Major Project-I

Report on

Classification and detection of power quality disturbances using deep learning algorithms

Submitted towards partial fulfillment of the requirements for the degree of

BACHELOR OF TECHNOLOGY

in

ELECTRICAL AND ELECTRONICS ENGINEERING

by

Mohammed Shebil 191EE135

Somyashree Pradhan 191EE155

P Vishal 191EE139

Hruthwik C Vijayakumar 191EE218

Ashwin J 191EE206

under the guidance of

Dr Debashisha Jena

DEPARTMENT OF ELECTRICAL AND ELECTRONICS ENGINEERING NATIONAL INSTITUTE OF TECHNOLOGY KARNATAKA SURATHKAL, MANGALORE - 575025

November, 2022

CERTIFICATE

This us to certify that the project report work entitled Classification and detection of power quality disturbances using deep learning algorithms which is being submitted by Mohammed Shebil 191EE135, Somyashree Pradhan 191EE155, P Vishal 191EE139, Hruthwik C Vijayakumar 191EE218 and Ashwin J 191EE206 in the department of Electrical and Electronics Engineering of NITK,Surathkal during the year 2017-18 even semester is a bonafide work carried out under my guidance.

Dr. DEBASHISHA JENA

Associate Professor

Department of Electrical and Electronics Engineering

NITK Surathkal

Chairman -DUGC

Department of Electrical and Electronics Engineering

NITK Surathkal

CONTENTS

ACKNOWLEDGEMENTS

We express heartfelt gratitude to our guide Dr. Debashisha Jena Assistant Professor, Department of Electrical and Electronics Engineering. National Institute of Technology Karnataka, Surathkal for providing us with guidance to complete our research and enabling us to achieve the goals that we had set out to accomplish with this project. His suggestions have been very helpful for the completion of this project.

We are equally grateful to Dr Dattatraya N Gaonkar, Head of Department of Electrical and Electronics Engineering. National Institute of Technology Karnataka, Surathkal for giving us an opportunity to undertake this project.

We would like to thank the Department of Electrical and Electronics Engineering including its entire faculty and support staff for having contributed to this project directly and indirectly and having made us able enough to complete the Project.

Mohammed Shebil 191EE135

Somyashree Pradhan 191EE155

P Vishal 191EE139

Hruthwik C Vijayakumar 191EE218

Ashwin J 191EE206

INTRODUCTION

The production, transmission, and distribution of energy to consumers make up the electric power system. Achieving a sufficient supply quality level is a challenging objective for distribution businesses due to the complexity of this system, fluctuations in generation, demand, external factors like weather, load heterogeneity, and other issues.

In IEEE dictionary, power quality is defined as “the concept of powering and grounding sensitive equipment in a matter that is suitable to the operation of that equipment.” Power quality (PQ) has come to be recognized as a significant issue in the electrical power system in recent years. One of the most significant study areas for the energy industry is power quality. It is important to spot harmonics in the energy as well as abruptly rising or falling voltages. When your equipment operates as intended and without disruption, there is a compatible environment between the power source and the end use.

The Power Quality (PQ) of electrical systems is affected by a set of electromagnetic phenomena, which can cause malfunctioning, failure of appliances and even monetary losses. Several factors may cause these electromagnetic phenomena (or PQ disturbances), as switching of capacitors bank, nonlinear charges connected to the network, atmospheric discharges, and the activation of large engines. Nowadays, PQ monitoring has become an important issue to the modem power industry, thus allowing it to protect the electrical, electronic equipment and identify the disturbances causes.

The problems associated with PQ mainly include voltage sag or swell with and without harmonics, interruption, pure harmonics, oscillatory transient, flicker, and so on. Large numbers of algorithms have been adopted to identify the categories of disturbances present in power systems.

For the identification of the PQ disturbances, there are different approaches focused on data-driven, in most of these approaches the disturbances analyzed are described in the norm. Commonly the approaches are based on fault detection identification methodologies. Besides, the pattern recognition techniques are involved in these methodologies and machine learning tools made the classification to identify the disturbances.

This problem has become a very important aspect for both energy distribution companies and end users due to different factors:

• Current loads are very sensitive to supply voltage conditions.

• Increased nonlinear loads cause harmonic disturbances that are on the rise in recent years.

• Increased knowledge of end users in terms of supply quality that forces companies to improve conditions.

• The distributed generation systems integration.

Problem Statement

Classification and detection of power quality disturbances using deep learning algorithms. In terms of PQ, any deviation from the ideal voltage or current can be considered as a disturbance. Therefore, the aim of this project is to implementa deep learning-based model that identifies PQ disturbances and analyses their performance.

Proposed Methodology

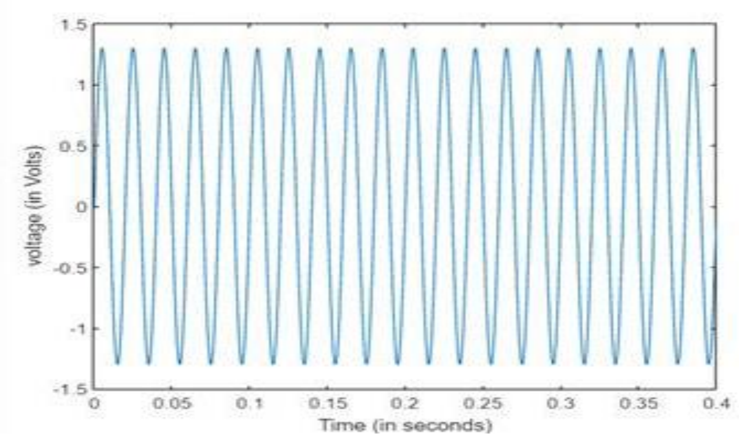
Synthetic signals with disturbances are generated using MATLAB and Signal processing techniques called S transform and statistical features will be used to extract the useful features from various power quality (PQ) disturbances. Deep learning is used for the PQ event classification. SVM is used after extracting statistical features.

Objectives of the project

- Protect the electrical, electronic equipment and identify the disturbances causes.
- Precisely detect the PQ disturbance with high accuracy.
- Classify the above detected disturbance using deep learning methods like SVM and CNN.
- Comparison between SVM and CNN models.

Signal Generation

Normal Signal: It is a sinusoidal waveform of voltage and time graph.



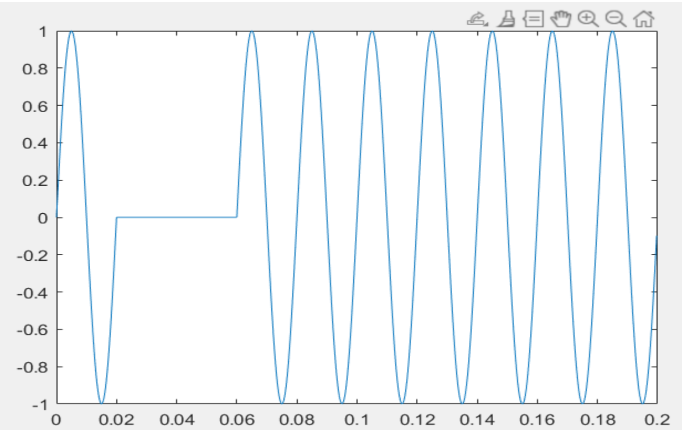
Normal Signal

$$s(t) = [1 \pm \lambda(u(t - t_1) - u(t - t_2))] \sin(2\pi f t - \psi)$$

Parameters : $\lambda \leq 0.1, T \leq t_2 - t_1 \leq 9T$

- Interruption: Interruptions occur when voltage levels drop to zero. Interruptions are classified as momentary, temporary or long-term. Momentary interruptions occur when service is interrupted but automatically restored in less than two seconds.

Temporary interruptions occur when service is interrupted for more than two seconds but is automatically restored in less than two minutes. Long- term interruptions last longer than two minutes and may require field workto restore service.



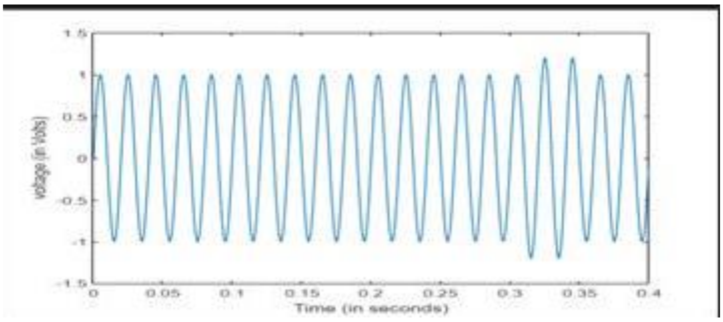
Interruption Signal

$$s(t) = [1 - \lambda(u(t - t_1) - u(t - t_2))] \sin(2\pi f t - \psi)$$

Parameters : $\lambda \leq 0.1, T \leq t_2 - t_1 \leq 9T$

- Swell: A voltage swell is a short-duration increase in voltage values. Voltage swells are commonly caused by large load changes and powerline switching.

If swells reach too high of a peak, they can damage electrical equipment. The utility's voltage regulating equipment may not react quickly enough to prevent all swells or sag



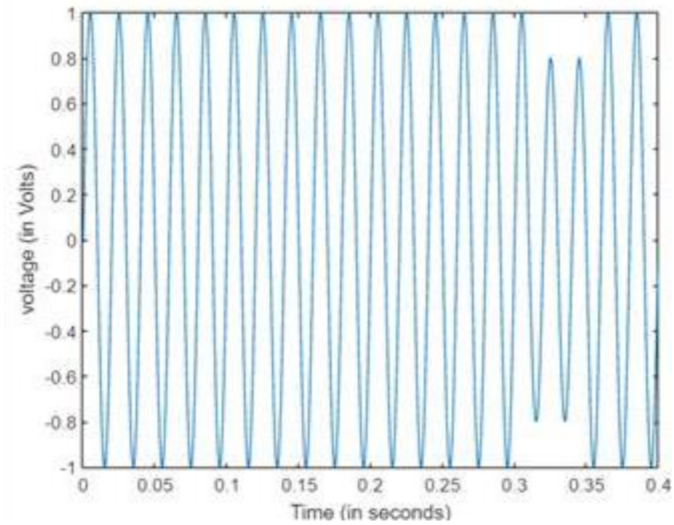
Swell Signal

$$s(t) = [1 + \lambda(u(t - t_1) - u(t - t_2))]\sin(2\pi f t - \psi)$$

Parameters: $0.1 \leq \lambda \leq 0.8, \quad T \leq t_2 - t_1 \leq 9T$

- Sag: A voltage sag is a short-duration decrease in voltage values Common causes of voltage sags are short circuits (faults) on the electric power system, motor starting, customer load additions, and large load additions in the utility service area.

Sags can cause computers and other sensitive equipment to malfunction or simply shut off.



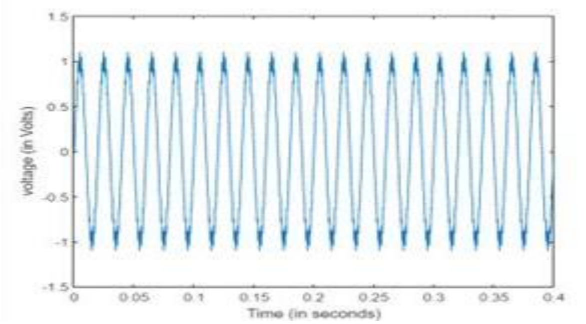
Sag Signal

$$s(t) = \left[1 - \lambda\left(u(t - t_1) - u(t - t_2)\right)\right]\sin(2\pi f t - \psi)$$

Parameters: $0.1 \leq \lambda \leq 0.9, \quad T \leq t_2 - t_1 \leq 9T$

- Flicker: Flicker can be defined as small amplitude changes in voltage levels occurring at frequencies less than 25 Hz. Flicker is caused by large, rapidly fluctuating loads such as arc furnaces and electric welders.

  Flicker is rarely harmful to electronic equipment but is more of a nuisance because it causes noticeable changes in lighting levels.

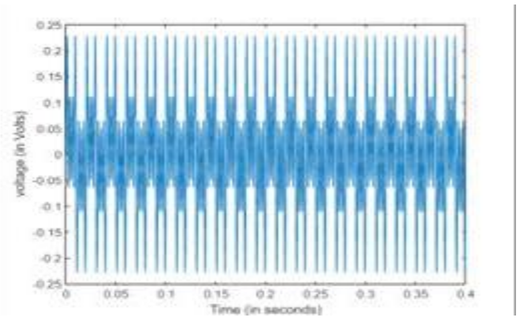


Flicker Signal

$$s(t) = [1 + \lambda \sin(2\pi f_1 t)]\sin(2\pi f t - \psi)$$
$$Parameters: 0.1 \leq \lambda \leq 0.2 \quad 5 \leq f_1 \leq 20\ Hz$$

- Harmonics (Distortion) : Distortion occurs when harmonic frequencies are added to the 60 hertz (Hz) voltage or current waveform, making the usually smooth wave appear jagged or distorted. Distortion can be caused by solid state devices such as rectifiers, adjustable speed controls, fluorescent lights and even computers.

  At high levels, distortion can cause computers to malfunction and cause motors, transformers and wires to heat up excessively. Distortion is probably the most complicated and least understood of all power disturbances.

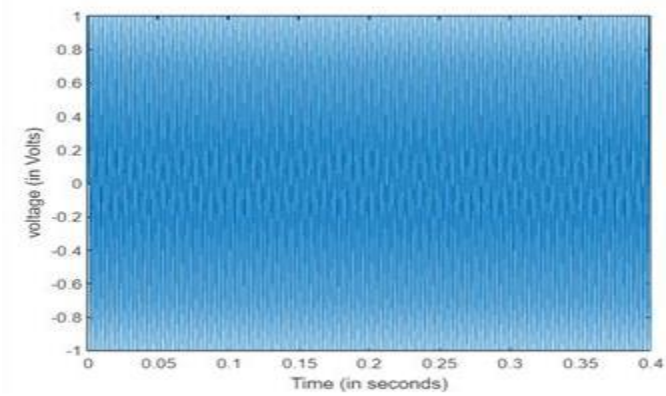


Harmonics Signal

$$s(t) = \sin(2\pi f t - \psi) + \sum \lambda_i \sin(2\pi f_c t - \psi_i)$$
Parameters: $0.05 \leq \lambda_i \leq 0.15$

Oscillatory Transient: Oscillatory Transient is described as a sudden, non–power frequency change in the steady-state condition of voltage, current, or both that has both positive and negative polarity values (bidirectional).

In other words, the instantaneous voltage or current value of an oscillatory transient varies its polarity quickly. It is described by its spectral content or predominant frequency, magnitude and duration.



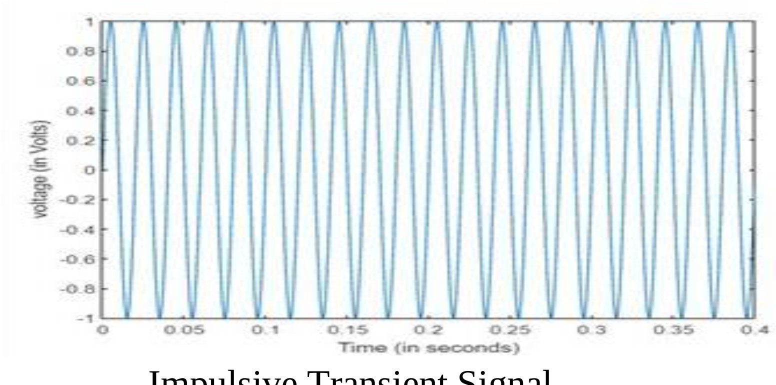
Oscillatory Transient Signal

$$s(t) = \left[\sin(2\pi f t - \psi) + \lambda \exp\left(-\frac{t - t_1}{\tau}\right) \times \big(u(t - t_1) - u(t - t_2)\big) \sin(2\pi f t)\right]$$

Parameters: $0.1 \leq \lambda \leq 0.8 \quad 0.5T \leq t_2 - t_1 \leq 3T$

$8 \leq \tau \leq 40ms \quad 300 \leq f \leq 3500\ Hz$

- Impulsive Transient: It is defined by IEEE 1159 as a sudden, non–power frequency change in the steady-state condition of voltage, current, or both that is unidirectional in polarity – either primarily positive or negative. It is normally a single, very high impulse like lightning.

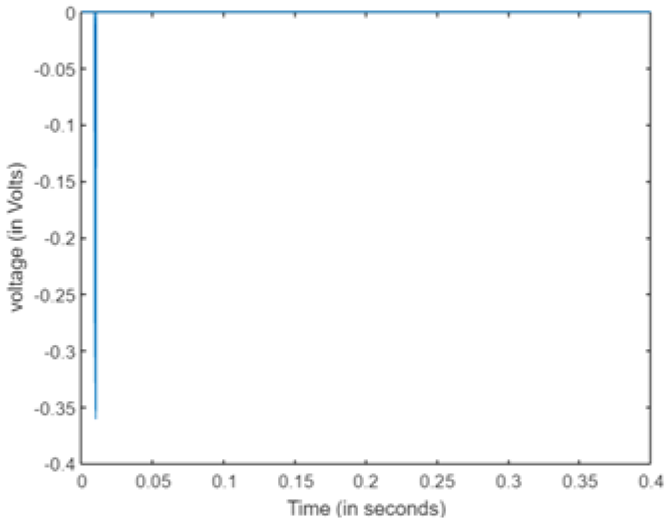


Impulsive Transient Signal

$$s(t) = \big[1 + \lambda\big(u(t - t_1) - u(t - t_2)\big)\big] \sin(2\pi f t - \psi)$$

Parameters: $1 \leq \lambda \leq 3 \quad 0.05T \leq t_2 - t_1 \leq 0.1T$

- Periodic notch: An unexpected intrusion or reduction of line width in patterned voltage-time graphs is called periodic notch



Periodic Notch Signal

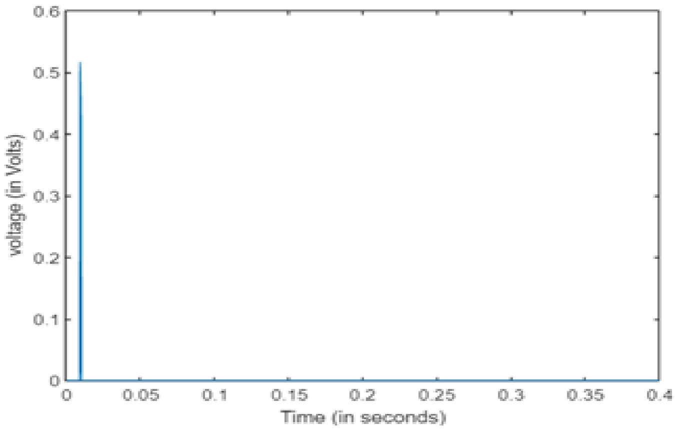
$$s(t) = \sin(2\pi f t - \psi) - sign(\sin(2\pi f t - \psi))$$
$$\times \left\{\sum \rho\left[u\left(t - (t_s - 0.02n)\right) - u\left(t - (t_s - 0.02n)\right)\right]\right\}$$

Parameters: $0 \leq t_2, t_1 \leq 0.5T$

$0.01T \leq t_2 - t_1 \leq 0.05T$

$0.1 \leq \rho \leq 0.4$

- Spike: Very fast variation of the voltage value for durations from a several microseconds to few milliseconds. These variations may reach thousands of volts, even in low voltage. Caused by lightning, switching of lines or power factor correction capacitors, disconnection of heavy loads. Have severe consequences like destruction of components (particularly electronic components) and of insulation materials, data processing errors or data loss, electromagnetic interference.



Spike Signal

$$s(t) = \sin(2\pi f t - \psi) + sign(\sin(2\pi f t - \psi))$$
$$\times \left\{\sum \rho\left[u\big(t - (t_s - 0.02n)\big) - u\big(t - (t_s - 0.02n)\big)\right]\right\}$$

Parameters: $0 \leq t_2, t_1 \leq 0.5T$

$0.01T \leq t_2 - t_1 \leq 0.05T$

$0.1 \leq \rho \leq 0.4$

A total of more than two thousand signals were generated using the above equations and parameters in MATLAB for synthetic signal generation.

Feature Extraction

The extraction of features is crucial in problems involving pattern classification since it aims to reduce the dimensionality of the data that are synthesized into a feature vector while highlighting and maintaining the relevant information from the original data. Thus, the classifier's computing workload can be kept to a minimum.
The features obtained in the time domain are standard deviation, mean deviation, kurtosis, rms value, skewness, entropy, and Shannon entropy.
Here dj refers to each data in the generated signal, j represents the index of each element contained in the window that varies in the range { 1 → N} (here N=1280).

1)Standard deviation
An indicator of how much variance there is from the mean is the standard deviation.

$$standard\ deviation\ value = \sqrt{\frac{\sum_{j=1}^{N} \left(d_j - mean\right)^2}{\mathrm{N} - 1}}$$

2)Mean deviation
The mean deviation is defined as a statistical measure that is used to calculate the average deviation from the mean value of the given data set.

$$mean\ deviation\ value = \frac{\sum_{j=1}^{N} |d_i - mean|}{N}$$

3)Kurtosis
It is the measure of the “peakiness” of a random signal. Signals that have a higher kurtosis value have more peak.

$$kurtosis\ value = \frac{\frac{1}{N}\sum_{j=1}^{N} \left(d_j - mean\right)^4}{\left(\frac{1}{N}\sum_{j=1}^{N} \left(d_j - mean\right)^2\right)^2}$$

4)RMS value
The RMS value of a set of values is the square root of the arithmetic mean of the squares of the values, or the square of the function that defines the continuous waveform.

$$rms\ value = \sqrt{\frac{1}{N}\sum_{j=1}^{N} |d_j|}$$

5) Skewness

Skewness is a measure of the asymmetry of a distribution. A distribution is asymmetrical when its left and right side are not mirror images.

6)Entropy

Entropy is the measure of randomness or disorder of a signal

$$Entropy\ value = \sum_{j=1}^{N} log\ \left(d_j^2\right)$$

7)Shannon Entropy

Shannon's entropy quantifies the amount of information in a variable, thus providing the foundation for a theory around the notion of information.

$$shannon\ entropy\ value = -\sum_{j=1}^{N} d_j^2 * log\left(d_j^2\right)$$

A 2d matrix of size 2000*7 is generated which is then passed to SVM and CNN models.

Model Training using SVM

One of the most robust supervised learning algorithms that analyze data and recognize patterns is Support Vector Machines (SVMs). It is used for solving both regression and classification problems. However, it is mostly used in solving classification problems. SVMs were first introduced by B.E. Boser et al. in 1992

Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic linear classifier.

The objective of applying SVMs is to find the best line in two or dimensions or the best hyperplane in more than two dimensions in order to help us separate our space into classes. The hyperplane (line) is found through the maximum margin, *i.e., the maximum distance between data points of both classes*.

Support Vector, Hyperplane, and Margin

The vector points closest to the hyperplane are known as the support vector points because only these two points are contributing to the result of the algorithm, and other points are not.
The dimension of the hyperplane depends upon the number of features. If the number of input features is 2, then the hyperplane is just a line. If the number of input features is 3, then the hyperplane becomes a two-dimensional plane. It becomes difficult to imagine when the number of features exceeds 3.

So our main goal is to maximize the distance between the two lines or hyperplanes passing through the support vectors.

For our work we used a simple SVM where we trained 10 different SVMs to classify the 10 different classes of the signals. We used the One v/s All method. In this method we classify each class separately from the other classes. Before feeding the data into the model,we filled the NaN values with 1and then the data is standardized using standard scalar. The data matrix was of size 2000 x 7.

For our SVM, we used Gradient descent algorithms for optimization. The hyperparameters that are set are -number of iterations (1000) , learning rate (0.001) and the lambda parameter (0.01) as regularization parameter. 70% of the data is taken as training data and 30% of the data is taken as test data.

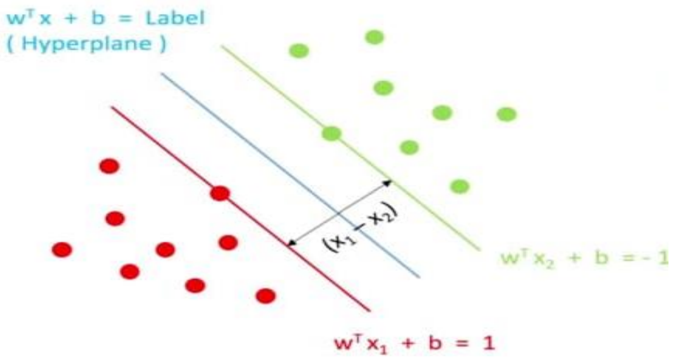


Fig. showing hyperplane equations and classifier hyperplane (*source: youtube)*

The equation of a hyperplane is given by : $w^T x + b = 0$

where w is weight vector; b is bias and x is feature matrix.

Distance between the support vectors is as below:

$$x_2 - x_1 = \frac{2}{\|w\|}$$

Where x1 and x2 are support vectors and $\|w\|$ is norm of vector w

So, we have to maximize this distance above i.e

$$max\left(\frac{2}{\|w\|}\right)$$

.

So essentially, we have to minimize its reciprocal as below

$$\frac{\|w\|}{2}$$

With the below constraint

$$y^i = -1, w^T x_1 + b \leq 1$$

$$y^i = 1, w^T x, +b \geq 1$$

Where y^i is label.

Loss Function: Hinge Loss $J = max\big(0{,}1 - y_i * (w^T x_i + b)\big) + \lambda * \big||w|\big| * \big||w|\big|$

Where λ is regularization parameter.

Gradient for SVM

If $(w^T x, +b \geq 1)$: $\frac{dJ}{dw} = 2\lambda w$ $\frac{dJ}{db} = 0$

If $(w^T x_1 + b \leq 1)$: $\frac{dJ}{dw} = 2\lambda w - y_i x_i$ $\frac{dJ}{db} = y_i$

The average training accuracy that we got after training 10 different SVMs is 93.39% and the average test accuracy is 93.38%.

Stock well-Transforms

Fourier analysis is the most common method to analyze signals. A major disadvantage of using Fourier Transform is that it just gives us information about global frequencies that are sensed over an entire signal. It doesn't give enough information about the instant at which respective frequencies are observed in the signal. Hence, processing of signals using Fast-Fourier transform (practical implementation of Fourier Transform) will not be able to serve all applications. The FFT of a signal *x(t)* is given as

$$X(f) \ = \ \int_{-\infty}^{+\infty} x(t)\,.\,e^{-i2\pi ft}.\,dt$$

The signal can be reconstructed from its FT using the synthesis equation

$$x(t) \ = \frac{1}{2\pi} \int_{-\infty}^{+\infty} X(f)\,.\,e^{i2\pi ft}.\,df$$

STFT evaluates the Fourier Transform over a short-time window, which gives information regarding the fluctuation of frequency contents over time. Hence, it is also called Time-dependent Fourier transform.

The continuous STFT of a signal *x(t)* is given by

$$X(t,f) = \int_{-\infty}^{+\infty} x(t)\,.\,w(t-\tau).\,e^{-i2\pi ft}dt$$

Where w(t) is the window function, a function in which the signal value is 0 outside a particular range. $(t - \tau)$is thought of as shifting the window to the right by τ seconds. However, major drawback is that the window functions used in STFT are of constant width and yield poor temporal resolution.

Another approach to signal decomposition is by using the Wavelet Transform, it decomposes a function into a set of wavelets. In wavelet transform, the idea is to compute how much of a wavelet is in a signal for a particular scale and space. We select a wavelet with known scale, slide it across the signal entirely, multiply the wavelet and signal at every instant of time. This multiplication gives us a coefficient for that wavelet scale at every instant. Then scales are varied and process is repeated. The major drawback of this was the lack of phase information, shift sensitivity and poor directionality. Hence a new transform which could overcome these drawbacks of both STFT and Wavelet Transform was introduced.

The S-Transform is a reversible time-frequency analysis method. It was proposed by Stockwell in 1996. S-transforms is one of the best techniques to process non-stationary signals. It is used to perform multi-resolution analysis on the signals. It performs the multi-resolution analysis process like a set of filters with constant bandwidths. In S-Transform, the window function used is a Gaussian function, whose window width changes with varying frequencies and also it retrieves the phase information.

Continuous S-Transform

For any given signal *x(t)*, the continuous S-Transformation is

$$S(t,f) = \int_{-\infty}^{+\infty} x(t).\, g(t-\tau,f).\, e^{-j2\pi f\tau}.\, d\tau$$

where *g* is the Gaussian window function, expressed as

$$g(t-\tau) = \frac{|f|}{\sqrt{2\pi}} e^{\frac{-(t-\tau)^2 f^2}{2}}$$

Here, $g(t-\tau)$ is used rather than $g(t)$, so it is to be thought of as shifting the window to the right by τ seconds.

Discrete S-Transform

The PQDs of signal *x(t)* can be discretized as *x(kT)*, where *T* is the Sampling time interval.

The discrete sampling signal in form of Fourier Transform can be written as

$$X\left[\frac{n}{NT}\right] = \frac{1}{N}\sum_{k=0}^{N-1} h(kT).\, e^{\frac{-i2\pi nk}{N}}$$

where $n = 0, 1, \ldots, N-1$.

Let

$$\tau \to jt, f \to \frac{n}{NT},$$

Discrete S-transform expression will be

$$S\left[jT,\frac{n}{NT}\right] = \sum_{m=0}^{N-1} H\left[\frac{m+n}{NT}\right].G(m,n)e^{\frac{i2\pi mk}{N}}, n \neq 0$$

$$S[jT,0] = \frac{1}{N}\sum_{m=0}^{N-1} h\left(\frac{m}{NT}\right), n = 0$$

Comparison between Fast-Fourier Transform (FFT), Short-Time Fourier Transform (STFT), Continuous Wavelet Transform (CWT) and S-Transform (ST)

S-transform is the extension of Short-time Fourier Transform and Wavelet transform. The S-Transform uses phase information of Continuous Wavelet Transform to correct phase of the original wavelet. Unlike CWT, FFT and STFT, S-Transform is unique as it has the frequency-related resolution, while positioning the complex spectra of the phase spectrogram.

The Fast-Fourier Transform cannot position time and frequency at the same instant. Hence, it is not helpful for time-frequency localization. So, Gabor proposed that we could adopt a moving and scalable localizing window as the base function, from which STFT came into picture.

The main disadvantage of STFT is that it is inefficient in the detection and resolution of low frequencies, and for high frequency events, it gives poor temporal resolution. These disadvantages are due to fixed width of the window used in STFT.

Compared to STFT, we can see an improved resolution in Continuous wavelet transform. It produces time-scale plots, which is not suitable for visual analysis. Also, the absolute referenced phase information cannot be deduced by CWT.

Hence, another Transform superior to these transforms, which could overcome these problems was deduced for better time-frequency analysis, S-Transform. We can say that S-Transform is the hybrid of CWT and STFT, as it can be derived either as an STFT, which has varying window width depending on the frequency, or as a phase-corrected CWT. ST uses sinusoidal basis functions multiplied with Gaussian window functions whose width varies with frequency, unlike the constant width of windows in STFT. Also, ST provides both true frequency and referenced phase of FT and

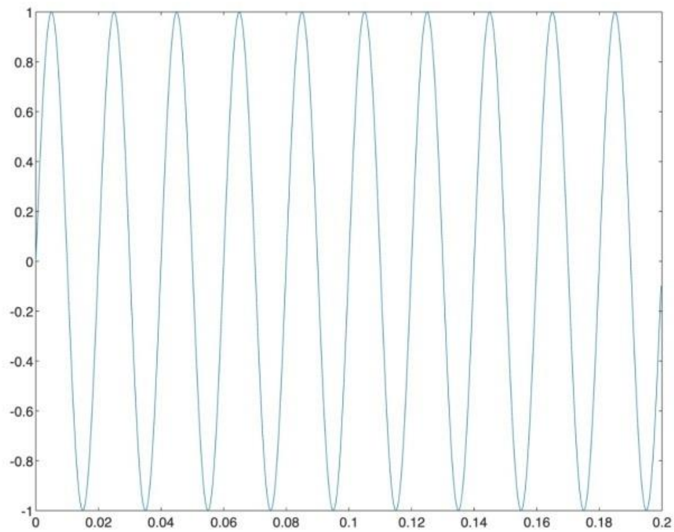
STFT. The Scaling Factor in ST is the reciprocal of the frequency.

Another important feature is that ST is a lossless method. Also, the inverse of ST is the Fourier Transform itself, i.e., it is closely related to the Fourier spectrum.

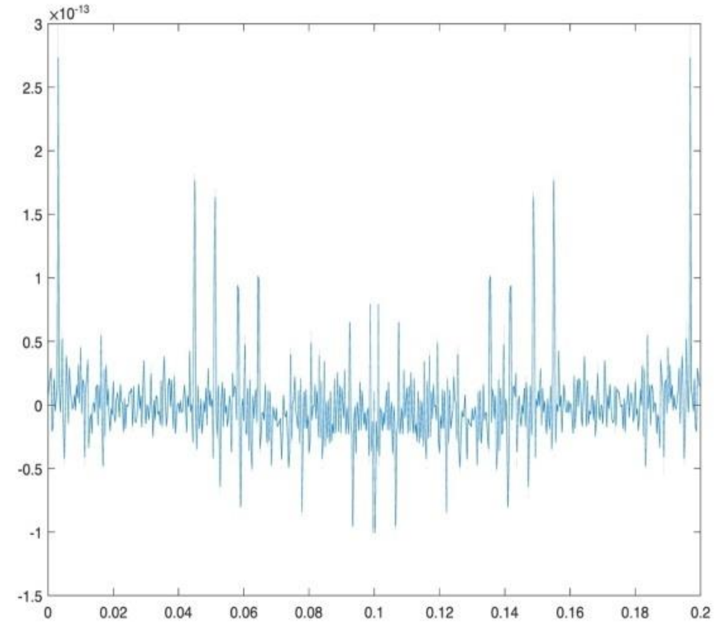
In detecting PQDs, harmonic disturbances and noisy signals are present, which makes it difficult for fast and accurate estimation of frequencies. S-Transform can characterize all harmonic components of the signal.

A signal without any disturbance, and with Sag and Swell disturbances are shown below with their FFT, CWT and ST representation respectively.

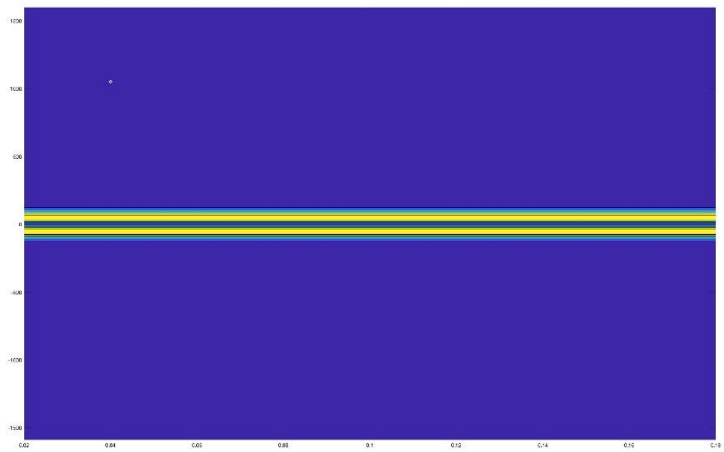
Signal with no disturbances:



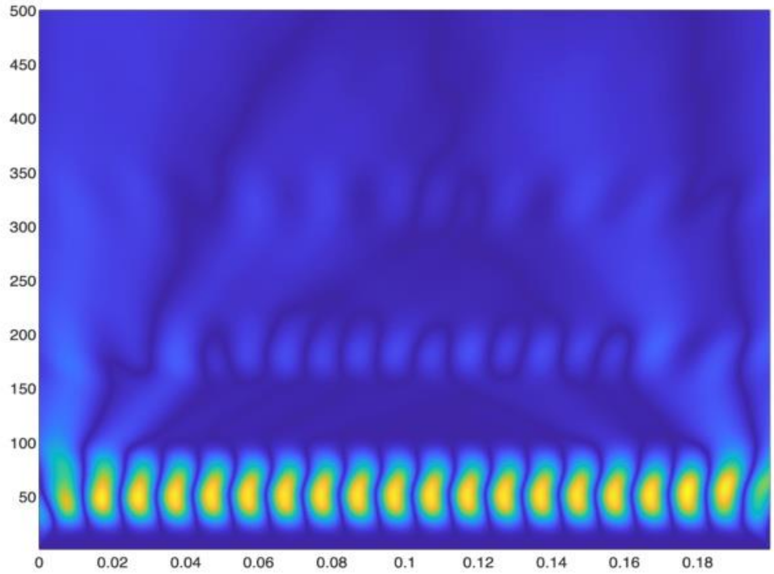
The above figure is a signal(a) with no disturbance.



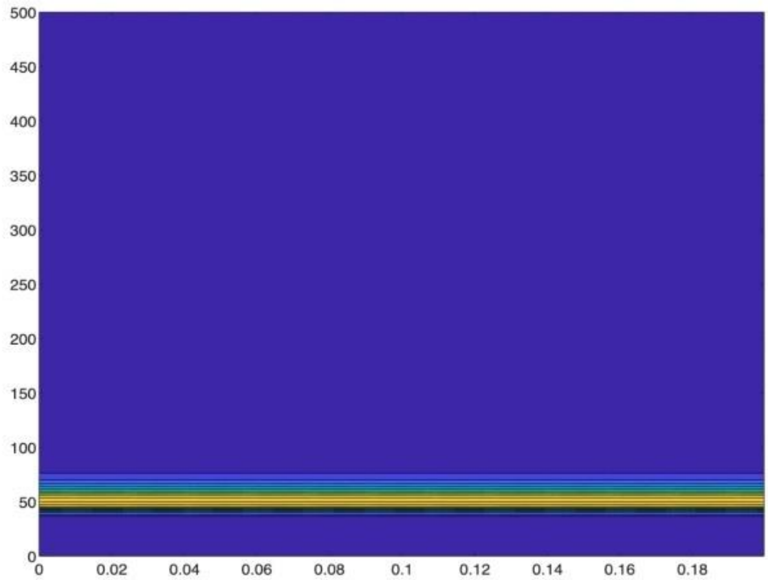
The above figure is the FFT of the signal(a). Only noise of very lower order can be seen in the representation. It is the magnitude spectrum of the signal.



The above figure is the STFT of the signal (a). The x-axis is time and y-axis is frequency. The window function used in simulation is Kaiser window.

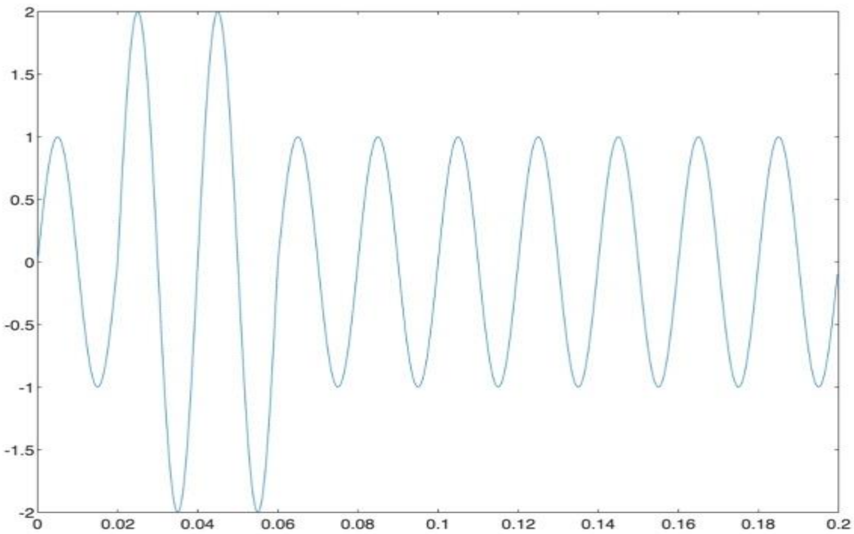


The above figure is the CWT of the signal (a). The x-axis is time and y-axis is frequency.

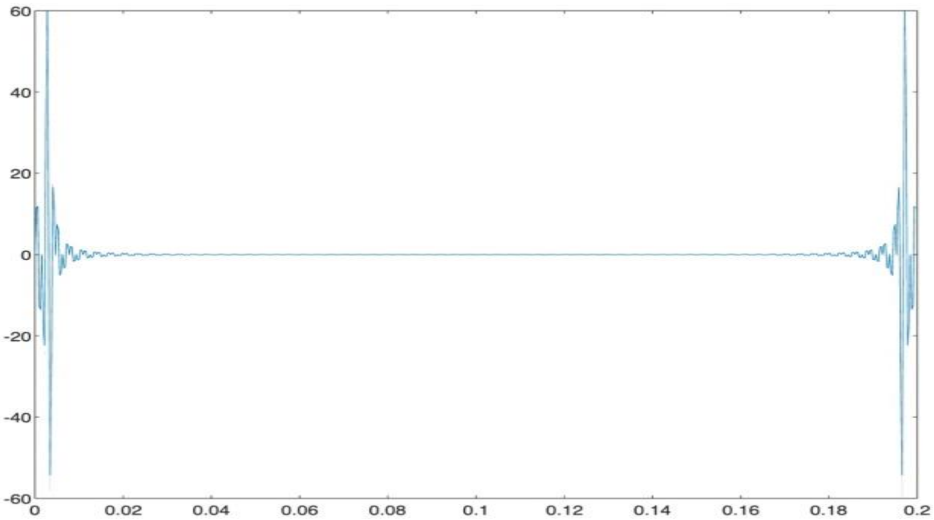


The above figure is the ST of the signal (a). The x-axis is time and y-axis is frequency. The ST image provides better resolution compared to STFT and CWT.

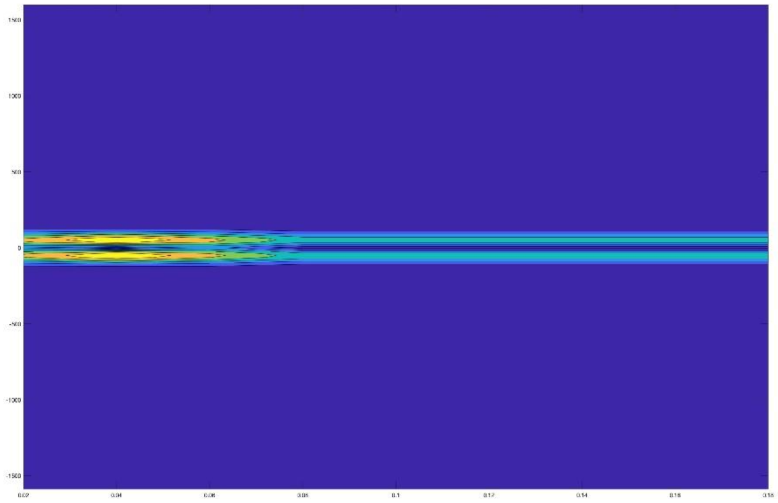
Swell:



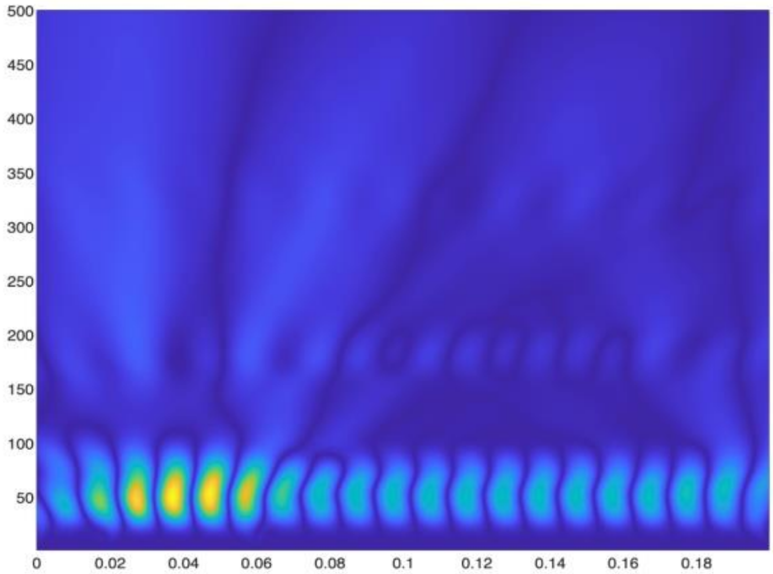
The above figure represents a signal(b) which has a swell disturbance.



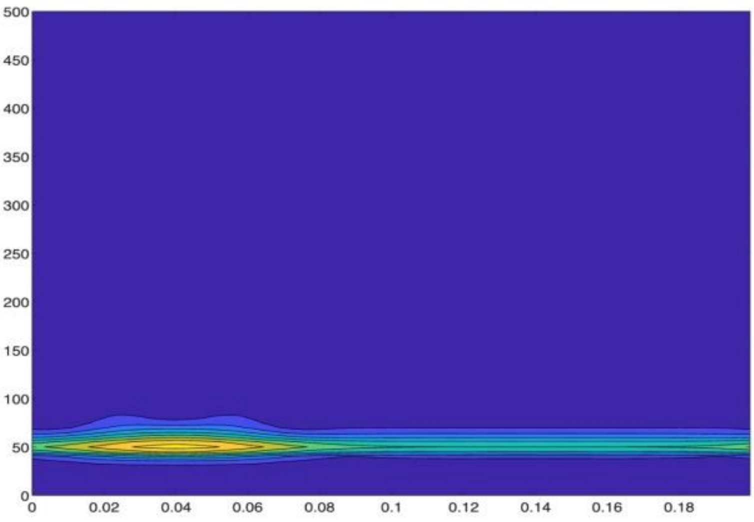
The above figure is the FFT of the signal(b). It just gives us a simple magnitude spectrum of the signal.



The above figure is the STFT of the signal(b). It is shown that there is a bright contour in the areas of swelling, but with poor temporal resolution. Kaiser window is used as the window function here. The x-axis and y-axis represent time and frequency respectively.

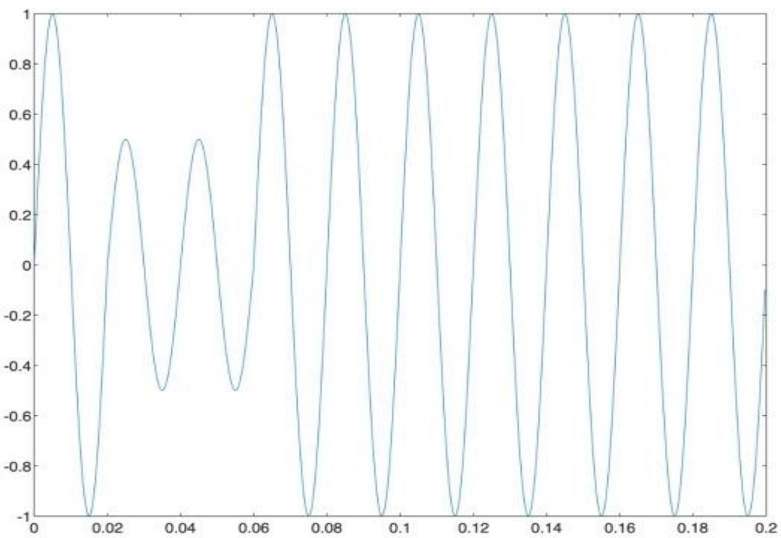


The above figure represents CWT of signal(b). In the image, the bright part indicates the swelling part. The x-axis and y-axis represent time and frequency respectively. But a poor frequency resolution is seen.

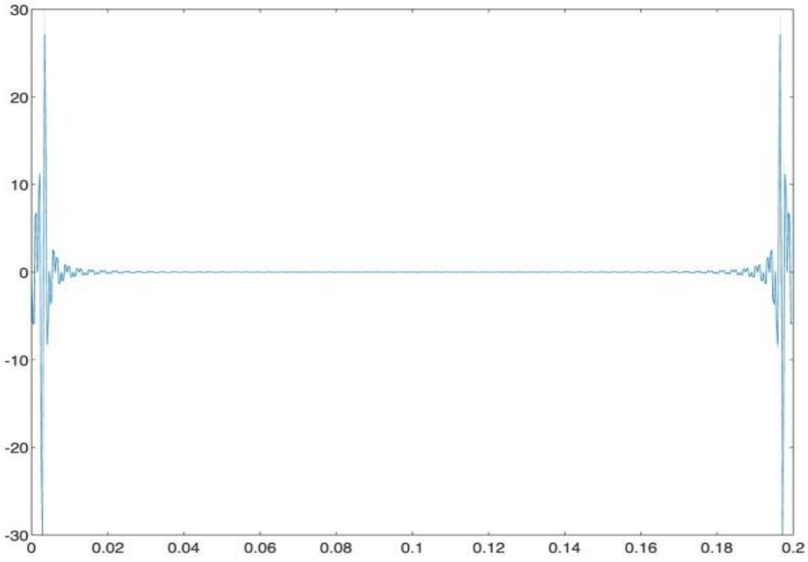


The above figure represents the ST of signal(b). In the image, the bright part indicates swell, with better frequency and temporal resolution as seen. The x-axis and y-axis represent time and frequency respectively

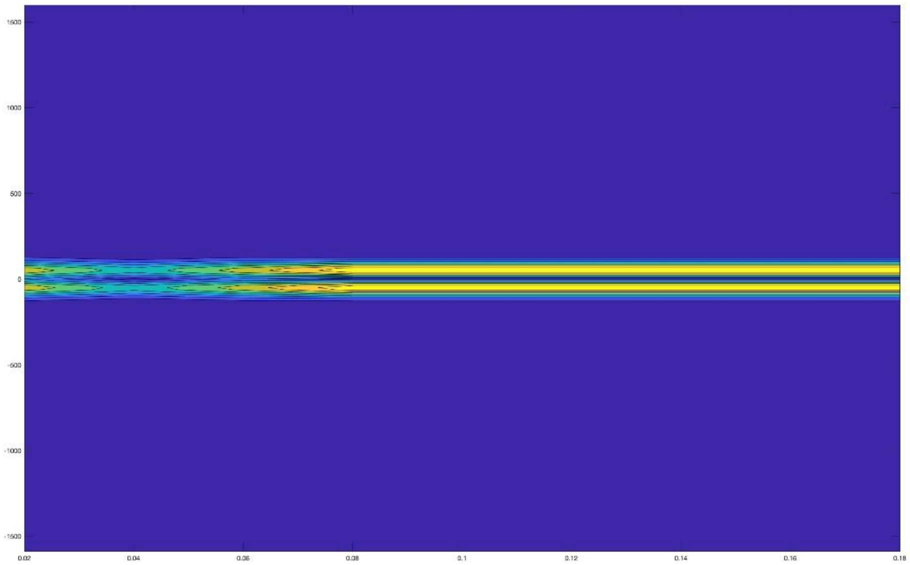
Sag:



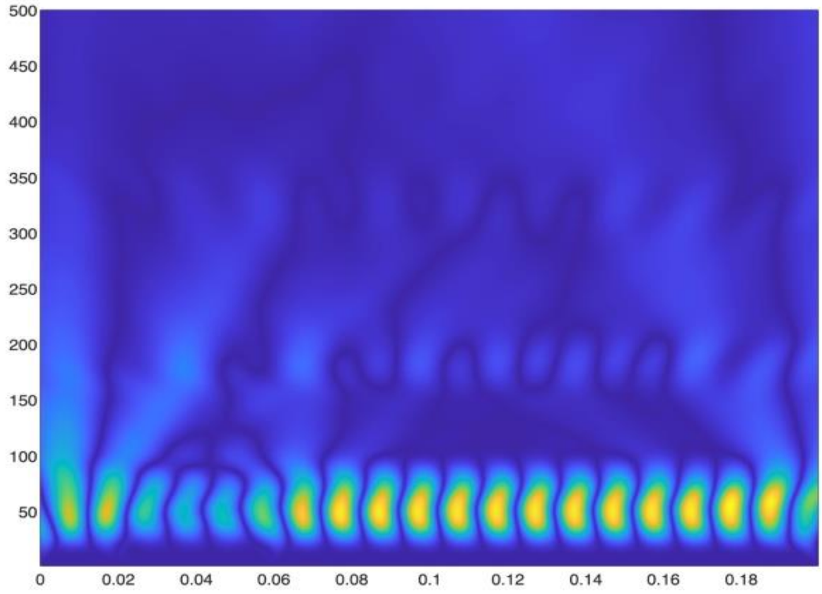
The above figure represents a signal(c) which has a sag disturbance.



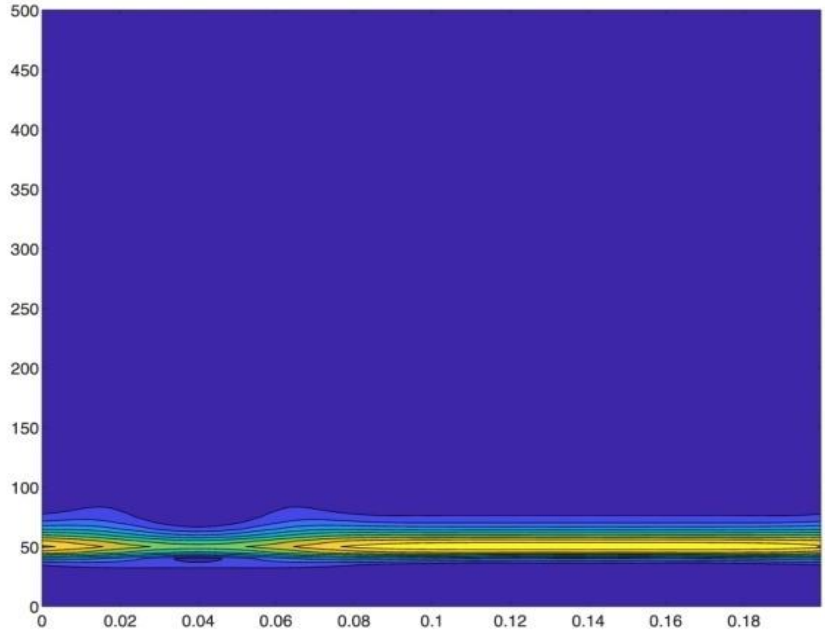
The above figure is the FFT of the signal(c). It just gives us a simple magnitude spectrum of the signal.



The above figure is the STFT of the signal(c). It is shown that there is a light contour in the areas of sagging, but with poor temporal resolution. Kaiser window is used as the window function here. The x-axis and y-axis represent time and frequency respectively.



The above figure represents CWT of signal(c). In the image, the light part indicates the sagging part. The x-axis and y-axis represent time and frequency respectively. But a poor frequency resolution is seen.



The above figure represents the ST of signal(c). In the image, the light part indicates sag, with better frequency and temporal resolution as seen. The x-axis and y-axis represent time and frequency respectively.

By the above figures and their Comparisions, it is evident that ST gives the best time and frequency resolutions.

S-Matrix:

The result of S-transform is a time-frequency matrix, called the S-Matrix. Its row vector is the representation of the frequency distribution of a certain time. Its column vector represents the changes of a certain frequency with varying time. The modulus of any element in the S-Matrix of a certain row and column is the amplitude of S-transform at that particular time and frequency.

Deep Learning: Introduction

Deep Learning is a subset of Machine Learning that is concerned with how the human brain works called artificial neural networks. It is visualized below:

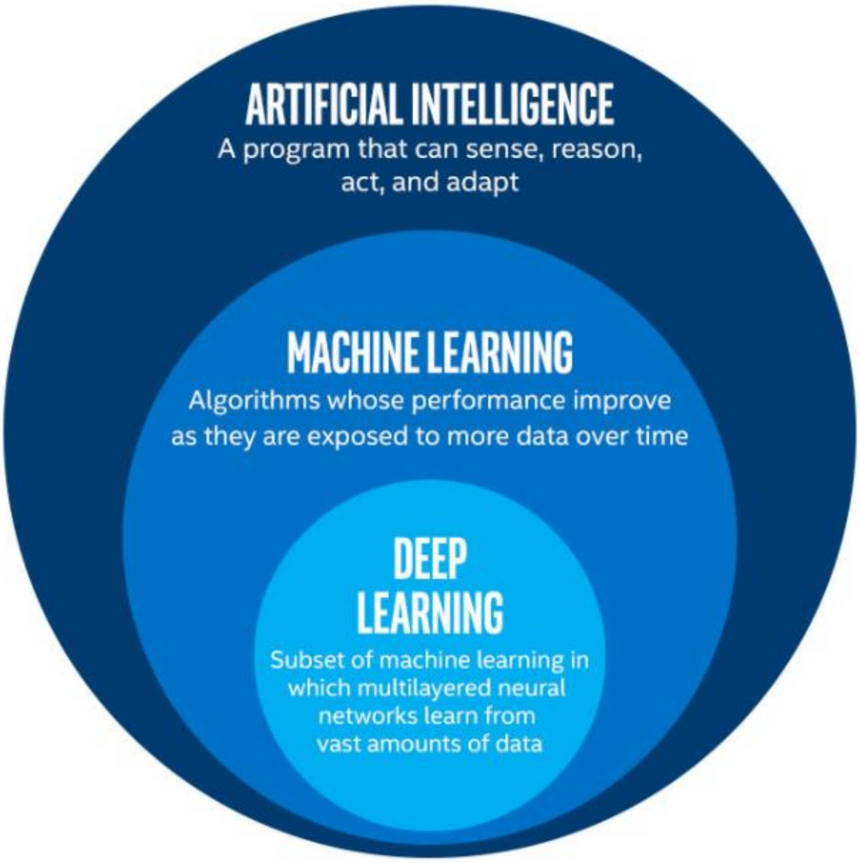


Fig: Relationship between Artificial Intelligence, Machine Learning, and Deep Learning
Source: google images

Introduction to CNN

Convolutional Neural Network (CNN), a subset of machine learning, is a type of artificial neural network that is specialized for image and signal classification and segmentation. Convolution is referred to as a mathematical function on CNN. It's a form of linear operation where you can multiply two functions to indicate how the shape of one function can be altered by the other by a third function. To put it simply, an output is produced by multiplying two images, which are represented as two matrices, and is then utilized to extract information from the image. Similar to other neural networks, CNN also uses a series of convolutional layers, which increases the complexity of the equation.

Description of the Deep Learning Model (CNN)

After the signals are converted to S-matrices, this report adopts the classifier based on CNN. CNN is a typical feed-forward deep neural network, and a large number of excellent CNN-based convolutional neural networks have emerged since the early years as LeNet. AlexNet, ResNet, GoogLeNet, etc. have contributed to the performance of image classification models. In this paper, based on AlexNet, we design a common CNN structure as shown in the figure below which consists of convolutional layer, pooling layer, dropout layer, pooling layer, and activation function between layers. The parameters of each layer are shown in the table.

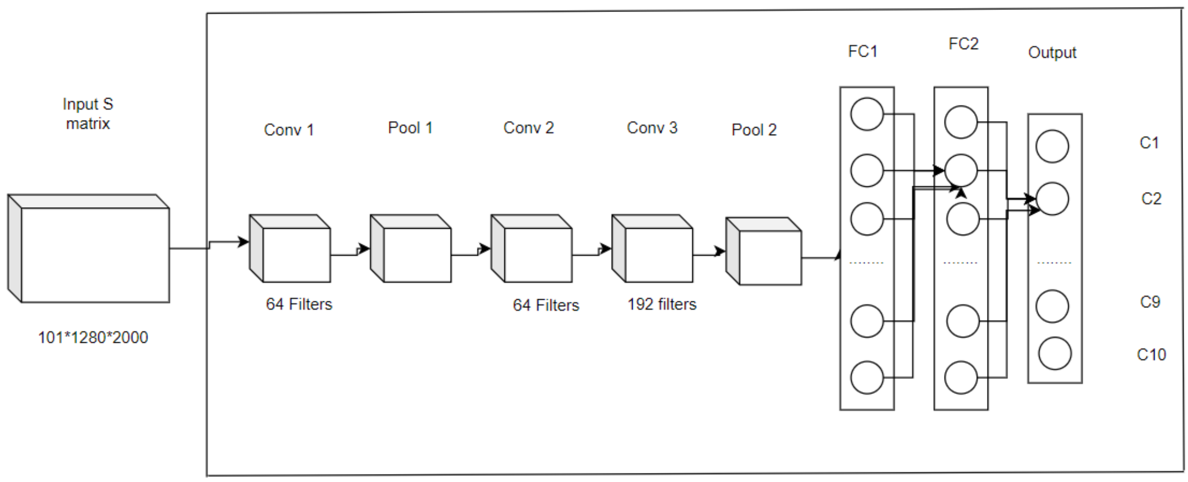


Fig: Proposed CNN-based Classifier

1) Input layer
The contour image, which has undergone S-transform and parameter optimization in the previous step, will be decomposed into S matrix as the input to the deep neural network with the input size of 101*1280.

2) Convolution layer 1
Through experiments, a Filter Size of 7×7 and stride of [2,2]for Convolution 1 is determined and a convolution kernel of 1×1 is used in the later convolution layers which increases the depth of the network and enables the network to learn to extract more diverse feature information. In convolution, the kernel traverses through the input matrix and performs dot product in each set of cells. The size of the output of this convolution layer is calculated using the formula as shown below:

$$O = \frac{I - K + 2P}{S} + 1$$

Here, O is the size of the output matrix, I is the size of the input matrix, K is the kernel size, P is the number of padding layers, and S is the stride. Thus, we obtain an output matrix of size 48*638.

3) Pooling layer 1
The operation of pooling layer is similar to that of convolution layer, except that the operation of convolution is replaced by the operation of averaging, or finding the maximum value and then filling as shown in figure below. Pooling layer can reduce the number of parameters and achieve the dimensionality reduction of features, focusing on the existence of certain features itself rather than their specific location. In this paper, we adopt the Max Pooling approach. The size of the output matrix after pooling is calculated using the formula as shown below:

$$O = \left\lfloor \frac{I}{K} \right\rfloor$$

Here, O is the size of the output matrix, I is the size of the input matrix, and K is the kernel size. By applying the above formula, we obtain an output matrix of size 16*212.

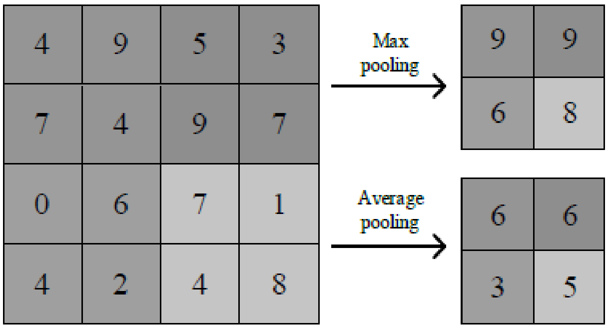


Fig: Pooling Method

4)Convolution layer 2: The input matrix for this layer is of size 16*212, kernel size is 3*3, and stride is [1,1]. The size of the output matrix is calculated using the formula given in layer (2) to be 16*212.

5) Convolution layer 3: The input matrix for this layer is of size 16*212, kernel size is 1*1, and stride is [1,1]. The size of the output matrix is calculated using the formula given in layer (2) to be 14*210.

6) Pooling layer 2: The input matrix for this layer is of size 14*210, kernel size is 3*3, and stride is [1,1]. The size of the output matrix is calculated using the formula given in layer (3) to be 4*70.

This output matrix is flattened to obtain a matrix of size 280*1.

7)Fully connected layers and dropout layers (FC1 and FC2)

The neurons in the fully connected layer are connected in the same way as a traditional ANN, where each neuron updates its weights by backpropagation and the incoming two-dimensional data needs to be tiled into one dimension before being fed into the fully connected layer. Dropout layer is a means to prevent overfitting of the model by turning off certain neurons to make the fully connected layer with a certain sparsity.

8) Activation functions

The activation function gives the set of given inputs and outputs between the layers. The commonly used activation functions are Sigmoid, Tanh, SoftMax, ReLU, and in this paper, we take a variant of ReLU named Leaky ReLU as shown in figure below. Both Relu and Leaky Relu activation functions are linear in the segmentation interval and faster in operation, and compared to ReLU, Leaky ReLU can mitigate the zero-gradient problem for negative-valued parameters.

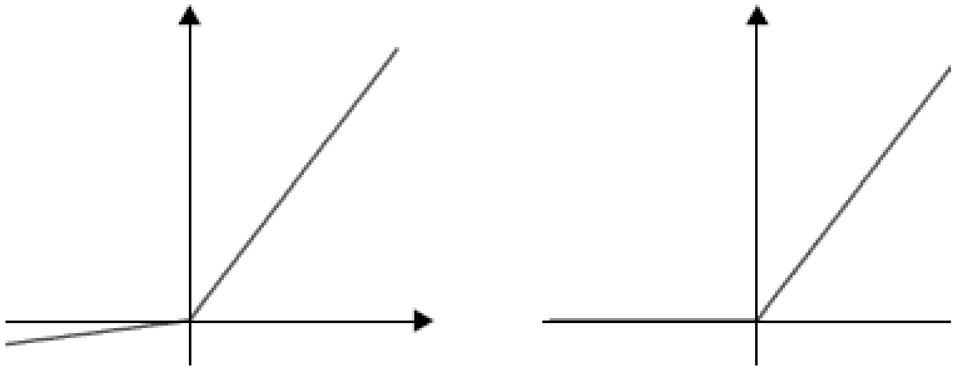


Fig: Leaky ReLU (left), ReLU (right)

After constructing the neural network structure as described above, the output of the last layer of the fully connected layer is 1 × 10, corresponding to the kinds of C1 to C10 in the labels. Unlike the other layers of Leaky ReLU, the output of this layer uses the SoftMax activation function shown in the equation below. The ratio of the exponent of a single output to the exponent of all outputs is used as the output of the SoftMax function, which is suitable for multi-classification problems.

$$S_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}}$$

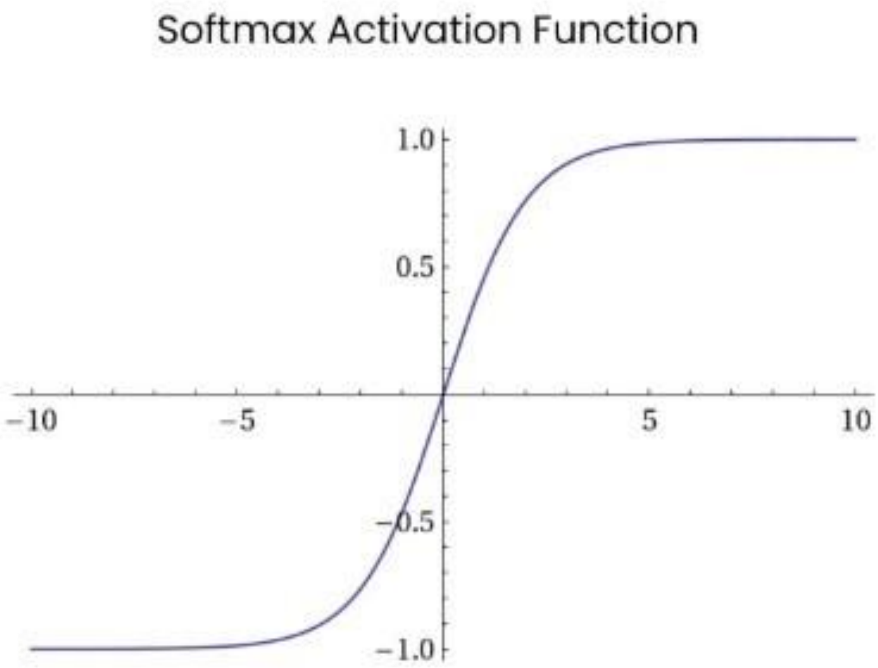


Fig: Softmax Activation Function
(source: google)

TABLE
NETWORK STRUCTURE

Layer	Filter Size/Stride	Filter Number
Convolution 1	7*7/[2,2]	64
Pooling 1	3*3/-	
Convolution 2	1*1/[1,1]	64
Convolution 3	3*3/[1,1]	192
Pooling 2	3*3/[1,1]	
Fully Connected 1	4096/-	
Dropout	0.5/-	
Fully Connected 2	4096/-	
Dropout	0.5/-	
Fully Connected 3	10/-	

CNN can automatically extract, select the features of the image and finally output through the output layer labels from Class 1 to Class 10, which represent different PQDs.

Results and Case Analysis

A) Experimental Setup:

The training sets are set to the normalized amplitude, and they were generated at an interval of 0.05 ms, the signal fundamental frequency is 50 Hz, the sampling frequency is 3200Hz, and the time duration is 0.4 s. 200 sets of data are randomly generated for each category in C1~C10, totalling 2,000 sets of data, which means there are 2,000 matrices. In the validation, 70 % of the images are set as the training set and 30 % of the images are the validation set and the outputs C1~C10 have been given in the previous section. The optimization algorithm uses stochastic gradient descent algorithm, and by experiment, the initial learning rate in the training parameters is set to 0.01.

B) S-Matrix:

The signals with PQDs generated in MATLAB, applying S-Transform on them to perform feature extraction on the sequence, a 2D matrix was obtained. This 2D matrix is the S-Matrix. The row information of the S-Matrix represents the frequency feature and the column information represents the time feature.

The S-Transform matrices of 2000 signals are generated and stored in a matrix of dimensions 101*1280*2000 in a .mat file of size 2.96GB. This data is then fed into the CNN model

C) Result:

The CNN model detects the power quality disturbances in the signals with an accuracy of 100%.

Comparison between SVM and CNN models

Models	Train Accuracy	Test Accuracy
SVM	93.39%	93.38%
CNN	99.57%	100%

CONCLUSION

This project report presents a compact methodology that employs a feature extraction step and classifiers based on Deep Learning technique and the traditional Machine Learning algorithm - SVM classifier. Thus, there was the need for a detailed study regarding the features extracted so that its efficiency could be guaranteed and the computational effort of the classifiers could be reduced. Thus we concluded that the result obtain by CNN classifier was proved to be better than the SVM classifier.

References

1. Cui, C., Duan, Y., Hu, H., Wang, L., & Liu, Q. (2022). Detection and classification of multiple power quality disturbances using Stockwell Transform and Deep Learning. *IEEE Transactions on Instrumentation and Measurement*.
2. Wang, Shouxiang, and Haiwen Chen. "A novel deep learning method for the classification of power quality disturbances using deep convolutional neural network." *Applied energy* 235 (2019): 1126-1140.
3. Shicheng, Xiong, Xia Li, and Bu Leping. "An effective S-transform feature extraction method for classification of power quality disturbance Signals." *2015 Chinese Automation Congress (CAC)*. IEEE, 2015.
4. Huang, Nantian, et al. "Power quality disturbances classification based on S-transform and probabilistic neural network." *Neurocomputing* 98 (2012): 12-23.
5. Wang, Yu-Hsiang. "The tutorial: S transform." *Graduate Institute of Communication Engineering, National Taiwan University, Taipei* (2010).
6. Feature Extraction and Power Quality Disturbances Classification Using Smart Meters Signals Fábbio A. S. Borges, Ricardo A. S. Fernandes, Member, IEEE, Ivan N. Silva, Member, IEEE, and Cíntia B. S. Silva