Power Grid Classification Through Electrical Network Frequency

Fall, 2017

Thesis submitted to,
The Department of Electrical and Electronics Engineering of BRAC University in partial fulfillment of Bachelor of Science Degree

Letter of Transmittal

24th December,2017
To,
Shoili Chakma,
Department of Electrical and Electronics Engineering,
Bangladesh University of Engineering and Technology.

Subject: Submission of our Thesis report to completing our graduation.

Dear Madam,

With high reverence I want to state that we have finished our thesis report on the **Power Grid Classification Through Electrical Network Frequency**. We put all description and details with elaborated research to compile our report in comprehensive manner. We faced some challenges in working as a team and to put people's perception. I would like to take this platform to thank you.

To end, I would like you to kindly accept our thesis report and to acknowledge our devotion and efforts.

Your sincerely,
MD.Arif Uz Zaman
On behalf of
Tasnim Mortoza
Zawad Hasan Abid
MD. Abu Musa

_____
Shoili Chakma(Supervisor)

_____
Dr. Hasanuzzaman Sagor(Co-Supervisor)

DECLARATION

We hereby declare that the thesis titled **Power Grid Classification Through Electrical Network Frequency** is our own work. The report is based on the result that we have done in our thesis. Content that we used from other research is mentioned clearly. This thesis has never been published or submitted anywhere else for any degree.

Acknowledgment

First and foremost we would like to express our deepest gratitude to Almighty Allah for always having made us feel so blessed. We also like to take the time to show our respect to our supervisor and co-supervisor for giving us the opportunity to do this research under their observation. We would also like to give special thanks to our supervisor **Shoili Chakma** madam for her contribution in our work to fulfill the research.

Abstract

A general review of certain techniques for power grid analysis and power signatures detection is presented in this paper. After some motivation and research on the existing trends on solving this issue of supply network system problems, particular techniques are described with proper analysis. Identifying such issues through a series of methodological steps, a solution can be predicted for designing better grids for a futuristic system. Such identifications can be done through software analysis of the power grid data. The main objective of this paper is the chronological overview and analysis of audio signals received from certain grid machines that can be utilized to detect errors or irregularities through pattern recognition technique and exploring feature detection algorithm. Performing a software analysis of electrical network frequency extraction, pattern recognition and accuracy measurement, certain information can be obtained. These information can be compared and matched in different ratios and percentages to get better accuracy of results. Finally, a solution for the existing issue can be predicted based on the analyzed results. The techniques for power grid analysis and detection followed in this research can be very useful for a number of other software based research works of similar interests.

Contents

List of Figures

List of Tables

1 INTRODUCTION

1.1 LITERATURE BACKGROUND

Detecting grid and analyzing data is a major work in recent power engineering. Being able to detect the power grid, researchers have opened a new field in data science of power engineering. Now-a-days we have big power plants in every country around the globe. These plants have many grids and these grids continuously supplying power to the every corner of the earth. But as like all the other fields these grids are also error prone.

For this the engineers are working day and night to detect and analysis the problem that might damage a grid. In our paper we are going to describe in details how anyone can detect the power grid if they know the electrical network frequency (ENF) (ENF will be discussed thoroughly in the next part of the paper). Before going to the next part let's study the research on similar topic that has been done by others.

In the paper titled "Electrical Network Frequency (ENF) maximum likelihood estimation via a multitoned Harmonic model" written by D. Bykhovsky and A. Coher they vividly described the possibilities and some ways to detect ENF. They found by the way the ENF is collected it add some noise in to it. Thus make the SNR low (SNR is the ratio between the signal and noise, SNR low means Noise is high). This leads to a problem. They then introduced harmonic signal model to reduce the noise and detect the ENF more precisely. They used several techniques to optimize the error and showed their estimated result. In another paper by Adi Hajj- Ahmed and Ravi Garg named Spectrum combining for ENF signal estimation" they carried out their experiment with two different environment settings. Thus they had two sets of recordings. They found that dominant frequency is present around 120 hz for residential recording and 360 hz for workplace recording. The harmonic varies across the recordings due to noise is their findings.

In our paper we are going to extend our work to find the ENF of power signal for the grids for various frequencies as 50 hz, 100 hz, (100/50) hz, (100+50) hz as well as for different power data also. We are expecting to find more robust solution to detect electrical grid from ENF.

1.2 ENF

1.2.1 What is Electrical Network Frequency

In power grids, the supply frequency is known as Electric Network Frequency or simply ENF. The ENF in most parts of the world is taken to be $50Hz$. The ENF value of North America is $60Hz$. The instantaneous ENF tends to stay around this nominal value. Due to the control systems and load change in grid, the ENF usually fluctuates around the destined frequency. These frequency variations occur mainly due to the differences in the production and consumption power. The grids are interconnected and hence, the changes and fluctuations are almost same all over the area comprising of the same grid. ENF signals can be collected and analyzed through extraction from power recordings. These power and audio recordings can be collected from various grids around the world. If these collected recordings are observed and analyzed trough FFT (Fast Fourier Transform), Spectrogram etc. a lot differences can be found in them in terms of properties, manners and nature. The ENF signals can be collected by recorders such as digital recorder, audio recorder, video camera, sound card and other devices. The $50/60Hz$ ENF is captured in these devices along with the intended sound recording. In certain experiments, the ENF signals are collected through a long period of time. The signals are then examined and analyzed from which it is evident that their patterning are not constant and the fluctuations in ENF are totally random and cannot be predicted. However, during some particular times in a day their presence can be detected throughout the week, but the shapes and values cannot be estimated. Variation of frequency can be about 0.6 Hz at an interval of 10s. The larger the time interval, the tighter is the frequency tolerance. The mathematical formula of ENF at any instant is given by:

$f = [50 \pm \Delta f]Hz$

Here,

f= Electric Network Frequency

Δf= Deviation between the instantaneous frequency and the set point frequency

Attributable to electromagnetic wave propagation, the whole electric system shaped by the interconnected frameworks, consisting of all control power stations and load i.e. shops, workplaces, houses, production lines etc. will have a similar frequency, regardless of point on the system from which it's values are taken. From a scientific perspective, it is the proceeding with variety in frequency over time together with its dependability over separations that give an effective asset to completing credibility examinations of digital recordings.

In accordance with the recommendations by the Union for the Coordination of Production and Transmission of Electricity (UCPTE), under typical conditions the ENF is kept up inside strict cutoff points and the set point frequency at around $50Hz$. There are three sorts of working conditions based on deviation between the instantaneous frequency and the set point frequency.

The working conditions are:

Serial	$\Delta f(mHz)$	Condition
1	$\Delta f(mHz) \leq 50Hz$	Normal
2	$50 \leq \Delta f(mHz) \leq 150Hz$	Impaired(No major risk)
3	$\Delta f(mHz) \geq 150Hz$	Impaired (Significant risk)

Table 1: Different working condition of ENF

All electric and electronic frameworks contain noise sources from an engineering instrumentation perspective (e.g. white noise: noise containing many frequencies with equal intensities, pink noise: random noise having equal energy per octave, and so having more low-frequency components than white noise etc.). Keeping in mind the end goal to break down the ENF, just high quality testing instruments with low levels of noise and distortion are to be prescribed. Practically speaking, a sound card with a signal to noise proportion under 94 dB and aggregate harmonic distortion of under 0.003% will enable the analyst to catch ENF without critical mistakes and contortions.

Using a BPF (Band Pass Filter), with the cut off frequencies lower than $49Hz$ and higher than $51Hz$, the signal can be captured. The BPF eliminates the signals which is not required and leaves the desired ENF component unchanged and genuine to serve the purpose of analyzing.

1.2.2 The effects of Noise

In the analysis of ENF, noise is taken to be any sound other than the roughly $50Hz$ voltage signal of the power network. Since every single significant recording don't plan to record the ENF, the sound to noise ratio (SNR) is usually low, which can bring about unusual ENF estimations. It is said that noise is the greatest restriction of the ENF criterion. Consequently, noise diminishment is the main subject in ENF writing. Numerous strategies have been proposed for noise reduction like a limit dependent median filtering, utilizing numerous ENF sounds or auto-correlation models and expanding the STFT window overlap.
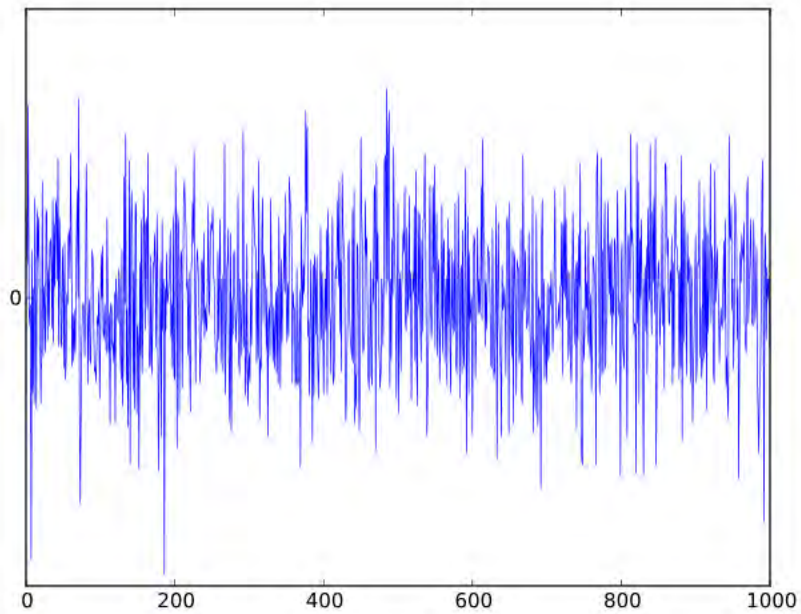

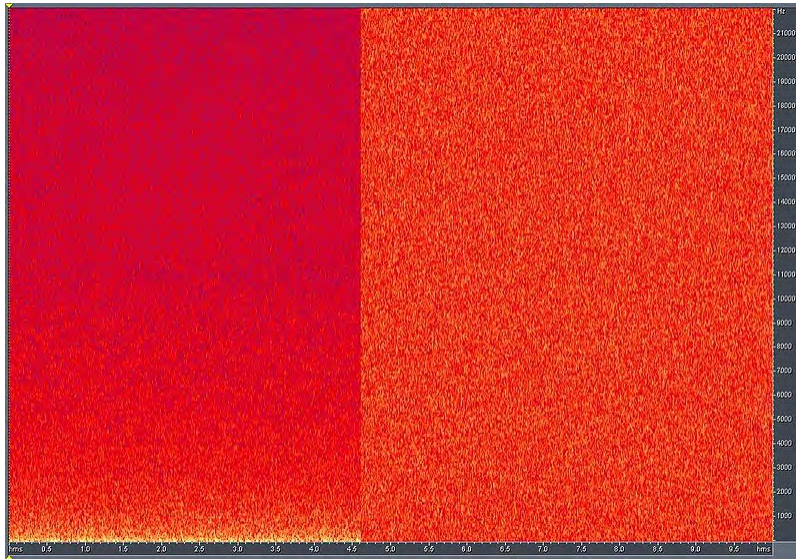Figure 1: Plot of a Gaussian white noise signal

Figure 2: Spectrogram of pink noise (left) and white noise (right), shown with linear frequency axis (vertical) versus time axis (horizontal).

1.2.3 Different ENF extraction process

1.2.3.1 A Procedure by Alan J. Cooper

A standout among the most essential contributing articles about the ENF foundation is composed by forensic researcher Alan J. Cooper. In his researches and article a standard methodology to extract the ENF signal from an audio is suggested. This system was then embraced by most others in the ENF research area. His methodology will be examined in this area and after that an altered technique is explained to build exactness and spare calculation time.

Coopers decision of signal dissemination to $300Hz$ isn't upheld by any hypothesis or analysis other than the way that $300Hz$ is well over the Nyquist frequency (double the frequency of interest). However, in the following segment it is demonstrated this prompts an ENF error of $0.1mHz$. This may impact the time stamping precision somewhat, despite the fact that $0.1mHz$ is little contrasted with the ENF errors caused by a low SNR. Yet, this report is concerned about an area estimation of a voltage signal and for that all frequencies should be exact up to $0.1mHz$. Consequently, every little error ought to be maintained and avoided.

The method of the band pass filter, trailed by the rectangular windowing with 9 seconds overlap has been taken into account. Since most authors utilize the FFT, there is a need to utilize zero-padding. Zero-padding by a factor 4 brings about $20mHz$ frequency determination, before the QI is connected. This section contains tests to test the precision of a $20mHz$ Fourier Transform determination. Utilizing the DFT instead of the FFT it is easily possible to modify the frequency resolution to acquire an adequate precision.

14
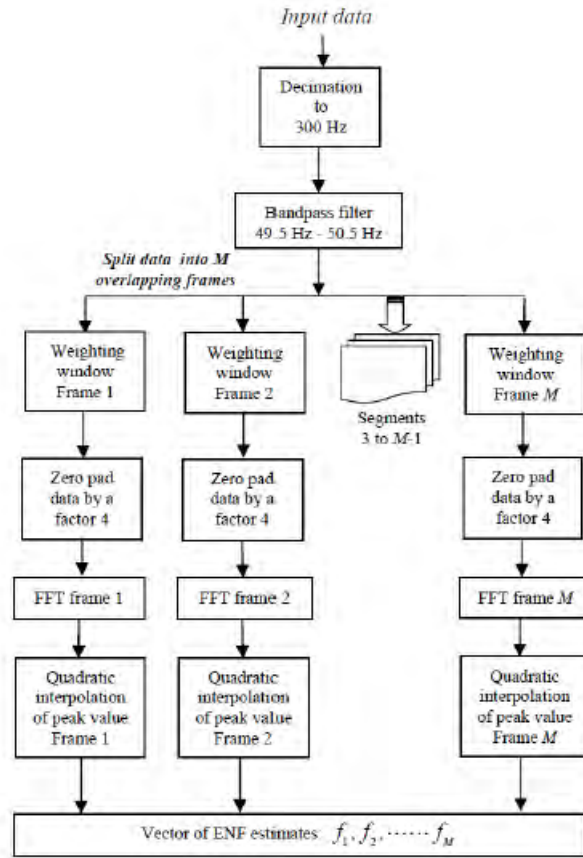
Figure 3: ENF Extraction procedure by Alan J. Cooper

1.2.3.2 Spectrogram method

One of the most known methods for ENF extraction is Spectrogram method. It is easy to execute and is considered an acknowledged a viable method by most researchers. A very useful feature of this method is that it can be done for recordings that are up to 15 minutes. This time/frequency domain spectrogram method, the existence of at least two ENF elements denotes to an unoriginal or not genuine recording. This method gives the ENF components number and hence this procedure is to be followed before utilizing any other methods.
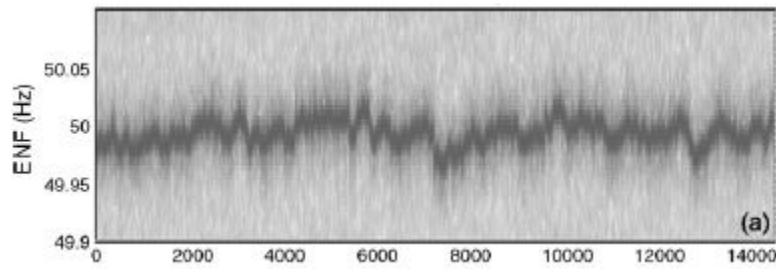
Figure 4: Spectrogram Method

1.2.3.3 Frequency Domain Method

This method can be applied on more than one ENF component recordings. Doing so, it can identify and separate the ENF components. For the searching of database and to find the date and time of the recording, this method is an effective one. Useful details can be detected through this method. It computes Fast Fourier Transform (FFT) over short time windows. It extracts around $50Hz$ as the maximum magnitude.
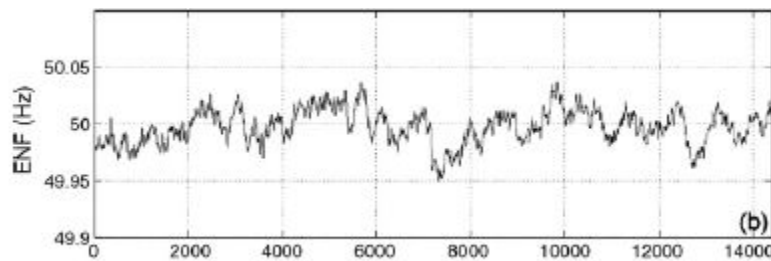
Figure 5: Frequency Domain Method

16

1.2.3.4 Time Domain Method

This method is applicable only in case of one ENF component recording. It extracts and analyzes the ENF at depth. In some cases, time domain method can provide better useful details and features than the frequency domain method. Normally, this method and analysis consists on zero-crosses value. Questioned versus database ENF contrast is possible through this method.
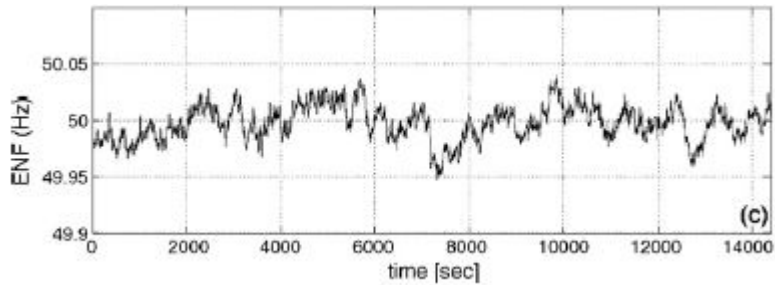
Figure 6: Time Domain Method

1.2.3.5 ENF Criterion and Analogue recordings

Right now, the ENF model can't promptly be connected to analogue tape recordings inferable from the nearness of both persevering long term and fluctuating short term varieties in tape transport speed. While the issue of moderately stable long term varieties may be overwhelmed by a general up shift or downshift in the frequency settings of the spectrograms for investigation, short and long term varieties exhibit a far more prominent issue and one which can bring about extreme thus far irremediable contortions of the original ENF signal

The ENF criterion is an apparatus that can be utilized to investigate digital sound and video recordings, checking their guarantee, confirming or deciding the point in time when they were made and considering the region they are originating from (50 or $60Hz$). This is refined by utilizing a reference frequency database recorded in a research facility or acquired from the electrical network system. The ENFC can likewise uncover the mains-controlled chronicle sort: ENF, UPS, inverter, and so on.

1.3 Data Base

We used 10 reference grids to do the research. We collected multiple power signal from same grid in different time. So it helps us to identify the grid much more easily. By collecting different time frame data we will be able to analyze more precisely.

we took 10 grids from 10 different location.This will help us to understand the effect of different location and time.

1.3.1 Location

Grid Label	Grid Location
A	Texas
B	Lebanon
C	Eastern U.S.
D	Turkey
E	Ireland
F	France
G	Tenerife
H	India
I	Western U.S.

Table 2: Location of Grids

1.3.2 Power signals details

Grid A				
Power Signal	Sample Rate	Total Sample	Duration	Bits Per Sample
1	1000	1200000	1200	16
2	1000	1800000	1800	16
3	1000	3600000	3600	16
4	1000	1200000	1200	16
5	1000	3600000	3600	16
6	1000	1200000	1200	16
7	1000	3600000	3600	16
8	1000	1800000	1800	16
9	1000	3600000	3600	16

Table 3: Grid A power signal details

Grid B				
Power Signal	Sample Rate	Total Sample	Duration	Bits Per Sample
1	1000	1800000	1800	16
2	1000	3000000	3000	16
3	1000	1800000	1800	16
4	1000	1800000	1800	16
5	1000	1200000	1200	16
6	1000	3600000	3600	16
7	1000	3600000	3600	16
8	1000	1200000	1200	16
9	1000	2400000	2400	16

Table 4: Grid B power signal details

Grid C				
Power Signal	Sample Rate	Total Sample	Duration	Bits Per Sample
1	1000	3600000	3600	16
2	1000	2100000	2100	16
3	1000	1200000	1200	16
4	1000	3000000	3000	16
5	1000	1800000	1800	16
6	1000	1200000	1200	16
7	1000	3000000	3000	16
8	1000	2400000	2400	16
9	1000	3600000	3600	16

Table 5: Grid C power signal details

Grid D				
Power Signal	Sample Rate	Total Sample	Duration	Bits Per Sample
1	1000	3600000	3600	16
2	1000	3000000	3000	16
3	1000	1200000	1200	16
4	1000	3000000	3000	16
5	1000	1200000	1200	16
6	1000	1800000	1800	16
7	1000	2400000	2400	16
8	1000	3600000	3600	16
9	1000	1800000	1800	16

Table 6: Grid D power signal details

Grid E				
Power Signal	Sample Rate	Total Sample	Duration	Bits Per Sample
1	1000	1800000	1800	16
2	1000	1200000	1200	16
3	1000	1800000	1800	16
4	1000	2400000	2400	16
5	1000	2400000	2400	16
6	1000	1200000	1200	16
7	1000	3600000	3600	16
8	1000	3000000	3000	16
9	1000	3600000	3600	16

Table 7: Grid E power signal details

Grid F				
Power Signal	Sample Rate	Total Sample	Duration	Bits Per Sample
1	1000	1800000	1800	16
2	1000	1800000	1800	16
3	1000	1800000	1800	16
4	1000	2400000	2400	16
5	1000	2400000	2400	16
6	1000	3600000	3600	16
7	1000	3600000	3600	16
8	1000	1200000	1200	16
9	1000	2100000	2100	16

Table 8: Grid F power signal details

Grid G				
Power Signal	Sample Rate	Total Sample	Duration	Bits Per Sample
1	1000	1200000	1200	16
2	1000	3600000	3600	16
3	1000	1800000	1800	16
4	1000	1200000	1200	16
5	1000	2400000	2400	16
6	1000	3600000	3600	16
7	1000	1800000	1800	16
8	1000	3600000	3600	16
9	1000	1200000	1200	16

Table 9: Grid G power signal details

Grid H				
Power Signal	Sample Rate	Total Sample	Duration	Bits Per Sample
1	1000	1200000	1200	16
2	1000	1800000	1800	16
3	1000	1200000	1200	16
4	1000	1200000	1200	16
5	1000	3600000	3600	16
6	1000	2400000	2400	16
7	1000	3600000	3600	16
8	1000	2400000	2400	16
9	1000	1800000	1800	16

Table 10: Grid H power signal details

Grid I				
Power Signal	Sample Rate	Total Sample	Duration	Bits Per Sample
1	1000	3000000	3000	16
2	1000	1800000	1800	16
3	1000	1200000	1200	16
4	1000	1200000	1200	16
5	1000	2400000	2400	16
6	1000	3600000	3600	16
7	1000	1800000	1800	16
8	1000	3600000	3600	16
9	1000	3000000	3000	16

Table 11: Grid I power signal details

2   Test Database

For testing the grids we have used 100 Test Signals. By this we can identify how our code is working. In this research we have collected the Test Database from a competition named **Signal Processing Cup 2016**. In the competition students from all over the world participated. The main work of the competition was to build a signal processing system that would recognize the ENF signatures in the audio and power data of a Grid.

IEEE Signal Processing Cup is a stage where undergraduate students can have a chance to solve a real life challenging problem. The SP cup first launched in $2013 - 2014$. We used the database from the 2016 competition. The competition is based on time varying and location dependent signature of power grids.

We have collected all the database from this competition as for our work. As it is a real life problem so worked on it vividly. We organized the Test data to have the best possible result.

We worked on the power data only. Here is the details of all the test data. To have a better understanding we put it on a table.

Test Database first 50*TestSignal*				
Test Signal no	Sample Rate	Total Sample	Duration	Bits Per Sample
1	1000	600000	600	16
2	1000	600000	600	16
3	1000	600000	600	16
4	1000	600000	600	16
5	1000	600000	600	16
6	1000	600000	600	16
7	1000	600000	600	16
8	1000	600000	600	16
9	1000	600000	600	16
10	1000	600000	600	16
11	1000	600000	600	16
12	1000	600000	600	16
13	1000	600000	600	16
14	1000	600000	600	16
15	1000	600000	600	16
16	1000	600000	600	16
17	1000	600000	600	16
18	1000	600000	600	16
19	1000	600000	600	16
20	1000	600000	600	16
21	1000	600000	600	16
22	1000	600000	600	16
23	1000	600000	600	16
24	1000	600000	600	16
25	1000	600000	600	16
26	1000	600000	600	16
27	1000	600000	600	16
28	1000	600000	600	16
29	1000	600000	600	16
30	1000	600000	600	16
31	1000	600000	600	16
32	1000	600000	600	16
33	1000	600000	600	16
34	1000	600000	600	16
35	1000	600000	600	16
36	1000	600000	600	16
37	1000	600000	600	16
38	1000	600000	600	16
39	1000	600000	600	16
40	1000	600000	600	16
41	1000	600000	600	16
42	1000	600000	600	16
43	1000	600000	600	16
44	1000	600000	600	16
45	1000	600000	600	16
46	1000	600000	600	16
47	1000	600000	600	16
48	1000	600000	600	16
49	1000	600000	600	16
50	1000	600000	600	16

Table 12: First 50 Test signals details

Test Database Second $50TestSignal$				
Test Signal no	Sample Rate	Total Sample	Duration	Bits Per Sample
51	1000	600000	600	16
52	1000	600000	600	16
53	1000	600000	600	16
54	1000	600000	600	16
55	1000	600000	600	16
56	1000	600000	600	16
57	1000	600000	600	16
58	1000	600000	600	16
59	1000	600000	600	16
60	1000	600000	600	16
61	1000	600000	600	16
62	1000	600000	600	16
63	1000	600000	600	16
64	1000	600000	600	16
65	1000	600000	600	16
66	1000	600000	600	16
67	1000	600000	600	16
68	1000	600000	600	16
69	1000	600000	600	16
70	1000	600000	600	16
71	1000	600000	600	16
72	1000	600000	600	16
73	1000	600000	600	16
74	1000	600000	600	16
75	1000	600000	600	16
76	1000	600000	600	16
77	1000	600000	600	16
78	1000	600000	600	16
79	1000	600000	600	16
80	1000	600000	600	16
81	1000	600000	600	16
82	1000	600000	600	16
83	1000	600000	600	16
84	1000	600000	600	16
85	1000	600000	600	16
86	1000	600000	600	16
87	1000	600000	600	16
88	1000	600000	600	16
89	1000	600000	600	16
90	1000	600000	600	16
91	1000	600000	600	16
92	1000	600000	600	16
93	1000	600000	600	16
94	1000	600000	600	16
95	1000	600000	600	16
96	1000	600000	600	16
97	1000	600000	600	16
98	1000	600000	600	16
99	1000	600000	600	16
100	1000	600000	600	16

Table 13: Last 50 Test signal details

From the above we can see that, The details is exactly the same for every test signal in this data set. This is why it will be tough for differentiate the data. This will also test the mathematical expertness of the coder.

3 Procedure

Our work is divided into different parts. We will explicitly describe each and every steps of our research.

Our research mainly divided into two steps. The steps are following

- Train and build the model

- Test the model

We have used **Matlab** to train and test the ***Support Vector Machine (SVM)*** for doing this research. We also used different frequency and power data to detect grids with that. In our research we have used $50hz$ frequency, $100hz$ frequency, Ratio of $\frac{50}{100}$ frequency, $50hz$ power data, $100hz$ power data, Ratio of $\frac{50}{100}$ power data, $(50 + 100)hz$ power data, $(100 - 50)hz$ power data for this research. We used total 26 features for the train and test for each of the different frequency and power data.

3.1 Train and Build the Model

To detect the grids using its power signals **electrical network frequency (ENF)**, we need to first build the model in Matlab using SVM. To do this we followed these steps.

3.1.1 Audio Signal Input

Nine grids ranging from Grid A to Grid I are taken for the desired outcome. These nine grids have audio data which mainly consist of audio recordings.

The nine grids are taken to train the MATLAB. Code is run for each data within the nine grids changing the required feature in the code and thus the training is completed. The audio data is read in MATLAB by using audioread.
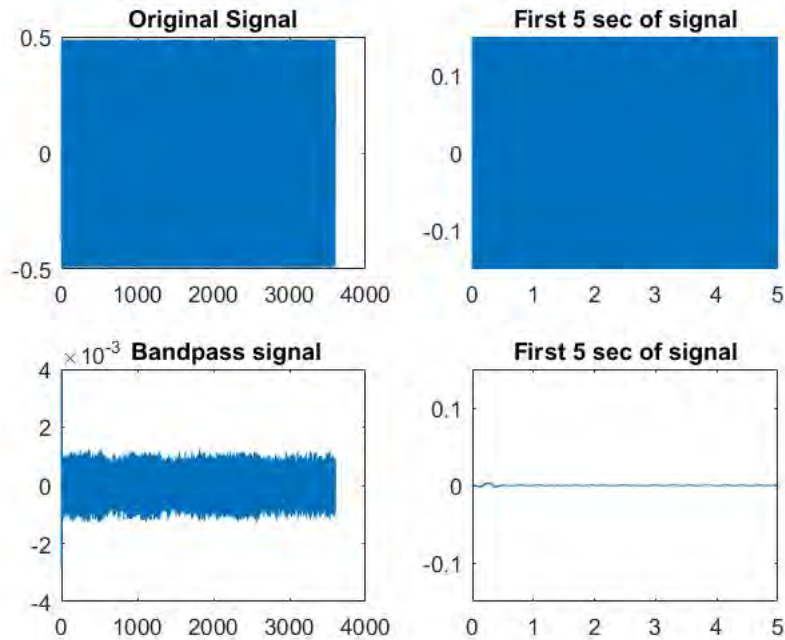

Figure 7: First 5 sec of Signal

3.1.2 Splitting of the audio recordings

The audio recordings from the nine grids are taken into account and each of these recordings is divided into segments of 5 seconds in duration. The splitting is done in MATLAB by using linespace. These segmented data are to be passed through a BPF (Band Pass Filter).

26

3.1.3   Band pass Filter

A band-pass filter is a device that passes frequencies within a certain range and rejects (attenuates) frequencies outside that range giving a filtered signal having the desired specifications.

These segmented data are then passed through a BPF (Band Pass Filter). In this case, data consisting or 50 Hz as well as 100 Hz are to be filtered. Band pass filter with exact specifications of 50 Hz or 100 Hz cannot be applied. Hence, BPF of certain limits are set for filtering as shown below.

Filtering of 50 Hz data: For filtering the 50 Hz data, a Band Pass Filter is taken which has the following features:

Lowest Frequency = 48 Hz
Highest Frequency = 51 Hz

Filtering of 100 Hz data:

For filtering the 100 Hz data, a Band Pass Filter is taken which has the following features:

Lowest Frequency = 98 Hz
Highest Frequency = 101 Hz

Digital filter is used to filter the frequencies. The filtering is done in MATLAB by using *designfilt*.

3.1.4   ENF Extraction

The filtered data are taken into account and each of these filtered data is extracted through ENF extraction method. ENF extracted data are obtained for each filtered data. A number of methods can be followed for ENF extraction purpose. In this case, **Time Domain Method** and **Frequency Domain Method** have been used as explained in this paper earlier. The extracted data are then saved.

We send the data and extract the ENF using Time Domain Method.

Here are some plots of ENF:
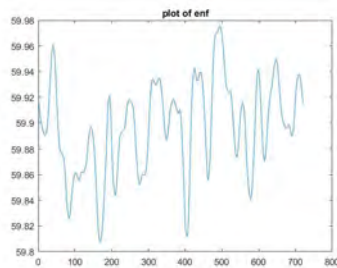
(a) Grid A Power Signal 1   (b) Grid A Power Signal 2
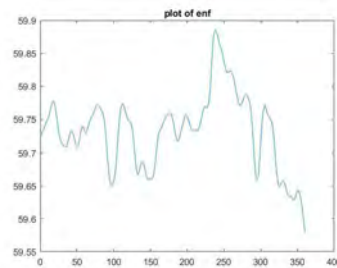
Figure 8: ENF of 50 hz

3.1.5 Extracting features from ENF

We used total 26 number of features to build the pattern of the grids electrical network frequency.
Below the details of all features we have taken for the research.

Name of the feature	Matlab function	Output
MEAN	mean(variable)	Output is a fixed value that represent the mean of enf
MEDIAN	median(variable)	It finds the median value of along the first non-singleton dimension of enf
MODE	mode(variable)	It returns the most frequency occurs in enf
STANDARD DEVIATION	std(variable)	Returns the standard deviation of elements of enf
CUMULATIVE MAXIMUM	cummax(variable)	It returns the single valued output with the cummulative maximum of enf
CUMULATIVE MINIMUM	cummin(variable)	A single valued output returns containing cummilative minimum of enf
MAX	max(variable)	finds the maximum value of enf
MEDIAN NORMALIZED FREQUENCY	medfreq(variable)	It calculate the median frequency of enf
ROOT MEAN SQUARE	rms(variable)	Calculate the rms value of enf

Table 14: Description of the Features

Name of the feature	Matlab function	Output
PERIOD OF SEQUENCE	seqperiod(variable)	Calculate the period of sequence of enf
MAXIMUM TO MINIMUM DIFFERENCE	peak2peak(variable)	Calculate the difference between maximum and minimum value of enf
MAXIMUM TO ROOT MEAN SQUARE DIFFERENCE	peak2rms(variable)	Calculate the difference between maximum and rms value of enf
BAND POWER	bandpower(variable)	Calculate the average power of the bandwidth
OCCUPIED BANDWIDTH	obw(variable)	Returns the 99% of occupied bandwidth of the enf
STATE LEVELS	statelevels(variable)	Calculate the low and high state levels in the enf
RISE TIME	risetime(variable)	Returns the vector of transition time of enf signal
FALL TIME	falltime(variable)	Returns a vector containing fall time of enf signal
SLEW RATE	slewrate(variable)	Returns the slew rate for all the transitions of enf signal
OVER SHOOT	overshoot(variable)	It returns the gratest absolute deviation of enf signal
UNDER SHOOT	undershoot(variable)	Returns the greatest deviation below the final state level of enf signal
PULSE WIDTH	pulsewidth(variable)	It returns the vector contains the time difference between the mid-reference instants of the initial and final transitions of enf
DUTY CYCLE	dutycycle(variable)	Returns the ratio of pulse period of each positive pulse of enf signal
TOTAL HARMONIC DISTORTION	thd(variable)	Returns the harmonic distortion of enf signal
SIGNAL TO NOISE RATIO	snr(variable)	Calculate the signal and noise power ratio of enf
SIGNAL TO NOISE AND DISTORTION RATIO	snr(variable)	Calculate the ratio of signal and noise power of enf

Table 15: Description of the Features

3.1.6 Pattern Recognition

We have used the feature to build the pattern to recognize the girds. As we know we had more than one power data for every grid, we combine all the enf of all the power data for a grid and then build the pattern for that grid. Building pattern means extracting the features from the enf. We collected all the enf from one power signal and then combine them and get the features. We saved the features for all grid to build the train model using **Support Vector Machine** in **MATLAB**.

After getting all the feature for a grid, we saved in a folder. We load it up latter when testing the test signals.

3.1.7 What is Support Vector Machine

SVM or Support Vector Machine is a tool that we can use in matlab to separate the data of exactly two classes. SVM classifies the data by drawing the best hyper plane between the data which can separate the all data point of both classes.
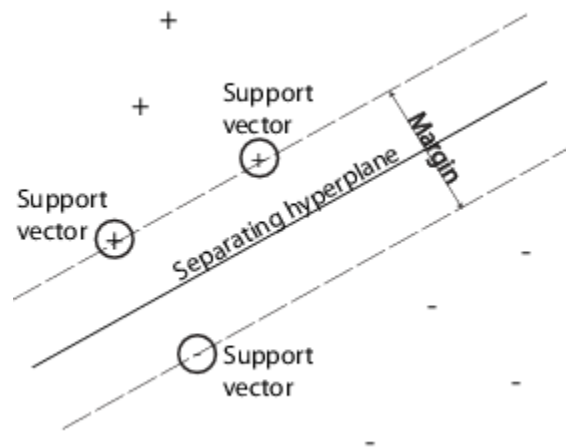
Figure 9: Support Vector Machine(Collected)

3.1.8 Build the SVM model

As we have discussed earlier SVM separates the data of exactly 2 classes. Build a SVM model is the last stage of Train Data. To have a closer look lets analyze the figure below.
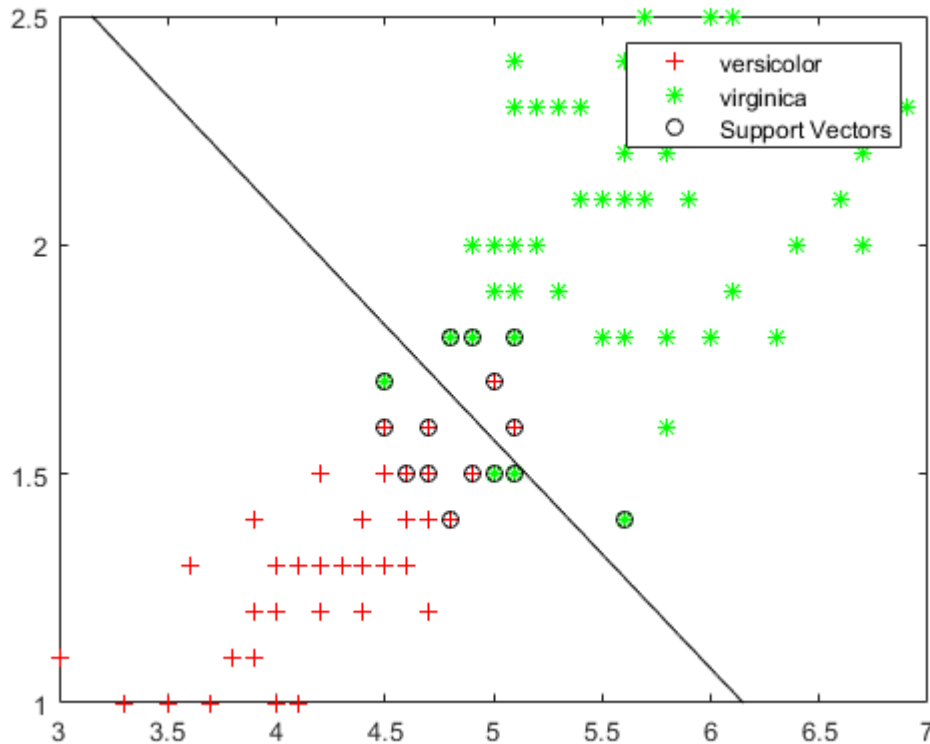
Figure 10: SVM Separates Data (Collected)

So, we can see that here **versicolor** and **virginica** are the two sets of data. Svm classifies the data in to two sets. We also see that some values are wrongly interpreted. This is the basic structure of svm in matlab.

In our case, we have 9 different grids with a lot of power signal which we have used for training. But we know SVM can only classify exactly two sets of data. So, we used a different approach. The approach we used is multisvm. A basic code of multisvm can be found in "https://www.mathworks.com/matlabcentral/fileexchange/39352-multi-class-svm". We have used this but we have modified it in order to do the research. Svmclassify returns a svmstruct and Index that is what we used to build models. To build the model we need the train data and groups. We defined the train data before. And the groups are the grids name. So we made the models using svm classify with the input **Train Data** and **Group Names**.

Training of SVM in matlab returns the models of the SVM.This models are used in the test data. This model will classify the test data in the matlab SVM classifier. The model is SVM struct. SVM struct can identify its kind by the pattern it have.

3.2   Testing the TestData

After getting all the Training our SVM is ready to test the data. As we mentioned earlier that SVM can identify only 2types at a given time, we will modify the original coding system of SVM and build the test environment. For this we set the multisvm code. Multisvm is a coding which takes all the groups at the same time but classify it by classifying two group at a time. It will take all the group names and train the svm. Then it will take the test signals features and will plot it on the svm struct and differentiate the data into the groups. Then it returns the group as well as the group name.

4   Results and Analysis

As we mentioned earlier we have done this research based on different frequency and power data. We have taken 50 Hz, 100 Hz, Ratio of (100/50) Hz, (100+50) Hz Power data, (100-50) Hz Power data, (100/50) Hz Power data.
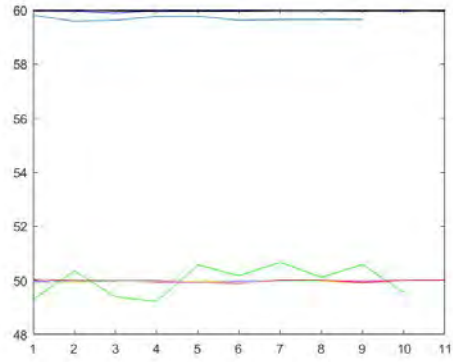
4.1   For 50 HZ

Here we took 50 Hz frequency of the power signal input. We collected the ENF of it. Then did the process. Let us see the graph of feature mean and mode of 50 Hz signal.
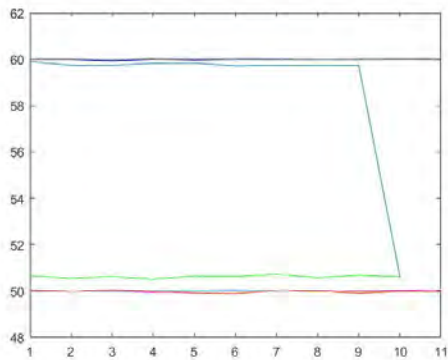
Beyond these features we took another 24 features. In the mean diagram we see the values are close. So it makes hard to make decision. And also for the mode.
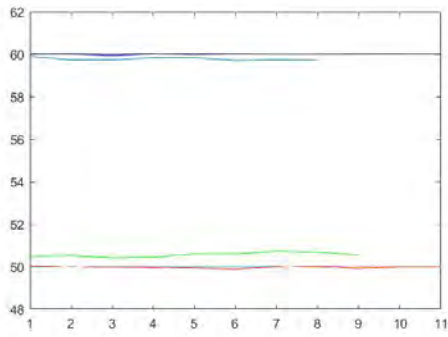
(a) Mean of 50 Hz

(b) Mode of 50 Hz

(c) Median of 50 Hz

Figure 11: Mean Median Mode for 50 Hz data

We can not plot graphically all the features as some of them have infinite and also imaginary values. To have a look we put only these three graph.

For linear type of feature we have this types of graph that shows overlap of points. And also it is not very good to use $50Hz$ as we got the least result from it. From the 100 test data it was able to detect only 87 of them.
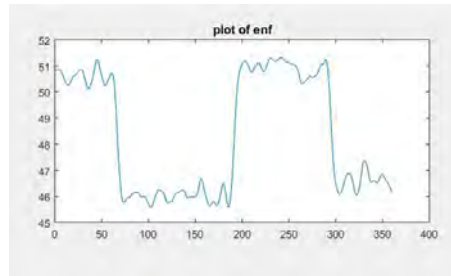
4.2 For 100 Hz

In the 100 hz frequency test we used the ENF of 100 hz test signal and also took the features of 100 Hz train data.
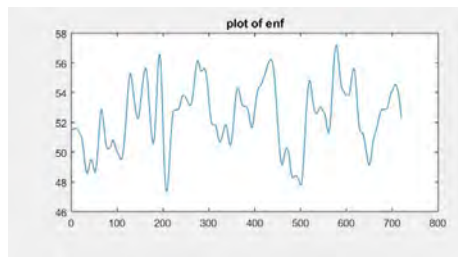
First let us see some ENF plot for 100 Hz signal.
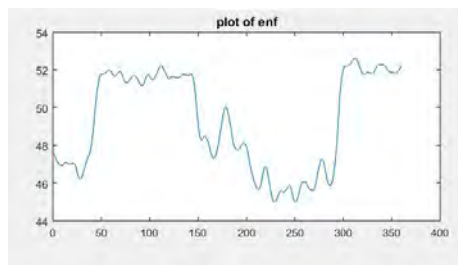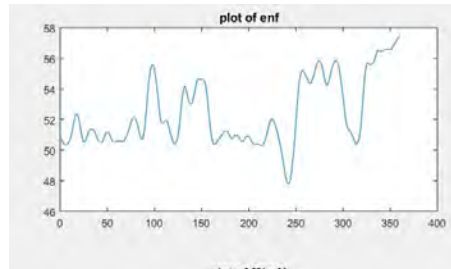

(a) ENF of 100 Hz   (b) ENF of 100 Hz

(c) ENF of 100 Hz   (d) ENF of 100 Hz

Figure 12: ENF plot of power signal of differen Grid

It clearly shows the ENF signal is varying from time and power signal change. We took the 26 features from the ENF and build the SVM to detect the grid.

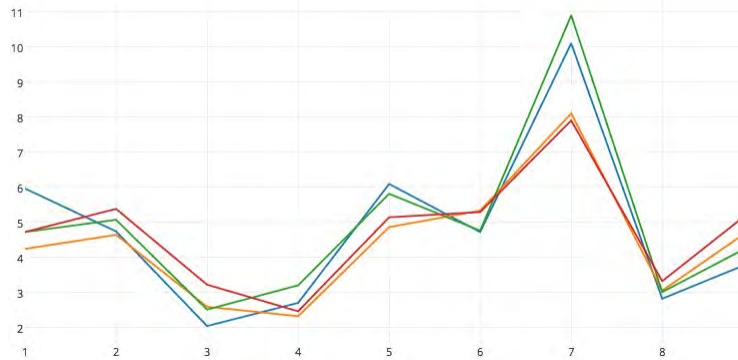Let us see the plot of mean of 4 grids in 100 hz.

Figure 13: Plot of a mean of 4 grids in 100 hz

Here we can see the mean of the 4 grids out of 9 Grids. we include only 4 because it will be easier to observe the details. We can see their in the of all the grids almost have the same mean but as the power signal change with time, the graph change. But we can clearly see the difference between them. To classify the data it is enough for matlab.
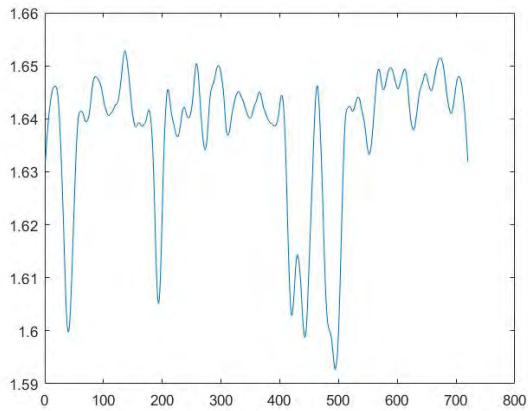
We used all the features and obtain a 92% correct result.

4.3 For ratio(100/50)Hz

At the first let us see some ENF plot of ratio signal.
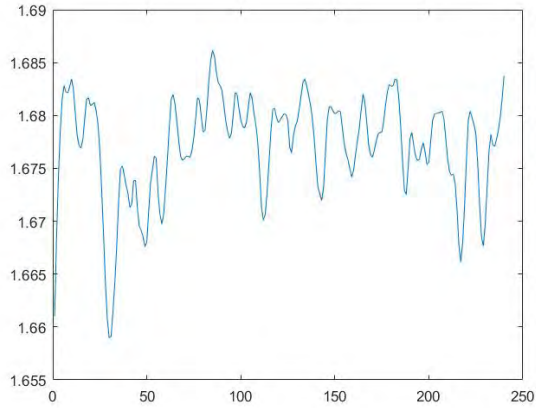
(a) ENF of 100/50 Hz

(b) ENF of 100/50 Hz

Figure 14: ENF plot of frequency ratio

So we can see all the ENF is very much different among them. This ENF then go for features collection and pattern recognition. In the analysis we found that this technique works better than the other two technique. We were able to track 96 out of 100 test data with this technique.

4.4 Final Result

So for the final, we found that all three of them works good and closely related. To visualize we made a graph.
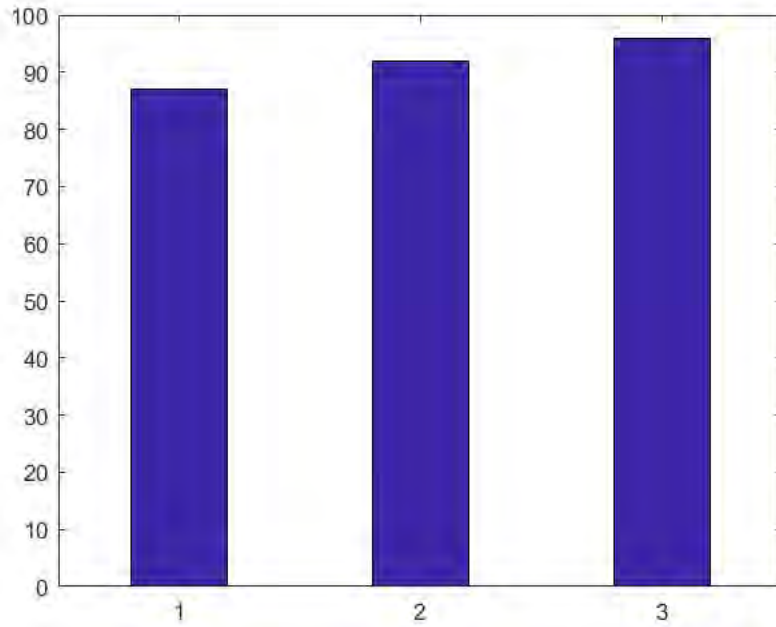

Figure 15: Final Result

5 Future Work

5.1 In our work

We have used all the 26 features for every case. But our knowledge says that by combining the features in to different groups we may find better result. And also we can take frequency other than these and also do the test. Which may end with different results.

5.2 In other works

5.2.1 Smart Grid

Smart Grid alludes to the electric framework, a system of transmission lines, substations and transformers that convey power from the power plant to the houses and desired work places in a more sophisticated, digitalized and smart way. The grid system comprises of more than 9,200 electricity producing units with more than 1 million megawatts of production limit associated with more than 300,000 miles of transmission lines. Despite the fact that the electric framework is viewed as a designing wonder, its interwoven nature is extended to its ability. To advance, another sort of electric grid is required that is developed from the base to deal with the groundswell of advanced and modernized gear and innovation reliant on itand one that can robotize and deal with the expanding unpredictability and necessities of power in modern society.

The researches and results of this paper can directly and indirectly influence the future of grid system through an organized synchronization and further development of the elements of this paper along with the idea of smart grid. The synchronization can be possible in cases of the following aspects:

- Smart meter development

- Efficient electricity transmission

- Security strengthening purposes

- Renewable energy systems

- Reduced electricity rates

- Reduced peak demands

- Faster power disturbance restorations

5.2.2 Power Quality Monitoring

Power quality can be ensured by monitoring the degree to which the power supply approaches the ideal case of unaltered, uninterrupted, distortion and disturbance free supply. Power quality is hindered due to a lot of reasons including variation in voltage magnitude, variation in frequency, sudden rise and fall of load, outages, repetitive varying loading pattern in rolling mills, power electronic converters, inverters, lightning etc. These mishaps can result in motor failing, transformer overheating, unexpected stoppage and failure of electronic machinery, failure of power factor capacitors and unreliable test results.

The researches of this paper on further improvement can be useful in monitoring the adversities of power quality cases. The detection of problem in grids and accuracy measurement can come in handy to solve the cases of power supply loss. Using similar procedures as followed in this paper, the components of power supply system can be represented graphically to identify the errors in power quality.

5.2.3 Cancer Cell Detection

Metastasis is the reason for most mortality due to cancer. Circulating tumor cells (CTC) are cells that are discharged from an essential tumor into the circulatory system and that are considered to go about as the principle promoters of metastasis. These cells are in this manner focuses for the comprehension of tumor science and a change of clinical administration of the infection. Early recognition of CTC and investigation of their cell surface markers give basic data to the analysis of malignancy and its objective treatment. CTC are shed into the circulatory system as tumors develop, and are accepted to start the spread of disease. The recognizable proof and examination of CTC is along these lines a critical goal for the advancement of noninvasive tumor determination. Techniques for capture and discovery of CTC have been confounded and exorbitant and the captured cells couldn't be refined and extended straightforwardly. Since CTC are uncommon, existing as just a couple of per one billion platelets, a profoundly productive strategy is required to catch and to culture CTC for further measure. Chip-based and conventional techniques have risen to

disengage, to recognize and to portray CTC. As CTC are an uncommon occasion, their examination requires multidisciplinary contemplation of both natural and physical properties. MATLAB based detection techniques are being brought into the innovation for the evaluation of CTC with promising outcomes. This audit portrays late advance in the distinguishing proof of CTC; the approach created has a straightforward work process and adaptable multiplexing, which makes it perfect for promoting examination of CTC biomarkers.

The researches and detection techniques followed in this paper can be very useful if synchronized for the detection of cancer cells informing the destination of the infected region.

5.2.4   Bio-Medical Engineering

Lately, the life sciences have held onto simulation as a vital tool in biomedical research. Specialists are additionally utilizing simulation as an effective advance in the plan procedure. In the two fields, MATLAB has turned into the highest quality level tool for biomedical engineering and research.

Following the same methods as followed in this research like extraction of ENF, training of MAT-LAB, filtering of signals, researches on biomedical engineering can also be done.

6   Conclusion

The ENF extraction methodology is a genuinely new field of research with just twelve years of history. On various subjects inside the ENF topic, the information is as yet expanding and there are numerous themes left to inquire about and advance in this field. Among those, noises filtering might be the most essential matter with a specific end goal to increase time stamping accuracy. The analyzing and detecting of power grids for solving grid based problems will become very vital and important in future designs. The software based techniques are likely to be the most fruitful in finding better solutions to the analysis and detection problems. Power system analysis data require accuracy in terms of power quality maintenance and efficiency.

ENF extraction accuracy can be ensured in future by a lot of techniques. It can be done by accelerating the use of DFT rather than FFT. In this paper, ENF extraction is made from audio recordings. It can also be done from video recordings following related procedures. Most video recordings are recorded at a frame rate of around 30 fps. This frame rate is below the Nyquist frequency comparing to the 100 Hz light glinting. This causes the alias impact which makes it difficult to evaluate the ENF effectively from the normal pixel intensity of the video frames. Usually, most cameras are furnished with CMOS sensors which compose frames one line at the time. This may bring about the wonder known as the moving screen impact. Considering every pixel row separately the ENF can be precisely extracted from the signal comprising the normal pixel intensity of each row. Just like audio recordings, the video must be recorded under particular conditions. Besides, numerous video recorders consequently filter out the fluorescent light flickering. Along these conditions the uses of video ENF extraction are as restricted as the uses of sound ENF extraction.

Due to a limitation in budget it is not always possible to set up the necessary databases and economical alternatives were not accurate as desired. It is found that estimating the geo-location of

audio- and video recordings demand very high and precise accuracy. The same ideology is held for the recording which needs to be localized. In real cases, digital videos capturing fluorescent light may result in the most accurate ENF values.

This paper states the research and analysis of the nine grids only, whereas the same can be done for more grid data using the same methods and steps. In this research, the detection of destination for the 50 Hz data is obtained at an accuracy of 87% which on further analysis can be increased to a higher percentage of accuracy. The same can be done for 100 Hz, ratio of the frequencies and all other possible combinations of the frequencies.

7   Reference

- Matthew,Min,Ray. "Information forensic: An Overview of the First Decade." IEEE access (2013).

- Catalin. "Application of ENF Criterion in Forensic Audio, Video, Computer and Telecommunication Analysis." Science direct. Forensic Science International 167 (2007) 136-145.

- Math,Irene. " Signal Processing of Power Quality Disturbances." IEEE Press Series on Power Engineering (2006).

- Adi, Ravi,Min. "Spectrum Combining for ENF Signal Estimation." IEEE Signal Processing Letters, VOL .20 ,NO .9(2013).

- Bykhovsky, Cohen. " Electrical Network Frequency (ENF) Maximum-Likelihood Estimation Via a Multitone Harmonic Model." IEEE Transactions on Information Forensics and Security, VOL. 8 , NO.5 (2013).

- Adi, Ravi, Min." ENF based Region-of- Recording Identification for Media Signals." IEEE Transactions of Information Forensics and Security, VOL.10,NO. 6 (2015).

- Philip,Mark,Ed,Oleg. " Observing the Power Grid." IEEE Signal Processing Magazine [24] (2012).

- Multi Class SVM. (2012, December 07) [online] Available: https://www.mathworks.com/matlabcentral/fileexchange/39352-multi-class-svm

8 Appendix

8.1 Train Period Coding Flow Chart

Figure 16: Flow Chart