650")

global filename

global cnn\_model

global X, Y

global le

global dataset

accuracy = []

precision = []

recall = []

fscore = []

global classifier

**1)Reading dataset:** using this module reading power consumption dataset

def uploadDataset():

global filename

global dataset

filename = filedialog.askopenfilename(initialdir = "Dataset")

text.delete('1.0', END)

text.insert(END,filename+' Loaded\n')

dataset = pd.read\_csv(filename)

text.insert(END,str(dataset.head())+"\n\n")

**2) Pre process dataset:** using this module we will normalize and clean dataset by removing missing dataset.

def preprocessDataset():

global X, Y

global le

global dataset

le = LabelEncoder()

text.delete('1.0', END)

dataset.fillna(0, inplace = True)

dataset['client\_id'] = pd.Series(le.fit\_transform(dataset['client\_id'].astype(str)))

dataset['label'] = dataset['label'].astype('uint8')

print(dataset.info())

dataset.drop(['creation\_date'], axis = 1,inplace=True)

text.insert(END,str(dataset.head())+"\n\n")

dataset = dataset.values

CMRCET B. Tech (CSE) Page No 30

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

X = dataset[:,0:dataset.shape[1]-1]

Y = dataset[:,dataset.shape[1]-1]

Y = Y.astype('uint8')

indices = np.arange(X.shape[0])

np.random.shuffle(indices)

X = X[indices]

Y = Y[indices]

Y = Y.astype('uint8')

text.insert(END,"Total records found in dataset to train Deep Learning : "+str(X.shape[0])+"\n\n")

**3)Train CNN Model:** using this module we will train CNN with dataset and then will extract trained features from CNN and then input this trained features to random forest algorithm to build theft prediction model. To remove irrelevant features we have added DROPOUT layer.

def runCNN():

global X, Y

text.delete('1.0', END)

global cnn\_model

accuracy.clear()

precision.clear()

recall.clear()

Y1 = to\_categorical(Y)

Y1 = Y1.astype('uint8')

if os.path.exists('model/model.json'):

with open('model/model.json', "r") as json\_file:

loaded\_model\_json = json\_file.read()

cnn\_model = model\_from\_json(loaded\_model\_json)

json\_file.close()

cnn\_model.load\_weights("model/model\_weights.h5")

cnn\_model.\_make\_predict\_function()

print(cnn\_model.summary())

else:

counts = np.bincount(Y1[:, 0])

weight\_for\_0 = 1.0 / counts[0]

weight\_for\_1 = 1.0 / counts[1]

class\_weight = {0: weight\_for\_0, 1: weight\_for\_1}

cnn\_model = Sequential() #creating RNN model object

cnn\_model.add(Dense(256, input\_dim=X.shape[1], activation='relu', kernel\_initializer = "uniform")) #defining one layer with 256 filters to filter dataset

cnn\_model.add(Dense(128, activation='relu', kernel\_initializer = "uniform"))#defining another layer to filter dataset with 128 layers

CMRCET B. Tech (CSE) Page No 31

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

cnn\_model.add(Dense(Y.shape[1], activation='softmax',kernel\_initializer = "uniform")) #after building model need to predict two classes such as normal or Dyslipidemia disease

cnn\_model.compile(loss='categorical\_crossentropy', optimizer='adam', metrics=['accuracy']) #while filtering and training dataset need to display accuracy

print(cnn\_model.summary()) #display rnn details

hist = cnn\_model.fit(X, Y1, epochs=20, batch\_size=64,class\_weight=class\_weight)

cnn\_model.save\_weights('model/model\_weights.h5')

model\_json = cnn\_model.to\_json()

with open("model/model.json", "w") as json\_file:

json\_file.write(model\_json)

json\_file.close()

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, Y1, test\_size=0.2, random\_state=0)

y\_test = np.argmax(y\_test, axis=1)

predict = cnn\_model.predict(X\_test)

predict = np.argmax(predict, axis=1)

p = precision\_score(y\_test, predict,average='macro') \* 100

r = recall\_score(y\_test, predict,average='macro') \* 100

f = f1\_score(y\_test, predict,average='macro') \* 100

a = accuracy\_score(y\_test,predict)\*100

accuracy.append(a)

precision.append(p)

recall.append(r)

fscore.append(f)

text.insert(END,"CNN Precision : "+str(p)+"\n")

text.insert(END,"CNN FMeasure : "+str(f)+"\n")

text.insert(END,"CNN Accuracy : "+str(f)+"\n\n")

**4)Train CNN with Random Forest:** using this module will train random forest with CNN features and then calculate precision, recall, FSCORE and accuracy.

def runCNNRF():

global classifier

global X, Y

global cnn\_model

predict = cnn\_model.predict(X)

YY = []

for i in range(len(predict)):

val = np.argmax(predict[i])

YY.append(val)

YY = np.asarray(YY)

extract = Model(cnn\_model.inputs, cnn\_model.layers[-2].output)

XX = extract.predict(X)

CMRCET B. Tech (CSE) Page No 32

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

rfc = RandomForestClassifier(n\_estimators=200, random\_state=0)

rfc.fit(XX, YY)

classifier = rfc

X\_train, X\_test, y\_train, y\_test = train\_test\_split(XX, YY, test\_size=0.2, random\_state=0)

predict = rfc.predict(X\_test)

p = precision\_score(y\_test, predict,average='macro') \* 100

r = recall\_score(y\_test, predict,average='macro') \* 100

f = f1\_score(y\_test, predict,average='macro') \* 100

a = accuracy\_score(y\_test,predict)\*100

accuracy.append(a)

precision.append(p)

recall.append(r)

fscore.append(f)

text.insert(END,"CNN with Random Forest Precision : "+str(p)+"\n")

text.insert(END,"CNN with Random Forest Recall : "+str(r)+"\n")

text.insert(END,"CNN with Random Forest FMeasure : "+str(f)+"\n")

text.insert(END,"CNN with Random Forest Accuracy : "+str(f)+"\n\n")

**5)Train CNN with SVM:** using this module will train SVM with CNN features and then calculate precision, recall, FSCORE and accuracy.

def runCNNSVM():

global X, Y

global cnn\_model

predict = cnn\_model.predict(X)

YY = []

for i in range(len(predict)):

val = np.argmax(predict[i])

YY.append(val)

YY = np.asarray(YY)

extract = Model(cnn\_model.inputs, cnn\_model.layers[-2].output)

XX = extract.predict(X)

rfc = svm.SVC()

rfc.fit(XX, YY)

X\_train, X\_test, y\_train, y\_test = train\_test\_split(XX, YY, test\_size=0.2, random\_state=0)

predict = rfc.predict(X\_test)

p = precision\_score(y\_test, predict,average='macro') \* 100

r = recall\_score(y\_test, predict,average='macro') \* 100

f = f1\_score(y\_test, predict,average='macro') \* 100

a = accuracy\_score(y\_test,predict)\*100

CMRCET B. Tech (CSE) Page No 33

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

accuracy.append(a)

precision.append(p)

recall.append(r)

fscore.append(f)

text.insert(END,"CNN with SVM Precision : "+str(p)+"\n")

text.insert(END,"CNN with SVM Recall : "+str(r)+"\n")

text.insert(END,"CNN with SVM FMeasure : "+str(f)+"\n")

text.insert(END,"CNN with SVM Accuracy : "+str(f)+"\n\n")

**6)Train Random Forest without CNN:** Here we trained random forest on normal dataset without using CNN features and then calculate precision, recall, FSCORE and accuracy.

def runRandomForest():

global X, Y

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, Y, test\_size=0.2, random\_state=0)

rfc = RandomForestClassifier(n\_estimators=200, random\_state=0)

rfc.fit(X\_train, y\_train)

predict = rfc.predict(X\_test)

p = precision\_score(y\_test, predict,average='macro') \* 100

r = recall\_score(y\_test, predict,average='macro') \* 100

f = f1\_score(y\_test, predict,average='macro') \* 100

a = accuracy\_score(y\_test,predict)\*100

accuracy.append(a)

precision.append(p)

recall.append(r)

fscore.append(f)

text.insert(END,"Random Forest Precision : "+str(p)+"\n")

text.insert(END,"Random Forest Recall : "+str(r)+"\n")

text.insert(END,"Random Forest FMeasure : "+str(f)+"\n")

text.insert(END,"Random Forest Accuracy : "+str(f)+"\n\n")

**7)Train SVM model:** Here we trained SVM on normal dataset and then calculate precision, recall, FSCORE and accuracy.

def runSVM():

global X, Y

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, Y, test\_size=0.2, random\_state=0)

rfc = svm.SVC()

rfc.fit(X\_train, y\_train)

predict = rfc.predict(X\_test)

p = precision\_score(y\_test, predict,average='macro') \* 10

r = recall\_score(y\_test, predict,average='macro') \* 100

f = f1\_score(y\_test, predict,average='macro') \* 100

a = accuracy\_score(y\_test,predict)\*100

CMRCET B. Tech (CSE) Page No 34

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

accuracy.append(a)

precision.append(p)

recall.append(r)

fscore.append(f)

text.insert(END,"SVM Precision : "+str(p)+"\n")

text.insert(END,"SVM Recall : "+str(r)+"\n")

text.insert(END,"SVM FMeasure : "+str(f)+"\n")

text.insert(END,"SVM Accuracy : "+str(f)+"\n\n")

**8)Predict Electricity Theft:**

def predict():

global classifier

global cnn\_model

text.delete('1.0', END)

filename = filedialog.askopenfilename(initialdir = "Dataset")

test = pd.read\_csv(filename)

test.fillna(0, inplace = True)

test = test.values

data = test

extract = Model(cnn\_model.inputs, cnn\_model.layers[-2].output)

test = extract.predict(test)

predict = classifier.predict(test)

for i in range(len(predict)):

if predict[i] == 1:

text.insert(END,str(data[i])+" ===> record detected as ENERGY THEFT\n\n")

if predict[i] == 0:

text.insert(END,str(data[i])+" ===> record NOT detected as ENERGY THEFT\n\n")

**9)Comparison Graph**

def graph():

df = pd.DataFrame([['CNN','Precision',precision[0]],['CNN','Recall',recall[0]],['CNN','F1 Score',fscore[0]],['CNN','Accuracy',accuracy[0]],

['CNN-RF','Precision',precision[1]],['CNN-RF','Recall',recall[1]],['CNN-RF','F1 Score',fscore[1]],['CNN-RF','Accuracy',accuracy[1]],

['CNN-SVM','Precision',precision[2]],['CNN-SVM','Recall',recall[2]],['CNN-SVM','F1 Score',fscore[2]],['CNN-SVM','Accuracy',accuracy[2]],

['RF','Precision',precision[3]],['RF','Recall',recall[3]],['RF','F1 Score',fscore[3]],['RF','Accuracy',accuracy[3]],

['SVM','Precision',precision[3]],['SVM','Recall',recall[3]],['SVM','F1 Score',fscore[3]],['SVM','Accuracy',accuracy[3]],

],columns=['Parameters','Algorithms','Value'])

CMRCET B. Tech (CSE) Page No 35

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

df.pivot("Parameters", "Algorithms", "Value").plot(kind='bar')

plt.show()

font = ('times', 16, 'bold')

title = Label(main, text='Electricity Theft Detection in Power Grids with Deep Learning and Random Forests', justify=LEFT)

title.config(bg='lavender blush', fg='DarkOrchid1')

title.config(font=font)

title.config(height=3, width=120)

title.place(x=100,y=5)

title.pack()

font1 = ('times', 13, 'bold')

uploadButton = Button(main, text="Upload Electricity Theft Dataset", command=uploadDataset)

uploadButton.place(x=200,y=100)

uploadButton.config(font=font1)

preprocessButton = Button(main, text="Preprocess Dataset", command=preprocessDataset)

preprocessButton.place(x=500,y=100)

preprocessButton.config(font=font1)

cnnButton = Button(main, text="Generate CNN Model", command=runCNN)

cnnButton.place(x=200,y=150)

cnnButton.config(font=font1)

cnnrfButton = Button(main, text="CNN with Random Forest", command=runCNNRF)

cnnrfButton.place(x=500,y=150)

cnnrfButton.config(font=font1)

cnnsvmButton = Button(main, text="CNN with SVM", command=runCNNSVM)

cnnsvmButton.place(x=200,y=200)

cnnsvmButton.config(font=font1)

rfButton = Button(main, text="Run Random Forest", command=runRandomForest)

rfButton.place(x=500,y=200)

rfButton.config(font=font1)

svmButton = Button(main, text="Run SVM Algorithm", command=runSVM)

svmButton.place(x=200,y=250)

svmButton.config(font=font1)

CMRCET B. Tech (CSE) Page No 36

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

predictButton = Button(main, text="Predict Electricity Theft", command=predict)

predictButton.place(x=500,y=250)

predictButton.config(font=font1)

graphButton = Button(main, text="Comparison Graph", command=graph)

graphButton.place(x=800,y=250)

graphButton.config(font=font1)

font1 = ('times', 12, 'bold')

text=Text(main,height=20,width=120)

scroll=Scrollbar(text)

text.configure(yscrollcommand=scroll.set)

text.place(x=10,y=300)

text.config(font=font1)

main.config(bg='light coral')

main.mainloop()

CMRCET B. Tech (CSE) Page No 37

CHAPTER 4

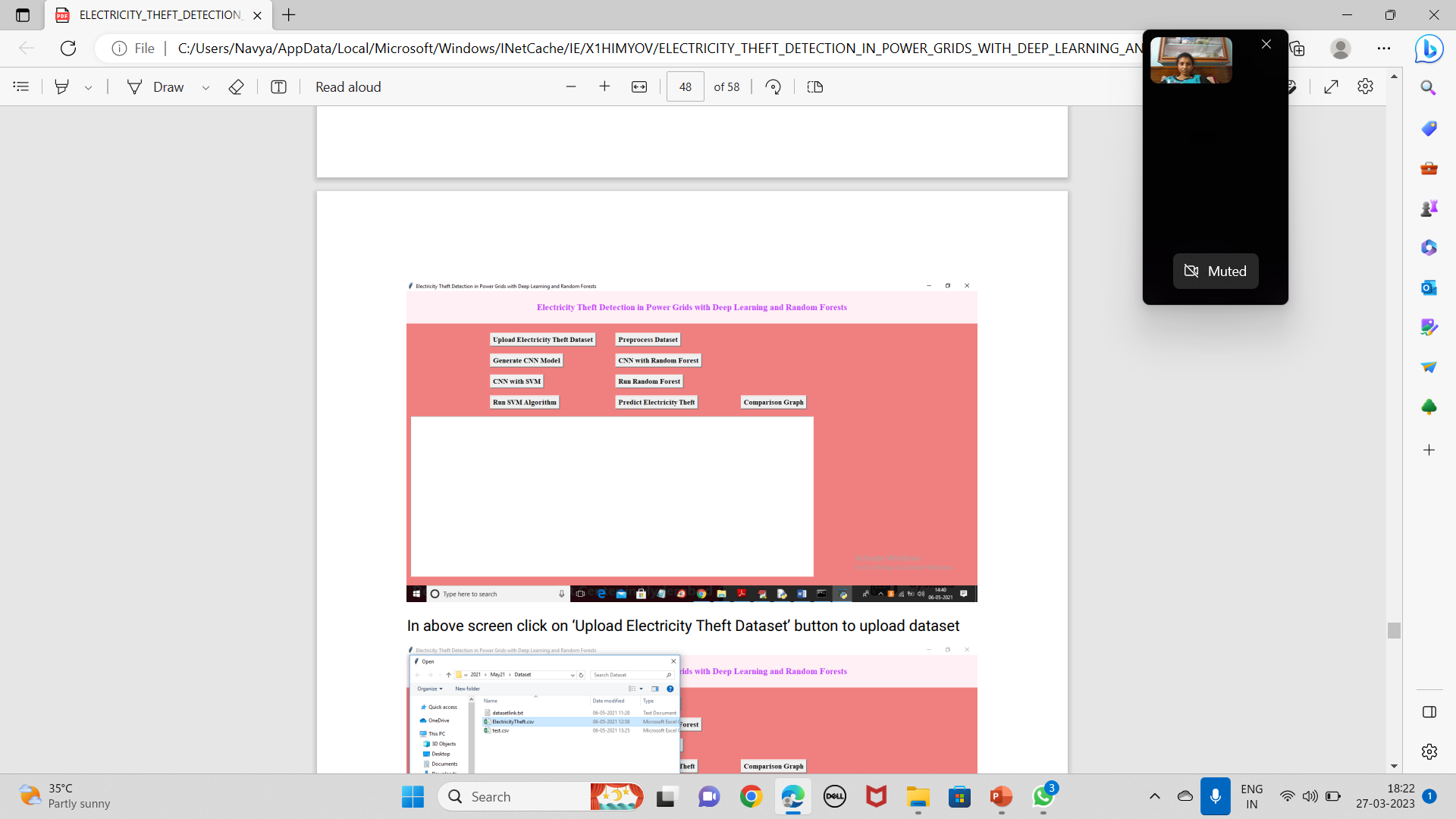
RESULTS AND DISCUSSION

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

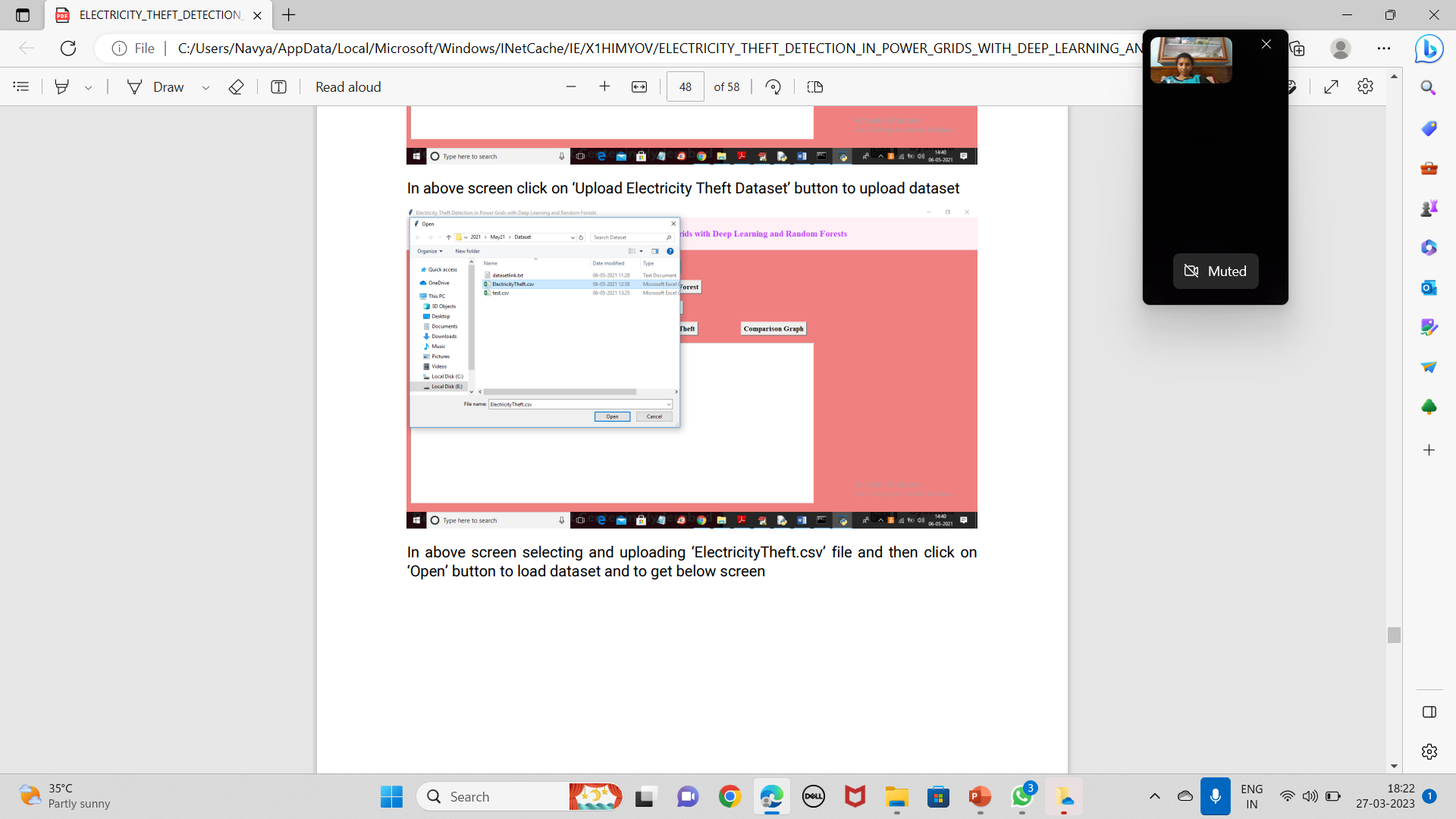
**CHAPTER 4**

**RESULTS AND DISCUSSION**

**4.1 PERFORMANCE METRICS**



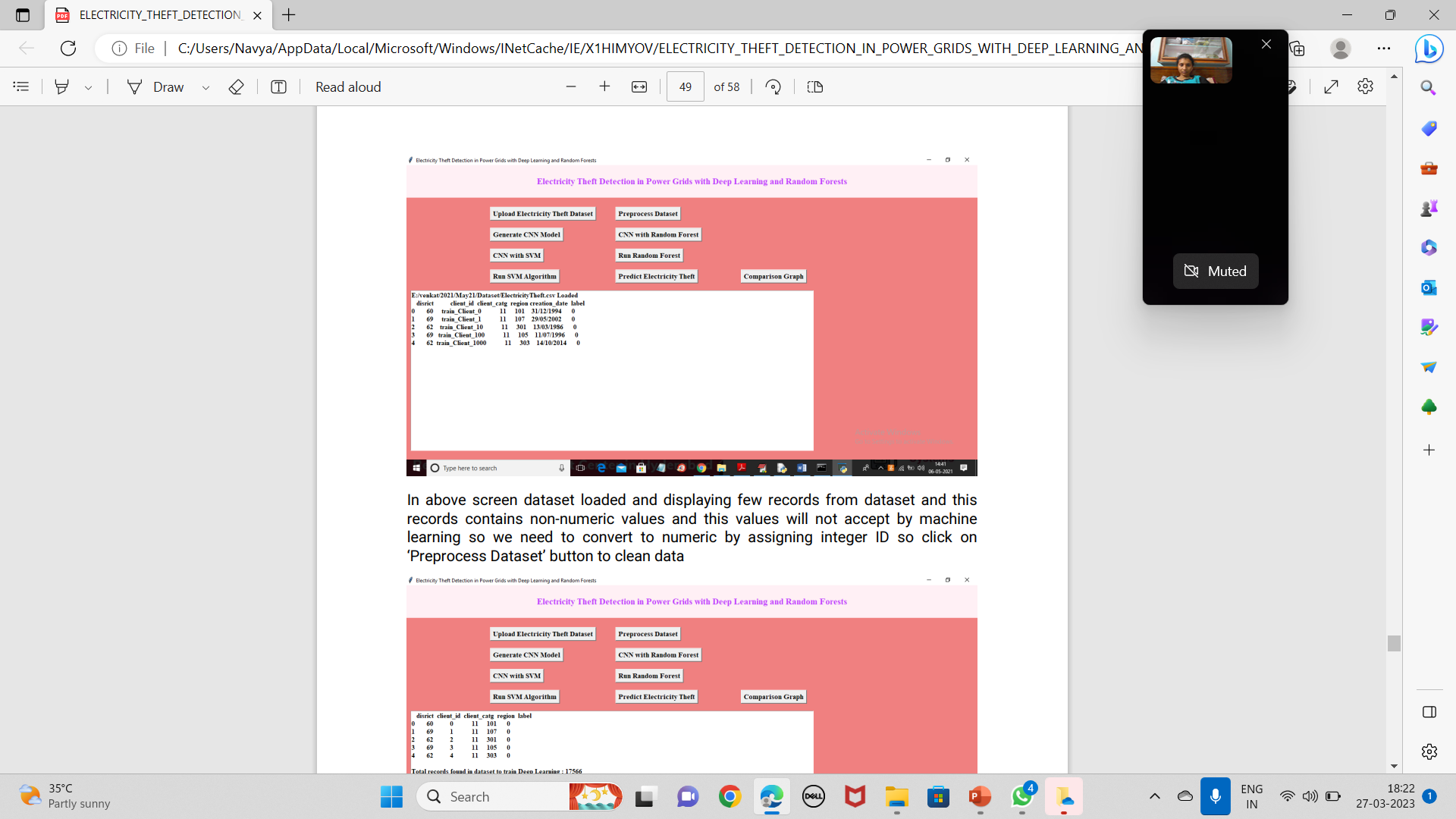
**Figure 4.1** Output



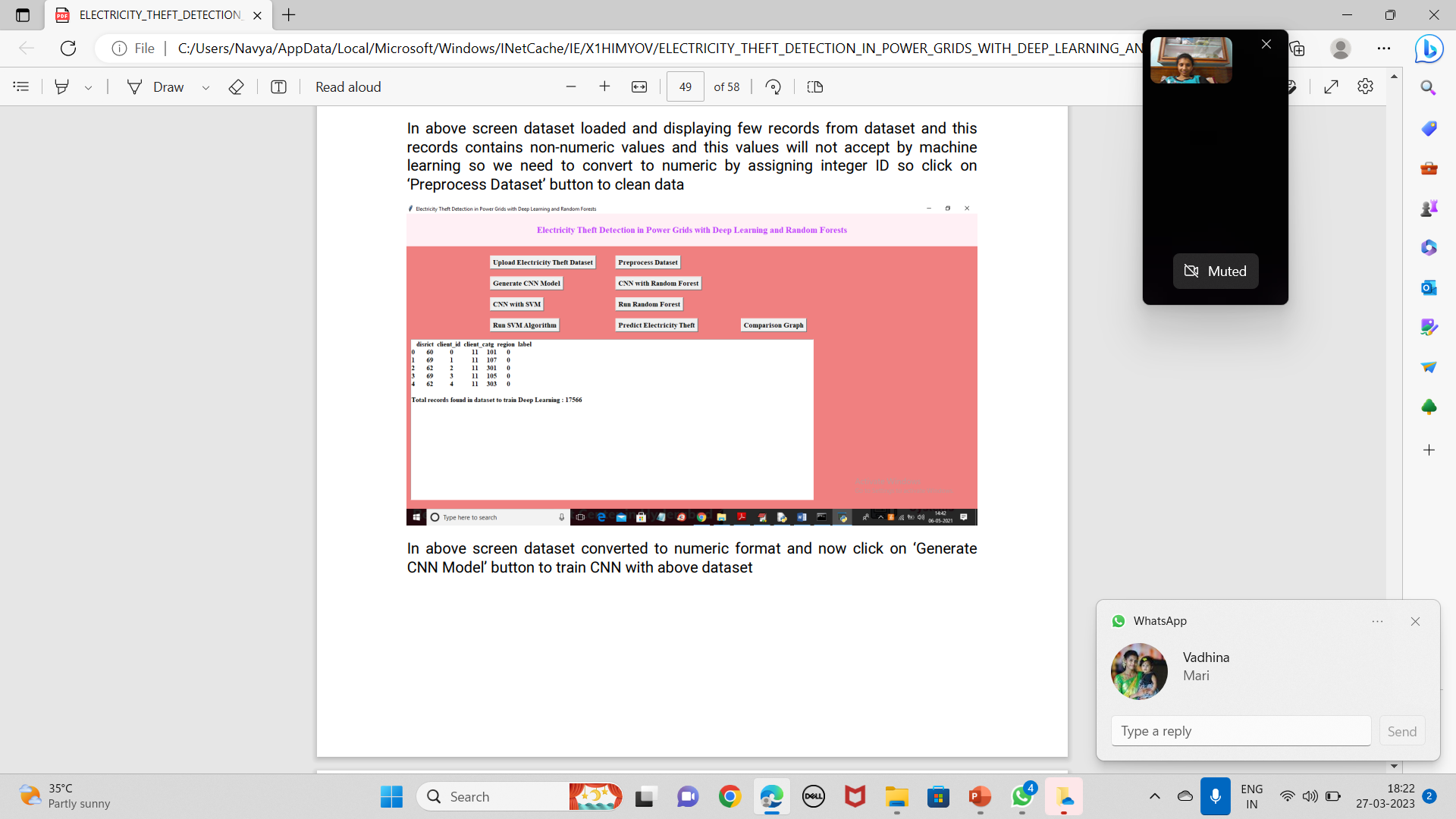
**Figure 4.2** Uploading Electricity Theft Dataset

CMRCET B. Tech (CSE) Page No 38

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



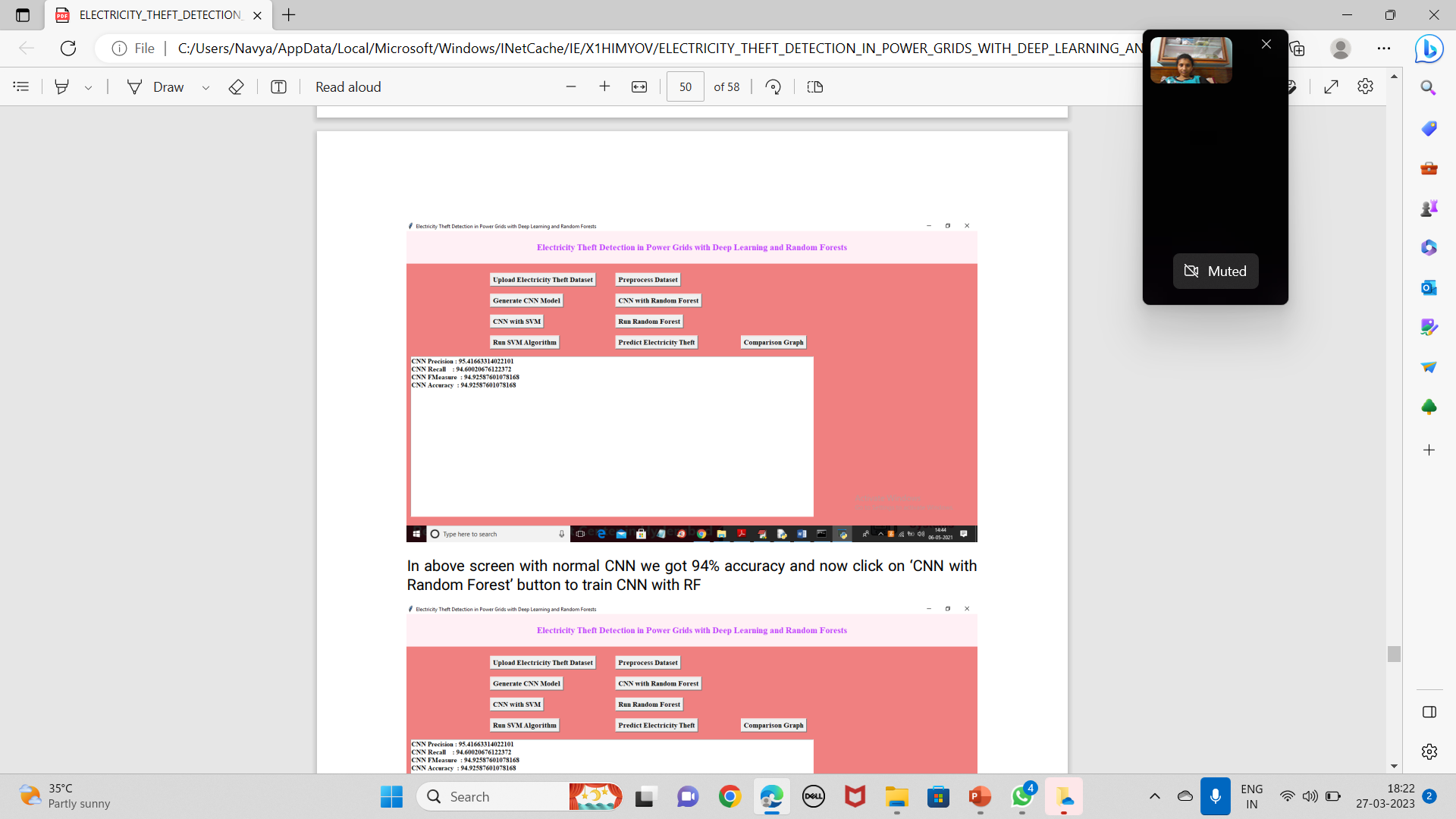
**Figure 4.3** Preprocessing Dataset



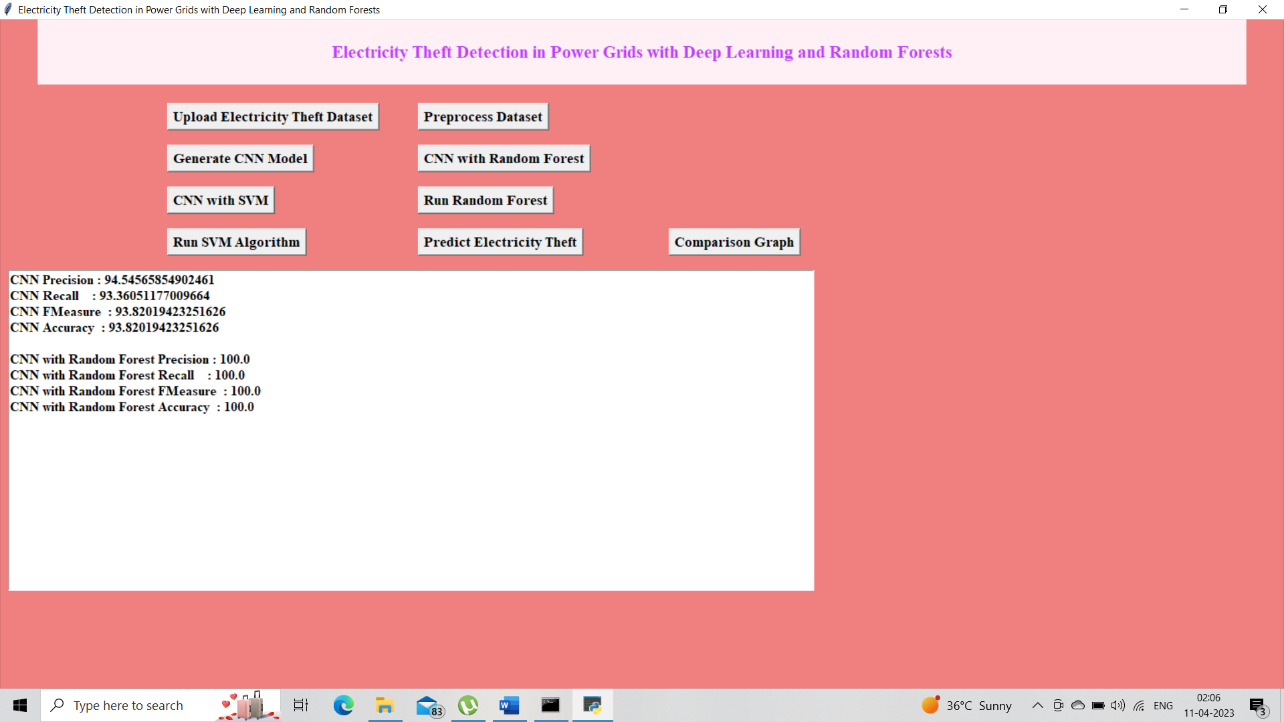
**Figure 4.4** Generating CNN Model

CMRCET B. Tech (CSE) Page No 39

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



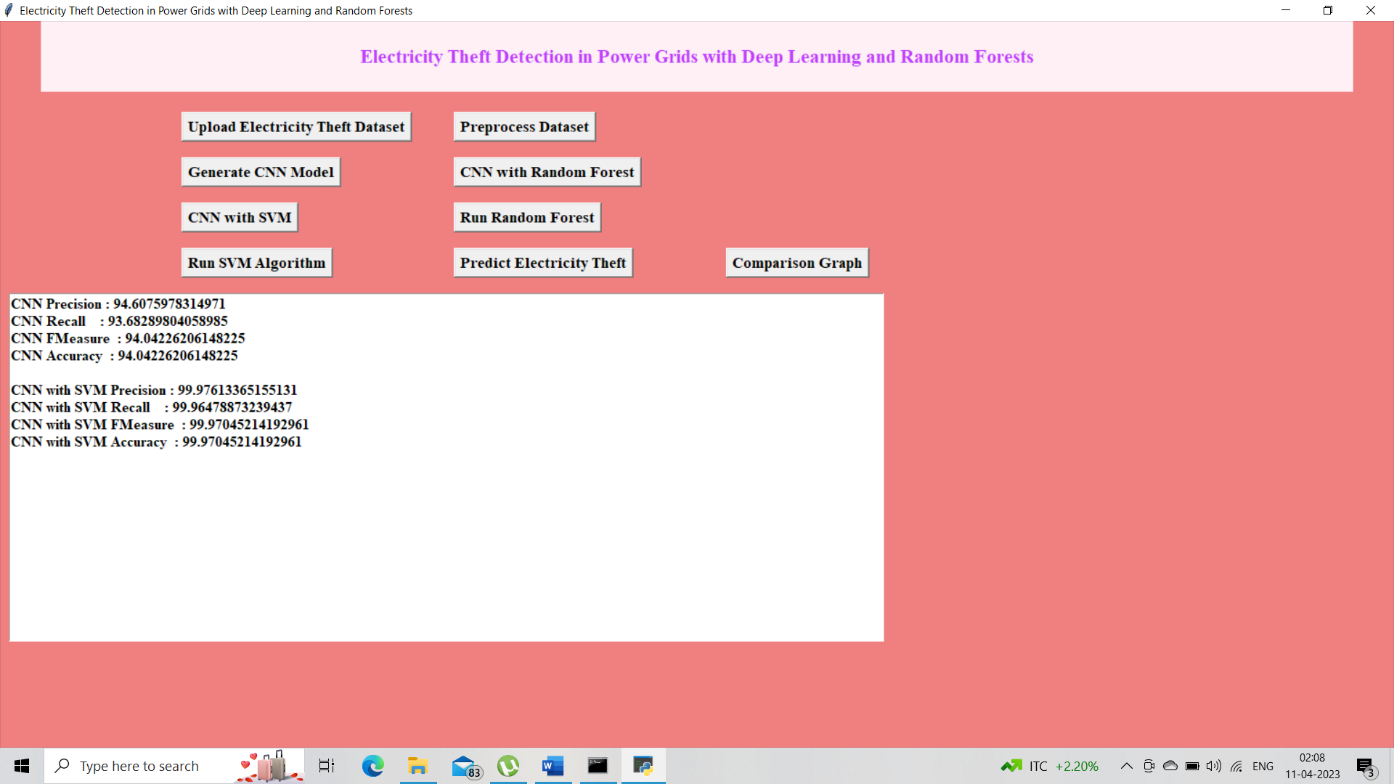
**Figure 4.5** Result Of CNN



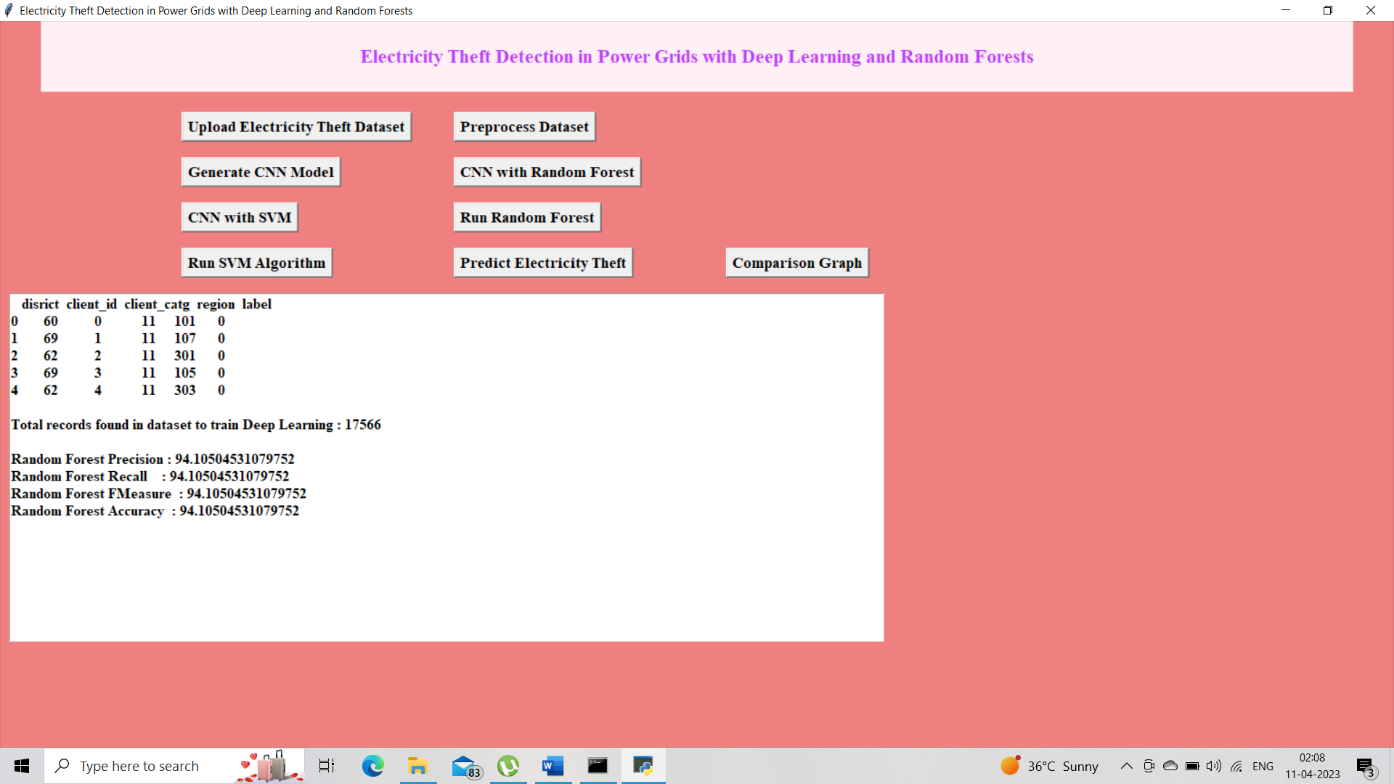
**Figure 4.6** Result of CNN With Random Forest

CMRCET B. Tech (CSE) Page No 40

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



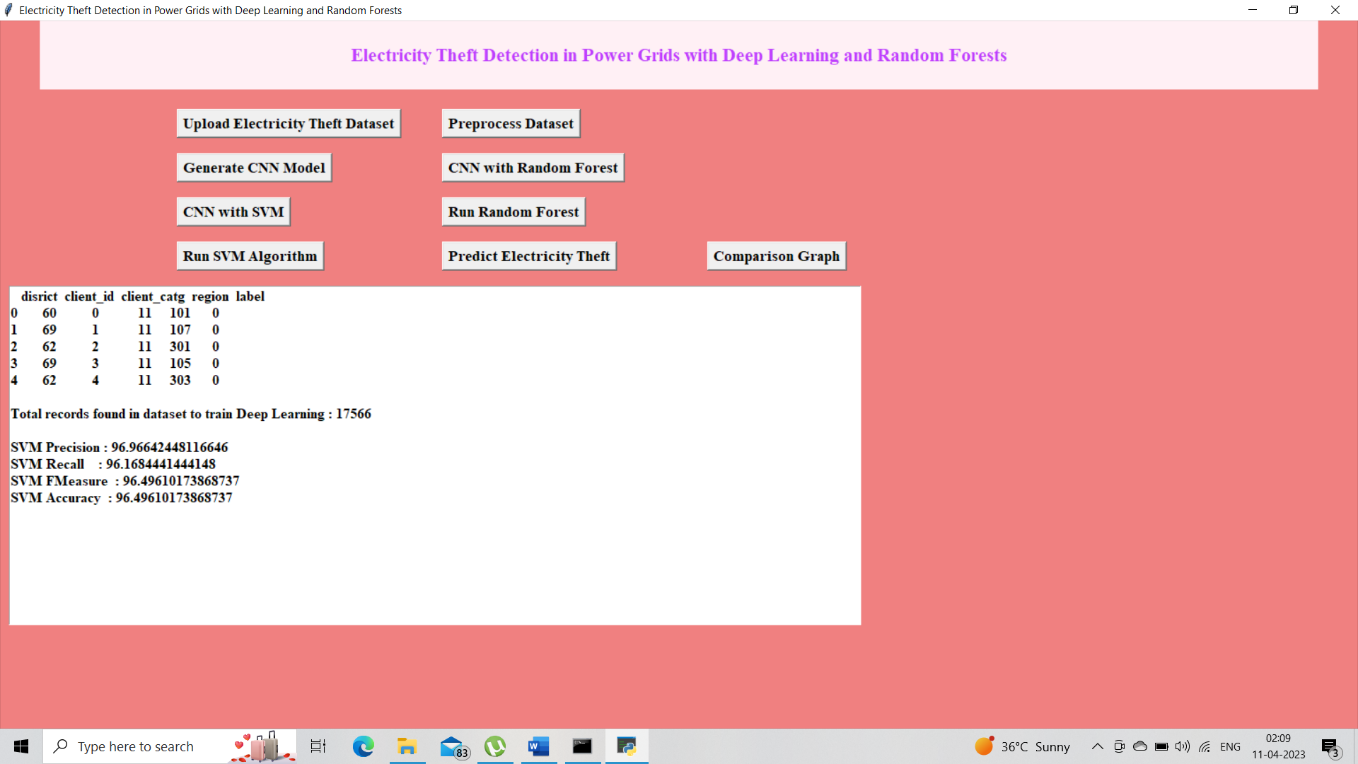
**Figure 4.7** Result of CNN With SVM



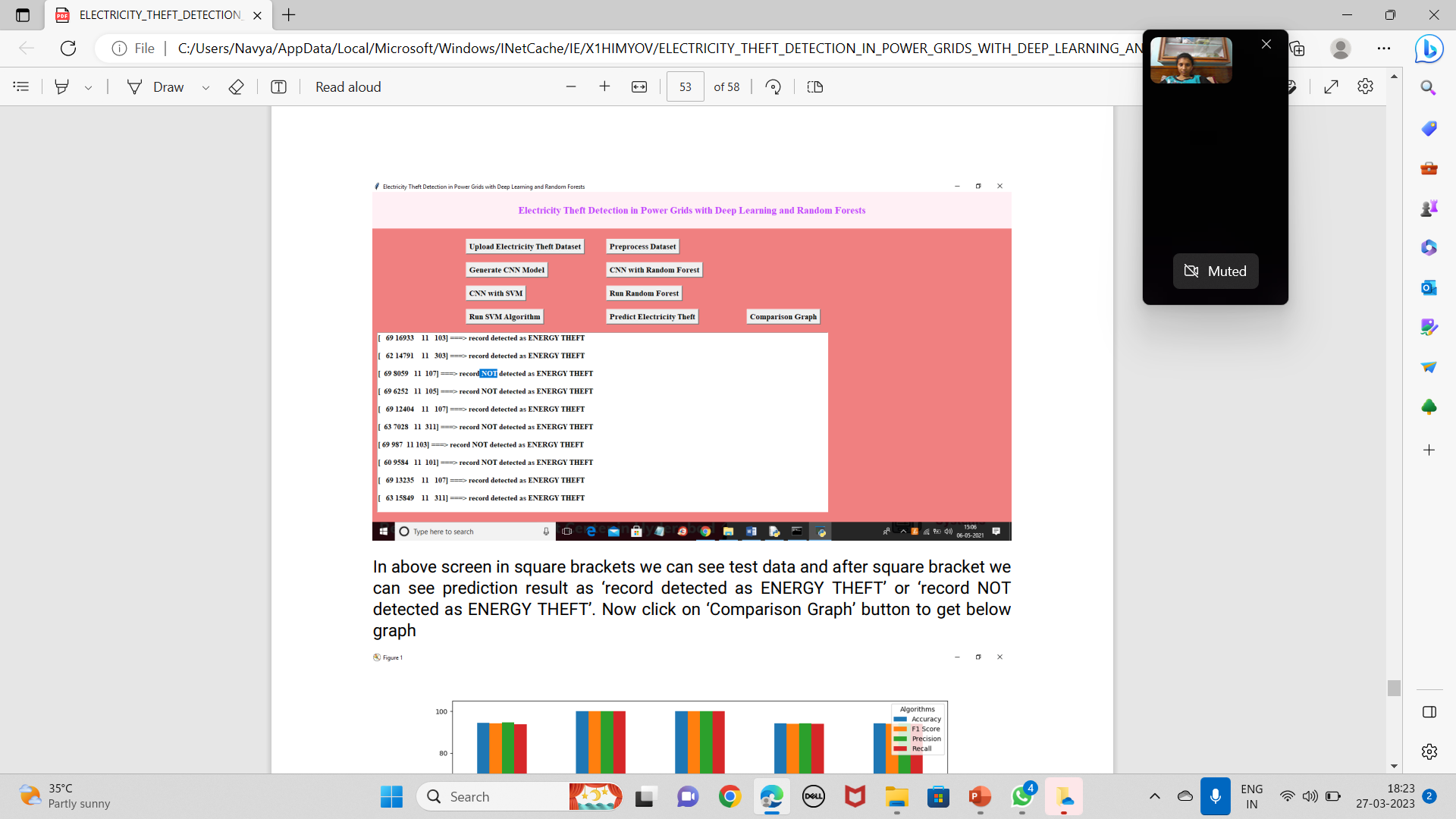
**Figure 4.8** Result of Random Forest

CMRCET B. Tech (CSE) Page No 41

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



**Figure 4.9** Result of SVM

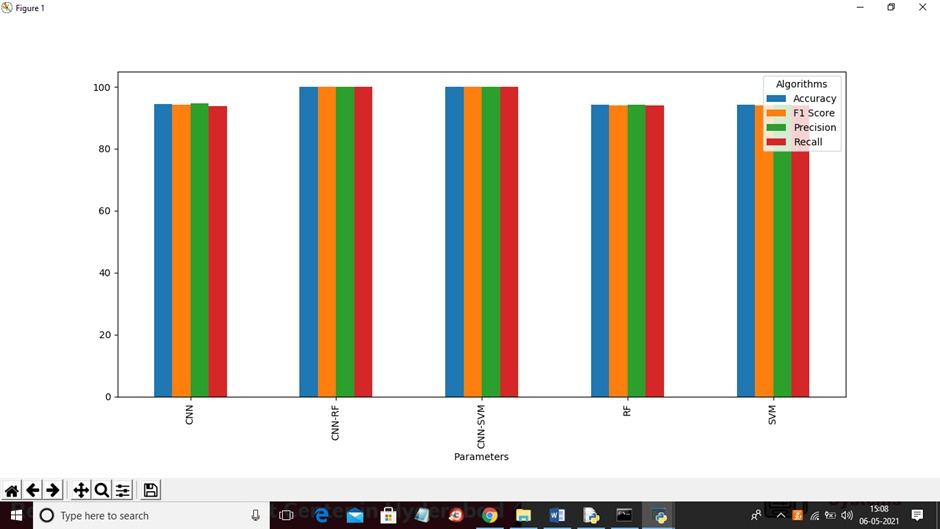


**Figure 4.10** Result of Predict Electricity Theft

CMRCET B. Tech (CSE) Page No 42

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

**Comparison Graph**



**Figure 4.11** Comparison Graph

CMRCET B. Tech (CSE) Page No 43

CHAPTER 5

**CONCLUSION**

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

**CHAPTER 5**

**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.

CMRCET B. Tech (CSE) Page No 44

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

**FUTURE ENHANCEMENT:**

* This project 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.

CMRCET B. Tech (CSE) Page No 45

**CHAPTER 6**

**REFERENCES**

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

**CHAPTER 6**

## 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:%20overview,%20issues,%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 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:%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 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,%20text%20mining,%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:%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)

CMRCET B. Tech (CSE) Page No 46

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

1. 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 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)
2. 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)
3. 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)
4. 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 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)

CMRCET B. Tech (CSE) Page No 47

GIT HUB LINK :-