Major Project-II
Report on

Power Quality Disturbance Classification using Kaiser Window based S-Transform and Deep Learning

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
April, 2023

DECLARATION

I/We hereby *declare* that the Major Project-II Work Report entitled ***” Power Quality disturbance classification using kaiser window based s-transform and deep learning”*** , which is being submitted to the **National Institute of Technology Karnataka, Surathkal**, for the award of the Degree of Bachelor of Technology in Information Technology, is a *bonafide report ofthe work carried out by me/us.* The material contained in this Major Project-II Report has not been submitted to any University or Institution for the award of any degree.

Name of the Student (Registration Number) with Signature

(1) Mohammed Shebil 191EE135

(2) Somyashree Pradhan 191EE155

(3) P Vishal 191EE139

(4) Hruthwik C Vijayakumar 191EE218

(5) Ashwin J 191EE206

Department of Electricals and Electronics Engineering
Place : NITK, Surathkal
Date: 24/04/2023

CERTIFICATE

This us to certify that the project report work entitled **Power Quality disturbance classification using kaiser window based s-transform and deep learning** 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 2023-23 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

ACKNOWLEDGEMENTS

We express heartfelt gratitude to our guide Dr. Debashisha Jena Associate 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

ABSTRACT

Power quality disturbances are a common occurrence in electrical systems and can result in various problems such as equipment damage, system failures, and safety hazards. Thus, it is essential to classify power quality disturbances to facilitate the implementation of effective mitigation measures. This project proposes a machine learning-based approach that utilizes Support Vector Machines (SVM) and Convolutional Neural Networks (CNN) to classify different types of power quality disturbances.

The proposed approach involves four main steps, including signal generation, data pre-processing, feature extraction, and classification. In the pre-processing step, the power signals are first normalized to remove the effect of the varying amplitude of the signals. Then, the kaiser window based s-transform is applied to extract the different frequency components of the signals. In the feature extraction step, statistical features such as mean, variance, and standard deviation are computed from the approximation coefficients to represent the power quality disturbances. Finally, the extracted features are used to train the SVM and CNN-based classifiers.

TABLE OF CONTENTS

TABLE OF FIGURES

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

1.1 What is Power Quality?

Power quality refers to the degree to which electrical power is delivered to electrical loads without causing any disturbances or interruptions that could affect their performance. Power quality can be affected by various factors, such as voltage fluctuations, frequency variations, harmonic distortion, and voltage sags or swells. Poor power quality can lead to equipment malfunction, damage, or failure, and can result in increased maintenance costs and decreased productivity. Therefore, ensuring high power quality is crucial for the reliable and efficient operation of electrical systems.

Power quality can also refer to the degree to which the voltage, frequency, and waveform of electricity supplied to a device or system conform to acceptable standards. In other words, it's a measure of how well the electrical power supplied to a particular device or system matches the ideal, steady-state sinusoidal waveform of AC power.

Issues with power quality can arise due to a variety of factors, including voltage sags, swells, spikes, and harmonics. These issues can lead to problems such as equipment damage, increased energy costs, and decreased system efficiency. As a result, power quality is an important consideration for many industries and applications, and measures are often taken to mitigate power quality issues and ensure that electrical systems operate reliably and efficiently.

1.2 Why do disturbances occur in signals?

There are several reasons why disturbances can occur in power quality. Some of the most common causes include:

i. Voltage fluctuations: Voltage fluctuations can occur due to changes in load demand, which can lead to voltage sags or voltage swells.

ii. Electrical noise: Electrical noise can be caused by a variety of factors, including lightning strikes, electromagnetic interference (EMI), and radio frequency interference (RFI).

iii. Harmonics: Harmonics are caused by non-linear loads that generate frequencies that are not related to the fundamental frequency of the AC power system. This can result in distorted waveforms that can cause issues with power quality.

iv. Transients: Transients are short-term voltage or current disturbances that can be caused by lightning strikes, switching operations, or other factors.

v. Power interruptions: Power interruptions occur when the supply of electrical power is completely lost due to faults or equipment failures.

vi. Power factor: Power factor is a measure of how efficiently electrical power is being used. A low power factor can cause issues with power quality and lead to increased energy costs.

By understanding the causes of power quality disturbances, it is possible to take steps to prevent them from occurring or minimize their impact on electrical systems.

1.3 How do disturbances affect signal?

Power quality disturbances can have a significant impact on electronic devices and systems, which can result in a degradation or disruption of the signals they transmit or receive.

For example, voltage sags or swells can cause fluctuations in the voltage supplied to electronic devices, which can result in unstable or unreliable performance. This can lead to issues such as data corruption, communication errors, and system malfunctions.

Harmonic distortion can also have an impact on signals. When non-linear loads generate harmonics, it can cause distortion of the waveform and introduce additional frequencies into the signal. This can lead to interference with other signals in the system, and result in a degradation of signal quality.

Transients, such as lightning strikes or switching operations, can also cause disturbances in signal quality. These disturbances can result in signal distortions or even complete signal loss.

In summary, power quality disturbances can have a range of negative effects on signals, including degradation, interference, and loss. As a result, it's important to ensure that electrical systems are designed and maintained to minimize the impact of power quality issues on signal quality.

1.4 Need for classification of signal disturbances

Classifying disturbances in a signal is important for several reasons:

i. Identify the source of the disturbance: By classifying a disturbance, it is possible to identify the source of the problem. This can help to determine the appropriate measures to take to mitigate the issue and prevent it from recurring in the future.

ii. Determine the impact of the disturbance: Different types of disturbances can have varying impacts on a signal. By classifying a disturbance, it is possible to determine the severity of the impact and how it might affect the performance of the system.

iii. Develop appropriate mitigation strategies: Different types of disturbances require different mitigation strategies. By classifying a disturbance, it is possible to develop appropriate strategies to mitigate the problem and restore signal quality.

iv. Ensure compliance with standards: In some industries, there are standards and regulations in place that specify acceptable levels of signal quality. By classifying disturbances and monitoring signal quality, it is possible to ensure compliance with these standards and avoid penalties or legal issues.

So classifying disturbances in a signal is important for identifying the source of the problem, determining the impact of the disturbance, developing appropriate mitigation strategies, and ensuring compliance with standards.

1.5 Literature Review

Li et al. (2007) presented one of the earlier publications on PQD classification using the S-transform. They used the S-transform to extract the features from the power signal and a radial basis function (RBF) SVM for classification. They got an accuracy of more than 99%. Their approach wasn't examined using a dataset from the real world, though.

In a different work, Zhang et al. (2012) suggested a hybrid method for PQD classification that combines Stockwell-transform, principal component analysis (PCA), and SVM. They extracted features using the S-transform, reduced the features using PCA, then classified the data using SVM. They tested their approach using a dataset from the actual world and got an accuracy of 98.62%..In this the features were manually selected and the power signal's time-frequency properties were ignored.

Li et al. (2014) suggested a PQD classification approach employing the kaiser-window based S-transform and an optimised SVM to get beyond the drawbacks of the aforementioned methods. The kaiser-window was utilised to lessen spectral leakage and enhance the S-transform's time-frequency resolution. On a real-world dataset, they optimised the SVM using grid search and got accuracy of over 99%. In this approach, both time and frequency information were considered which increased the classification's precision.

In the field of PQD classification, deep learning approaches like Convolutional Neural Networks (CNNs) have been more well-liked in recent years. A CNN-based approach for PQD classification utilising the S-transform was put forth by Gao et al. (2018). The power signal was transformed using the S-transform into a time-frequency representation, which was then input into the CNN for classification. They tested their approach using a dataset from the actual world and got an accuracy of 99.03%. Their approach performed better than conventional approaches like decision tree.

Wang et al. (2019) suggested a hybrid approach integrating the kaiser-window based S-transform, PCA, and a deep neural network (DNN) to further increase the accuracy of PQD classification. They employed a DNN for classification, PCA for feature reduction, and the Kaiser-window to reduce spectral leakage. They tested their approach using a dataset from the actual world and got an accuracy of 99.94%. Their approach fared better than conventional approaches like SVM and CNNs.

In conclusion, the kaiser-window based S-transform, SVM, and CNN techniques have been widely used for PQD classification. The kaiser-window based S-transform can improve the time-frequency resolution of the power signal, and SVM and CNN techniques can effectively classify PQDs. Hybrid methods combining these techniques have achieved high accuracy in PQD classification. However, more research is needed to develop robust and efficient methods for PQD classification in practical applications.

1.6 Previous Work

Previously, time domain features were extracted from the signals generated. Firstly, the statistical features were extracted. The following were the features extracted :

i. **Standard deviation**

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

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

ii. **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.

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

iii. **Kurtosis**

It is the measure of the "peakiness" of a random signal. Signals that have a higher kurtosis value have more peak.

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

iv. **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.

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

v. **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.

vi. **Entropy**

Entropy is the measure of randomness or disorder of a signal.

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

vii. **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.

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

Here d_j refers to each data in the generated signal, j represents the index of each element contained in the window that varies in the range $\{1 \rightarrow N\}$ (here N=1280).

After the features were extracted, it was then arranged in a matrix form as dataset. The size of the matrix was 1800 x 7 where this 1800 imply the total number of signals and this 7 imply the number of features.

As a part of data preprocessing, the null and NaN values were replaced with the mean values. Then the dataset was standardized with standard scalar by using the sklearn library.

Then finally this dataset was fed into the SVM model for training.

The training accuracy achieved was 93.39% and the testing accuracy was 93.38%.

Also, the S matrices of the signals were generated and fed into a CNN model for classification shown below.

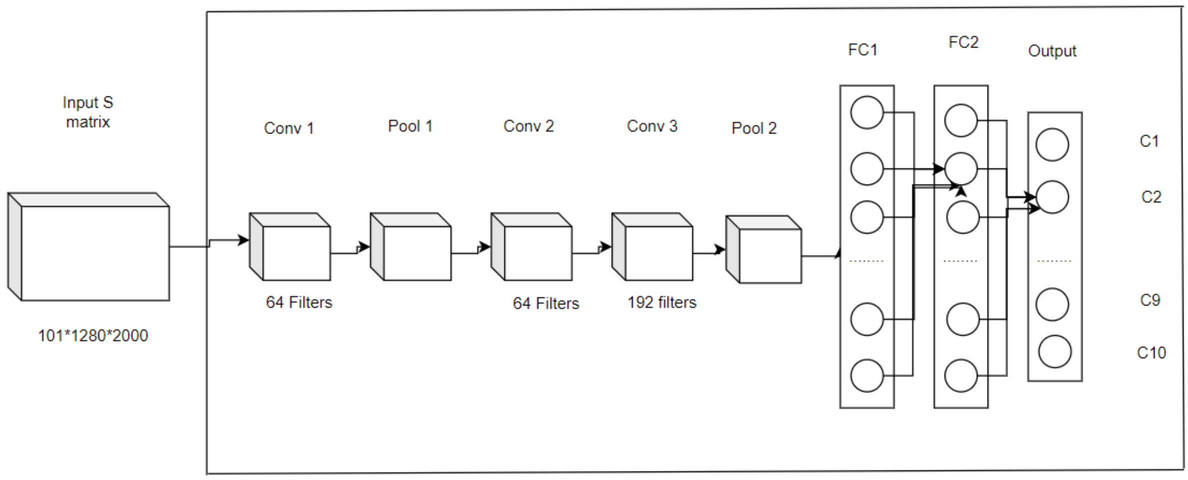


Fig: Proposed CNN-based Classifier

The above model detected and classified power quality disturbances in the signal with a training accuracy of 99.57% and a testing accuracy of 100%.

CHAPTER 2

PROPOSED METHODOLOGY

2.1 Signals

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

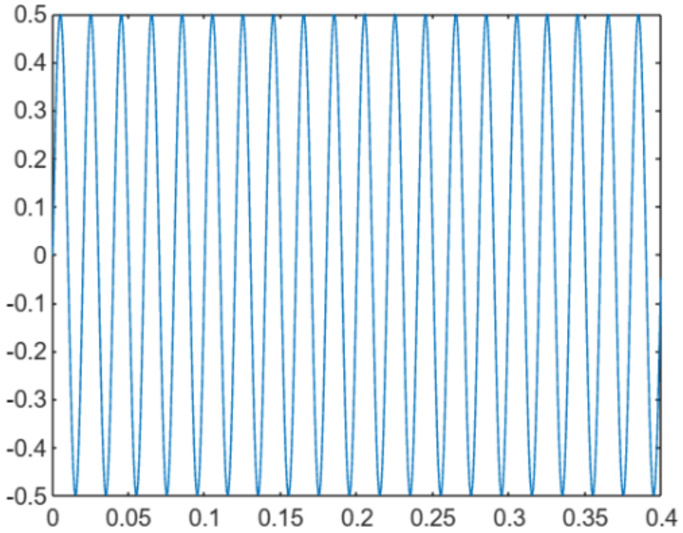


Figure 1 Normal Signal

$$s(t) = A*\sin(2\pi f t - \psi)$$

2.1.2 Swell signal: An abrupt rise in voltage levels is known as a voltage swell. Powerline switching and significant fluctuations in load are two prominent causes of voltage swells. Swells can harm electrical equipment if they reach an excessive peak.

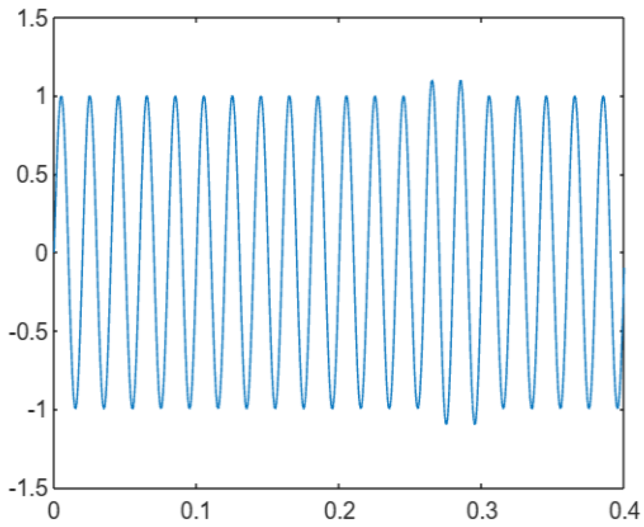


Figure 2 Swell Signal

$$s(t) = [1 + \lambda(u(t - t1) - u(t - t2))]*\sin(2\pi f t - \psi)$$

Parameters: $0.1 \leq \lambda \leq 0.8, \qquad T \leq t2 - t1 \leq 9T$

2.1.3 Sag Signal: A short-term drop in voltage values is known as a voltage sag. Short circuits (faults) in the electric power system, motor starting, customer load additions, and huge load additions in the utility service region are common causes of voltage sags.

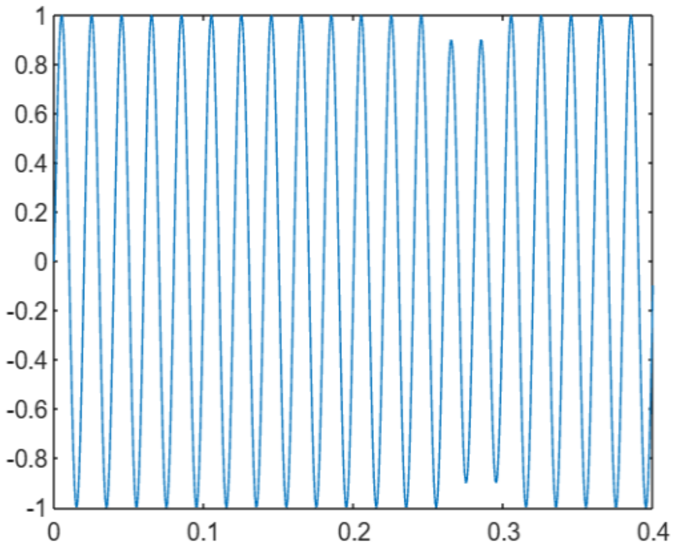


Figure 3 Sag Signal

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

$$\text{Parameters: } 0.1 \le \lambda \le 0.9, \qquad T \le t2 - t1 \le 9T$$

2.1.4 Flicker: Small amplitude variations in voltage levels that take place at frequency lower than 25 Hz are known as flicker. Large, quickly varying loads like arc furnaces and electric welders are what generate flicker. Electronic devices are rarely harmed by flicker, but it is more of an annoyance because it results in apparent variations in lighting levels.

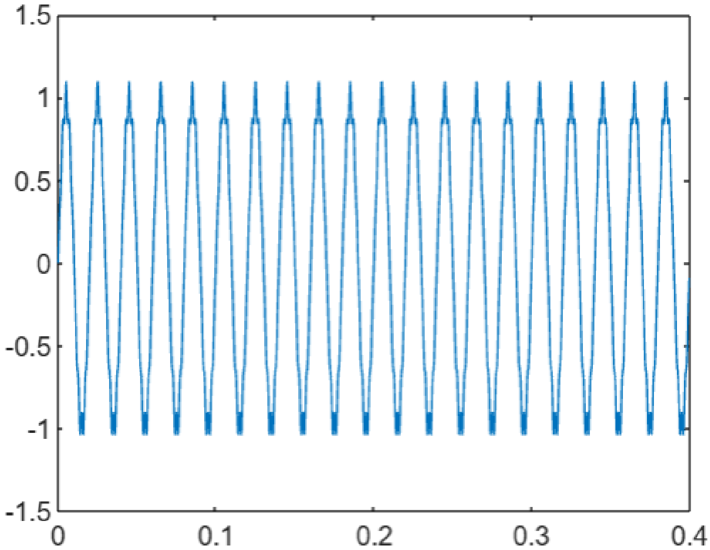


Figure 4 Flicker

$$s(t) = [1 + \lambda \sin(2\pi f_1 t)]*\sin(2\pi f t - \psi)$$

$$Parameters: 0.1 \le \lambda \le 0.2\ ; 5 \le f_1 \le 20\ Hz$$

2.1.5 Interruption: At zero voltage levels, interruptions happen. There are three types of interruptions: brief, short-term, and long-term. Service outages that are momentarily disruptive but automatically resume within two seconds are known as momentary interruptions.

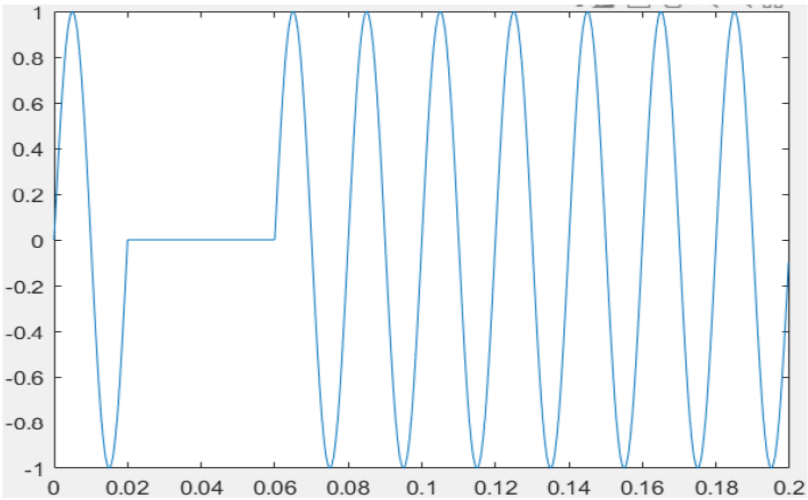


Figure 5 Interruption

$$s(t)=[1-\lambda(u(t-t1)-u(t-t2))]\sin(2\pi f t-\psi)$$

$$\text{Parameters : } \lambda\leq 0.1;\ T\leq t2-t1\leq 9T$$

2.1.6 Oscillatory transient: A abrupt, non-power frequency shift in the steady-state condition of a voltage, current, or both that has both positive and negative polarity values is referred to as an oscillatory transient.

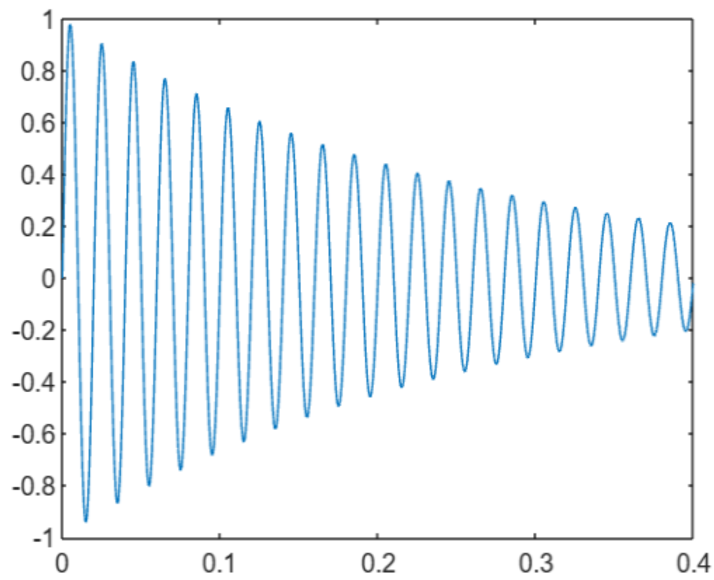


Figure 6 Oscillatory Transient

$$s(t)= \sin(2\pi f t-\psi)+\lambda*\exp(-t-t1\tau)\times(u(t-t1)-u(t-t2))\sin(2\pi f t)]$$

$$\text{Parameters: } 0.1\leq\lambda\leq 0.8;0.5T\leq t2-t1\leq 3T$$

2.1.7 Impulsive transient: A quick, non-power frequency change in the steady-state condition of voltage, current, or both that is unidirectional in polarity—either mostly positive or negative is referred to as a transient event. In most cases, it comes as a single, lightning-like impulse.

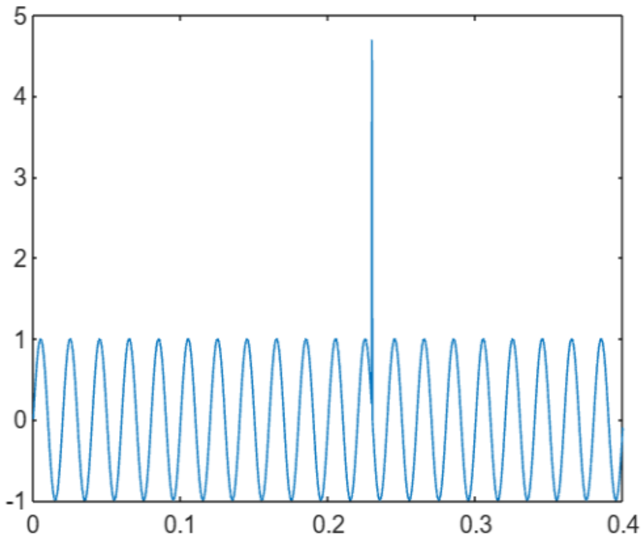


Figure 7 Impulsive Transient

$$s(t)=[1+\lambda(u(t-t1)-u(t-t2))]\sin(2\pi f t-\psi)$$

Parameters: $1\leq\lambda\leq3$; $0.05T\leq t2-t1\leq0.1T$

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

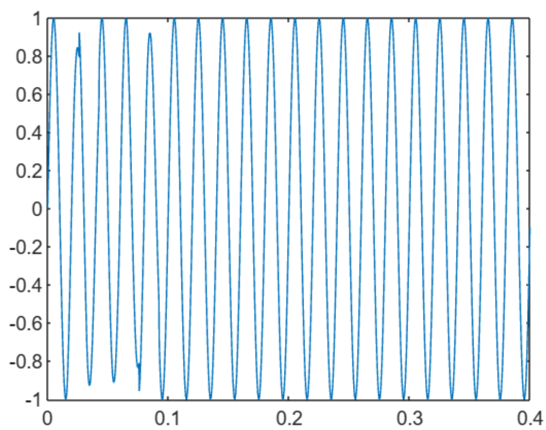


Figure 8 Periodic Notch

$$s(t)=\sin(2\pi f t-\psi)-sign(\sin(2\pi f t-\psi))\times\{\Sigma\rho[u(t-(ts-0.02n))-u(t-(ts-0.02n))]\}$$

Parameters: $0 \leq t2,\ t1 \leq 0.5T$; $0.01T \leq t2-t1 \leq 0.05T$;$0.1 \leq \rho \leq 0.4$

2.1.9 Harmonics (Distortion): When harmonic frequencies are added to a voltage or current waveform at 60 hertz (Hz), distortion results, making the typically smooth wave appear jagged or deformed. Solid state components like rectifiers, variable speed controllers, fluorescent lights, and even computers are capable of introducing distortion.

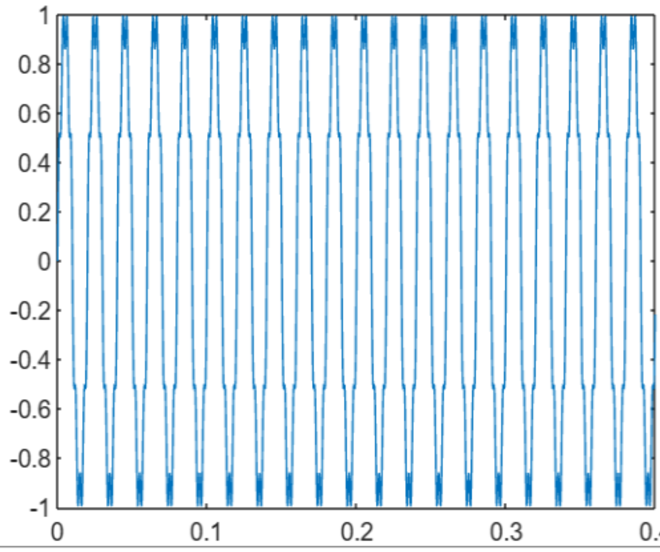


Figure 9 Harmonic Distortion

$$s(t)=\sin(2\pi f t-\psi)+\Sigma\lambda i*\sin(2\pi f c t-\psi i)$$

Parameters: $0.05 \leq \lambda i \leq 0.15$

200 signals of each class were generated. The sampling frequency is set as 3200 Hz and the signal’s fundamental frequency is set 50 Hz, and signal duration is 0.4 s. A gaussian noise of 30 dB is then added to the signals. So a total of 1800 signals were generated.

2.1.10 Signals after adding noise

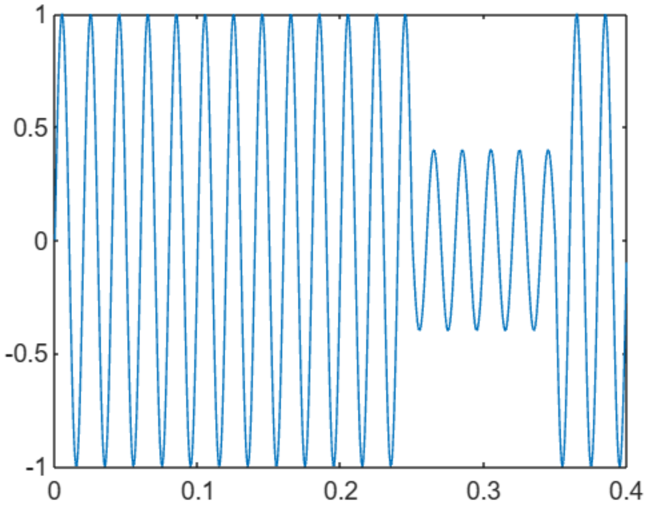


Fig. Sag Signal without noise

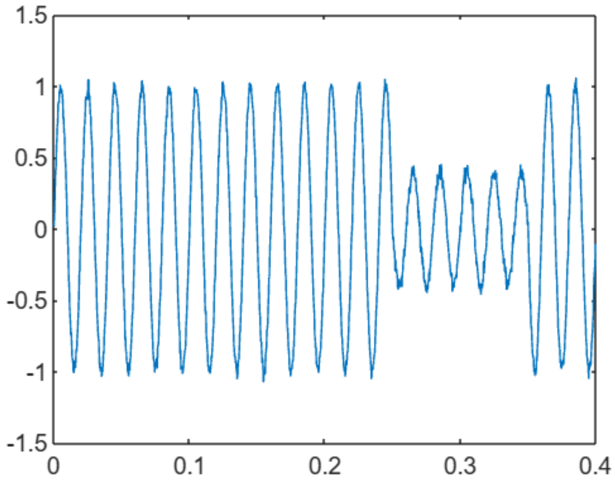


Fig. Above sag signal with noise

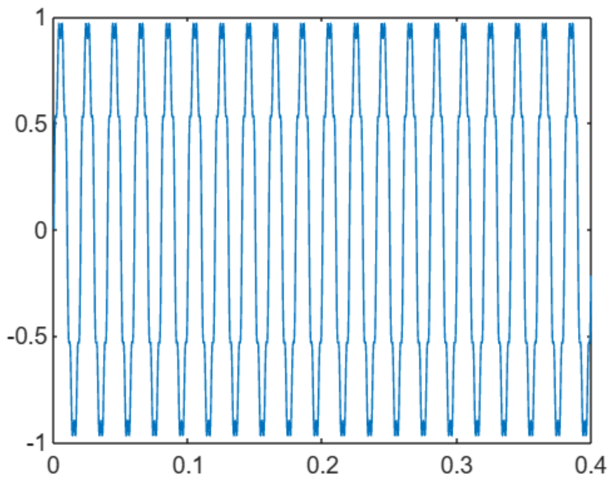


Fig. Harmonic Signal without noise

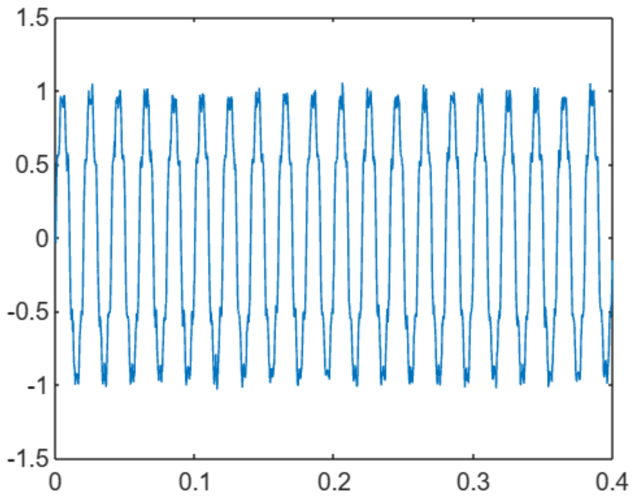


Fig. Above Harmonic signal with noise

2.2 Stockwell-Transforms

The Stockwell-Transform or commonly referred as 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.

2.3 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

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

2.4 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,…,N-1*.

Let

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

Discrete S-transform expression will be

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

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

2.5 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 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 the 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 are 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 :

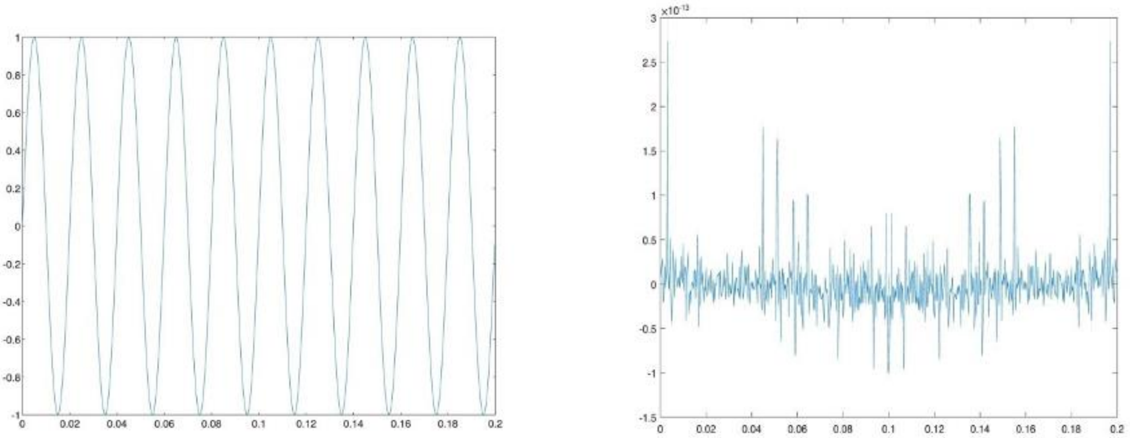


Figure 10 Signal a and FFT of Signal a

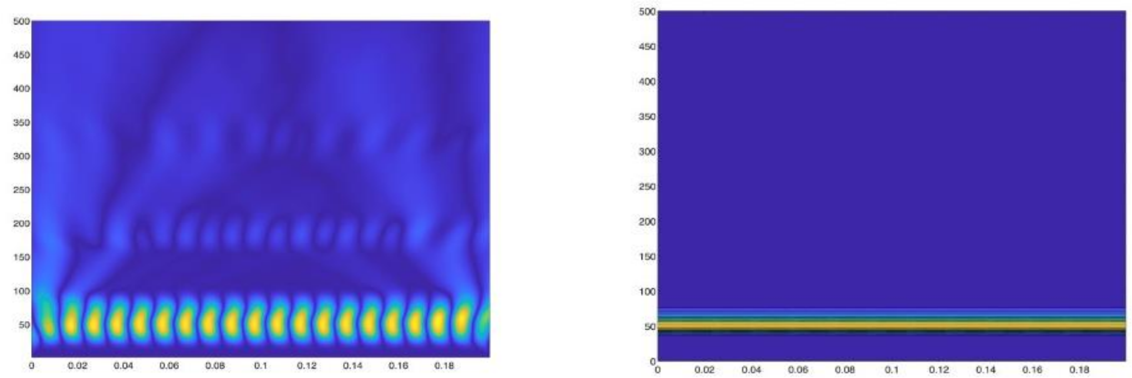


Figure 11 CWT of Signal a and ST of Signal a

Swell:

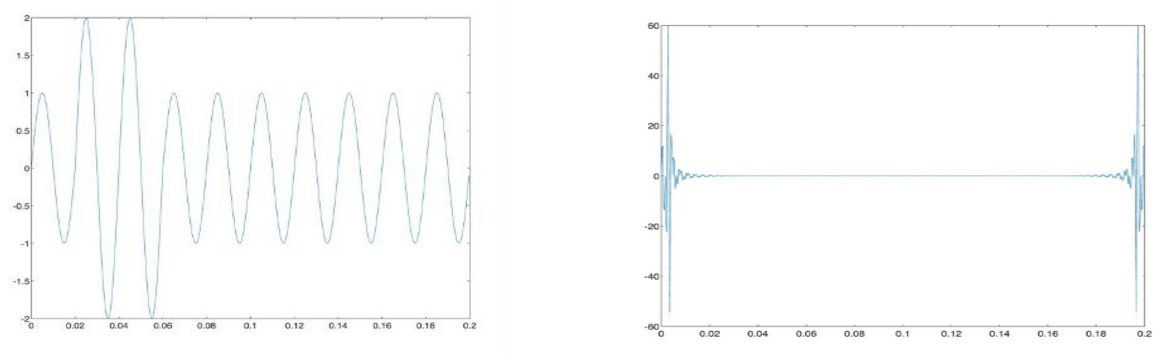


Figure 12 Signal b and FFT of Signal b

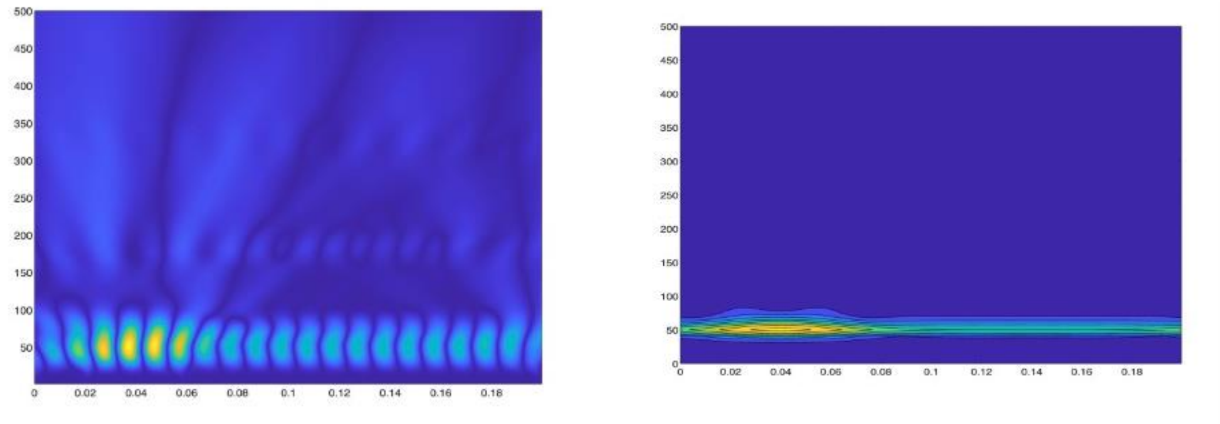


Figure 13 CWT of SIgnal b and ST of Signal b

Sag:

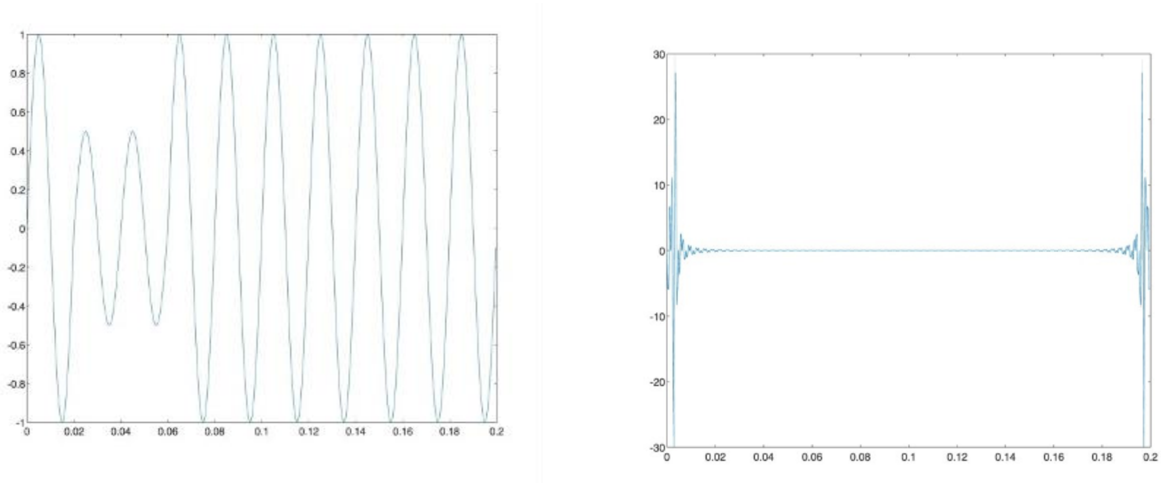
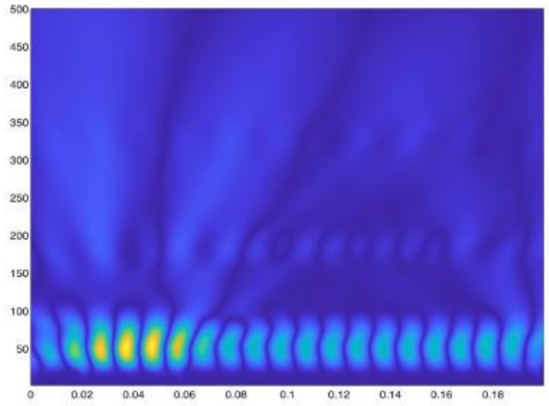


Figure 14 Signal c and FFT of Signal c



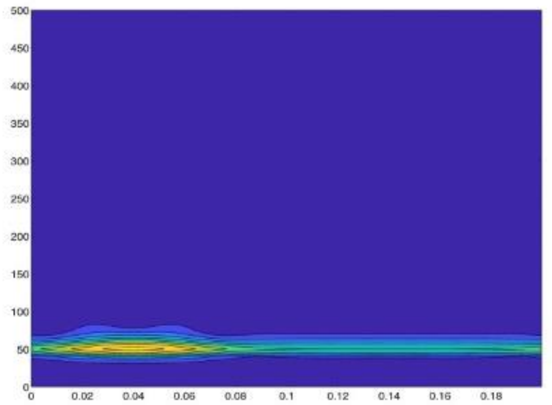


Figure 15 CWT of Signal c and ST of Signal c

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

2.6 Kaiser Window Based S-Transform

The Kaiser window is a family of window functions that are parameterized by a single parameter, called the beta parameter. The beta parameter controls the tradeoff between the main lobe width and the side lobe level of the window. A larger beta value results in a narrower main lobe and lower side lobe levels, while a smaller beta value results in a wider main lobe and higher side lobe levels.

The Kaiser window is defined as:

$$w_k(t,f) = \frac{I_0\left[\alpha(f)\sqrt{1-\left(\frac{T}{t}\right)^2}\right]}{I_0[\alpha(f)]},\ |t| \le T$$

Where $I_0(.)$ is the modified Bessel function of the first kind of order 0, $\alpha(f)$ is a function related to frequency and beta is the parameter that controls the shape of the window.

$$I_0(.) = 1 + \sum_{m=1}^{\infty} \left[\frac{(\cdot/2)^m}{m!} \right]$$

The Kaiser window has several desirable properties, such as good stopband attenuation, good spectral resolution, and adjustable main lobe width and side lobe levels. It is widely used in digital signal processing applications where high performance is required.

The Kaiser window is often used in the S-Transform as the window function because of its desirable properties, such as adjustable main lobe width and side lobe levels. The window length is typically chosen to be an odd number to avoid phase shifts in the time-domain representation.

The S-Transform with the Kaiser window has the form:

$$KS(\tau, f) = \int_{-\infty}^{+\infty} x(t) \cdot w_k(\tau - t, f) e^{-i2\pi f t} \cdot dt$$

where *x(t)* is the input signal, $w_k(t)$ is the Kaiser window function, t is the time index, f is the frequency index. The Kaiser window function is parameterized by a single parameter, β, which controls the tradeoff between the main lobe width and the side lobe levels of the window. The parameter β is typically chosen based on the desired main lobe width and side lobe levels.

The S-Transform with the Kaiser window has several advantages over other time-frequency analysis methods. It has high time and frequency resolution, good localization in both time and frequency, and good sensitivity to frequency changes. It is useful for analyzing non-stationary signals with rapidly changing frequency content, such as signals with transient or impulsive components.

2.6.1 Spectrograms with Kaiser Window

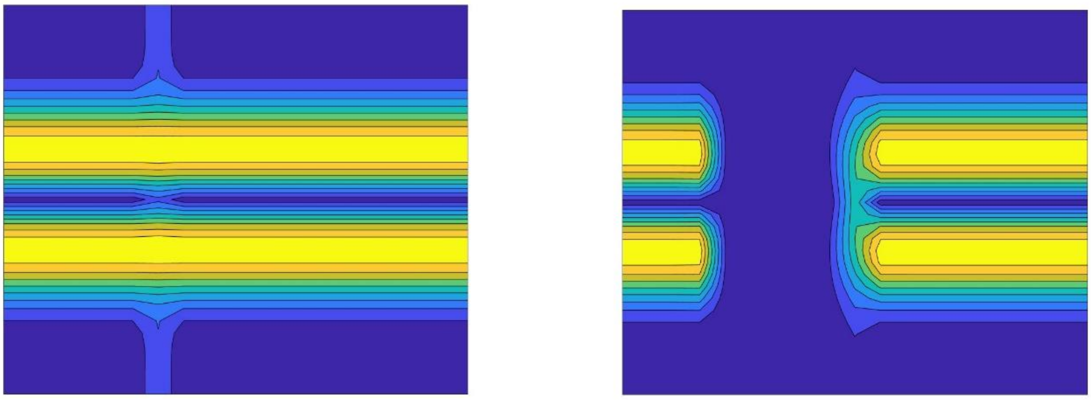


Figure 16 Impulsive Transient and Interruption

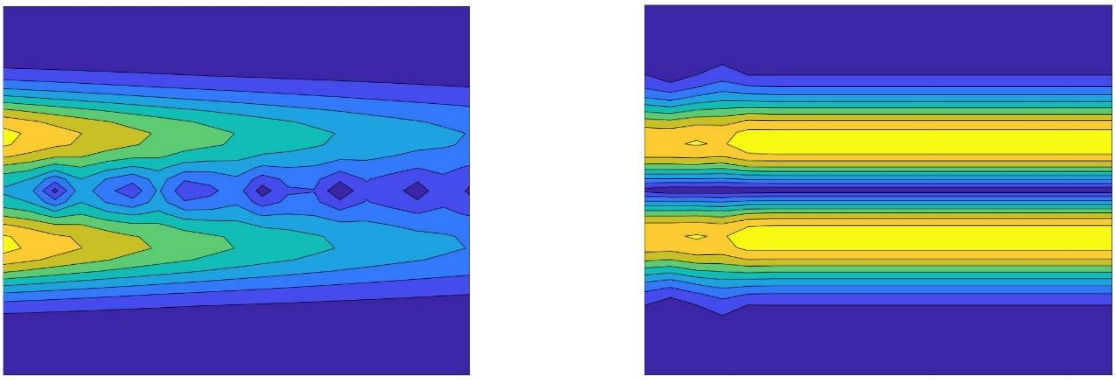


Figure 17 Oscillatory Transient and Periodic Notch

2.6.2 Comparison between S-Transform and Kaiser Window based S-Transform

S-Transform and Kaiser window based S-Transform are both time-frequency analysis techniques that can be used to analyse signals with time-varying spectral content.

The main difference between the two methods is the way in which they handle the trade-off between time and frequency resolution.

S-Transform uses a Gaussian window function to transform the signal into the time-frequency domain. The width of the Gaussian is proportional to the instantaneous frequency of the signal, which allows for excellent time-frequency resolution. However, the Gaussian window function is not ideal for signals with sharp transitions or discontinuities, as these can cause spectral leakage.

On the other hand, Kaiser window based S-Transform uses a Kaiser-Bessel window function, which has better spectral resolution and less spectral leakage than the Gaussian window. The Kaiser-Bessel window also has a tuneable parameter called the beta value that allows the user to adjust the trade-off between time and frequency resolution. However, the Kaiser-Bessel window has a wider main lobe than the Gaussian window, which can result in lower time resolution.

Overall, the choice between S-Transform and Kaiser window based S-Transform depends on the specific application and the characteristics of the signal being analysed. S-Transform is suitable for signals with smooth spectral content, while Kaiser window based S-Transform may be more appropriate for signals with sharp transitions or discontinuities.

2.7 SUPPORT VECTOR MACHINE

A well-liked supervised learning technique for classification and regression applications is the Support Vector Machine (SVM). The technique divides the data into various classes by locating a hyperplane in a high-dimensional space. B.E. Boser et al. originally presented SVMs in 1992. SVM's capacity to handle high-dimensional data with a limited number of training samples is one of its primary advantages. SVM is a good choice for mathematical analysis and optimisation because it also has a solid theoretical base.

The hyperplane is the line that divides the two classes.Because only these two vector points and no other points contribute to the algorithm's output, the vector points closest to the hyperplane are known as the support vector points. Finding the hyperplane that maximises the margin between the support vectors is the fundamental tenet of SVM. To help us divide our space into classes, we therefore identify the optimal line in two dimensions or the best hyperplane in more than two dimensions. The distance between the hyperplane and the nearest data points for each class is referred to as the margin. SVM can discover a solid decision boundary that applies well to new data by maximising the margin.

Both linear and non-linear classification tasks can be done with SVM. The SVM identifies a hyperplane that divides the data into two groups in the case of linear classification. SVM employs a method known as the kernel trick for non-linear classification, which transforms the input data into a higher-dimensional space where a linear hyperplane can be utilised to divide the data.

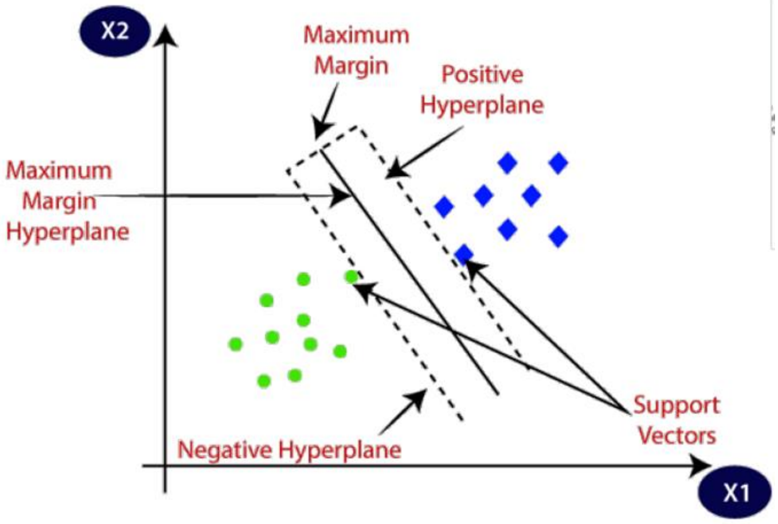


Figure 18 showing the support vectors and hyperplanes (source: google)

SVM can be extended for multiclass classification as well. There are three methods – one v/s one, one v/s all and Directed Acyclic Graph (DAG) SVM. For our work we chose the one v/s all method for the classification. In this method we classify each class separately from the other classes. So we have to train as many SVMs as there are number of classes.

For our work we used a simple SVM where we trained 9 different SVMs to classify the 9 different classes of the signals. Before feeding the data into the model, the data is standardized using standard scalar.

2.7.1 Mathematics behind SVM

As it can be seen in the fig. 18 , our goal is to maximize the distance between the positive hyperplane and negative hyperplane.

The equation of positive hyperplane is given by

$\omega^{T}x_1 + b = 1$

The equation of positive hyperplane is given by

$\omega^{T}x_2 + b = -1$

where ω is weight vector; b is bias and x_1 and x_2 are the respective support vectors.

Subtracting eq.2 from eq.1. So the equation will be as below:

$$\omega^{\mathrm{T}}(x_1 - x_2) = 2$$

Now the distance between x_1 and x_2 is given by

$$(x_1 - x_2) = \frac{2}{||\omega||}$$

So now essentially, we have to minimize the reciprocal of the right-hand side of the eq. 4. with the following constraint:

$$(\omega^{\mathrm{T}} x_i + b) \geq -1 , y_i = 1$$

$$(\omega^{\mathrm{T}} x_i + b) < -1, y_i = -1$$

where y_i is the label.

The loss function is given by

J = $max\left(0,1 - y_i * (\omega^{\mathrm{T}} x_i + b)\right) + \lambda * ||\omega|| * ||\omega||$

where λ is the regularization parameter.

Gradient for the optimization is given by

$\frac{dJ}{d\omega} = 2\lambda\omega$ and $\frac{dJ}{db} = 0$ if $y_i \cdot (\omega^{\mathrm{T}} x_i + b) \geq 1$

or $\frac{dJ}{d\omega} = 2\lambda\omega - y_i x_i$ and $\frac{dJ}{db} = 2\lambda\omega - y_i$ if $(\omega^{\mathrm{T}} x_i + b) < 1$

For our SVM, we used Gradient descent algorithms for the quadratic optimization. The hyperparameters that are set are -number of iterations (1000) , learning rate (0.001) and the lambda parameter (0.01) as regularization parameter.

2.8 Feature Extraction

Feature extraction is performed in order to give the input data a better representation that the SVM algorithm can utilise to discover the underlying patterns in the data, There are numerous methods for extracting features, and the selection of features can have a big impact on how well the SVM model performs. The following are the statistical features extracted from the S-Matrix obtained from the output of kaiser window-based S-Transform as used in [1].

i. **Feature-1: The standard deviation derived from median of S -matrix.**

   A set of values' variance or dispersion is measured by the standard deviation. To calculate this feature, we first calculate the median of each row of the S-matrix. We then take the standard deviation of the medians across all rows. The resulting value represents the amount of variation in the median values of the S-matrix.

ii. **Feature-2: The standard deviation derived from maximum amplitude of S-matrix.**

   To calculate this feature, we first calculate the maximum amplitude of each row of the S-matrix. We then take the standard deviation of the maximum values across all rows. The resulting value represents the amount of variation in the maximum values of the S-matrix.

iii. **Feature-3: The maximum amplitude derived from kurtosis of S -matrix.**

   To calculate this feature, we first calculate the kurtosis of each row of the S-matrix. We then take the maximum amplitude across all rows.

iv. **Feature-4: The standard deviation derived from kurtosis of S -matrix.**

To calculate this feature, we first calculate the kurtosis of each row of the S-matrix. We then take the standard deviation of the kurtosis values across all rows. The resulting value represents the amount of variation in the kurtosis values of the S-matrix.

v. **Feature-5: The standard deviation derived from mean of S-matrix.**

The mean of each row of the S-matrix is first calculated in order to calculate this feature. We then take the standard deviation of the mean across all rows. The resulting value represents the amount of variation in the mean values of the S-matrix.

vi. **Feature-6: The standard deviation derived from skewness of S-matrix.**

The mean of each row of the S-matrix is first calculated in order to calculate this feature. We then take the standard deviation of the skewness values across all rows. The resulting value represents the amount of variation in the skewness values of the S-matrix.

2.9 Description of the Deep Learning Model (CNN)

This report employs a CNN-based classifier after converting the signals to S-matrices. CNN is a typical feed-forward deep neural network. LeNet, AlexNet, ResNet, GoogLeNet, and other great CNN-based convolutional neural networks have been developed since the early years and have improved the performance of image classification models. Based on AlexNet, we created a CNN structure for this research that comprises of a convolutional layer, a pooling layer, a dropout layer, a pooling layer, and an activation function between layers, as shown in the picture below. The table displays the parameters for each layer.

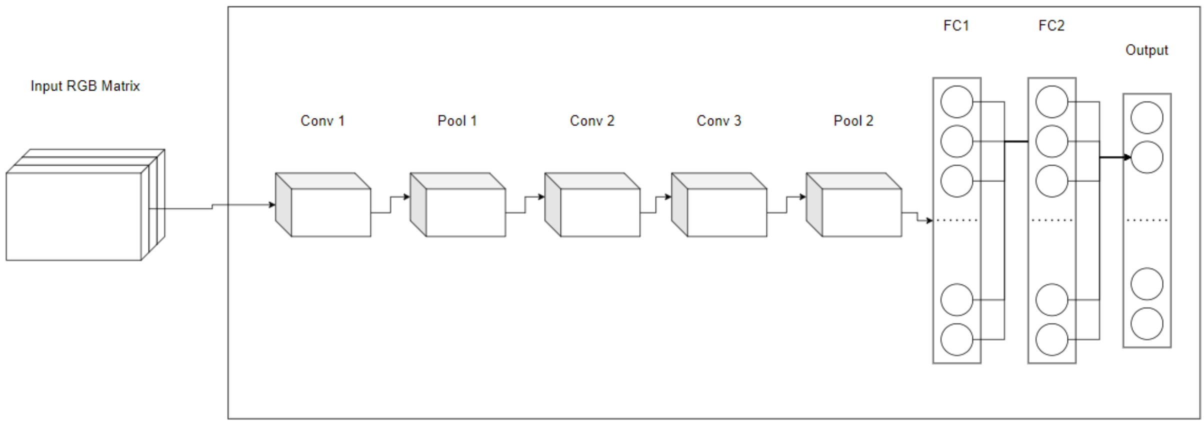


Figure 19 Proposed CNN-based Classifier

2.9.1 Input layer

The previous step's S-transform and parameter optimisation on the contour image will be used to break it down into an S matrix, which will be used as the input to a deep neural network with an input size of 224*224*3.

2.9.2 Convolution layer

A filter size of 3 for Convolution 1 is finalized through experimentation. In the later convolution layers, a 1*1 convolution kernel is used, which enhances the network's depth and enables it to extract diversified feature details.

2.9.3 Pooling layer

The pooling layer operates analogous to the convolution layer, with the exception that averaging or finding the maximum value is performed instead of convolution, as shown below. The pooling layer minimises the number of parameters and achieves feature dimensionality reduction by concentrating on the existence of particular characteristics rather than their precise position. We use the Max Pooling technique in this project.

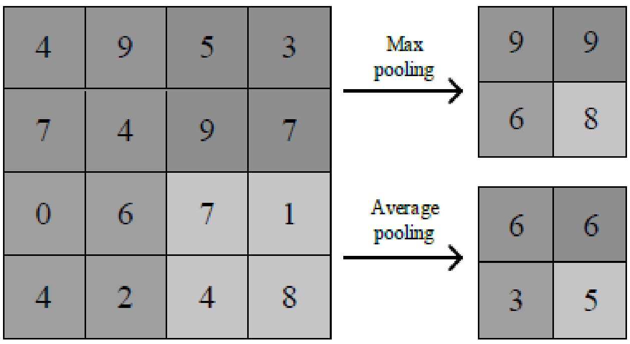


Figure 20 Pooling Method

2.9.4 Activation functions:

The set of specified inputs and outputs between the layers is provided by the

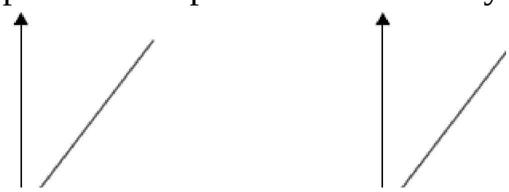


Figure 21 Leaky ReLU (left), ReLU (right)

activation function. The activation functions that are most frequently used are Sigmoid, Tanh, SoftMax, ReLU. Here, we use the Leaky ReLU variant of ReLU, which is depicted in the image below. The activation functions of Relu and Leaky Relu are both linear in the segmentation interval and operate faster than others.

Leaky ReLU can solve the zero-gradient issue for negative-valued parameters better than ReLU.

After constructing the neural network structure as described above, the output of the last layer of the fully connected layer is 1 × 9, corresponding to the kinds of C1 to C9 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.

The output of the final layer of the fully connected layer after building the neural network structure as previously mentioned is 1*9, which corresponds to the kinds of C1 to C9 in the labels. The output of this layer uses the SoftMax activation function, as shown in the equation below, in contrast to the other Leaky ReLU layers. The output of the SoftMax function is equal to the exponent of one output divided by the exponent of all outputs, which is apt for multi-classification issues.

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

2.9.5 Fully connected layer and dropout layer

Similar to a typical ANN, the fully connected layer's neurons are interconnected in a way that allows backpropagation to update each neuron's weights. Additionally, two-dimensional input data must be tiled into one dimension before being fed into the completely connected layer. By turning off specific neurons to create a fully connected layer with a specific degree of sparsity, the dropout layer can be used to prevent overfitting of the model.

Table 2.9.1 CNN NETWORK STRUCTURE

Layer	Filter Size/Stride	Filter Number
Convolution 1	3/[2]	32
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	9/-	

CNN has the ability to automatically extract, choose, and output the image's features. The output layer labels, which range from Class 1 to Class 9, indicate various PQDs.

CHAPTER 3

RESULTS AND CASE STUDY AND ANALYSIS

3.1 Experimental Setup:

There are 9 classes of signals C1 to C9 to be classified. 200 signals of each class were generated. The sampling frequency is 3200 Hz, the signal's fundamental frequency is 50 Hz, and signal duration is 0.4 s. 80% of the generated spectrograms were used for training and the remaining 20% for testing. Stochastic gradient descent is used for optimization and the learning rate is set as 0.01.

3.2 Kaiser Window based S-Matrix:

The signals with PQDs generated in MATLAB, applying Kaiser window based S-Transform on them to perform feature extraction on the sequence, an RGB image of the spectrogram was obtained.The X-axis of the spectrogram represents the frequency feature and the Y-axis information represents the time feature.

The Kaiser window based S-Transform spectrograms of the 1,800 signals are generated, plotted and saved as images. This data is then fed into the CNN model for classification.

3.3 Results:

- The SVM model gives an accuracy of 100%.
- The CNN model gives an accuracy of 99.14% when spectrograms of Kaiser window based S-transforms of signals are given as input.
- The CNN model gives an accuracy of 98.85% when spectrograms of S-transforms of signals are given as input.

3.4 Comparison of results between SVM and CNN models

Models	Train Accuracy	Test Accuracy
SVM (taking statistical features)	100%	99.7%
CNN (taking Kaiser window based S-transform as input)	97.13%	99.14
CNN (taking S-transform as input)	96.78%	98.85%

Table 3.4.1 Comparision of Accuracy

	C1	C2	C3	C4	C5	C6	C7	C8	C9
C1	197	0	0	0	0	0	1	0	2
C2	0	196	0	0	0	0	0	0	4
C3	39	0	161	0	0	0	0	0	0
C4	0	0	0	200	0	0	0	0	0
C5	0	0	0	0	199	0	0	1	0
C6	4	0	0	0	0	196	0	0	0
C7	0	0	0	0	0	0	200	0	0
C8	0	0	0	0	0	0	0	200	0
C9	0	0	0	0	0	0	0	0	200

Table 3.4.2 Confusion matrix obtained from CNN model

	C1	C2	C3	C4	C5	C6	C7	C8	C9
C1	200	0	0	0	0	0	0	0	0
C2	0	200	0	0	0	0	0	0	0
C3	0	0	200	0	0	0	0	0	0
C4	0	0	0	200	0	0	0	0	0
C5	0	0	0	0	200	0	0	0	0
C6	0	0	0	1	0	199	0	0	0
C7	0	0	0	0	0	0	200	0	0
C8	0	0	0	0	0	0	0	200	0
C9	0	0	0	0	0	0	0	0	200

Table 3.4.3 Confusion matrix obtained from SVM model

C1 -Flicker C2 -Harmonics C3 -Impulsive Transient C4 -Interruption C5 -Normal C6 -Oscillatory Transient C7 -Periodic Notch C8 -Sag C9 -Swell

In confusion matrix, the rows are the predicted classes where as the columns are the actual classes.

CHAPTER 4

CONCLUSION AND FUTURE WORK

The results obtained from our project, titled "Power Quality Disturbance Classification using Kaiser-window based S-transform, Support Vector Machine and CNN," have shown promising results with an SVM accuracy of 100% and a CNN accuracy of 99.14%. The classification of power quality disturbances is crucial for the efficient functioning of electrical power systems. Accurately identifying the type of disturbance can help in taking corrective measures and ensuring the smooth operation of the system.

The combination of Kaiser-window based S-transform, SVM, and CNN proved to be an effective method for classifying power quality disturbances. The S-transform provides excellent time-frequency localization, allowing for the precise detection of disturbances in the signal. The SVM algorithm is a powerful machine learning algorithm that has been widely used for classification tasks. It works by finding the best hyperplane that separates the data into different classes. The CNN, on the other hand, is a deep learning algorithm that has been shown to perform well in image and signal classification tasks. It works by learning features at different levels of abstraction, making it well-suited for complex classification tasks.

The high accuracy obtained from both SVM and CNN models suggests that the proposed method is reliable and effective in classifying power quality disturbances. The SVM model, in particular, showed perfect classification accuracy, indicating that it can accurately distinguish between different types of power quality disturbances. The CNN model also showed high accuracy, with only a small number of misclassifications.

In conclusion, the proposed method combining the Kaiser-window based S-transform, SVM, and CNN has shown excellent performance in classifying power quality disturbances. The results obtained in this project can be used in the development of more efficient and reliable power systems. The high accuracy obtained from the SVM and CNN models suggests that these methods can be used in real-world applications for the classification of power quality disturbances. Future work can focus on improving the computational efficiency of the proposed method and exploring its application in real-world power systems.

REFERENCES

[1] Mohd Saiful Najib b. Ismail et.al , “Single and Double PQ Disturbances Classification Based on Stockwell transform and Support Vector Machine”, *2021 IEEE International Conference in Power Engineering Application (ICPEA), 8-9 March 2021.*

[2] Wang, X., Ma, L., Chen, Y., & Yang, G. (2019), “Power quality disturbance classification using convolutional neural network based on S-transform”. *Journal of Physics: Conference Series, 1283(2), 022060.*

[3] Gao, S., Xie, J., Zhou, Y., & Yao, Y. (2018),” Power quality disturbance classification using convolutional neural networks with feature transfer learning.” *Energies, 11(10), 2759.*

[4] Li, J., Yang, H., Li, H., & Zhang, J. (2014), “Power quality disturbance classification based on improved SVM and S-transform”. *Journal of Electrical Engineering & Technology, 9(4), 1317-1324*

[5] Zhang, Y., & Jiang, L. (2012). “Classification of power quality disturbances using combined SVM and wavelet packet transform”. *IEEE Transactions on Power Delivery, 27(2), 866-873.*

[6] Li, Y., Chen, Z., & Wang, L. (2007). “SVM-based classification of power quality disturbances using S-transform*”. Electric Power Systems Research, 77(4), 356-362.*