Power Quality Disturbance Feature Extraction

EE 234: Smart Grid Sensors and Data-Driven Applications
Dan O'Connor

June 13, 2024

Abstract

This report analyzes a dataset of synthetically generated electrical signals. There are 15 unique classes of power quality disturbances. Focusing on the techniques learned throughout this course, meaningful features are extracted: root mean square, crest factor, total harmonic distortion, phasor harmonic index, and total interharmoic distortion. Finally, a classification model is used to test the effectiveness of the features.

1 Introduction

Power quality disturbances (PQDs) are deviations in electrical power signals that can lead to equipment malfunction, reduced efficiency, and potential damage. Accurate detection and classification of these disturbances from the waveforms is an important aspect of maintaining situational awareness of the power system. Correct identification of PQD can aid in preventative maintenance and identifying root cause of the disturbance.

2 Data Description

The dataset used in this analysis was produced by Uvesh Sipai, Rajendrasinh Jadeja, Nishant Kothari, and Tapan Kumar Trivedi from Marwadi University. The dataset can be found at : `https://ieee-dataport.org/documents/synthetic-power-quality-disturbances-dataset-single-and-combined-disturbances-generated`. The DOI is `10.21227/035e-rx20`.

The dataset consists of 7,500 synthetically generated electrical signals. Each signal contains 10 cycles at 100 samples per cycle, totaling 1000 samples per signal. It is sampled at 5 kHz. Although not explicitly stated, the data is likely voltage signals given an amplitude of 1 p.u. and the fundamental frequency of 50 Hz.

There are 15 different classes, each containing 500 samples. Although the authors did not provide class descriptions, each class is a single (or combination of) power quality disturbance denoted by IEEE 1159 guidelines - including sags, swells, harmonics, interharmonics, and flickering.

3 Features Extracted out of Each Signal

3.1 Root Mean Square (RMS)

The RMS value across each signal was calculated using Eq. 1. RMS calculated across the signal (10 cycles) provides a broad view of the average amplitude. Figure 1 shows the dispersion of RMS across the different classes. It appears to be a fairly descriptive feature as it separates the classes into three distinct groups. For an RMS value less than 0.65, the signal must belong to classes 2,4,10,12, or 14. An RMS value of 0.65 - 0.8 corresponds to classes 1,5,6,7,8 and 9. Finally, any RMS value above 0.8 belongs to classes 3, 11, 13, or 15.

$$\mathrm{RMS} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2} \tag{1}$$

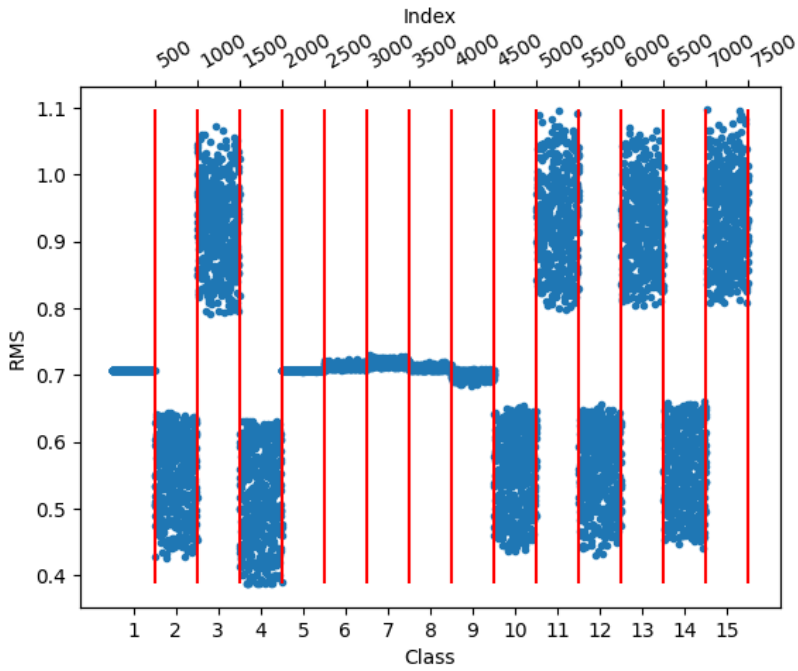

Figure 1: RMS Values across all signals. Each dot is the RMS of a signal. The vertical red lines separate each of the 15 distinct PQD signal classes. The RMS alone can be informative about the waveform. Class 1 (purely sinusoidal) has a uniform value of 0.707. Classes 5, 6, 7, and 8 have near sinusoidal RMS due to small signal variations caused by interharmonics, harmonics, or small fluttering. The other classes have power sags and swells, which directly cause an RMS decrease or increase, respectively.

3.2 Crest Factor (CF)

Crest factor is the ratio of the maximum amplitude of the signal to the signal's RMS value. CF was calculated across each signal using Eq. 2. A high CF indicates high peaks (transient) within the signal, meaning that the amplitude is higher than the average power level of the signal. Lower CF implies a more consistent waveform. Figure 2 shows that most classes display a wide range of values, making it less descriptive of a feature for classification purposes.

$$\mathrm{CF} = \frac{\max(|x_i|)}{\mathrm{RMS}} \tag{2}$$

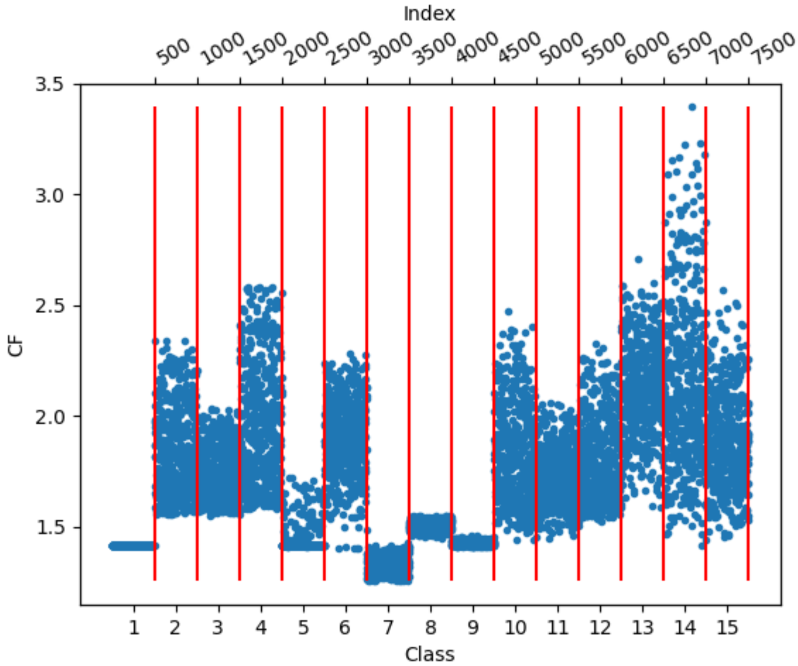

Figure 2: CF Values across all signals, separated by class. The signals with high CF in class 14 exhibit large transient power spikes with several cycles of power sag, creating a lower RMS and a high max amplitude.

3.3 Total Harmonic Distortion (THD)

Harmonics are frequency components of a signal that are integer multiples of the fundamental frequency. For the signals in this dataset, the fundamental frequency is 50 Hz, so the harmonics are 100 Hz, 150 Hz, etc. Total Harmonic Distortion measures the distortion caused by harmonics by taking the ratio of the amplitudes of the 2nd harmonic and onward to the fundamental frequency (harmonic 1). THD is calculated by using Eq. 5. In my analysis, THD was calculated up to the 50th harmonic (2500 Hz).

Eq. 3 shows the general form for a Discrete Fourier Transform (DFT). Essentially, DFT is a technique used to convert a signal from the time domain to the frequency domain. Eq. 4 shows the special case for harmonic frequency extraction, where $\Delta t = \frac{1}{\text{sr}} = \frac{1}{5000}$, f_0 is 50 Hz, and n is the number of samples. In short, Eq. 4 compares each discrete point within the measured signal to a basis function at the harmonics and computes the similarity between the two. Figure 3 shows that most classes have signals that contain harmonics, however, THD does not contain information about which specific harmonics are present.

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-j\frac{2\pi}{N}kn} \tag{3}$$

$$X(h) = \sum_{n=0}^{N-1} x(n) e^{-j2\pi f_0 hn\Delta t} \tag{4}$$

$$\text{THD} = \sqrt{\sum_{h=2}^{\infty} \left(\frac{X_h}{X_1}\right)^2} \tag{5}$$

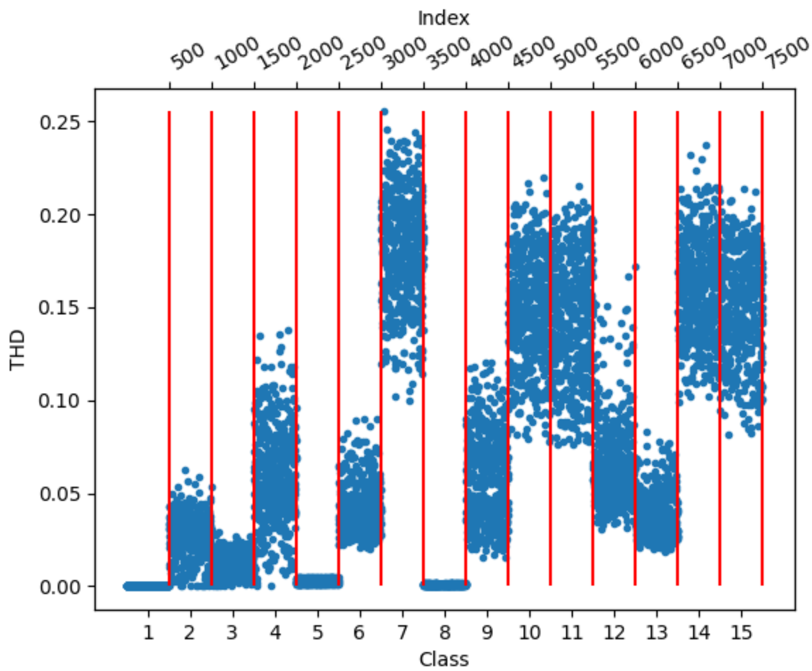

Figure 3: THD Values across all signals, separated by class. Classes 1,5 and 8 contain no harmonics. However, this does not conclude the signal is purely sinusoidal - see 3.5 on interharmonics.

3.4 Phasor Harmonic Index (PHI)

Eq. 4 provides information about the real and imaginary components of the harmonics, and from this we can find the phase angle for the hth harmonics using Eq. 6. PHI is then calculated using 7. PHI is a weighted sum of the cosine difference between the harmonic phase and the fundamental phase normalized by the total harmonic magnitude. The maximum value of PHI is 1, which is the case when all harmonics are in phase with the fundamental harmonic, or when there are no harmonics so the amplitude of the hth harmonic is zero.

$$\theta_h = \tan^{-1}\left(\frac{\text{Im}(X(h))}{\text{Re}(X(h))}\right) \tag{6}$$

$$\text{PHI} = \left(\frac{\sum_{h=1}^{\infty} X_h \cos(\phi_h - \phi_1)}{\sum_{h=1}^{\infty} X_h}\right) \tag{7}$$

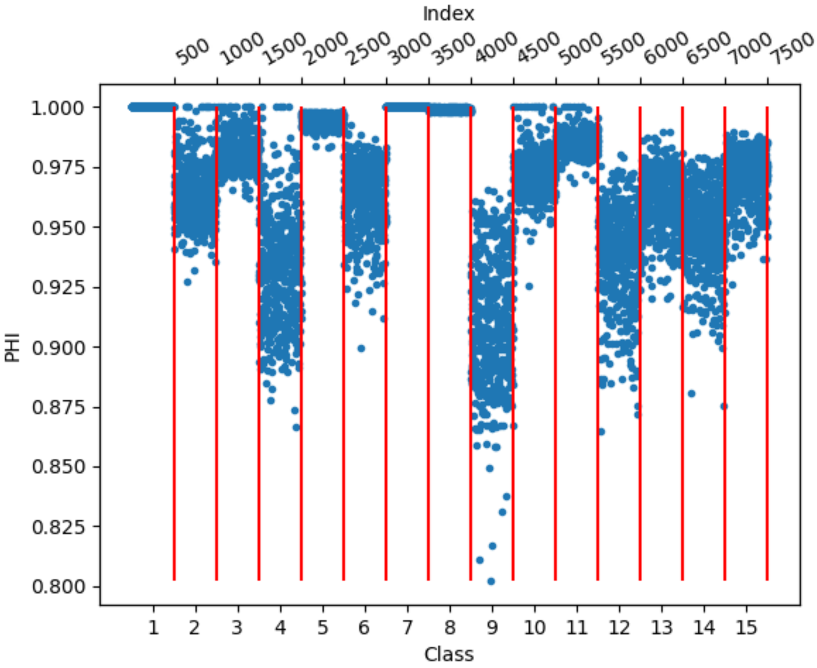

Figure 4: PHI values calculated up to the 50th harmonic, separated by class.

3.5 Total Interharmonic Distortion (TIHD)

Total Interharmonic Distortion (TIHD) is similar to THD, except TIHD includes only frequencies that are not harmonics, see Eq. 8. There is a minimum frequency resolution Δf in DFT, which is defined as $\Delta f = \frac{f_s}{N}$. For this data: $N = 1000 Samples$ and $f_s = 5000Hz$, $\Delta f = \frac{5000}{1000} = 5$ Hz. I calculated TIHD up to 2500 Hz for each signal. In Figure 5, we can see that TIHD values exceed 1 in some samples, indicating strong interharmonics.

$$\text{TIHD} = \sqrt{\sum_{k \neq n \cdot f_1} \left(\frac{X_k}{X_1} \right)^2} \tag{8}$$

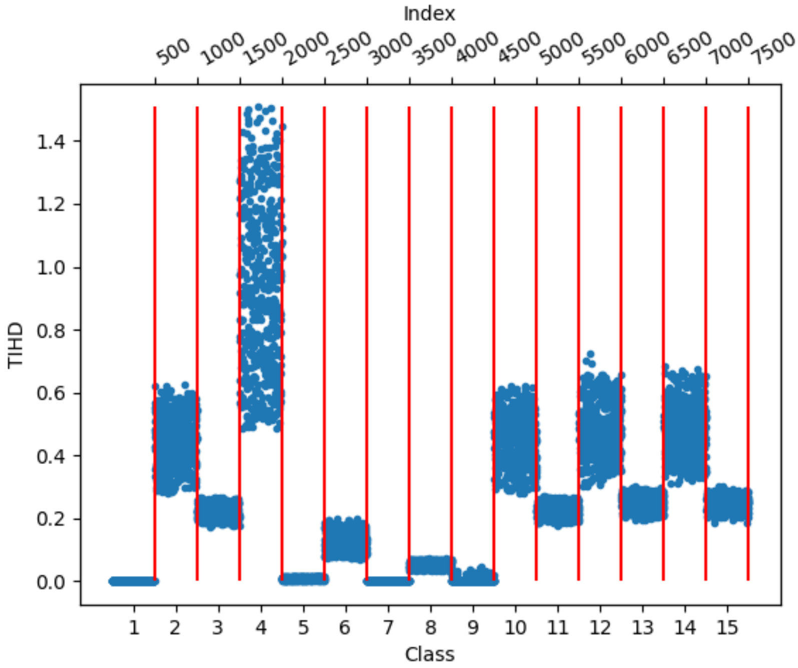

Figure 5: TIHD calculated up to 2500 Hz per sample, separated by class. Note that some classes with 0 values for THD, have non-zero TIHD. The classes mostly have a tight range of TIHD values, except for class 4.

4 Modeling

The purpose of this report was to extract features, however, it is worth seeing how informative these features are for classifying the PQD.

The model selected was Random Forest Classifier from the Sci-kit Learn library in Python. The data was split into 80% training and 20% testing, stratifying to ensure equal class representation in both training and testing. The default hyperparameters were used. The model had 94% accuracy on the testing data, further details about the performance are shown in the confusion matrix in Figure 6.

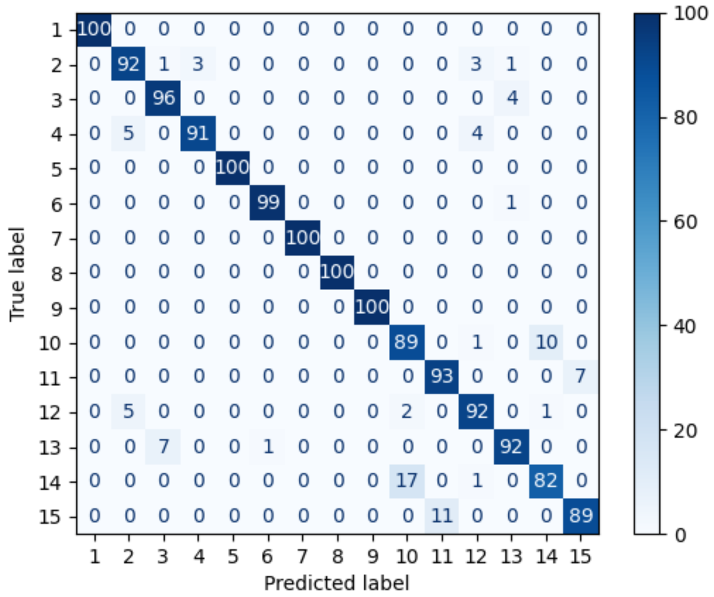

Figure 6: Confusion Matrix. The model did extremely well on classes 1-9. Classes 10-15 contained mixed disturbances; i.e. sags, harmonics, and interharmonics, which resulted in less consistent performance.

5 Conclusion

In this report, synthetically generated electrical signals were analyzed to extract features: RMS, Crest Factor, THD, PHI, and TIHD. These features were then used to classify the signals into 15 distinct classes. The Random Forest Classifier achieved a 94% accuracy on the testing data, demonstrating the effectiveness of the extracted features. This analysis shows the importance of informative features, 5 features can successfully distinguish a PQD from 14 others.