NTNU

Norwegian University of
Science and Technology

A study of fault analysis in power grids using machine learning

TDT4501 Computer Science, Specialization Project

Vegard Hellem
Vemund Mehl Santi

Department of Computer Science

Supervisor: Helge Langseth

Autumn 2018

Abstract

In this report, we summarize the current state-of-the-art methods within power grid analysis. Based on our summary we propose machine learning-based methods for voltage disturbance prediction in the Norwegian power grid that should be further explored.

The power grid is one of the most critical parts of a country's infrastructure. Providing a stable power distribution network is of utmost importance, ensuring that the industry and households have a predictable source of energy. While most power grid infrastructure is old, recent advances in machine learning and data collection volumes allow grid operators to have more fine-grained control of the current state of the power grid than what was previously possible.

With the recent advances in smart grid technology, actors that used to be pure consumers are now becoming both producers and consumers of energy. This change leads to a more rapid change in load on the electrical grid, with variable load at uneven intervals. The changes call for better monitoring and analysis tools, to cope with the added dynamicity from thousands of new producers.

We have reviewed the current state-of-the-art within machine learning techniques used to monitor, analyze and predict the state of power grids. By combining these results with sensor datasets from installed power quality analyzers in the Norwegian power grid, we look the potential and applicability of these machine learning methods for use in power analysis, monitoring and prediction systems in Norway. Built on this analysis, we have come up with recommendations for further research opportunities, which we hope will lead to better monitoring and prediction tools for power grids in the future.

Keywords Norwegian Power grid, Power Quality Analyzers, Machine learning, Voltage analysis, Fault prediction

Acknowledgements

We would like to extend our gratitude to our supervisor Helge Langseth for valuable insight and feedback in the process of writing this report. His weekly feedback has been very helpful for us, as well as steering us in the right direction when we were unsure of which direction to go.

We would also like to extend our thanks to Christian Andresen and Bendik Torsæter at SINTEF Energi, for allowing us access to SINTEF's dataset and previous work, as well as guidance for where to go with our work and a vision for the end result. They have been great partners to work with for discussion of power related issues, as well as narrowing down the scope of our research.

A big thanks also goes out to Abakus, our student association, for a constant supply of coffee at their office, as well as excellent opportunities for procrastination and good conversation.

Table of Contents

List of Tables

List of Figures

Abbreviations

Symbol	=	Definition
AC	=	Alternating Current
AI	=	Artificial Intelligence
CENS	=	Cost of Energy Not Supplied
DC	=	Direct Current
DDG	=	Dynamic Dataset Generator
DSO	=	Distribution System Operator
DT	=	Decision Trees
FFT	=	Fast Fourier Transform
HIF	=	High Impedance Faults
HMM	=	Hidden Markov Models
I	=	Current
LSTM	=	Long Short-Term Memory
NN	=	Neural Networks
NTNU	=	Norwegian University of Science and Technology
PMU	=	Phasor Measurement Units
PQ	=	Power Quality
PQA	=	Power Quality Analyzer
PQI	=	Power Quality Indices
ReLU	=	Rectified Linear Unit
RMS	=	Root Mean Square
RNN	=	Recurrent Neural Network
STFT	=	Short Time Fourier Transform
SVM	=	Support Vector Machines
TSO	=	Transition System Operator
V	=	Voltage
WT	=	Wavelet Transform

Chapter 1

Introduction

1.1 Motivation

The power grid is one of the most important parts of infrastructure in the modern world and it is hard to imagine a world without electricity. The total use of electricity per capita in Norway is one of the highest in the world, primarily because electric power is commonly used to heat residential and industrial buildings [1]. Researchers have recently looked at how to make the grid more robust and secure and are moving towards a Smart Grid architecture to enable more flexible solutions [2]. In Norway, the grid includes over 130 000 km of power lines, divided between the transmission, regional and distribution grid [3]. The net power consumption in Norway in 2016 was 116.6 TWh [4], and the energy consumption is expected to increase in the future [5].

The first part of the Norwegian power grid was built in 1892 and powered a single light bulb in Oslo [6]. Construction thereafter expanded, and large parts of the current infrastructure were built in the 1960s and 1970s. These parts are now reaching the end of their expected lifetime, and the grid now requires substantial renovation, upgrades, and replacements [7].

Since the middle of the 2000s, large investments have been made to expand and improve the power grid infrastructure. In 2017, the total investment in the power supply infrastructure was 32.7 billion NOK, and it is estimated a 25.8% increase to 41.1 billion NOK for 2018 [8]. The recent years' investment cost can be seen in Figure 1.1.

With the combination of aging infrastructure, Norway's hazardous weather, and the sheer size of the power grid, interruptions naturally occur and come with a cost to repair. In 2017, 63% of end-users in Norway were affected by faults, experiencing one or more

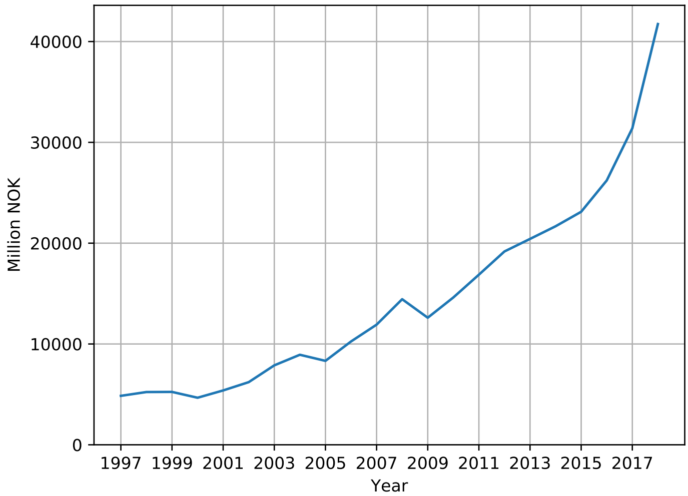


Figure 1.1: Total cost per year for investments in the Norwegian power grid 1997-2018 [7, 8].

interruptions lasting longer than 3 minutes [4]. An average of 793 faults occurred every year between 2008 and 2017 in the Norwegian power grid [9], causing a mean 5 042 MWh of power not to be delivered to consumers. This amounts to only 0.12‰ of the total delivered power, but a cost of 800 million NOK per year in maintenance and repairs [10]. This cost is known as the CENS cost (Cost of Energy Not Supplied). Figure 1.2 shows an overview of CENS costs per county in Norway. If one could anticipate some of these faults, it could have a substantial socioeconomic impact in reducing the costs of maintaining the power grid.

Recent advances in sensors and monitoring of the power grid make such predictions possible. In Norway, there are Power Quality Analyzers (PQA) and Phasor Measurement Units (PMU) in place at strategic sites around the grid, that monitor and send data to a centralized server. This data can be combined with domain knowledge of disturbances in the power grid and machine learning techniques. The result could make it possible to give an early warning of upcoming faults.

1.2 Research Goals

This report is part of the EarlyWarn project, which aims to develop monitoring systems that can detect and identify problems in the high voltage grid in Norway before any significant damage or interruption occur.

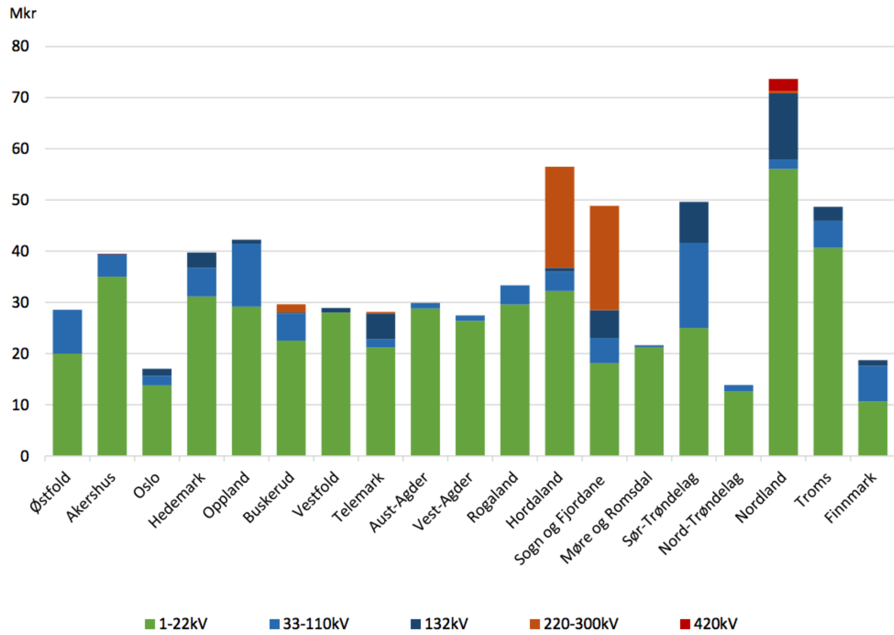


Figure 1.2: An overview of CENS costs for counties in Norway. All amounts in million NOK.

The goal of this report is to explore the techniques of machine learning and power quality analysis that is used in power grids today. We will look into what has worked in the past and what has been done in previous work when it comes to prediction of faults before they occur. We also explore what type of data is needed to have a dataset for training a machine learning algorithm to do fault prediction.

Our work will focus on voltage disturbances and voltage signal analysis. First, we evaluate how to combine previous methods with data from PQA sensors in Norway. Secondly, we investigate their potential and applicability and how they can be used for monitoring and prediction of voltage disturbances. Lastly, we recommend what should be done as further work to explore the application and testing of these methods on data in Norway.

We have formulated the following research questions:

1. How can machine learning be used to predict single line failures in the Norwegian Power grid using data from Power Quality Analyzers?

2. What is the current state of the art in predicting voltage disturbances in power grids?

3. What kind of large-scale monitoring is done in Norwegian power grids today?

1.3 Structure of report

Our report is structured as follows: in Chapter 2 we introduce basic concepts in power analysis. We continue in Chapter 3 with an introduction to machine learning methods. More specifically, we look at machine learning methods used in dynamic systems, such as power grids. Chapter 4 gives a thorough analysis of previous work in the field of machine learning in power analysis, discussing multiple methods and results. In Chapter 5 we present the ongoing field of research in power analysis, monitoring, and prediction in the Norwegian power grid. We consequently look at which methods from Chapter 4 that might be interesting to look at in the case of the Norwegian power grid. In Chapter 6 we explore further work, providing substantial suggestions on future research that might prove useful for developing better power analysis, monitoring and prediction systems in the Norwegian power grid. Finally, in Chapter 7 we conclude both on our research on previous work and on our analysis of how it might be useful for the Norwegian power grid.

Chapter 2

Background - Power Grids

The high voltage power grid in Norway is responsible for the transmission of electric energy across vast distances inside the country as well as distribution to other countries. It is divided between the main transmission grid and a regional grid, The regional grid connects the main transmission grid to the lower voltage distribution grids. The high voltage power grid in Norway is made up 30 000 km of cables [3]. This chapter describes common terms and methods used for describing power and power quality. Section 2.1 describes current and voltage metrics, methods for power analysis and three-phase electric power, the most common method for power transmission. Section 2.2 looks at the most common faults and disturbances in power grids around the world and the reason for faults in the Norwegian power grid.

2.1 Power

2.1.1 Current and Voltage

Definitions

Current (I) is the rate of flow of electric charge past a given point [11]. Current is measured in ampere (A), where 1 ampere is defined as a flow of 1 coulomb of charge past the given point.

Voltage is the cost in energy required to move a unit of positive charge from a point with a lower electric potential to a point with higher electric potential [11]. The unit of measure is the volt (V).

Direct and alternating currents (DC and AC)

Alternating current (AC) is the type of current most commonly used in electric grids throughout the world [12]. In contrast to direct current (DC), alternating current switches the direction of the current flow multiple times per second, while direct current maintains a constant direction of current flow. In Norway, the standard AC frequency is 50Hz. A plot comparing direct and alternating current is shown in Figure 2.1.

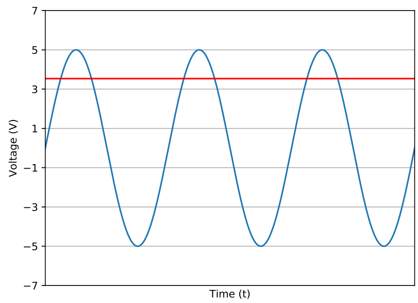


Figure 2.1: Example of a direct voltage level (red) and alternating voltage level (blue).

The advantage of using systems with alternating current compared to direct current is the ability to change the voltage level by the use of transformers. Everyday household electronics in Norway needs a voltage level of 230V, but a much higher voltage level is required for transmitting electricity across large distances. The ability to increase the voltage level before transmission and decrease the voltage level before distribution is a crucial advantage of using AC.

Mathematical functions for voltage A periodic, continuous alternating current can be described by a sinusoidal function of the following form [11].

$$v(t) = \alpha \sin(\omega t + \phi) \tag{2.1}$$

where α is the maximum amplitude, ω is the angular frequency, t is the time and ϕ is the initial phase of the signal.

A sinusoidal function is often described using another mathematical concept called a phasor. A phasor is a complex number used to describe the amplitude and the phase angle of a sinusoidal function. By using Euler's identity [11]:

$$re^{it} = r\cos t + ir\sin t \tag{2.2}$$

where i is the imaginary unit, r is a real number, e is Euler's number and t is time, the real part of the identity becomes the phasor of the signal [13]. Equation 2.1 can then be transformed to:

$$\begin{aligned} v(t) &= \alpha \cos(\omega t + \phi) \\ &= Re(\alpha e^{i(\omega t+\phi)}) \\ &= Re(e^{i\omega t} \cdot \alpha e^{i\phi}) \end{aligned} \tag{2.3}$$

where Re is the real part of a complex number. Using phasor notation, this can be further shortened:

$$\mathbf{\Phi} = \alpha e^{i\phi} = \alpha \underline{/\phi} \tag{2.4}$$

where $\underline{/\phi}$ is the polar form representation of a complex number. A bold typeface is used to symbolize a phasor variable. The shorthand notation thus becomes:

$$Re(\alpha e^{i\phi} \cdot e^{i\omega t}) = Re(\mathbf{\Phi} e^{i\omega t}) \tag{2.5}$$

The relationship between phasor notation and sinusoidal notation is illustrated in Figure 2.2.

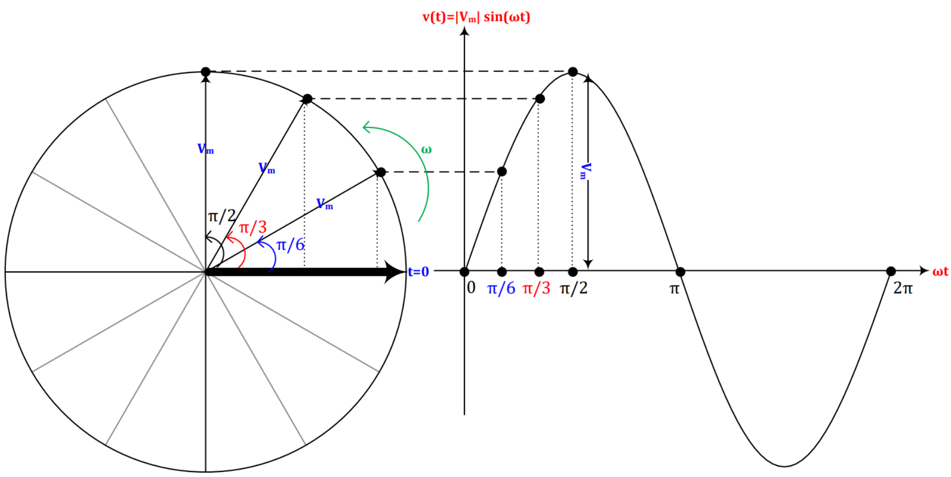


Figure 2.2: The relationship between a phasor representation (left) and a sinusoidal function (right) [13].

2.1.2 Power analysis

When conducting research and analysis of electric circuits, a wide variety of methods and measurements are used. In this section we will describe the following methods and measurements, which are commonly used in signal processing today:

1. Steady state periods
2. Root Mean Square (RMS)
3. Fourier analysis
4. Discrete Fourier analysis
5. Harmonics
6. Short Time Fourier Transform
7. Wavelet Transform
8. S-Transform

Steady state periods

The steady state of a sinusoidal power function is a state where the sinusoidal signal repeats with a constant period T and does not change frequency nor amplitude between periods. If the sinusoidal representation of the signal fulfills the following equality, we say that the signal is in a steady state [14]:

$$\alpha \sin(\omega t) = \alpha \sin(\omega t + nT) \tag{2.6}$$

where n is an arbitrary integer.

A power system should generally be in a steady-state condition, but could after a disturbance, or when starting up, be in a transient condition. A healthy power line should after a short time return to steady-state condition to prevent errors and damages to the equipment [14].

Root Mean Square (RMS)

The Root Mean Square is a standard measurement used for power quality in AC systems [15]. The value is a way of characterizing sinusoidal functions through a mean value, and thus reducing the complexity of comparison functions. The RMS voltage of an AC signal is equal to the voltage of a DC system that transfers the same amount of power [16]. The RMS value of a sinusoidal function is defined as

$$V_{RMS} = \sqrt{\frac{1}{T}\int_0^T v(t)^2 dt} \tag{2.7}$$

where $v(t)$ is a sinusoidal function with period T. By expanding the sinusoidal function and using the trigonometric formula [16]

$$\cos^2(\omega t) = \frac{1}{2}(1 + \cos(2\omega t)) \tag{2.8}$$

we can simplify the equation to a simple expression:

$$\begin{aligned} V_{RMS} &= \sqrt{\frac{1}{T}\int_0^T v(t)^2 dt} \\ &= \sqrt{\frac{1}{T}\int_0^T \alpha^2 \cos^2(\omega t) dt} \\ &= \sqrt{\frac{\alpha^2}{T}\int_0^T \frac{1}{2}(1 + \cos(2\omega t)) dt} \\ &= \frac{\alpha}{\sqrt{2}} \end{aligned} \tag{2.9}$$

where α is the amplitude of the signal and $\int_0^T \frac{1}{2}(1 + \cos(2\omega t))$ is zero, because T is the period of the sinusoidal function [16].

Figure 2.3 shows the relationship between the amplitude of an AC signal and its RMS value. Deviations in the RMS value of a signal is often used to detect faults in the power grid. One often assumes that as long as the RMS voltage remains constant, the system is in a steady state. Note, however, that not all faults and disturbances in the grid can be detected by only monitoring the RMS value [17].

Fourier analysis

Fourier analysis is the study of approximating functions with a finite sum of sinusoidal functions. It is often used in signal processing to decompose signals into separate frequencies, through the use of a Fourier Transform.

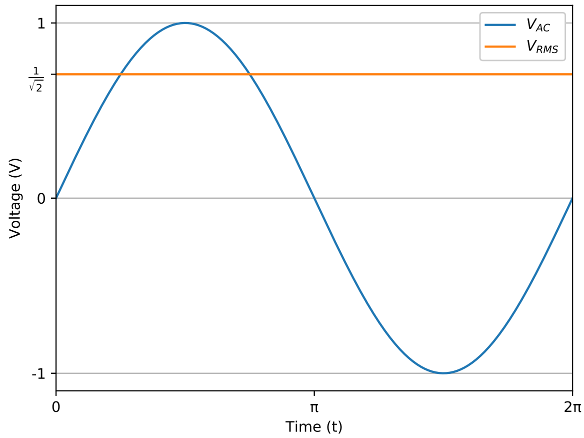


Figure 2.3: AC voltage level (blue) and corresponding RMS voltage level (orange) of a signal with amplitude 1 and period 2π.

A Fourier Transformation is a function $\mathscr{F}$ that takes as input an integrable function, usually a time-based signal, and outputs a function which maps frequencies to Fourier coefficients. By adding together all frequencies multiplied with their Fourier coefficients, we can restore the original function.

A continuous Fourier Transform is defined as [18]:

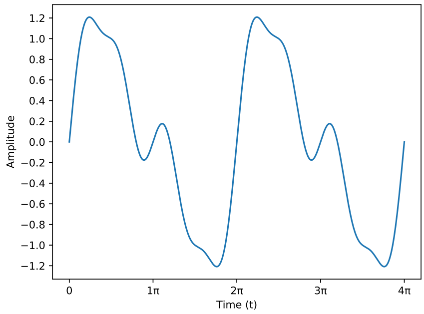
$$\hat{x}(f) = \mathscr{F}(x(t)) = \int_{-\infty}^{+\infty} x(t)e^{-i2\pi ft}dt \tag{2.10}$$

where $\hat{x}(f)$ is a function that outputs the Fourier coefficient for the input frequency f, the original signal function is given by $x(t)$ which takes as input time t, and i is the imaginary unit. One can describe the Fourier Transform as transforming a function from the time domain to a function in the frequency domain.

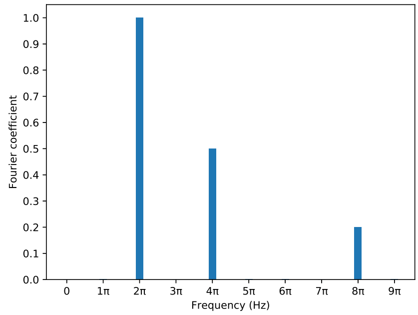
The Fourier Transform is invertible and can thus be used to map frequencies into a time-dependent signal again. The following equation defines the inverse Fourier Transform [18]:

$$x(t) = \mathscr{F}(\hat{x}(f)) = \int_{-\infty}^{+\infty} \hat{x}(f)e^{i2\pi ft}df \tag{2.11}$$

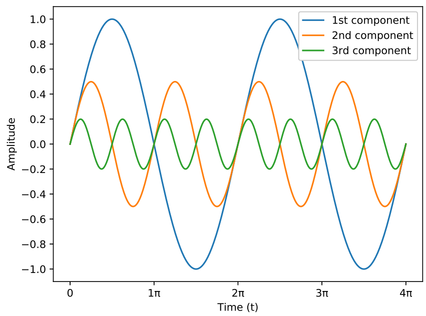
The relationship between a periodic, sinusoidal signal and it's Fourier Transform is shown in Figure 2.4 and 2.5



(a) Sinusoidal signal with a period of 2π.

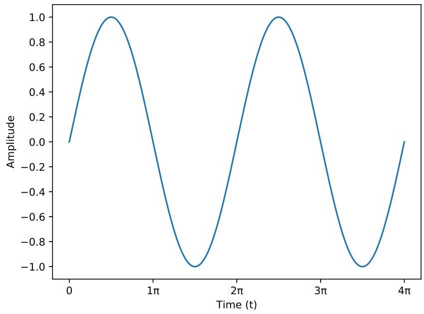


(b) The Fourier coefficients of the periodic signal in 2.4a.

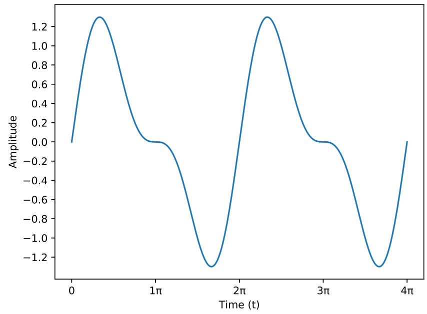


(c) The individual sine components that make up the signal in 2.4a.

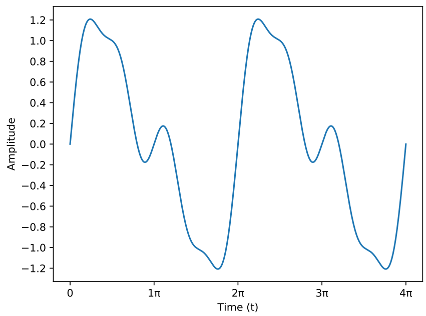
Figure 2.4: Three plots of a periodic signal, it's Fourier coefficients and its corresponding sine components. The periodic signal is given by $\sin(2\pi t) + 0.5\sin(4\pi t) + 0.2\sin(8\pi t)$. Note that the Fourier coefficients are 1, 0.5 and 0.2 at the frequencies 2π, 4π and 8π respectively.



(a) The first sine component of signal 2.4a.



(b) The sum of the first and second sine components of signal 2.4a.



(c) The sum of the first, second and third sine components of signal 2.4a.

Figure 2.5: Three plots showing the gradual composition of individual sine components from signal 2.4a. Notice that the sum of all the components is equal to the original signal.

Discrete Fourier analysis

When performing signal analysis in the real world, the signal $x(t)$ is not recorded as a continuous function, but rather as a series of samples with a given interval T. Each sample k can be regarded as an impulse, with area $x(k)$. The area will then be zero between each sample, and we can transform the definition of the Fourier Transform from an integral into a sum

$$
\begin{aligned}
\mathscr{F}(x(t)) &= \int_{-\infty}^{+\infty} x(t)e^{-i2\pi ft}dt \\
&= x(0) \cdot e^{-i2\pi f0} + ... + x(k) \cdot e^{-i2\pi fk} + ... + x(N-1) \cdot e^{-i2\pi f(N-1)} \\
&= \sum_{k=0}^{N-1} x(k)e^{-i2\pi fk}
\end{aligned}
\tag{2.12}
$$

where N is the number of samples [19].

Harmonics

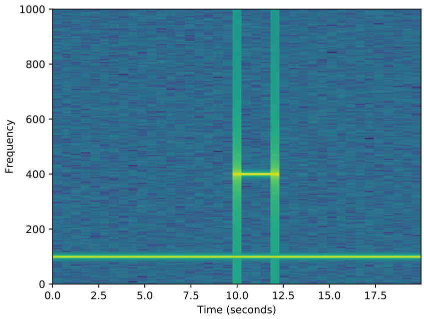
The fundamental frequency of a signal is the frequency where the signal appears to repeat at a rate f, which is usually referred to as the first harmonic. The second, third and further harmonics are then defined as integer multiples of the fundamental frequency, e.g., $2f$, $3f$ and so on. These harmonics are captured by the discrete Fourier Transform where the assumption is that the given signal is periodic.

As the signal is sampled with time interval T, the discrete Fourier Transform captures all the harmonic frequencies up to the $\frac{N}{2}$th frequency, where N is the number of samples per period.

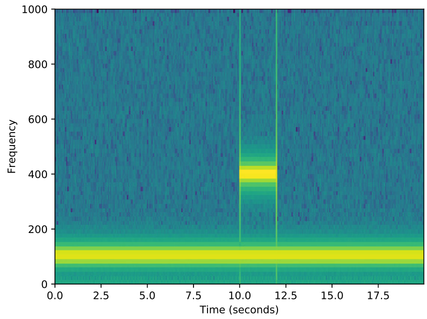
Short Time Fourier Transform

The drawback of using the Fourier Transform to analyze periodic signals is that it completely obscures the temporal locality of the signal. In some applications, it might be beneficial to retain some of the information from the time domain, with a trade-off of some loss in frequency resolution. The Short Time Fourier Transform introduces such a compromise.

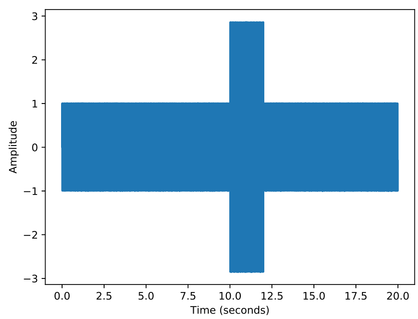
By defining av window of length L along the time dimension of the signal, and computing the Fourier Transform of the signal within this window only, we transform the signal from the time domain into the frequency domain, while still keeping the transformation bound to an interval in the time domain. By moving the window along the time dimension of the original signal, we can transform each segment into its frequency domain, and the result would be a 3-dimensional tensor with time segments, frequencies, and amplitude as its dimensions. The transformation for the signal given by $x(t)$ is defined as



(a) The frequency space created by computing the Short Time Fourier Transform on the signal in 2.6c with a window length of 0.5 seconds.



(b) The frequency space created by computing the Short Time Fourier Transform on the signal in 2.6c with a window length of 0.05 seconds.



(c) A signal given by the function $\sin(100t)$ with an added signal of $2\sin(400t)$ in the interval $[10.0, 12.0]$.

Figure 2.6: Three plots showing the trade-off between long and short window length in Short Time Fourier Transform. A blue color indicates a low amplitude, while a more yellow color indicates a higher amplitude. Plot 2.6a has a long window length, which results in a higher resolution in the frequency dimension at the expense of a loss in precision in the time dimension. Plot 2.6b has a shorter window length, which results in a lower frequency resolution, but a higher precision in the time dimension [20].

$$STFT(\tau, f) = \int_{-\infty}^{\infty} x(t)g(\tau - t)e^{-i2\pi ft}dt \tag{2.13}$$

where τ denote the position of the window in the time domain, f denote the frequency of which to extract the amplitude, i is the imaginary unit and $g(t)$ is the window length function.

Only operating on a smaller segment of the signal reduces the dimensionality of the out-

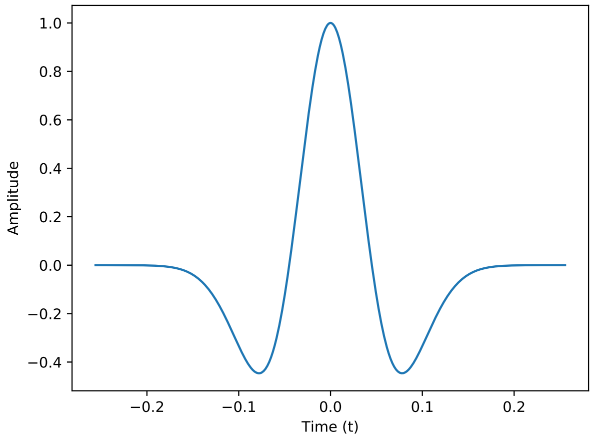


Figure 2.7: An example of a wavelet function. The function is given by $\frac{1-2\pi^2 f^2 t^2}{e^{\pi^2 f^2 t^2}}$ with $f = 5$ and referred to as the Ricker wavelet [21].

put frequency domain, and important information may be lost in this trade-off. One may address this issue by adjusting the window length, where a larger window will lead to a higher frequency resolution, but at the same time reduce the precision in the time dimension. Equivalently, a shorter window will lead to higher time precision, but at the cost of frequency resolution. This trade-off between time and frequency is illustrated in Figure 2.6.

Wavelet Transform

A wavelet is a rapidly-decaying, wave-like oscillation that has zero mean, and where the wave begins and ends with an amplitude of 0. An example of a wavelet function is illustrated in Figure 2.7.

Wavelet Transform is often used in signal analysis. It is similar to the Fourier Transform in that it approximates a given signal by fitting a number of functions to the signal, but differs in which functions it uses for approximation. While the Fourier Transform approximates functions by fitting a finite sum of sine-functions to the original signal, a Wavelet Transform uses wavelets to transform the signal from its time domain into the frequency domain.

The transformation uses a given wavelet, often called the analyzing function, and a varying window length, which is translated along the time dimension and scaled in size. This enables the transformation to capture frequency information at different window sizes, thus varying the time-frequency trade-off depending on the frequency. High-frequency signals will get a high precision in time, while low-frequency signals will get a lower precision in time. The output of a Wavelet Transform is a 3-dimensional tensor describing the frequency space with time translation, scale and amplitude as its dimensions.

S-transform

The S-transform is a generalization of the Short Time Fourier Transform, where the Short Time Fourier Transform is modified to enable varying window sizes, to extract a frequency dependent resolution [22].

The following Gaussian function is used to define the window size:

$$g(t) = \frac{|f|}{\sqrt{2\pi}} e^{-\frac{t^2 f^2}{2}} \tag{2.14}$$

By replacing the window function $g(t)$ in 2.13, we get the following definition for the S-transform [22]:

$$S(\tau, f) = \int_{-\infty}^{\infty} x(t) \frac{|f|}{\sqrt{2\pi}} e^{-\frac{(\tau - t)^2 f^2}{2}} e^{-i2\pi f t} dt \tag{2.15}$$

As the relation between time window length and frequency is inversely proportional, the transform will use a wider window width on low frequencies, and narrower window width on high frequencies. This ensures a better time resolution for higher frequencies and better frequency resolution for lower frequencies.

2.1.3 Three phase power

In single-phase systems, instantaneous power dissipation is changing along with the amplitude of the sinusoidal signal, which results in an uneven power supply. To deal with this problem, one has devised a balanced three-phase system, where the instantaneous power is constant over time. A three-phase power generator is built by placing three coils $\frac{2}{3}\pi$ radians away from each other on a circle, with a rotating magnet in the center. This leads to three different voltage waves, equal in magnitude and frequency but out of phase from each other by $\frac{2}{3}\pi$ radians, with the same properties applying to the generated current. Figure 2.8 illustrates the offset in phase between the three waves. While there are multiple benefits to this system, it also leads to some more possible faults. If one of the phases go out of sync, it will lead to an unbalanced supply [23].

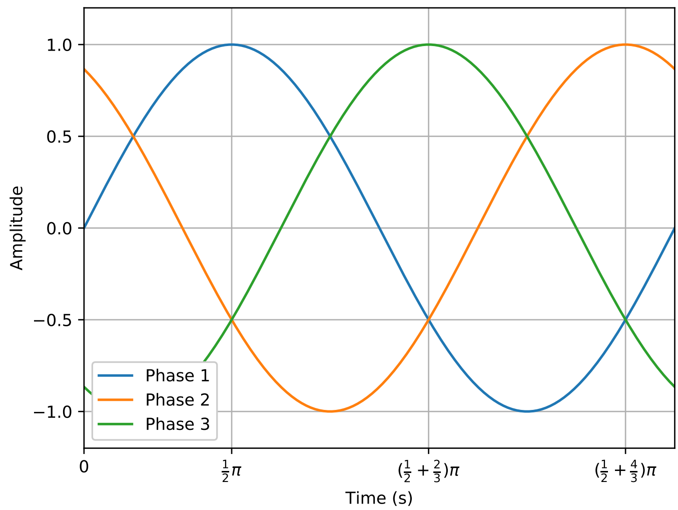


Figure 2.8: Three phase power. The three signals are phase offset by $\frac{2}{3}\pi$ from each other.

2.2 Faults and disturbances

In Norway 2017, a total of 895 disturbances occurred in the electrical power transmission grid. 32.4 % of these consisted of faults on power lines [4]. Of these again, nearly half led to interrupted delivery, meaning power was not delivered to customers, while 13 % caused an interruption duration of more than 30 minutes. On average, interruptions like these cause a socioeconomic cost of 800 million NOK each year, also known as the CENS cost [10]. In this section, we look at the reason for some of these faults, as well as the different types of voltage disturbances that occur in a power grid.

2.2.1 Cause of faults

In a complex power grid, there can be multiple reasons leading to a fault in the electrical network. A power line can go down because of weather, e.g., a tree falling on a line due to the wind, a thunder strike, or icing caused by cold weather. In Statnett's report from 2016 [9], they have mapped the causes of each fault in the Norwegian grid. The results are summarized in Table 2.1.

The surrounding environment alone amounts to 36 % of all faults in the grid for the years 2009-2016, while it is the reason behind 74.7 % of all undelivered power. The surroundings are further categorized into more fine-grained categories, which are listed in Table 2.2.

	Antall driftsforstyrrelser				ILE pga. driftsforstyrrelser			
	Antall		Andel		MWh		Andel	
Utløsende årsak (hovedgruppe)	2016	Årsgj.snitt 2009-2016	2016	Årsgj.snitt 2009-2016	2016	Årsgj.snitt 2009-2016	2016	Årsgj.snitt 2009-2016
Driftspåkjenninger	63	49	7,8 %	6,0 %	214	160	10,4 %	3,0 %
Konstr./montasje mm	48	56	5,9 %	6,8 %	60	148	2,9 %	2,7 %
Mennesker (andre)	7	11	0,9 %	1,3 %	106	57	5,2 %	1,0 %
Mennesker (personale/innleid)	111	91	13,7 %	11,1 %	52	111	2,5 %	2,0 %
Omgivelser	292	345	36,0 %	42,1 %	694	4 051	33,7 %	74,7 %
Teknisk utstyr	178	173	21,9 %	21,0 %	852	623	41,4 %	11,5 %
Tidligere feil	9	4	1,1 %	0,5 %	0	3	0,0 %	0,1 %
Årsak ikke klarlagt	104	92	12,8 %	11,2 %	77	273	3,7 %	5,0 %
Sum	**812**	**821**	**100 %**	**100 %**	**2 057**	**5 426**	**100 %**	**100 %**

Table 2.1: Overview of causes of faults in the Norwegian power grid in 2016 [9].

	Antall driftsforstyrrelser				ILE pga. driftsforstyrrelser			
	Antall		Andel		MWh		Andel	
Utløsende årsak: Omgivelser	2016	Årsgj.snitt 2009-2016	2016	Årsgj.snitt 2009-2016	2016	Årsgj.snitt 2009-2016	2016	Årsgj.snitt 2009-2016
Fugl/dyr	8	8	2,7 %	2,3 %	1	10	0,1 %	0,2 %
Salt/forurensing	15	10	5,1 %	2,9 %	7	21	1,0 %	0,5 %
Snø/is	38	36	13,0 %	10,5 %	223	145	32,1 %	3,6 %
Tordenvær	77	138	26,4 %	39,9 %	120	232	17,4 %	5,7 %
Vegetasjon	43	46	14,7 %	13,4 %	156	589	22,4 %	14,5 %
Vind	86	78	29,5 %	22,4 %	127	2 900	18,4 %	71,6 %
Øvrige	17	20	5,8 %	5,9 %	34	123	4,9 %	3,0 %
Årsak ikke klarlagt	8	9	2,7 %	2,6 %	26	32	3,7 %	0,8 %
Sum	**292**	**345**	**100 %**	**100 %**	**694**	**4 051**	**100 %**	**100 %**

Table 2.2: Faults in the Norwegian power grid caused by surroundings in 2016 [9].

For failures on power lines, fraction of failures caused by surrounding is even higher, as can be seen in Figure 2.9. The surroundings cause nearly all faults that occur on lines in the Norwegian power grid, and the lines are often placed in areas with none nearby to fix potential faults. This shortage of on-site personnel makes it even more beneficial to be able to detect or predict faults ahead of time, so that personnel can be deployed before a potential power outage.

We can see in Table 2.2 that animals, birds and lightning amount to 45 % of all faults related to surroundings in recent years. These causes are out of the scope of this report to predict. We are hypothesizing that other causes, namely vegetation, wind, snow, and

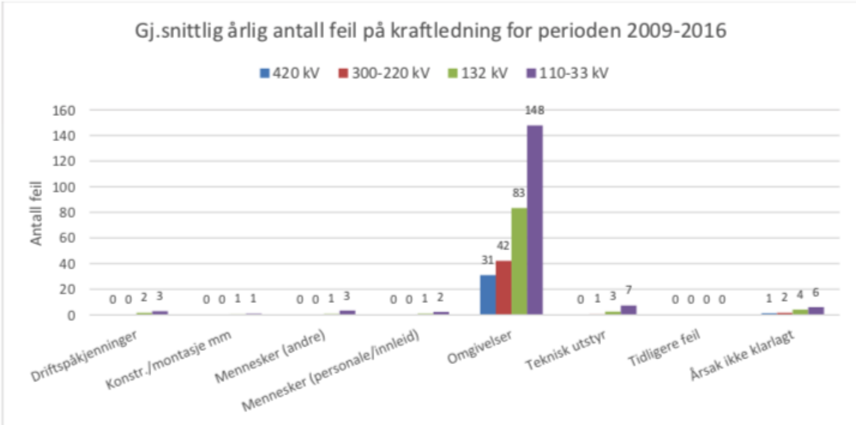


Figure 2.9: Faults in power lines categorized by cause [9].

salt will lead to disturbances in the electric signal that will be detectable before a power failure occurs. A single failure may not be detectable in advance, but the cascading effects on other power lines might be detectable still, and further damage to equipment or loss of power can be avoided by taking the appropriate actions.

2.2.2 High impedance faults

High impedance fault (HIF) is a group of power system disturbances that happens when a conductor makes unwanted electrical contact with another surface element, for example a road, a tree or other vegetation. These faults restrict the flow of current by a level that is usually lower than what is reliably detectable by regular devices [24]. HIF is particularly dangerous since it is not only harmful to the electronic equipment but can also be dangerous to animals and people around it, as HIF can generate inflammable gases resulting in explosions or fires. The detection of these faults have therefore sparked an interest in the research community, and a number of algorithms have been applied to the problem[25]. The presence of high impedance faults can typically be detected when analyzing the waveform of a signal, with asymmetry and extra high- or low-frequency components added to the usual waveform [26].

2.2.3 Voltage disturbances

A power line can experience different disturbances which may eventually lead to a fault. In this section, a list of the most common known voltage disturbances is presented. Each disturbance is described as how it affects a sinusoidal wave signal and some of the possible causes. We will differ between the following 7 categories of disturbances based on

Seymour [27]:

1. Transients
2. Interruptions
3. Sag / Undervoltage
4. Swell / Overvoltage
5. Waveform distortion
6. Voltage fluctuations
7. Frequency variations

Transients

Transients can be categorized into two types of faults: impulsive and oscillatory transients.

Impulsive transients are sudden peaks or surges in voltage level and can be the result of lightning or faults in the equipment used.

Oscillatory transients are changes in the steady-state of a voltage signal, typically causing an increase in voltage, and then a sudden loss, which causes the voltage level to fluctuate back and forth.

Transients are illustrated in Figure 2.10.

Interruptions

An interruption is the complete loss of voltage in the system and can be further categorized dependent on its duration. It is often caused by some damage to the line itself, e.g. from lightning strikes, animals, trees falling on the line, extreme weather or equipment failure. An interruption is easy to spot if it happens at home, as it typically causes all lights and electronic equipment to go black, only to come back shortly after. Voltage readings are useful for detecting interruptions, as the output will be 0 for a period of time, as illustrated in Figure 2.11.

Sag / Undervoltage

A sag is a reduction of the peak AC voltage, where the maximum amplitude drastically lowers for a few periods. This can occur during system faults or from heavy load machinery starting up in the power system. A sag with a duration longer than 1 minute is

(a) Impulsive transient. **(b)** Oscillatory transients.

Figure 2.10: Example waveforms of transient disturbances [27].

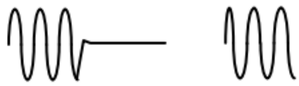


Figure 2.11: Example waveform of an interruption disturbance [27].

(a) Sag. **(b)** Swell.

Figure 2.12: Example waveforms of sag/swell disturbances [27].

classified as an undervoltage disturbance. The effect of sag on a signal is illustrated in Figure 2.12a

Swell / Overvoltage

A swell is the opposite of a sag, with an increase in peak AC voltage for a given duration. If this duration is longer than 1 minute, it is called overvoltage. This often occurs as a result of load switching where the system is too weak to handle a needed voltage regulation. The effect of a sag on a signal is illustrated in Figure 2.12b

Waveform distortion

Waveform distortions are disturbances in the sinusoidal wave and have a variety of causes. The five most common distortions are:

1. DC offset
2. Harmonics
3. Interharmonics
4. Notching

(a) DC offset. **(b)** Harmonics. **(c)** Interharmonics.

Figure 2.13: Example waveform distortions [27].

(a) Notching. **(b)** Noise.

Figure 2.14: Example waveform distortions (2) [27].

5. Noise

DC offset DC offset is an offset in the sinusoidal wave such that the average value is not zero. This often causes unwanted current to devices that may already be operating at their maximum capacity and can cause overheating.

Harmonics Harmonics are corruptions in the sinusoidal wave at specific frequencies which are multiples of the fundamental frequency of the wave.

Interharmonics Interharmonics are waveform corruptions where a periodic signal which is not an integer multiple of the fundamental frequency of the signal is mixed with the original signal.

Notching Notching is a disturbance in the voltage level that is periodic in demand and could be seen as a periodic impulse problem, with constantly fluctuating voltage.

Noise Noise is an unwanted voltage or current imposed on the system from the outside and can be caused by poor grounding or other devices such as radio transmitters.

The different types of waveform distortions can be seen in Figures 2.13 and 2.14

(a) Voltage fluctuation. **(b)** Frequency Variations.

Figure 2.15: Example waveforms of frequency variations and voltage fluctuations [27].

Voltage fluctuations

Voltage fluctuation is a variation in the sinusoidal waveform that is systematic in its form, where the voltage differs between 95 % and 105% of its target voltage. This is typically due to a load on the system that has great variations in its demand. Voltage fluctuations are illustrated in Figure 2.15a

Frequency Variations

Frequency variations are the rarest type of problem occurring in an electrical grid, but is, as the name suggests, a variation of the frequency in the voltage, which can be seen in Figure 2.15b

Chapter 3

Background - Machine Learning

3.1 Machine learning

A machine learning algorithm is an algorithm that can **learn** from data. The definition of learning is widely discussed, but a commonly used definition by T. Mitchell is :

> A computer program is said to **learn** from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E [28].

3.1.1 Notation

We will use the following notation throughout this chapter:

Function estimators will be denoted as the function name with an added hat, e.g. an estimator of the function F will be denoted as $\hat{F}$. Functions that are defined by a set of parameters θ will be denoted as F_{θ}.

Vectors will have a bold font with lowercase letters, e.g. $\mathbf{x}$, while scalars will have a normal weighted font, x. Matrices will be written in uppercase, with a bold font, e.g. $\mathbf{W}$.

For series of variables, we will use the notation $x_{1:n}$ to denote the series of variables x_1, x_2, x_3 ... up to and including x_n.

3.1.2 Introduction to machine learning

A common goal for machine learning is to approximate a function $\hat{F}_\theta$ to an unknown function F, which takes as input a scalar or vector $\mathbf{x} \in \mathcal{X}$, where $\mathcal{X}$ is the domain of possible inputs, and outputs a scalar or vector $\mathbf{y}$. The goal is to maximize the performance of a model, with respect to some loss measure L. This is done by finding the parameters θ of the model, so that they minimizes the loss L.

Problems in the machine learning domain are often divided into three categories; supervised learning, unsupervised learning and reinforcement learning. In this report we will focus on the problem category supervised learning, where a dataset of labeled samples is provided. The assumption is that given enough training samples, the machine learning method can create a model that will be able to generalize to new, unseen inputs as well. In the instance of fault prediction in power grids, the goal of the machine learning model could be to predict the probability of a fault occurring within a specified time interval. The loss measure L can be defined as the difference in predicted probability of a fault and the ground-truth of the provided sample, whether or not a fault will occur within a specified time interval.

Moving back to the definition of learning, in the context of fault prediction in power grids, the task T is recognizing faults, the experience E is the input of sensor data, while Performance Measure P is the negative value of the loss defined by L. A machine learning method would then be said to learn if the loss of the model decreases, given more samples of input sensor data.

Machine learning differs from some other Artificial Intelligence (AI) methods, in that feature extraction is performed by the method itself. There is often no human intervention required for rule generation, as the method is capable of recognizing general patterns in the dataset.

Overfitting and underfitting

Two of the most well-known problems in machine learning are overfitting and underfitting. As the goal of supervised learning is to approximate an unknown function by using a dataset of samples, it is a common problem that the model either adapts too well to the input data or is unable to approximate the unknown function because of lack in model capacity. This is undesirable, as we want the model to learn the general patterns found in the input space, and not adapt too much to the noise in the data samples. If the model is unable to approximate the function due to lack of model capacity, we call it underfitting. If the model adapts too well to the training dataset, and ends up memorizing the data samples, we call it overfitting.

There are multiple ways to deal with overfitting and underfitting of models, which can roughly be divided into two categories: data augmentation and model tuning. Methods

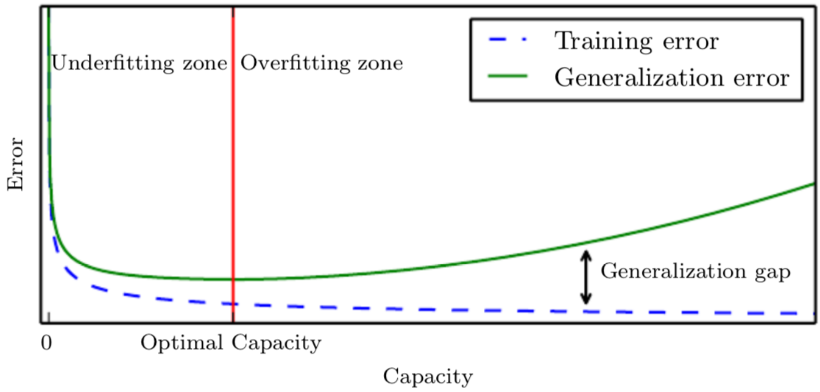


Figure 3.1: Overfitting and underfitting effects on accuracy. [29]

from both categories are often used to reduce the risk of overfitting or underfitting. The process of making the model more robust is called generalization.

To estimate the generalization error, the error in prediction accuracy when presented with new, unseen samples, one usually divides the dataset into three categories; a training set, validation set and test set. The training set is used as input during model fitting, and the validation set is used to estimate the generalization error when faced with new samples. By tuning the hyperparameters to achieve a best possible generalization error estimate on the validation set, we can thereafter test the model on the test set, to get a final generalization score. It is important not to use the test set during model fitting or hyperparameter tuning, as it will make the model conditionally dependent on the test data, and introduce a bias into the final generalization score. The same can be said about using the validation set as training samples during model fitting, as this will also introduce a bias in the generalization estimator.

Figure 3.1 shows the relationship between over-/underfitting and the generalization error. Figure 3.2 illustrates the relationship between the approximated function complexity and over-/underfitting.

Data augmentation By augmenting the samples in the dataset with label-preserving transforms, one is able to increase the size of the dataset without explicitly collecting new data samples. There is a multitude of transforms that are label-preserving in nature, but the label-preserving property is dependent on the problem at hand. An example is the problem of classifying images into a number of classes. Label-preserving transforms can then be translating the image, changing the lighting or rotating the image. The label will still be the same, as the image still expresses the same concept of an object.

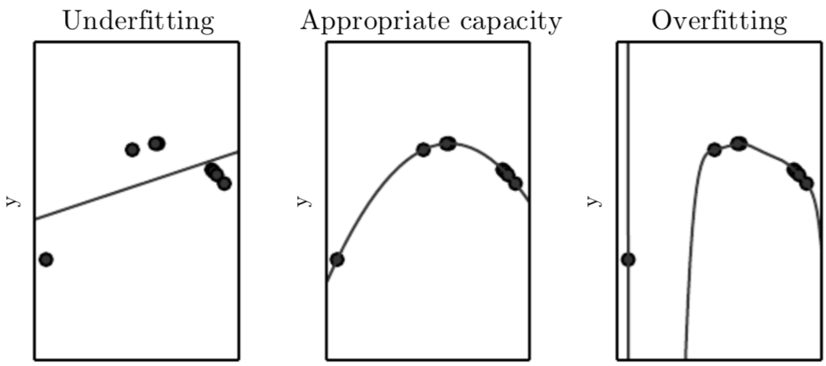


Figure 3.2: How a function could look with over and underfitting. Underfitting creates a too simple model, while overfitting fits a too complex function to a simple dataset. [30]

Model tuning A machine learning model usually consists of an architecture defined by a set of hyperparameters. Learning rate, the number of layers, input weighting and a number of trainable parameters are only a few examples of the many parameters that might be changed to create a better model.

By systematically testing all possible combinations of hyperparameters, it should in theory be possible to create the model that is best suited for the problem at hand. In practice, however, this approach is unfeasible for models with a large number of hyperparameters. The number of possible combinations increases exponentially with the number of parameters, and most machine learning models use considerable time to learn a new model given parameters and input data. This makes the approach too time-consuming for practical use.

An approach that is often used in practice is regularization, which is a term for multiple methods used to tune models. An example of regularization is is adding a penalty for large weights as part of the loss function in the model's target function. Putting a penalty on the magnitude of weights in the model favours function approximations with sparse weight matrices. This in turn helps increasing the interpretability of the model, which is favourable, as a common goal in machine learning is to answer the question of why the model behaves the way it does, not just that it behaves correctly.

Two common regularization penalties are the L_1 and L_2 norm. The definitions are given below,

$$L_1 = \lambda \sum_{i=1}^{m} |w_i| \tag{3.1}$$

$$L_2 = \lambda \sum_{i=1}^{m} w_i^2 \tag{3.2}$$

where m is the total number of weights in the model, λ is a hyperparameter determining the weight of the loss term and w_i is a model weight. Increasing λ increases the penalty for large weights, and will result in sparser models.

Another common approach to reduce error is a method called ensemble learning. The method consists of training an ensemble of multiple models rather than training a single model to perform prediction. Each individual model calculates an output given the input, and the final output is determined by aggregating all of the outputs.

There are multiple ways to train an ensemble of models, where Bagging [31] and Boosting [32] are two common algorithms. Bagging is the process of training multiple models on different subsets of the dataset, where all the outputs are combined by taking the mean value of the individual outputs. Boosting is an iterative process of model training, where the next model in the iterative process increases the loss contributed by previously misclassified samples, which creates a model that better predicts outliers in the output space.

3.1.3 Machine learning methods

In this section, we introduce specific machine learning methods that have been proven to be effective on the type of problems that arise in dynamic systems. We will explain:

- Deep Learning
- Hidden Markov Models
- Decision Trees
- Support Vector Machines

Deep Learning

Deep Learning has seen huge advancements and rises in popularity the recent years, mostly due to an increase in computation power. It was initially inspired by the way the human brain works, and consists of layers of nodes and weights combined by a linear function, with an added non-linear function applied element-wise for each output of a layer. Deep Learning is often mentioned along with neural networks, as it is a subcategory of neural networks. Deep Learning is mostly used when there are multiple layers in a model, thus being a deep neural network [29]. Figure Figure 3.3 illustrates the difference between a deep learning and a simple neural network.

The input layer of a neural network consists of a set of nodes where data in the form of a vector is ”fed” into the network, and forwarded to the next layer. The vector is multiplied with a weight matrix, symbolized by the edges between nodes in Figure 3.3 before applying an element-wise non-linear function, often called an activation function. The resulting

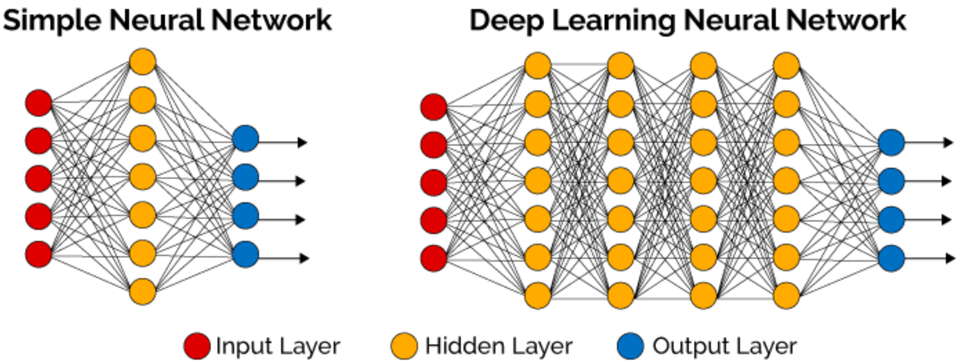


Figure 3.3: Deep learning vs simple neural networks, from [33].

vector is then forwarded in the same manner to the next layer in the network. The final output is the output of the final layer in the network.

To improve the performance of a neural network, a set of labeled samples in run through the network to generate predictions. The output of the network is then compared to the labels of the samples, which is regarded as the ground-truth label for each sample. A loss measure can then be calculated by the use of a comparison function, to calculate the total error of the network for the given samples. By using the gradient descent algorithm, the weights in the network can be adjusted to reduce the total loss by propagating the error layer-by-layer backwards through the network and calculating the individual contribution to the loss for each weight in the network [29]. Deep Learning has been widely successful, achieving a high accuracy on a multitude of problems. While it may take a long time to learn and update the weights of the network, it is quite fast in producing values once it is fully trained [30].

A neural network, and consequently a deep network, consists of a number of layers which defines the mapping

$$\mathbf{y} = \hat{f}_\theta(\mathbf{x}) \tag{3.3}$$

where $\mathbf{y}$ is the output of the network, $\mathbf{x}$ is the input, and θ is parameters of the network which are optimized to the best function approximation with respect to some loss measure L. For each layer l, each individual node x_i outputs the weighted sum of its inputs, with a non-linear function applied to the sum of inputs, defined as

$$x_i^l = \sigma(\sum_{j=1}^{N}(x_j^{l-1} \cdot \mathbf{W}_{j,i}^l) + b_i^l) \tag{3.4}$$

where x_i^l is the output of the node i in layer l, N is the number of nodes in the previous layer, $\mathbf{W}^l_{j,i}$ is the weight from node j in the previous layer to node i in the current layer, b^l_i is an added bias and σ is a non-linear function.

One of the more commonly used non-linear activation functions is the Rectified Linear Unit (ReLU), which is shown in equation 3.5

$$f(x) = \max(0, x) \tag{3.5}$$

Further details on how the gradient descent algorithm is implemented through backpropagation in a neural network is thoroughly described in the Deep Learning book by Goodfellow [29].

Hidden Markov Models

Hidden Markov Models (HMMs) are probability based machine learning methods for unobserved variables in time series data. It is one of multiple algorithms within the class of Bayesian classifiers. Given a series of observable variables, we want to predict the state of an unobservable variable, on which the observable variable is conditionally dependent on.

The observable variables are usually denoted $\mathbf{y}_t$ for the observed value at time t, and the unobservable variable is denoted as $\mathbf{x}_t$. In literature, $\mathbf{x}_t$ is often referred to as the belief state [34].

An important assumption in hidden Markov Models is that x_t satisfies the *Markov Property*, defined as

$$P(\mathbf{x}_t|\mathbf{x}_{1:t-1}) = P(\mathbf{x}_t|\mathbf{x}_{t-1}) \tag{3.6}$$

which means that the variable at time t is only conditionally dependent on the state of the variable at the previous time step, and not any other previous states. The observable variable $\mathbf{y}_t$ must also satisfy the following conditional dependency

$$P(\mathbf{y}_t|\mathbf{x}_{1:t}, \mathbf{y}_{1:t-1}) = P(\mathbf{y}_t|\mathbf{x}_t) \tag{3.7}$$

which means the observed variable $\mathbf{y}_t$ is only conditionally dependent on the state of the unobserved variable at the current time step [28]. A Markov Model must also be a stationary process, which is the property that the transition probabilities between states and the conditional probabilities for the observed variable $\mathbf{y}_t$ does not change over time.

Hidden Markov Models was one of the earliest algorithms to show great potential for fault detection in dynamic systems [35], and is visualized in Figure 3.4.

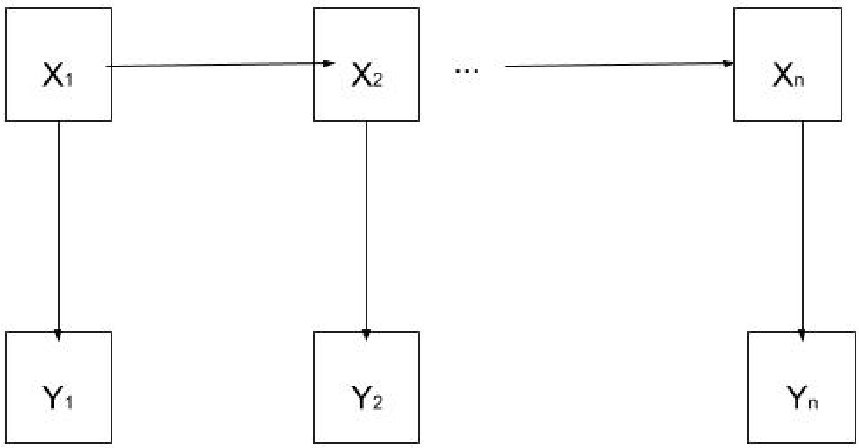


Figure 3.4: An illustration of a Hidden Markov Model, X_n is the observed variable, while Y_n is the output of the hidden state.

Decision Trees

Decision trees is a technique for classification problems that involves inferring rules by splitting the dataset into different labels based on the properties of the data. It works quite like Expert Systems in that it creates rules by finding the best way to split the data according to the information gained, measured as information entropy. Once a certain threshold of information gained by splitting is reached, the sample is classified according to what category is the most frequent category for that branch of the tree [36].

Decision trees give a very logical view of how the rules are created, and makes it easy visualize the model and how the decisions are made. Decision trees are universal function approximators, but to reduce overfitting, pruning of lower branches are usually done to help reduce the generalization error [28]. An example of a decision tree can be seen in Figure 3.5.

An extension to decision trees called Random Forest algorithm has lately gotten a surge of popularity in the AI sector. This due to the simplicity of its implementation combined with its often high accuracy. The Random Forest algorithm produces an ensemble of decision trees on different subsets of the total dataset and aggregates their outputs to get a more accurate result. This is also helpful in finding which features produces are the most important features for prediction [37].

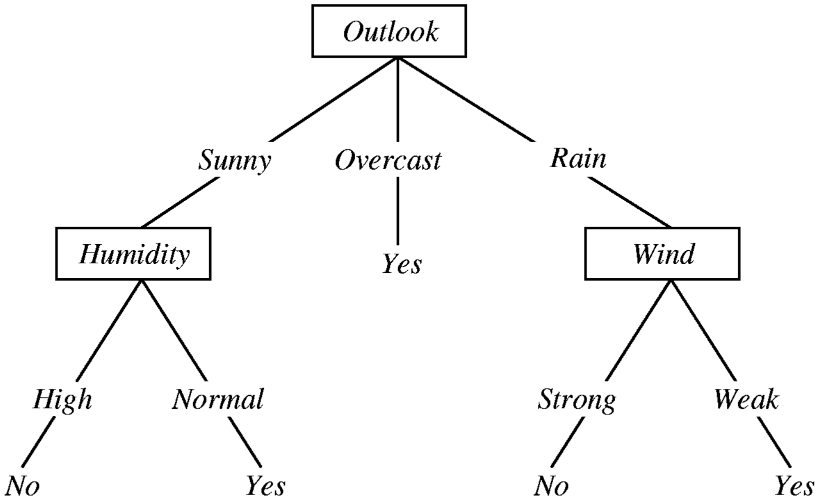


Figure 3.5: An example of a decision tree for whether or not to play tennis [28].

Support Vector Machines

Support Vector Machines (SVM) is a method that has been widely used in classification problems. The SVM works by finding the hyperplane in a multidimensional space that best separates the classes that one wants to predict. A 2D example is illustrated in Figure 3.6.

The hyperplane defined by the SVM is the line or plane with the longest distance to the samples in each category that are closest to each other, according to some comparison measure. This can be done in any multidimensional feature space, which makes SVMs an excellent tool for inputs with high dimensionality.

If the data is not linearly separable, an SVM applies a method popularly called the kernel trick, which transforms the features in the input space into a higher dimensional representation, and finds a separating hyperplane in the higher dimensional space instead [28]. Classification is done by calculating which side of the dividing hyperplane a sample appears on, and labels the sample accordingly.

3.1.4 The usage of machine learning in dynamic systems

> A dynamical system is a system whose state evolves with time over a state space according to a fixed rule [39].

The notion of time in dynamical systems adds complexity to many of the existing machine learning methods that were described in the previous section. While most basic machine

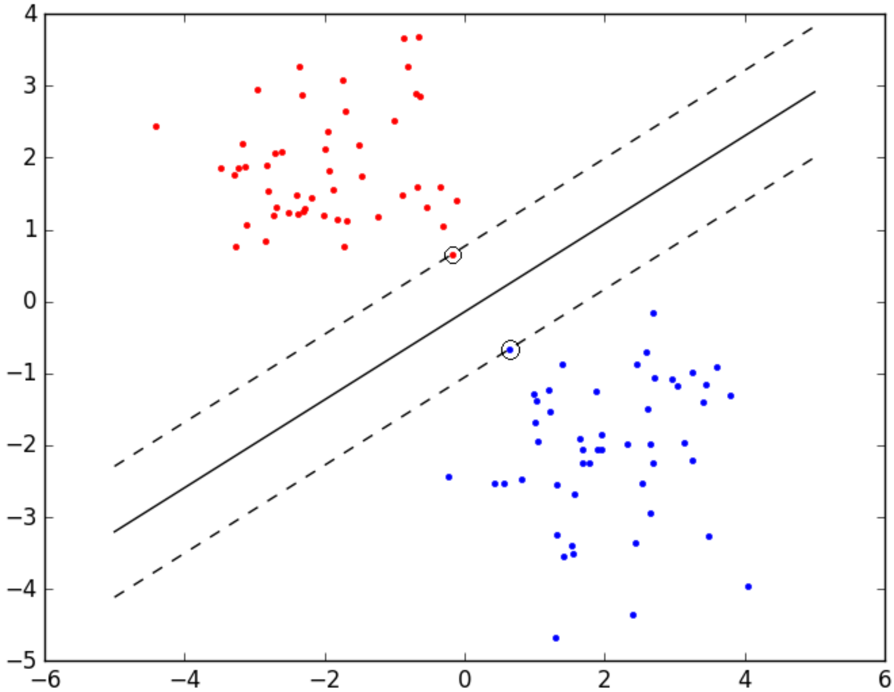


Figure 3.6: Illustration of Support Vector Machines in a 2D space [38].

learning methods classify their data according to the current input, the added time dimension makes the output dependent on not only the *current state* of the data, but also the *change over time*. This adds another dimension to the input data, which can make the dimensionality of a single sample computationally intractable for classical methods.

Algorithms that work on time series data has to take into consideration the number of previous steps to look at by considering the value of more data versus the added time complexity of the algorithm itself. The more data to look at, the higher the complexity, and the more time is required to train and run the algorithm.

One of the more common usages of such sequential data involves Natural Language modelling, which is predicting the next word, given a sequence of words [40]. Hidden Markov Models, as described in Section 3.1.3 is a popular way to deal with this, as is an extension to neural networks, which we describe in the next section.

3.1.5 Recurrent Neural Networks

One of the most popular ways to deal with sequential data is Recurrent Neural Networks (RNN). Recurrent Neural Networks are Neural Networks that not just feed the values forward, but also propagates the value itself back to use in the next calculation, as shown in Figure 3.7

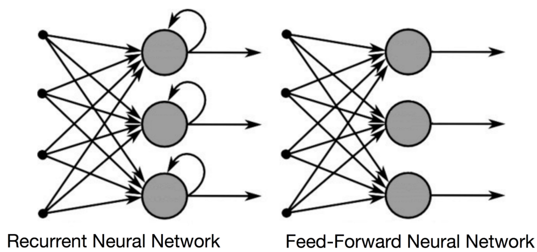


Figure 3.7: An illustration of the differences between a Recurrent Neural Network and a Feed Forward Network. [41]

Recurrent Neural Networks and more specifically its implementation Long Short-Term Memory (LSTM) networks are responsible for much of the success in Deep Learning, in areas where sequential data is used [41]. RNNs is one of the few models with internal memory, making it capable of remembering previously seen input from a sequence, thus increasing the accuracy of predicting what comes next, because RNNs can then use the entire sequence of state changes to predict the next value.

The internal memory makes it capable to learn from experiences that may have a long time delay between a specific pattern and the actual state. This makes it especially useful for fault prediction in the power grids, where disturbances sometime earlier could be the only telling of a fault that will happen, and it has been shown to be successful in classifying voltage disturbances [42].

Chapter 4

Previous work

Much of the previous work done in the field of voltage analysis seems to be done on simulated data and not on real data from sensors, which naturally contains more noise. Previous work has also focused on the area of classification of voltage faults while predicting if faults are going to happen has is lacking research. This classification is a natural first step, as this job has been done, and still is being done, manually by operators after a fault occurs. An improvement in this process would help in generating a decent test dataset for predicting disturbances.

4.1 Goal of review

The goal of this chapter is to present and explain some of the history and advancements in power disturbance prediction throughout the last decades. We have looked at what type of machine learning methods that seem to show most potential, and what kind of feature engineering and data types that are used in previous work. To increase the area of research, we have looked not only at the prediction of voltage disturbances but also at the steps coming after it, namely classification and stability. These steps are shown in the timeline in Figure 4.1. In this figure a fault occurs during the time interval $10 - 20$. The time leading up to this fault, here exemplified in the time between $0 - 10$ is the prediction stage of voltage analysis. In prediction, one tries to look at the waveform and look for disturbances that could signify a fault is coming. When the fault is happening, between $10 - 20$, the classification of the fault occurs. In classification, there is a disturbance in the voltage, and one wants to find out what type of disturbance it is. After the fault has occurred, one wants to find out whether the voltage will be stable, or if the fault will potentially lead to an unstable network and a blackout. This process we will refer to as voltage stability and is done after a fault has occurred, exemplified in the figure in the time

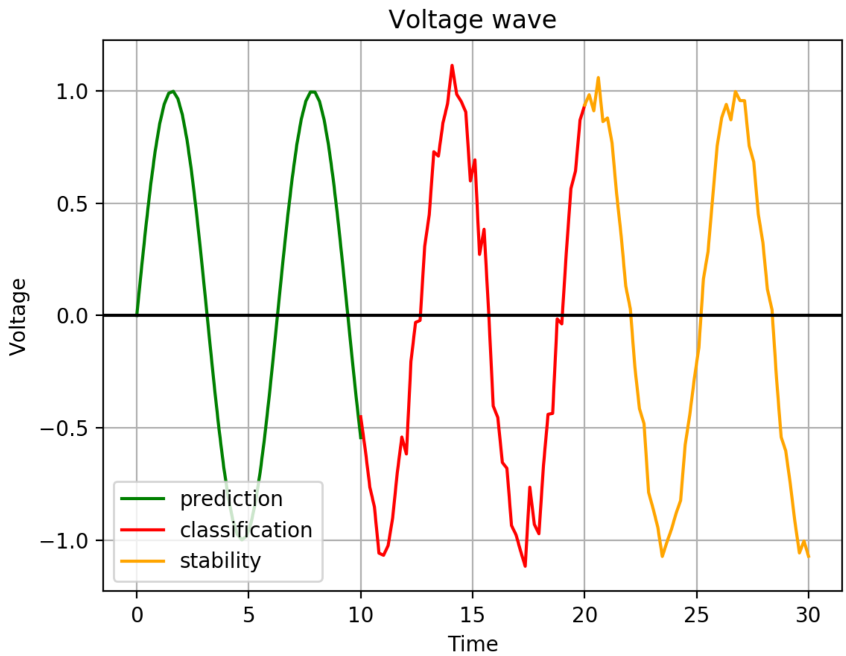


Figure 4.1: The different steps of voltage analysis.

interval $20 - 30$.

Two recent breakthroughs have been fundamental to the problem of applying machine learning to power quality prediction; the lack of real data, and an increase in computation power. The problem with lack of data is partly solved by the wide-scale availability of power quality (PQ) monitoring equipment, where sensors are applied throughout the power grid. These sensors are continually monitoring important variables like voltage and current and sending sensor readings to a centralized server.

The performance of power quality computations is still a significant issue, as the data may be in a resolution as high as 50 000 data points per second, while still in need of real-time analysis. There is not much research on the frequency that is needed to gain a good result, but it is shown that some faults cannot be detected unless one has a high time resolution of data, up to 25 KHz [43]. We will explore further with the assumption that a higher time resolution is more beneficial when it comes to predicting faults.

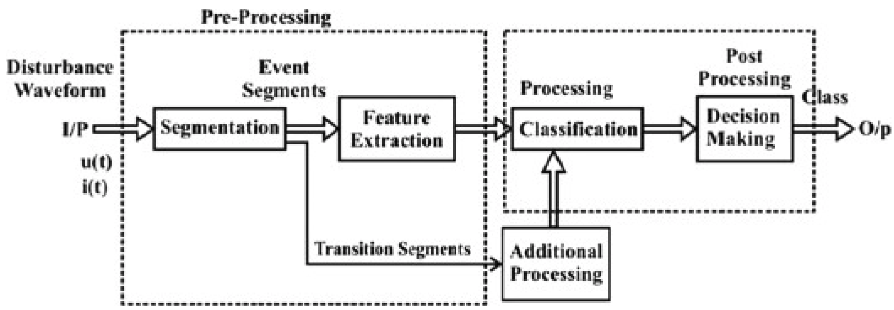


Figure 4.2: The classification of voltage disturbances process [45].

4.2 Classification of voltage disturbances

Since there are few publications on the prediction of voltage disturbances, we explore a similar field that has been more thoroughly researched over the years, namely the classification of voltage disturbances. While the classification of disturbances is not the overall goal of this report, a good classification method would be of immense use to a machine learning algorithm trying to predict faults, using missed fault detections as training examples to learn constantly. Classification of voltage disturbances has been an important issue, as all classification of disturbances on the power grid normally is done through visual inspection of the disturbance by an operator. This procedure is susceptible to human errors, as well as being a very time-consuming practice [44].

Most research about classification is divided into two parts. First, we have feature engineering of the data. Feature engineering involves transforming the data to keep as much as possible of the information it contains. This is done because monitoring data at a 50 KHz rate and using all these data points is still close to unfeasible for a real-time computer system. The second part is the processing itself, the classification methods used to obtain a result. Figure 4.2 shows the general steps of this process. First, we have the waveform coming in; it is divided into different segments to look at; one or multiple feature extraction methods are used to gain further insight into the data and reduce the frequency domain. Lastly, classification is used to classify which fault class the disturbance belongs to, or whether there is any fault at all. The last step, decision making, might be further used by a computer system to act upon data automatically — for example, a decision to shut off power delivery to a faulty line. We will not explore this action further and will assume that all actions will still have to be taken by a manual operator receiving suggestions or data from a prediction system.

Some of the earliest work done on the classification of power disturbances using machine learning was done by in 1995 by Ghosh and Lubkeman [46]. They describe a neural network approach for classification of waveforms, in part as a preprocessing step to use for data collection of recorders. They used sensor data that automatically triggered on certain thresholds of disturbance and used this data to further classify the type of disturbance experienced. As seen in Section 2.2.3 most of the common voltage disturbances have a particular characteristic in their waveform. It should be noted that this classification was done on sensor data that had already gone over some automatic set threshold and thus was made to classify disturbances that were already discovered. This was done to reduce manual work done by engineers, so while they achieved an accuracy of over 90 % with a neural network, this was on simulated data, and data that was already classified as having a disturbance.

Also of early interest is a paper produced in 1999 on Power Quality Detection and Classification using Wavelet-Multiresolution Signal decomposition [47]. Wavelet Transforms have the ability to analyze problems both in time and frequency domains, and have been one of the most popular feature engineering tools used in power analysis and learning. Using this transform, and then calculating the standard deviation at different resolution levels of the signal, they used this to classify voltage sag, swell, harmonic distortion and transient distortion. The classification was done through mathematical rules and definitions, and no machine learning method of learning from the dataset was applied. However, the results of this study were unclear, and no benchmarked comparison to other solutions was performed. Although inevitably correct, as they used definitions of these interruptions for mathematical calculation, some disturbances that have unclear definitions or is very close to a disturbance could have been missed.

In 2011 a review on different techniques and methodologies for disturbances classification was published [48]. The traditional Power Quality Indices (PQI) includes peak values and total harmonic distortion, properties that have remained as standard in later work. The standard in feature engineering at that time for was Fast Fourier Transform, Goertzel's algorithm, Zoom FFT and Welch's algorithm among others. The information these algorithms provide is sometimes insufficient, and not all problems can be resolved by these algorithms, especially for detecting short spikes or transients [49]. The Short Time Fourier Transform fixes many of these issues as long as the window is short enough, along with the Wavelet Transform and S-transform.

For classification algorithms, one of the most widely used has been Neural Networks, while others have also shown success such as Fuzzy Logic, which has been successfully combined with expert-based systems [50].

Support Vector Machines (SVM) and their variants have also shown to be useful in the classification of voltage disturbances [51]. SVMs have been tested on data leading up to and including the actual disturbance and even done very well on data from another power line than it was being tested on, leading to the theory that a pre-trained factory SVM could be deployed in multiple surveillance grids [52].

The results of the mentioned review [48] is summarized in Table 4.1 and 4.2 with their mentioned advantages and disadvantages. For feature extraction methods, many have already been successfully implemented in different experiments. Support Vector Machines and Neural Networks have been successfully applied, and dependent on the dataset can

Method	Advantage	Disadvantage
Short Time Fourier Transform	Simplicity of implementation.	Limited time and frequency resolution
Wavelet Transform	Time-frequency resolution	Strong computational burden, looses some information regarding transients
Phase Locked Loops	Accurate frequency measurement and synchronization	Inconvenience to harmonic or interharmonic component measurement
S-Transform	Phase correction	Block processing, may not be suitable for real-time requirement

Table 4.1: Table of promising feature extraction methods, adapted from [48].

provide excellent results for classification. A popular method that has been widely used but was not thoroughly mentioned in the review was Decision Trees, and their use will be discussed next.

In 2013 a real-time power quality disturbance classification method based on extensive feature extraction using a hybrid of known methods [53] by He, Li and Zhang was published. To transform their input data they used a modified version of the S-transform. They argue that for real-time computation, many of the more computational heavy machine learning methods are undesirable, and instead use Decision Trees to classify the different type of disturbances. The system was tested and validated real-world data and performed well (>95 %) on classification on these data, with a sample rate of 52.2 kHz. Decision trees have also been successfully applied to the problem in multiple instances [54], also using the S-transform as a feature extraction method, with decision trees outperforming SVM in a comparison done by Ray[55].

In 2015, to deal with the problem of adapting over time, Barros proposed an algorithm that should continuously learn from new input. It is based on the Wavelet Transform (2.1.2), and a neural network that was able to do continuous learning. Accuracy rates with continuous learning rose, going from an 83.51 % accuracy on classifying voltage swells to 100 % [44].

Other more advanced models have also been proposed recently. Bauloji proposed using an LSTM deep learning approach for classifying voltage dips, using strictly training data, and no other learning or definitions of voltage disturbances. He achieved a accuracy of 93.5% on a test set [56].

Method	Advantage	Disadvantage
Neural Networks	Mathematical flexibility, learns only based on input	Can get stuck in local minimum. High computational cost.
Fuzzy Logic	Reduce difficulty of modelling and analyzing, deals with uncertainty	Not adaptable, need expert to generate rules.
Support Vector Machines	Handle large feature space, very good training process	Often dependent on a large training sample.
Bayesian Classifiers	Based on existing probabilities.	Must know probability density function of each event.
Decision Trees	Based on existing probabilities	Must know probability density function of each event.

Table 4.2: Table of promising classification methods for classifying power disturbance, adapted from [48].

4.3 Voltage stability

Very closely related to voltage disturbance prediction is the field of voltage stability. Voltage stability is a problem class about detecting whether any instability is present in the waveform of the voltage to reduce blackouts. This instability is usually a result of some of the faults mentioned in Section 2.2.3, gradually leading up to an interruption. The main difference separating it from voltage prediction is that security monitoring is done after an instability has already happened, and therefore has to look at a shorter period to see whether the power quality is stabilizing or worsening. In voltage stability one therefore already knows that something is wrong, and wants to predict the consequences of the fault.

In 2009, Decision Trees were used to predict voltage stability following blackouts. Synchronized PMU data was used and predicted the signal to be either secure or insecure. By using decision trees they were also able to find out which type of variables that were effectively splitting the dataset. The cost of classifying an insecure case as a secure case was more costly than classifying a secure case as insecure, as the consequences of that option would be higher in a real system [57]. To deal with different types of data giving different values, they proposed extending the decision tree to a Random Forest algorithm to increase accuracy. This method that has since been successfully implemented and tried out, using a security index and degree of alertness instead of a boolean output [58].

Different types of neural networks and versions of them have also been successful in monitoring voltage stability. In 2015, Zhukov proposed a hybrid architecture consisting of different types of neural networks using aggregated data from sensors [59].

In recent advancements, Ibrahim, Amr & El-Amary [60] proposed a recurrent neural network to deal with voltage instabilities, and tried it out on simulated data, to help prevent a network from collapsing. They used simulated PMU data as input, as well as other power variables calculated, like active and reactive power.

Guo and Milanović [61] proposed decision trees as a method of classifying the state of a system after a fault has occurred and whether it was a critical fault or if the system could keep operating. They achieved a 95 % prediction rate using data 1 second after a fault on PMU data on predicting whether the system was stable.

4.4 Voltage disturbance prediction

Little work has been done in the field of voltage disturbance prediction. The prediction process is very similar to the classification process as seen in Figure 4.2, however, the classification step is now a prediction of future disturbances that might occur and not a classification of current disturbance. Since the work is done on the same data, we work under the assumption that the same methods of segmentation and feature extraction as used in the classification process can be used here. Also a lot of the methods used in classification can be reused to solve this problem. One has to take into account that prediction might have to look at at a larger period of data than the classification process and thus is more dependant on the time series and the change over time than many classification methods.

Some of the earliest work seen on voltage fault prediction was in 1999 on a large transmission system, with Qader, Bollen and Allan predicting voltage sags using stochastic prediction with the usage of phasor data generated by on-site sensors. This was a purely probabilistic prediction based on earlier faults and time between faults and used statistics to predict when the next voltage sag might happen based on historical sensor data [62].

In 2016, researches developed a computer system for the detection of high impedance fault in distribution lines, using power quality meters installed at feeders that generated high time resolution data. As features, they used the set of harmonics present in the current signal. For detection, they used a neural network to discover if it was a disturbance present in the system [63].

Similar high impedance fault prediction was also proposed by German researches in 2013, to detect trees falling on the power lines. This was done through the use of sensors, and a manually created algorithm for detecting an abnormal change in impedance due to disturbances on the power line. This was detected before the incident could cause more severe damage to equipment [64].

Early prediction of voltage sags by various methods was explored by Weckesser using real-time-measurements from PMU sensors in generators, using signal analysis. While they were able to predict the voltage falling below a critical value ahead of time using statistical analysis, their definition of ahead of time was 1 second before the fault [65]. While this may be enough for a computer system response, a manual operator would have no chance of responding to such a prediction.

The Texas A&M's Distribution Fault Anticipation project has long been trying to anticipate faults in power lines through real-time intelligent monitoring. In their paper describing their efforts, they showed numerous cases of power outages that had incipient faults leading up to the outage as far as one month before, where subtle anomalies were produced in advance of the outage [66]. In the following report the year after, they showed that intelligent algorithms could detect some of these cases [67]. The researchers behind this paper also state voltage harmonics at a sufficient sample rate as the most important metric to widespread deployment of this system. We note that no testing was done on whether these algorithms would incorrectly detect disturbances in signals without a fault as well. Recently, the same researchers published a paper proposing a method to find recurrent disturbances in power lines, which could signify a more extensive outage happening shortly [68].

In 2018, Xiao and Qian [69] used a Hidden Markov Model to predict power quality disturbances, which included weather conditions, using real data from Chinese cities and local distribution grids. The reason for their choice of model is to be more able to derive the relation that exists in the data and make predictions. Recording the transient voltage and current waveforms of the event with a resolution of 256 samples per cycle, they could correctly predict a coming voltage disturbance in many of the cases [69].

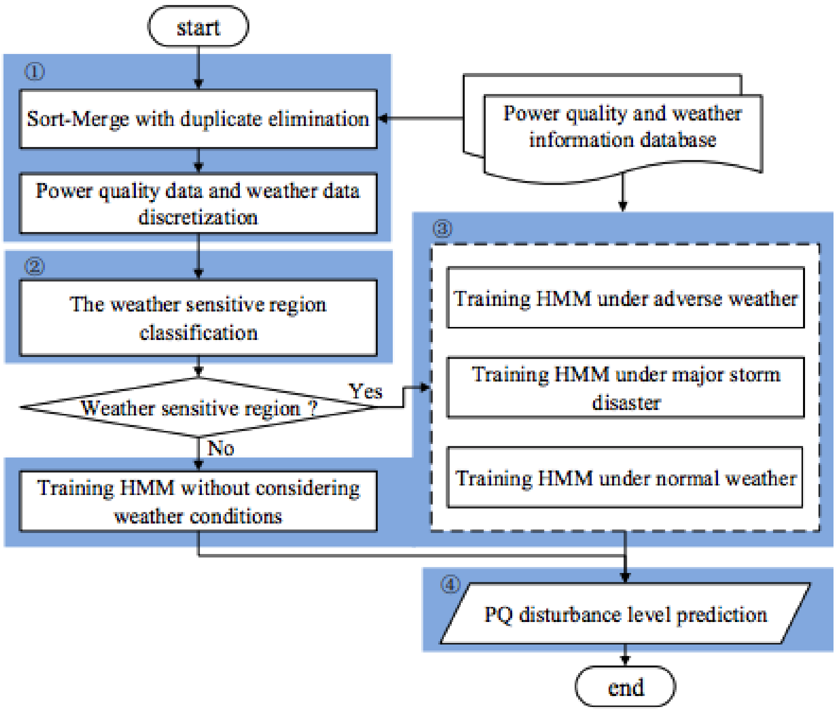


Figure 4.3: Xiao and Qian power quality disturbance prediction levels [69].

Chapter 5

Prediction of voltage disturbances in the Norwegian Power Grid

In this chapter, we present the EarlyWarn project, a project that aims to predict power failures in Norway's power grid. We explain the data that we are focusing on, how it is collected, and what parameters it contains. Further, we look at how the results of this project could be used in an application by power grid companies. Lastly, we combine the information about available data with our research in machine learning to look at which methods may be the most suitable for further research by Early Warn.

5.1 EarlyWarn

EarlyWarn is a collaboration project between SINTEF, NTNU, Statnett, and numerous operators in the Norwegian power grid. Their ultimate goal is proactive detection and early warning of incipient power system faults. This will achieve this by using sensor data placed at strategic places in Norway [70]. In combination with the recent advances within the field of machine learning, they hope to detect faults by learning the early patterns of some of the voltage disturbances described in Section 2.2.3.

Figure 5.1 shows the overall information, methods to produce the outcome, and value that this will bring to the Norwegian power grid. They are experimenting with different sensors, to see what value they deliver when it comes to predicting power faults early.

Using recently advanced methods of machine learning and big data they want to achieve a system capable of detecting upcoming faults, and give a warning either to a manual

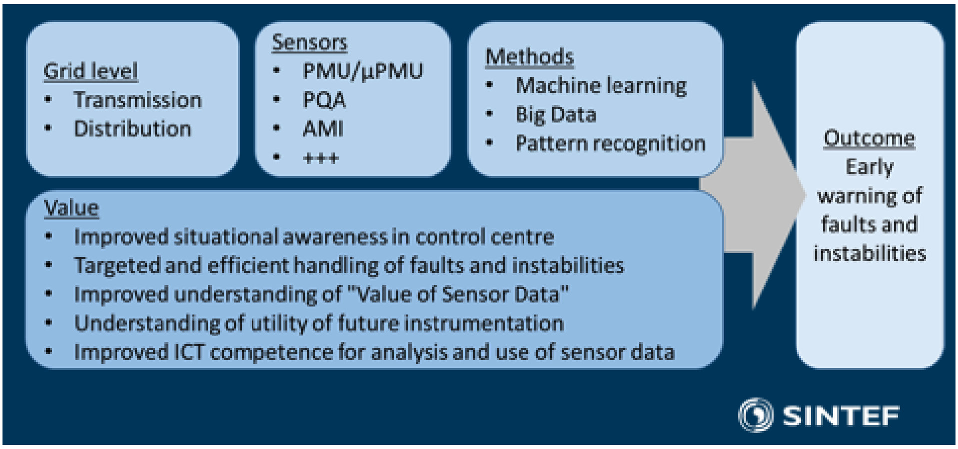


Figure 5.1: EarlyWarn overview [71].

operator, or in the foreseeable future, the computer should automatically do what is needed to fix the system, if possible.

5.2 PQA data

PQA data is produced by Power Quality Analyzers, an example of which is seen in Figure 5.2.

These sensors are installed at some locations in the Norwegian power grid and give real-time insights into the power quality at these sites. PQA data is one of the primary sensors in the Norwegian grid, as well as Phasor Measurement Units. Compared to PMU data, PQA data has a superior sampling rate, and are thus able to obtain more detailed data on the power quality than other sensors [43]. This higher resolution, up to 50 kHz versus PMUs' 50 Hz makes it possible to detect faults that a method using other sensors would be unable to detect due to the lower sampling rate. An example of this is seen in Figure 5.3 where the distortion in the signal is obvious when looking at the waveform, but inconspicuous at a lower sampling frequency.

PQA data gives valuable information on different voltage quality parameters, including harmonic distortion, transients, and voltage variation. The disadvantage of PQA compared to PMU is that they are not synchronized with each other, which makes it harder to compare data from different sensors to aid in pinpointing the location of faults.

PQA data is collected by power grid companies in Norway, and SINTEF has received access to some of it for research purposes. An important aspect to remember is that this data needs to be analyzed real-time or with the shortest delay possible to have any value

Figure 5.2: An example of a Power Quality Analyzer. [72]

in the prediction of faults. Figure 5.4 shows the location of PMU sensors planned and installed in Norway's power grid in 2015. No similar map exists for PQA sensors.

5.3 Learning - false negatives and false positives

When it comes to learning to predict a problem like voltage disturbance, it is essential to take into consideration the consequences of false positives and false negatives. We will give a quick explanation of both in the interest of voltage disturbance prediction.

A false positive error would be to predict a data sample as a disturbance, even when the

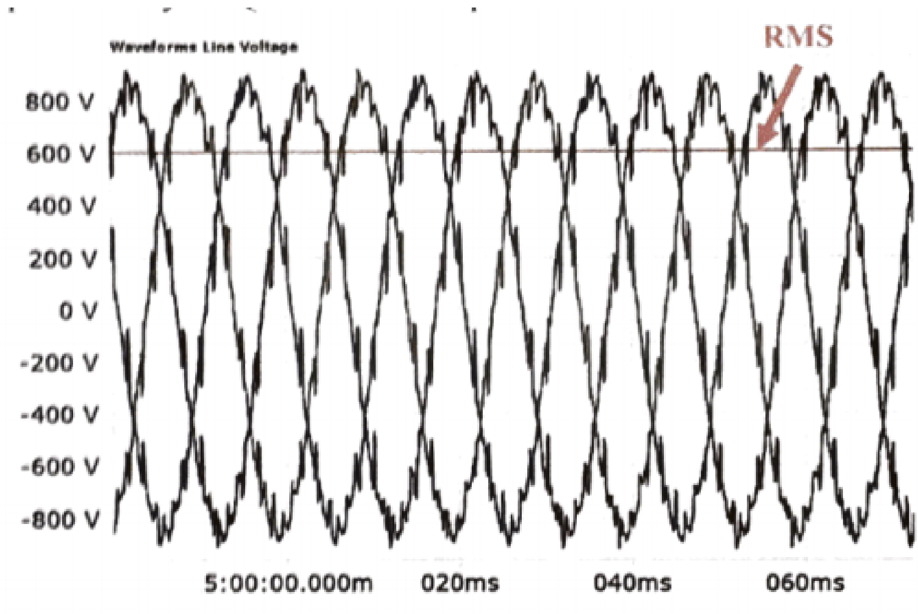


Figure 5.3: RMS and waveform of a voltage with high level of distortion [43].

correct answer is that no disturbance will happen.

False negative error would be to predict a data sample with an upcoming disturbance as a non-faulty dataset, not advising any action for the operator.

The consequences of a false positive are some extra time used for the operator, as long as they have the competence to say that no action needs to be taken. The consequence of a false negative, however, might be catastrophically for the power grid, causing a potential blackout and a more significant cost to society [73]. It is obvious then, given the extent of the consequences, that a machine learning method that predicts fewer disturbances could be inferior to an algorithm that predicts more disturbances even though the aforementioned algorithm had a higher total accuracy. Any algorithm that is going to do prediction on the Norwegian power grid would have to consider this.

A simple example of this is shown in Table 5.1. Here we compare two methods against each other, with a fictive cost of 100 000 NOK for each false negative, and an estimated 5 minutes of extra work if an operator has to consider a false positive. As one can see, even though method 1 has higher accuracy, it also has a vastly higher cost attached to it, primarily due to its higher number of false negatives.

It should be noted that there must be a balance between these two. For an operator, having to manually sift through thousands of extra non-faulty detections to find the disturbance would be too much. A precise number of faults that is acceptable is hard to say, but anything as long as it does not hinder the operator's job by giving them too many warnings

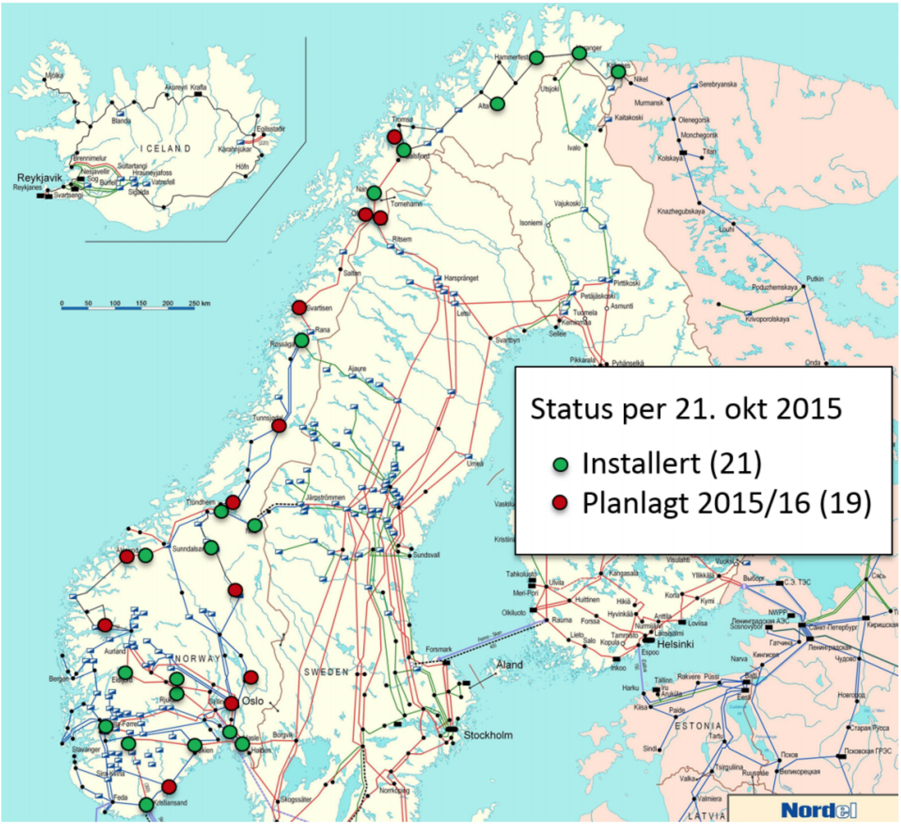


Figure 5.4: Geographical locations of planned and installed PMU sensors, courtesy of Statnett.

would be acceptable, and as described, more beneficial than missing some of the actual faults. It should be noted that this section has been written in mind that this kind of system is the primary system in place. The reality in power systems monitoring is that errors from various system may flash all the time, and it may be more critical for them to focus only on the warnings they know will lead to faults.

5.4 Dataset

5.4.1 Generating a dataset

To generate a suitable dataset for voltage disturbance prediction, one can take advantage the Norwegian regulation on system responsibility in power grids [74], which states that

	Classified positives/ correct positives/ true positives	**Predicted negatives/ correct negatives/ true negatives**	**Accuracy**	**Estimated cost (false negatives * 100 000)**	**Estimated extra time used (5 min * false positives)**
Method 1	95/95/100	405/400/400	99 %	1 000 000	0
Method 2	115/98/100	385/383/400	96 %	200 000	85 minutes

Table 5.1: Example of false positive and false negative consequences.

all power suppliers in Norway have to analyze, document and report faults that occur in the grid. This generates a suitable dataset for classifying errors, with the date of the fault, duration of the fault and what type of error that occurred. SINTEF has access to PQA sensor data from many sensors across Norway, which is stored in a proprietary database.

SINTEF has created a script that uses the mathematical definitions of voltage disturbances to dynamically create a dataset listing the faults that have occurred. Using this dataset, one can extract a suitable training set using the date and duration of the fault and the sensor data from the sensor that observed the fault. Information about the faults is additionally used to extract a non-faulty dataset, defined as any period where a fault did not occur. This program is called Dynamic Dataset Generator (DDG), which extracts a dataset by taking as input a list of parameters that are explained in Table 5.2

Parameter	**Description**
Buffer	The time duration to include in the sample, after the fault occurred
Transient	The minimum duration of time that should pass between a fault, and a non-faulty data sample
Resolution	The sampling frequency of the signal in the generated sample
Duration	The duration of time to include in the sample, before the fault occurred

Table 5.2: Table of parameters for Dynamic Dataset Generator.

The dataset can include a number of parameters for power quality that are retrieved from the database using specific queries. Some of the parameters that can be extracted are listed in Table 5.3.

The original script only supported fetching RMS and waveform data when generating the dataset. During this project, we have further developed the script to support fetching harmonic data from the signal.

The result is a dataset of samples for each fault and an equivalent amount of non-fault samples that can be used for further feature engineering and machine learning. Figure

Parameter	Description
RMS	The RMS value of the given signal, computed cycle-by-cycle.
Waveform	The amplitude of the signal at each sample point.
Harmonics	The harmonics computed cycle-by-cycle by the Discrete Fourier Transform, up to and including the 512^{th} harmonic.

Table 5.3: Table of parameters available to extract when generating a dataset.

5.5 illustrates the entire workflow required for the data generation process. The dataset is saved as a comma separated list of values (.csv), with added metadata for each sample. The metadata included in each sample is described in Table 5.4.

Name	Description
Fault Detection	Whether the data contains a fault or not
Fault type	What type of fault it was
Fault time	When the fault occurred
Start time	Start time of the first sensor reading
End time	End time of the last sensor reading
Total duration seconds	How many seconds of data the dataset contains
Total duration days	How many days of data the dataset contains
Resolution ms	How often the data points are sampled
Time buffer seconds	How many seconds after the fault occurred to sample data
Time transient seconds	The minimum amount of seconds that should pass between a faulty sample and a non-faulty sample
N points	Total amount of data points for each parameter
Node	The name of the node from which the sensor data is accessed
Nominal voltage	The voltage of the line

Table 5.4: Metadata included per sample in the datase.

An example of an anonymized data sample is shown in Figure 5.6.

5.5 End goal

While our goal is to look at whether PQA data could be used to predict voltage disturbance in the power grid, the overall goal and idea of such predictions and the EarlyWarn project

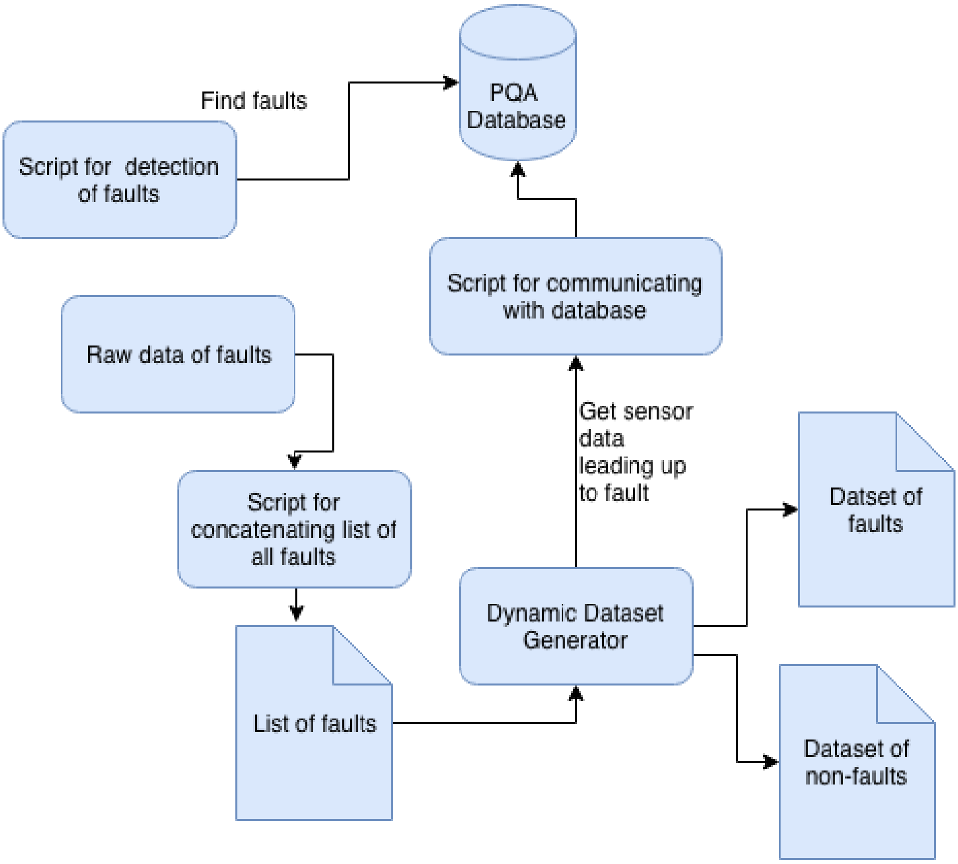


Figure 5.5: The process of generating a suitable dataset.

would have to be an application for real-time monitoring, here explained in steps.

1. Have hardware in place for intelligent monitoring of conditions of power systems. (PQA and PMU analyzers) .
2. Use this data to be able to predict faults to the power system.
3. Predict failures on the active line and send this data to an operator station for further action.
4. Act upon data and recommendation received, first through a manual operator, but ultimately through an advanced computerized system with minimal delay.
5. Deploy necessary personnel needed to avoid component failure, clear out fallen trees or other damages that could have caused the disturbance.

	A	B	C	D	E
1	fault_detection	True			
2	fault_type	Avbrudd			
3	fault_time	<fault_time>			
4	start_time	<start_time>			
5	end_time	<end_time>			
6	total_duration_sec	10			
7	total_duration_days	0.000115741			
8	resolution_ms	20			
9	Time_buffer_sec	0			
10	Time_transient_sec	0			
11	N_points	500			
12	node	<node>			
13	Nominal voltage	15 kV			
14					
15	Time [s]	HARMONIC_RMS_V1_AVG_3	HARMONIC_RMS_V1_AVG_2	HARMONIC_RMS_V1_AVG_1	HARMONIC_RMS_V1_AVG_0
16	0	0	0	8943.546	20.93071
17	0.02	0	0	8943.546	20.93071
18	0.04	0	0	8943.546	20.93071
19	0.06	0	0	8943.546	20.93071
20	0.08	0	0	8943.546	20.93071
21	0.1	0	0	8943.546	20.93071
22	0.12	0	0	8947.105	20.93071
23	0.14	0	0	8949.389	20.93071
24	0.16	0	0	8949.389	20.93071
25	0.18	0	0	8949.389	20.93071
26	0.2	0	0	8949.389	20.93071
27	0.22	0	0	8949.389	20.93071

Figure 5.6: An example of a data sample with the first four harmonic values for phase 1.

5.6 Machine Learning approach

By combining feature extraction methods and machine learning models from different fields within power quality analysis, it should be possible to develop models aimed at fault prediction. In this section, we look at feature extraction methods and machine learning models that have shown potential at similar tasks, and that combined with the available data may be suitable for EarlyWarn's project.

5.6.1 Feature extraction methods

Taking the studied literature in Chapter 4 as our basis for conclusions, the feature extraction methods that have shown the most potential are Wavelet transform, a modified S-transform and Short Time Fourier Transform. These methods have been successfully applied in systems showing good results on real-world data for voltage disturbance classification analysis, which proves that they can extract information from the signal about its current state. If there are disturbances in the signal before a fault occurs, we hypothesize that these extraction methods should be able to pick up these disturbances.

5.6.2 Machine learning models

Focusing on the literature studied in Chapter 4, we see that decision trees and deep neural networks show the most promise as models for processing dynamic power quality data. Both deep recurrent neural networks using an LSTM as cell architecture and Random

Forest models have shown to be successful for classification and stability monitoring, in combination with the listed feature extraction methods. Out of the few studies on voltage disturbance prediction, neural networks have also shown to be successful in this field, at least for high impedance faults.

Other models that have been tested in practice are fuzzy logic and Bayesian classifiers, where especially Hidden Markov Models have shown good potential. In Table 5.5 we summarize our conclusions about what has shown promising results, and what we suggest exploring further.

5.6.3 Conclusion voltage disturbance prediction

The most successful attempts in voltage disturbance prediction has been on the detection of high impedance faults, where neural networks have been successful in usage, and the data is also from the same sensors and sampling frequency that are installed in Norway. Based on this, it shows potential for predicting faults, based purely on sensor data and not including any other type of data, like weather data.

Machine learning method	**Explore further**	**Justification**
Decision Trees	Yes	Promising results on classification, can extract rules and see how it has come to its conclusion
Neural Networks	Yes	Very promising results, LSTM can handle time series data very well, works for most cases
Support Vector Machines	No	Other better solutions, have not been used specifically on our type of data
Fuzzy Logic	No	Requires domain knowledge
Bayesian Classifiers	Yes	Performed well on time series data

Table 5.5: Machine learning methods for future experimentation.

Most other studies referenced in this report have used time-synchronized data from PMU sensors as their data source. As the data made available by SINTEF in this report is PQA data with a higher sampling frequency than PMU sensors, we hypothesize that the same methods used on PMU data can be successfully combined with the dataset available. Some extra preprocessing of the high-frequency data may be required in order to make the models computationally tractable.

Looking at the machine learning models used in previous applications, we conclude that neural network and random forest approaches in combination with the feature extraction methods wavelet transform, modified S-transform and Short Time Fourier Transform show great potential for voltage disturbance prediction.

Chapter 6

Further work

In this chapter, we explore and describe the next steps in evaluating whether a computer system can predict voltage disturbances in the Norwegian power grid. Section 6.1 explains what can be done to improve the dataset of faults is already available, and how existing scripts can be modified to improve the quality of the data. Section 6.2 summarizes our finds in what machine learning methods and extraction methods have shown most potential in the past and gives recommendations for which methods to try out in the future.

6.1 Expanding the dataset

The dataset containing a list of faults in the Norwegian grid currently used is generated automatically by a script that SINTEF has created using sensor data. This script is subject to faults, as it may not happen to find all the errors, and labeling of the data is strictly done by mathematical definitions. To improve upon this process, we suggest that an official list of faults, composed by each producer or using Statnett's statistics, [9], to find faults that have happened and the type of fault. Clearing out data on planned interruptions as mentioned in Section **??** should also be done to avoid excessive noise in the dataset. The same data generation script that is used today can then be used to generate sensor data samples leading up to the faults.

Having generated a training set for fault prediction labeled from historical faults, this could also be used to create a machine learning system able to classify faults that have already happened, producing a dynamically larger dataset of faults, which the fault prediction methods can use for training, as described in Section 4.2.

The DDG script and the connection to the Elspec database can also be improved to include more parameters and data than what is given today. As Section 5.4.1 explained, it is possible to extract more parameters from the database, and this can be expanded to include a wider variety of calculated data. Further research to understand which parameters have the best chance of being descriptive of upcoming faults should be undertaken to find out which parameters that should be used and which parameters that work well together.

Further research should consider accessing the raw waveform data and compute calculations and transformations directly. However, there are some technical limitations here. The data is automatically compressed and in a proprietary format, which can be hard to access.

A script to concatenate the available data should be developed, as each fault and non-fault is curently written to a separate file. A dataset, with each row containing information about the parameters and the label (- type of fault or no fault -) at the end will be useful for further experimentation.

Further research into what type of faults that are possible to predict would also be advantageous when constructing a dataset. What type of faults have incipient faults leading up to them, and deviate from a standard pattern. Research observing whether any specific faults are more detectable than others will be helpful to determine if a machine learning algorithm is able to detect a pattern in a dataset.

6.2 Machine learning

Chapter 4 looked at what has been previously done in the field of power disturbance prediction and classification, and lists the most successfully used methods both in feature engineering and machine learning.

For feature engineering, we first recommend using what is possible to extract from the database without having to do calculations on the raw data and transform these manually. The data that is possible to get included in the current version of the script includes RMS and up to the 512^{th} degree of harmonics.

Machine learning methods that have shown most potential as summarized in Section 5.6.2 are neural networks, decision trees, and their extensions, e.g. LSTM and Random Forest. While other approaches have been successful, they often require domain knowledge, and may thus require a deeper understanding of the problem to implement. Based on our findings we recommend trying out NN and DT and evaluate their performance before other methods are tried out. Decision Trees can also be helpful in finding out which type of parameters that are most descriptive of incipient faults. A Bayesian probability classifier like a Hidden Markov Model could also be tested, as their work on time series data has proven fruitful previously.

We suggest an initial test of multiple methods to test out the potential of each model before further experiments are designed for models that seem most successful. This would require further tuning of parameters and feature engineering.

Chapter 7

Conclusion

Providing a stable power distribution network is of utmost importance, but is becoming more challenging with an aging infrastructure. Maintaining and repairing the power grid is very expensive for the maintainers, with the total cost averaging more than 800 million NOK per year in the Norwegian power grid, as introduced in Section 1.1

In this report, we have looked at the state of the Norwegian power grid, what type of monitoring is done, and how machine learning can be applied to help provide stability and safety in power grids.

The following research questions have been answered:

How can machine learning be used to predict single line failures in the Norwegian Power grid using data from Power Quality Analyzers?

We have looked at previous methods involving machine learning in power quality analysis. By using data from sensors monitoring the voltage level in power lines in combination with feature extraction methods such as Wavelet Transform, S-transform and Short Time Fourier Transform one can obtain suitable datasets. Using these datasets with machine learning models based on neural networks and decision trees, applications have successfully been implemented for monitoring and classification of voltage disturbances. The same methods show promise for being able to predict power disturbances ahead of time.

What is the current state of the art in predicting voltage disturbances in power grids?

There has been very little research done in the field of predicting voltage disturbances, and nominating a state of the art method has proven to be difficult. The most promising results, as seen in Section 4.4 was neural networks, decision trees and HMM. Using neural networks to detect high impedance faults before they can cause any severe damage to equipment or personnel has been successful in some circumstances. In combination with weather data, a Hidden Markov Model has shown promising potential and been successful when used on data from Chinese cities. Most research done using power quality data is centered on classification, where neural networks and decision trees seem to be the most promising choice of machine learning model. Texas A&M's Distribution Fault Anticipation project showed that incipient failures would numerous time give early warning of future outages, and it remains to be seen if there is a pattern to such incipient faults that a machine learning algorithm can recognize.

What kind of large-scale monitoring is done in Norwegian power grids today?

The Norwegian power grid have mainly two types of sensors monitoring the state of the power grid through real-time analysis. These are Phasor Measurement Units (PMU), and Power Quality Analyzers (PQA). PMU sensors have the advantage of being synchronized with each other through GPS, while PQA sensors have a superior sampling rate, and are thus able to give further insight into the data.

A combination of using both types of sensors for power quality monitoring is advantageous. PMUs synchronization can be used for monitoring the state of the power grid as a whole, and follow the consequences of faults that may cascade between lines in the grid. Meanwhile, PQA data can be used for further analysis and disturbance prediction where PMU data proves to be insufficient sampling rate, and focus on predicting single line failures and faults to equipment.

Final remarks

There exist massive amounts of data from long-time monitoring of the Norwegian power grid. There is still significant work remaining in creating a suitable dataset for a machine learning application, but it is in theory possible to extract all the information needed. While not much work has been done in the field of predicting voltage disturbances, it has been shown that one can detect incipient faults in advance. The assumption is that enough faults have incipient faults to be used for differentiating between standard operating procedures. Utilizing state start of the art methods from voltage disturbance classification and adapt these to suit a longer timespan and detecting patterns occurring before a disturbance should be within reach of a well-implemented machine learning method.

Bibliography

[1] Norges vassdrags-og energidirektorat. Overview of Norway's electricity history. http://publikasjoner.nve.no/rapport/2017/rapport2017_15.pdf, 2017. Accessed: 2018-09-27.

[2] Nur IqtiyaniIlham, M Hasanuzzaman, and M Hosenuzzaman. European smart grid prospects, policies, and challenges. *Renewable and Sustainable Energy Reviews*, 67:776–790, 2017.

[3] Olje og energidepartementet. Fakta energi- og vannressurser i Norge. https://www.regjeringen.no/contentassets/fd89d9e2c39a4ac2b9c9a95bf156089a/1108774830_897155_fakta_energi-vannressurser_2015_nett.pdf. Accessed: 2018-12-05.

[4] Norges vassdrags-og energidirektorat. Avbrotsstatistikk 2017. http://publikasjoner.nve.no/rapport/2018/rapport2018_64.pdf, 2018. Accessed: 2018-10-26.

[5] International Energy Agency. *World Energy Outlook*, chapter 1, pages 38–43. IEA Publications, 2018.

[6] Statnett SF. Strømnettets historie. http://www.statnett.no/nettutvikling/nettplan-stor-oslo/om-nettplanen/stromnettets-historie/. Accessed: 2018-10-26.

[7] Statistisk sentralbyrå. Kraftinvesteringer i støtet. https://www.ssb.no/energi-og-industri/artikler-og-publikasjoner/kraftinvesteringer-i-stotet/. Accessed: 2018-10-26.

[8] Statistisk sentralbyrå. Investeringer i olje og gass, industri, bergverk og kraftforsyning. https://www.ssb.no/energi-og-industri/statistikker/kis. Accessed: 2018-10-26.

[9] Statnett. Årsstatistikk 2017. http://www.statnett.no/global/dokumenter/kraftsystemet/systemansvar/feilstatistikk/%C3%85rsstatistikk%202017%2033_420%20kv.pdf. Accessed: 2018-10-26.

[10] Gerd Kjølle. KILE. https://www.sintef.no/projectweb/kile/. Accessed: 2018-10-09.

[11] Paul Horowitz and Winfield Hill. *The Art of Electronics*, pages 1–2, 14, 1099. Cambridge University Press, New York, NY, USA, 3 edition, 2015.

[12] MJ Matabaro and G Atkinson-Hope. Efficiency of an alternating current transmission line converted into a direct current system. In *Industrial and Commercial Use of Energy Conference (ICUE), 2013 Proceedings of the 10th*, pages 1–6. IEEE, 2013.

[13] Vijay Venu Vadlamudi. Fundamentals of Power Systems Refresher for TET4115 (Power System Analysis), 2018.

[14] Sergey Makarov, Reinhold Ludwig, and Stephen Bitar. *Practical Electrical Engineering*. Springer International Publishing, 2016.

[15] Mihaela Albu and Gerald T. Heydt. On the use of RMS values in power quality assessment. *IEEE Transactions on Power Delivery*, 18(4):1586–1587, 2003.

[16] Richard Dorf and James Svoboda. *Introduction to Electric Circuits*, pages 502, 871. John Wiley and Sons, Inc, 8 edition, 2010.

[17] J. Barros and E. Pérez. Limitations in the Use of R.M.S. Value in Power Quality Analysis. In *IEEE Instrumentation and Measurement Technology Conference Proceedings*, pages 2261–2264, 2006.

[18] Lucas Illing. Fourier analysis. https://www.reed.edu/physics/courses/physics331.f08/pdf/fourier.pdf, 2008. Accessed: 2018-10-14.

[19] Julius O. Smith. *Mathematics of the Discrete Fourier Transform (DFT)*, chapter 1. W3K Publishing, 2007.

[20] Juhan Nam. Super-resolution spectrogram using coupled PLCA. https://ccrma.stanford.edu/~juhan/super_spec.html, 2010. Accessed: 2018-12-06.

[21] Hai-Man Chung and Don C. Lawton. Amplitude responses of thin beds: Sinusoidal approximation versus Ricker approximation. *GEOPHYSICS*, 60(1):223–230, 1995.

[22] Yu-Hsiang Wang. The Tutorial: S Transform, 2010.

[23] Edith Clarke. Simultaneous faults on three-phase systems. *Transactions of the American Institute of Electrical Engineers*, 50(3):919–939, 1931.

[24] Yong Sheng and Steven M. Rovnyak. Decision tree-based methodology for high impedance fault detection. *IEEE Transactions on Power Delivery*, 19(2):533–536, 2004.

[25] M. Sedighizadeh, A. Rezazadeh, and Nagy I. Elkalashy. Approaches in high impedance fault detection a chronological review. *Advances in Electrical and Computer Engineering*, 10(3):114–128, 2010.

[26] Amin Ghaderi, Herbert L. Ginn III, and Hossein Ali Mohammadpour. High impedance fault detection: a review. *Electric Power Systems Research*, 143:376–388, 2017.

[27] Joseph Seymour and Terry Horsley. The seven types of power problems. *APC, USA*, 2005.

[28] Thomas M Mitchell. *Machine Learning*. McGraw-Hill, Inc., New York, NY, USA, 1 edition, 1997.

[29] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. http://www.deeplearningbook.org.

[30] Yann LeCun, Yoshua Bengio, and Geoffrey E. Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.

[31] Leo Breiman. Bagging predictors. *Machine Learning*, 24(2):123–140, 1996.

[32] Robert E. Schapire. The strength of weak learnability. *Machine Learning*, 5(2):197–227, 1990.

[33] Favio Vázquez. Deep learning made easy with deep cognition. https://becominghuman.ai/deep-learning-made-easy-with-deep-cognition-403fbe445351. Accessed: 2018-12-07.

[34] Antonio Salmerón, Rafael Rumí, Helge Langseth, Thomas D Nielsen, and Anders L Madsen. A review of inference algorithms for hybrid Bayesian networks. *Journal of Artificial Intelligence Research*, 62:799–828, 2018.

[35] Padhraic Smyth. Hidden Markov models for fault detection in dynamic systems. *Pattern recognition*, 27(1):149–164, 1994.

[36] Sreerama K Murthy. Automatic construction of decision trees from data: A multi-disciplinary survey. *Data mining and knowledge discovery*, 2(4):345–389, 1998.

[37] Andy Liaw, Matthew Wiener, et al. Classification and regression by random forest. *R news*, 2(3):18–22, 2002.

[38] Andreas Marouchos. DS Fundamentals: Support Vector Machines. https://dataincognita.net/2017/12/06/ds-fundamentals-p2-support-vector-machines/. Accessed: 2018-10-12.

[39] Duane Q. Nykamp. Dynamical system definition. https://mathinsight.org/definition/dynamical_system. Accessed: 2018-09-27.

[40] Elizabeth D Liddy. Natural language processing. In *Encyclopedia of Library and Information Science*, volume 18. Marcel Decker, 2 edition, 2001.

[41] Niklas Donges. Recurrent Neural Networks and LSTM. https://towardsdatascience.com/recurrent-neural-networks-and-lstm-4b601dd822a5. Accessed: 2018-09-28.

[42] Dong Chan Lee. Automatic power quality monitoring with recurrent neural network. Master's thesis, University of Toronto, 2016.

[43] Christian Andre Andresen, Bendik Torsæter, Hallvar Haugdal, and Kjetil Uhlen. Fault detection and prediction in smart grids. *IEEE International Workshop on Applied Measurements for Power Systems*, 2018.

[44] Ana Claudia Barros, Mauro S Tonelli-Neto, José Guilherme Magalini Santos Decanini, and Carlos Roberto Minussi. Detection and classification of voltage disturbances in electrical power systems using a modified euclidean artmap neural network with continuous training. *Electric Power Components and Systems*, 43(19):2178–2188, 2015.

[45] Om Prakash Mahela, Abdul Gafoor Shaik, and Neeraj Gupta. A critical review of detection and classification of power quality events. *Renewable and Sustainable Energy Reviews*, 41:495–505, 2015.

[46] Atish K Ghosh and David L Lubkeman. The classification of power system disturbance waveforms using a neural network approach. *IEEE Transactions on Power Delivery*, 10(1):109–115, 1995.

[47] AM Gaouda, MMA Salama, MR Sultan, and AY Chikhani. Power quality detection and classification using wavelet-multiresolution signal decomposition. *IEEE Transactions on power delivery*, 14(4):1469–1476, 1999.

[48] D Granados-Lieberman, Rj Romero-Troncoso, Ra Osornio-Rios, A Garcia-Perez, and E Cabal-Yepez. Techniques and methodologies for power quality analysis and disturbances classification in power systems: a review. *IET Generation, Transmission & Distribution*, 5(4):519–529, 2011.

[49] H Siahkali. Power quality indexes for continue and discrete disturbances in a distribution area. In *Power and Energy Conference, 2008. PECon 2008. IEEE 2nd International*, pages 678–683. IEEE, 2008.

[50] Florence Choong, MBI Reaz, and F Mohd-Yasin. Advances in signal processing and artificial intelligence technologies in the classification of power quality events: a survey. *Electric Power Components and Systems*, 33(12):1333–1349, 2005.

[51] Przemysław Janik and Tadeusz Lobos. Automated classification of power-quality disturbances using SVM and RBF networks. *IEEE Transactions on Power Delivery*, 21(3):1663–1669, 2006.

[52] Peter GV Axelberg, Irene Yu-Hua Gu, and Math HJ Bollen. Support vector machine for classification of voltage disturbances. *IEEE Transactions on power delivery*, 22(3):1297–1303, 2007.

[53] Shunfan He, Kaicheng Li, and Ming Zhang. A real-time power quality disturbances classification using hybrid method based on S-transform and dynamics. *IEEE transactions on instrumentation and measurement*, 62(9):2465–2475, 2013.

[54] Raj Kumar, Bhim Singh, DT Shahani, Ambrish Chandra, and Kamal Al-Haddad. Recognition of power-quality disturbances using S-transform-based ANN classifier and rule-based decision tree. *IEEE Transactions on Industry Applications*, 51(2):1249–1258, 2015.

[55] Prakash K Ray, Soumya R Mohanty, Nand Kishor, and João PS Catalão. Optimal feature and decision tree-based classification of power quality disturbances in distributed generation systems. *IEEE Transactions on Sustainable Energy*, 5(1):200–208, 2014.

[56] Ebrahim Balouji, Irene YH Gu, Math HJ Bollen, Azam Bagheri, and Mahmood Nazari. A LSTM-based deep learning method with application to voltage dip classification. In *Harmonics and Quality of Power (ICHQP), 2018 18th International Conference on*, pages 1–5. IEEE, 2018.

[57] Ruisheng Diao, Kai Sun, Vijay Vittal, Robert J O'Keefe, Michael R Richardson, Navin Bhatt, Dwayne Stradford, and Sanjoy K Sarawgi. Decision tree-based online voltage security assessment using PMU measurements. *IEEE Transactions on Power Systems*, 24(2):832–839, 2009.

[58] Michael Negnevitsky, Nikita Tomin, Victor Kurbatsky, Daniil Panasetsky, Alexey Zhukov, and Christian Rehtanz. A random forest-based approach for voltage security monitoring in a power system. In *PowerTech, 2015 IEEE Eindhoven*, pages 1–6. IEEE, 2015.

[59] Aleksey Zhukov, Nikita Tomin, Denis Sidorov, Daniil Panasetsky, and Vadim Spirayev. A hybrid artificial neural network for voltage security evaluation in a power system. In *Energy (IYCE), 2015 5th International Youth Conference on*, pages 1–8. IEEE, 2015.

[60] Amr M Ibrahim and Noha H El-Amary. Particle swarm optimization trained recurrent neural network for voltage instability prediction. *Journal of Electrical Systems and Information Technology*, 5(2):216–228, 2018.

[61] T. Guo and J. V. Milanović. On-line prediction of transient stability using decision tree method — sensitivity of accuracy of prediction to different uncertainties. In *2013 IEEE Grenoble Conference*, pages 1–6, June 2013.

[62] Mohammed R Qader, Math HJ Bollen, and Ron N Allan. Stochastic prediction of voltage sags in a large transmission system. *IEEE Transactions on Industry Applications*, 35(1):152–162, 1999.

[63] Nelson Kagan, Nelson M Matsuo, Eduardo L Ferrari, Valdomiro Vega-García, Juan C García-Arias, Tiago P Souza, Alberto B Neto, Leonardo B Chaves, Fernanda S Ide, Luciano AF Carvalho, et al. Computerized system for detection of high impedance faults in MV overhead distribution lines. In *Harmonics and Quality of Power (ICHQP), 2016 17th International Conference on*, pages 992–997. IEEE, 2016.

[64] A. M. El-Hadidy and C. Rehtanz. Early detection of tree faults. In *22nd International Conference and Exhibition on Electricity Distribution (CIRED 2013)*, pages 1–5, June 2013.

[65] Tilman Weckesser, Hjörtur Jóhannsson, and Thierry Van Cutsem. Early prediction of transient voltage sags caused by rotor swings. In *PES General Meeting— Conference & Exposition, 2014 IEEE*, pages 1–5. IEEE, 2014.

[66] B Don Russell, Carl L Benner, Robert M Cheney, Charles F Wallis, Thomas L Anthony, and William E Muston. Reliability improvement of distribution feeders through real-time, intelligent monitoring. In *Power & Energy Society General Meeting, 2009. PES'09. IEEE*, pages 1–8. IEEE, 2009.

[67] B Don Russell and Carl L Benner. Intelligent systems for improved reliability and failure diagnosis in distribution systems. *IEEE Transactions on Smart Grid*, 1(1):48–56, 2010.

[68] Karthick Manivinnan, Carl L Benner, B Don Russell, and Jeffrey A Wischkaemper. Automatic identification, clustering and reporting of recurrent faults in electric distribution feeders. In *Intelligent System Application to Power Systems (ISAP), 2017 19th International Conference on*, pages 1–6. IEEE, 2017.

[69] Fei Xiao and Qian Ai. Data-Driven Multi-Hidden Markov Model-Based Power Quality Disturbance Prediction that Incorporates Weather Conditions. *IEEE Transactions on Power Systems*, 2018.

[70] Christian Andre Andresen. Earlywarn. https://www.sintef.no/en/projects/earlywarn/. Accessed: 2018-09-26.

[71] Christian Andresen. EarlyWarn Presentation.

[72] Elspec Ltd. Power Quality Analyzer. https://cuthbertsonlaird.co.uk/power-quality/328-blackbox-g4500.html. Accessed: 2018-10-04.

[73] Matthias Hofmann, Helge Seljeseth, Gro Holst Volden, and GH Kjølle. Study on estimation of costs due to electricity interruptions and voltage disturbances. *SINTEF Energy Rearch*, 2010.

[74] Forskrift om systemansvaret i kraftsystemet. https://lovdata.no/dokument/SF/forskrift/2002-05-07-448. Accessed: 2018-11-06.