Henrik Vassdal
Yuval Abraham Regev

Machine Learning Methods for Anomaly Detection on Phasor Measurement Unit Data

Masteroppgave i Energi og Miljø
Veileder: Ümit Cale
Medveileder: Ugur Halden
Juni 2022

Masteroppgave

NTNU
Norges teknisk-naturvitenskapelige universitet
Fakultet for informasjonsteknologi og elektroteknikk
Institutt for elkraftteknikk

NTNU
Norwegian University of
Science and Technology

Henrik Vassdal
Yuval Abraham Regev

Machine Learning Methods for Anomaly Detection on Phasor Measurement Unit Data

Masteroppgave i Energi og Miljø
Veileder: Ümit Cale
Medveileder: Ugur Halden
Juni 2022

Norges teknisk-naturvitenskapelige universitet
Fakultet for informasjonsteknologi og elektroteknikk
Institutt for elkraftteknikk

NTNU

Kunnskap for en bedre verden

Abstract

In a world of increased digitization and digitalization together with the increased demand of power and renewable energy sources, the cyber-physical interaction in the power system is expanding at a fast pace. This expansion requires new and improved techniques to ensure a stable and protected power system, able to deliver the necessary power demanded from the grid. Phasor Measurement Units (PMUs) provide the ability to improve monitoring and protection capabilities in power systems. These measurement units allow for precise high frequency measurements throughout the power system, increasing the situational awareness for system operators. However, these units require multiple components and systems working in unison, as well as data transmission over long distances. This makes the measurement system vulnerable for anomalies originating in the system itself, as well as external malicious cyber attacks. It is therefore crucial for rapid and effective anomaly detection schemes to protect and hinder instability in the power system.

This thesis investigates and proposes an anomaly detection method using machine learning and artificial intelligence models. Prediction based machine learning models together with error thresholding is used to locate and label anomalous data provided by PMUs. Two different data sets with PMU measurements are analyzed. The data sets contain real measurements from the power system in Norway and Texas, USA. The hybridization of machine learning models based on Convolutional Neural Network (CNN) and Long Short Term Memory (LSTM) architecture proved more effective in detecting anomalies in the tests completed, with a higher percentage of anomalies detected together with fewer false positives. The models lacked in performance detecting drift anomalies, however showed promising results for anomaly detection for injected noise, spikes and offset anomalies, laying a good foundation for further model development and real-time model deployment.

Sammendrag

I en verden av økende digitalisering, kombinert med en økende etterspørsel etter strøm fra rene energikilder, vokser den cyber-fysiske interaksjonen i kraftsystemet stadig fortere. Denne veksten krever nye og bedre teknikker for å sikre et stabilt og beskyttet kraftsystem, med evnen til å møte etterspørselen i samfunnet. Phasor Measurement Umits (PMUs) gir nettselskapene muligheten til å forbedre deres overvåknings- og sikkerhetskapasitet. Disse måleenhetene tilbyr presise høyfrekvente målinger gjennom hele kraftsystem, og øker dermed nettselskapenes systemoversikt. Samtidig krever disse enhetene at flere komponenter og systemer fungerer feilfritt, blandt annet dataoverføring over lange avstander. Dette gjør måledataen sårbar for feil som har sin opprinnelse i selve systemet, samt eksterne nettangrep. Det er derfor avgjørende for raske og effektive feildeteksjonsmodeller for å beskytte og hindre ustabilitet i kraftsystemet.

Denne masteroppgaven undersøker og foreslår en metode for feildeteksjon ved bruk av maskinlæring og modeller for kunstig intelligens. Prediksjonsbaserte maskinlæringsmodeller sammen med metoder for terskelverdier brukes til å lokalisere og merke unormale måleverdier levert av PMUer. To forskjellige datasett med PMU målinger har blitt analysert. Datasettene inneholder ekte målinger fra transmisjonsnettet i Norge, og i Texas, USA. Maskinlæringsmodeller basert på en kombinasjon av Convolutional Neural Networks (CNN) og Long Short Term Memory (LSTM) arkitektur viste seg å være mest presise i å finne feil under testene som ble gjennomført. Disse modellene fant en høyere prosent av feilene, samtidig som de markerte færre falske positive. Alle modellene manglet evnen til å detektere avdriftsavvik, men viste derimot lovende resultater for deteksjon av injisert støy, hopp og kortvarige topper. Dette legger et godt grunnlag for videre modellutvikling, og til slutt implementering av modellen i sanntid.

Preface

This thesis is the conclusion of our MSc in Electrical Energy Technology and Smart Grids, carried out at the Department of Electrical Power Engineering, at the Norwegian University of Science and Technology.

We would like to express our deepest gratitude towards our supervisor Prof. Ümit Cali for excellent guidance the last year. Also, a special thanks to our co-supervisor Ugur Halden for contributions and good discussions during the thesis work.

Lastly, we would like to thank all our classmates at Energi og Miljø, and especially at Satsesalen, for making this 5-year period the best we have had so far.

Norwegian University of Science and Technology
Trondheim, June 2022

Table of Contents

List of Figures

List of Tables

1 Introduction

Operating and maintaining a functioning and reliable power grid is becoming increasingly difficult with exceedingly complex power systems [1]. The increasing electricity demand over the last century has forced the power systems to become vastly greater in size and complexity. The future power system will have to deal with a higher share of Variable Renewable Energy Sources (VRES), which are introduced in order to reach the climate goals of 2050 [2]. A robust power grid is necessary to accommodate for the expansion, requiring precise measurement and analysis techniques to ensure the reliability and availability of power that society demands.

Conventional power grid State Estimation (SE) techniques using Supervisory Control and Data Acquisition (SCADA) systems monitors and collects data from connected Remote Terminal Units (RTUs) stationed in the grid [3]. These RTUs measure voltage magnitude as well as active and reactive power at their locations. The system works by collecting data over a time interval, and as such the technique provides an overview of the grid and a SE. However, because the different RTUs are not time synchronized, the measurements might not be relevant for the power grid state if a change or an anomalous event has occurred during or after the time interval [4] [5].

The introduction of Phasor Measurement Units (PMUs) has increased the potential for more frequent grid measurements and therefore, also increased the potential for grid SE [4]. The time synchronous measurements allows for a clearer and more immediate picture of the state of the power grid, increasing the situational awareness for System Operators (SOs). The PMUs are able to provide real-time online measurements across the grid, using a measurement frequency close to the grid frequency. This allows for rapid anomaly detection and continuous SE, however the immense amount of data can cause issues. With a high measurement frequency, where each PMU can measure voltage magnitude, angle, frequency and current, millions of data points are accumulated quickly. Processing and analyzing such vast amounts of data requires special techniques and methods [5].

This thesis will look into the cyber-physical interaction between power systems and SOs. More specifically what types of anomalies and cyber attacks occur in power systems and how to most effectively detect them, using artificial intelligence (AI) and machine learning (ML) techniques. Prediction based models based on AI and ML architecture allows for the use of error-threshold based anomaly detection methods [6]. The last few decades, a huge amount of work and research has gone into developing ML models, allowing for specific models for specialized use cases.

This thesis will test the strengths and weaknesses of single-output prediction based error-thresholding models comparing pure and hybridized ML models based on Convolutional Neural Networks (CNN) and Long Short Term Memory (LSTM) architecture. Investigating the possibility of improving the cyber-physical resilience of PMU integrated power systems.

The scope of this thesis can be summarized into these points: (1) Presentation of different anomalies commonly found in the power system. (2) Investigate the impacts of anomalies on power systems and their political and social consequences. (3) Comprehensive review of different AI/ML methods for anomaly detection purposes. (4) Development of pure and hybrid AI-based anomaly detection models using PMU data. (5) Testing the developed models for anomaly detection purposes with various false data injection attacks. (6) Comparative evaluation of the models and their anomaly detection results as a sensitivity analysis.

2 Background Information

This section is partly based on previous work done by the same authors [5].

The power system is a complex system that requires precise analysis and operation. Introducing PMUs to the power system provides new possibilities for power system analysis and state estimation. This section will look into and explain the workings of PMUs and how they are integrated in the power system. Further, the focus will shift towards the cyber-physical interaction that the PMUs require. With this interaction follows some drawbacks in the form of vulnerabilities. Anomalies in the power system will be explained further, these include both physical and cyber related. The impacts of these anomalies will be presented, as well as methods for anomaly detection that has previously been utilized. The methods discussed are mainly methods within machine learning, both supervised and unsupervised. Statistical anomaly detection methods are also presented.

2.1 The Power System

A power system can be simplified and explained by the coexistence of load centers and generation units connected by transmission lines. For power systems on regional, national and international levels, thousands of different components and installations work together to bring power from the generation source to the end user. These components include generators, transmission lines, transformer stations, loads as well as other electrical installations [7]. These components work in harmony to satisfy the main function of the system - to reliably supply electrical power to the consumer while still being economically efficient [8].

For the purpose of this thesis, a closer look at the transmission system is necessary. As the generation sites and load centers are usually spaced far apart, a transmission system is needed to connect both points. The transmission system also allows for a more interconnected grid on an international scale introducing power as a commodity that is available to be bought and sold on a power marked, which is essential for the balance between supply and demand [8].

To most effectively transmit the power the transmission system is usually divided into different grids. These grids have their own use, either it is for long distance transmission or distribution to different smaller loads, like buildings and other infrastructure. The Norwegian transmission system is divided into three main grids [9]:

- **Transmission Grid**: $\geq 132kV$
- **Subtransmission Grid**: $22kV - 132kV$
- **Distribution Grid**: $230V - 22kV$

Operating with different grids allows for efficiently balancing safety, losses and ease of use. Figure 1 shows the structure of a power system with the three different grids represented. The following subsections will explain each grid in more detail.

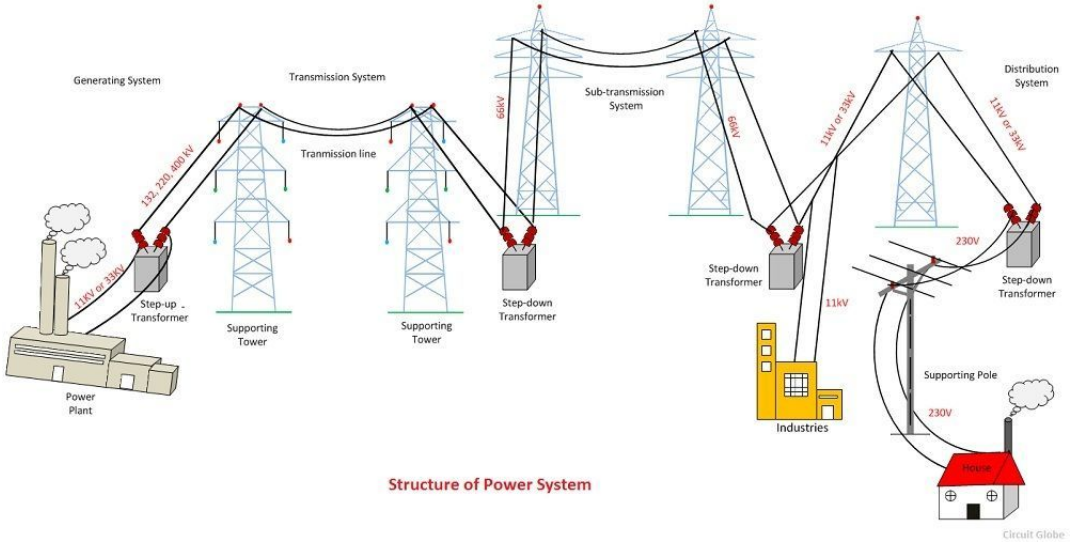


Figure 1: Illustration of the power system [10].

2.1.1 Transmission Grid

The different power grids can be compared to the road network. The transmission grid being the high speed highways of the power system, moving large amount of power from one location to another through high voltage lines and cables. The transmission grid connects the power system on national and international scales by providing high capacity transmission between locations. The transmission grid does not supply individual areas with power, but operates in a point-to-point fashion, where the subtransmission grid and distribution grid further distributes the power to the consumers.

Each countries' transmission grid is usually operated by a single Transmission System Operator (TSO). These TSOs are responsible for the transmission grid and operates under strict regulatory control [11]. In Norway the TSO is Statnett which is owned by the Norwegian state through the Ministry of Petroleum and Energy (MPE) with its supervisory authority The Norwegian Water Resources and Energy Directorate (NVE). Statnett's responsibilities as a TSO are divided into three categories [12]:

- **System Operator**: Ensure instantaneous balance between consumption and production of electrical power.
- **Grid Owner**: Own and maintain the Norwegian transmission grid and its connections to other countries' transmission grids.
- **Grid Planner**: Planning and building the transmission grid.

With the green shift in full motion, the TSOs play a key role in the development and expansion of VRES. In the EU, the increase in VRES has led to the ambitious goal of creating a large interconnected European grid to withstand the varying power production across borders expected in the future. The TSOs will be responsible for this expansion, where innovation and investments into infrastructure is needed to accomplish the goal, as well as restructuring and development of new technology and digital tools [11].

2.1.2 Subtransmission Grid

The subtransmission grid is the next step down in the transmission system. This grid is the connection between the transmission grid and the distribution grid. The subtransmission grid brings high voltage power closer to load centers, like city centers, where it is further downscaled to match the voltage of the distribution grid. The grid also supplies power intensive industries, production facilities and buildings. Both the subtransmission grid and distribution grid are operated by Distribution System Operators (DSOs). In Norway, these companies span and are owned by multiple Norwegian counties [8].

2.1.3 Distribution Grid

The distribution grid is the main supplier for end users. This grid is operated by the DSOs and runs at a voltage level of 22 kV and below. This grid is by far the largest grid of the three, spanning around 100 000 km of lines and cables in Norway [8]. The grid is separated and categorized into low and high voltage distribution, having 1 kV as the upper and lower limit for the low and high voltage grids respectively.

2.2 Phasor Measurement Unit

PMUs are units capable of measuring positive sequence voltages and currents, the phase angle and Rate Of Change Of Frequency (ROCOF) can then be calculated with high accuracy. Using Global Positioning System (GPS) communication to synchronize the measurements across the power grid allows for anomalies to be detected [13]. In the 1980s the first PMU devices were introduced, these were, however, not utilized to a great degree. Further advancements in its technology has made the PMUs more attractive to be used by the TSOs, and only now their true potential is starting to be exploited. The occurrence of blackouts in major power systems globally has driven the industry towards implementing Wide-Area Monitoring Systems (WAMS). PMUs are central to the implementation of WAMS since the data gathered can be used to locate anomalies quickly and initiate preventive measures to avoid larger cascading faults [14].

Traditional SCADA measurement techniques do not allow for a high measurement rate. The higher frequency of measurements, which the PMU delivers, provide an excellent opportunity to get a clearer and more accurate picture of the state of the power grid [15]. Best practices for power grid state estimations have been lacking in real time, having estimates take tens of seconds, and sometimes minutes [3]. Using PMUs along with GPS communication allows for the ability to make syncrophasors. A block diagram presents the PMU components in figure 2. Syncrophasors are time syncronized phasor measurements at different locations which provides information on both the supply and demand side of the grid within the same time frame. This gives vital information about the correlation of the demand and supply side of the system [16]. The synchrophashor will be discussed in detailed manner in subsection 2.1.2.

PMUs are designed to make high frequency measurements, typically around the grid frequency, however newer technology using µPMUs allow for a much higher sampling rate exceeding 10 000 samples per second. The µPMUs are used on the distribution grid levels [17], and therefore not relevant for this thesis.

The vast amount of data that PMUs collects is considered a potential big resource for the system operators. However, tapping in to this potential has proven to be a considerable challenge [18]. Developing good and effective machine learning algorithms is one way to extract the most important information from large data sets.

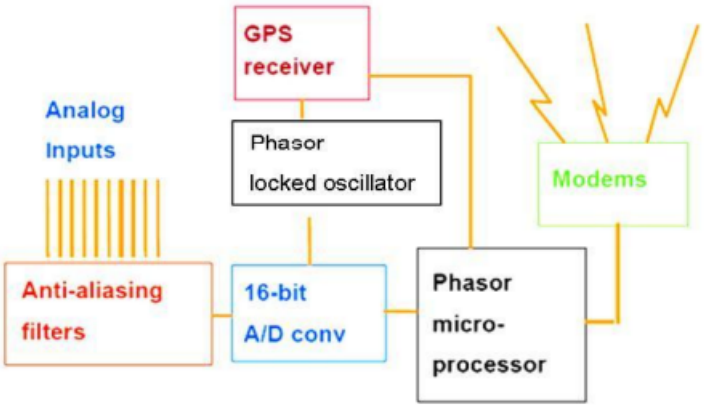


Figure 2: PMU block diagram [16].

2.2.1 Phasor Representation

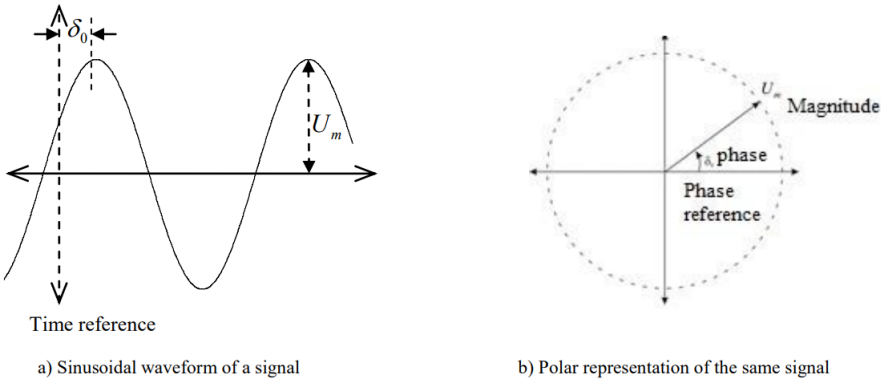


Figure 3: Sinusoidal and polar representation of a phasor [16].

The measurements from PMUs come in the form of phasors. A phasor is a vector representation of a voltage or current. In electrical engineering, the phasor represents the magnitude of current or a voltage and its angle with respect to a reference point. In this representation, the positive angles are counted in a counter-clockwise direction. The mathematical expression for a sinusoidal wave is given by:

$$u(t) = U_m cos(\omega t + \delta_0) \tag{1}$$

The corresponding synchrophasor representation is given by:

$$U := \frac{U_m}{\sqrt{2}} e^{j\delta_0} = \frac{U_m}{\sqrt{2}}(cos(\delta_0) + jsin(\delta_0)) = Re\{U\} + jIm\{U\} \tag{2}$$

Where $\frac{U_m}{\sqrt{2}}$ represents the root mean square value (RMS) of the sinusoid, the angular frequency, $\omega = 2\pi f$, where f is the frequency and j is the imaginary number. The frequency is either 50 or 60 Hz depending on what frequency is used in the power system, where 60 Hz is mostly used in North and South America, while 50 Hz dominates the rest of the world. δ_0 is the phase angle of the signal with regards to the reference point, and U_m is the magnitude; both are represented in Figure 3.

This type of technology is considered to become an important tool in the future power system [16]. The advantage lies in the PMUs ability to monitor voltages and currents over large areas with the use of GPS technology. Synchronizing the different signals from different locations on the grid is done with respect to the same phasor reference [16]. This is visualised in Figure 4:

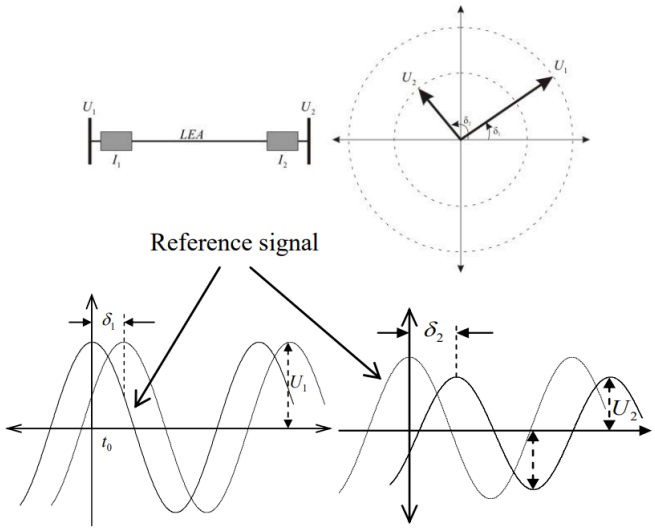


Figure 4: Phasors at different locations has the same signal as reference [16].

In an ideal environment the measured phasor is constant. However, in a real power system, the input signals are not constant and their frequency may be variable. Since all phasors that share a diagram in a WAMS, shown in Figure 4, must have the same frequency, the issue of varying signals needs to be dealt with [14].

It is further explained in [14] that the PMUs job is to monitor these changes, but due to differing frequencies this is not a straightforward task. To deal with this issue there is a need for considering the input signal data within a limited time period. The time period, which is used is usually one period of the fundamental frequency of the input data. The data considered in this time period is called a sample. If the input signal's frequency differs from the nominal frequency of the power system (which it often does when at least two decimals are considered), the PMU estimates the fundamental frequency component. This is done using a frequency-tracking step which gives a desired estimation for the phasor.

The input signal usually has harmonic and non-harmonic components. The PMU needs to separate the fundamental frequency from these signals to find the phasor representation. Discrete Fourier Transform (DFT) is the most common way of dealing with this issue. DFT is applied to the data sample which the PMU considers, and its phasor is calculated. The sampled data represents the input signal which needs to be filtered for noise such that the results will be more accurate. This process is shown as the analog anti-aliasing filter in Figure 2. Due to the Nyquist criterion, the

pass band of the anti-aliasing filter is limited to less than half of the sampling frequency of the PMU [14]. PMUs measure 50/60 Hz AC voltages and currents usually at 48 samples per cycle (2400/2880 samples per second) [16]. The phasor, X, is then given by [19]:

$$X = \frac{1}{\sqrt{2}} \frac{2}{N} \sum_{k=0}^{N-1} x_k \epsilon^{-jk\frac{2\pi}{N}} \quad \forall k \in N = \{0, 1..., N-1\} \tag{3}$$

Where N is the number of samples over a period of the input signal and x_k is the measured value of sample k.

The summation is multiplied by $\frac{1}{\sqrt{2}}$ since X is converted to its RMS value. Each frequency ω component of the signal calculated from DFT has a complex conjugate and thus appear at $\pm\omega$ in the graph. Combining these and dividing by the number of measurements in the sample (N) to obtain the average value, explains the factor $\frac{2}{N}$ in front of the summation in Equation 3 [19]. The fundamental frequency's peak value is thus obtained and the harmonics of the sample is eliminated by the DFT. However, there is an error in the estimation of the phasor which is caused by non-harmonics and other random noise in the sample [14].

2.2.2 Synchrophasor Definition

The PMUs ability to synchronize each measurement over a large area using GPS deems it a promising technology for the future power systems [16]. The term synchrophasor describes the phasors that have been estimated at a given time stamp. For WAMS it is essential to synchronize the time stamps over a large area, to make sure that the measurements happened at exactly the same time. However, the anti-aliasing filter causes an issue for the synchronization due to applying a phase delay on the input signal. This delay depends on the characteristics of the filter as well as the frequency of the signal. For the synchronization to be true, the PMU needs to compensate for this delay as the measurement is performed after the filter [14], as is shown in Figure 5

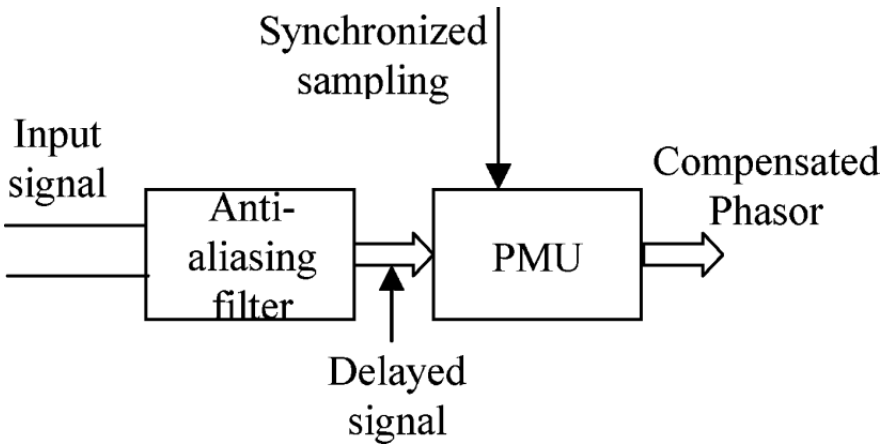


Figure 5: Block diagram of how the PMU compensates for the phase delay created by the anti-aliasing filter [14].

To ensure correct synchronization, the PMUs take advantage of the internal clock of the GPS. This clock is very accurate due to its connection to satellites in orbit around Earth. The GPS conveys this information by sending out a one-pulse-per-second signal [20]. The PMUs synchronization is based on using a sampling clock which is phase locked to the signal from the GPS. The time

stamps are created by the PMU at multiples of the power systems' nominal frequency [14].

An Analog to Digital converter digitizes the analog wave forms for each phase. The GPS receiver and phase-lock oscillator [21] synchronizes each sample with a location and a time stamp within an accuracy of one microsecond. Now that the phasors are time tagged and located the samples are transferred to a receiver at up to 60 Hz [16].

2.2.3 Phasor Data Concentrators

Synchrophasors only exist if there are multiple phasors to time align and synchronize. Therefore it is necessary to compare the data collected from multiple PMUs in one system. To connect and aggregate the data provided by different PMUs in multiple different locations across a country, Phasor Data Concentrators (PDCs) are used. PDCs collect data from multiple PMUs and processes it for time alignment, and therefore synchronization. This process includes aggregating and validating before the data is archived and further distributed for storage or processing [22]. Due to the PMUs being located at different locations, sometime spanning hundreds of kilometers, delays are inevitable. Depending on the use of PMU data, forwarding data from the PDCs at a specific time interval might be more important than data completeness. The PDCs can therefore also have the function to filter out data points collected from the PMUs with excessive delays [22].

2.3 Anomalies in Power System

The power grid is a vast and complex infrastructure which is prone to a number of different anomalies and faults. This section will give an understanding of the anomalies that might occur, and the causes behind them. When working with PMU data the different types of anomalies will reveal themselves differently. Therefore a deep understanding of these differences is essential for the work in this thesis [23].

This section discusses two different types of anomalies:

1. Bad data
2. Physical faults

Bad data refers to reported measurement values which does not correspond to the true values of the system. These might occur due to faulty measurement or issues with the communication between the PMU and the data receiver. If bad data is not detected, the SO might act and implement wrong measures as the operator might believe the system being in another state. Cyber attacks affecting the data received by the PMU can also go under the category bad data as they affect the reliability of the data. This is either done subtly, or by overloading the system with a lot of noise. The subtle attacks attempt not to get detected by simulating the normal behaviour of the power system, but perhaps in another state then the real one. The purpose of these attacks is to mislead the system operator and are called false data injection (FDI) attacks. The second type of cyber attacks, called Denial of service (DoS) attacks, disrupts the data flow by overloading the communication channels with information. The purpose of these attacks is to deny the SOs the opportunity to read the measurements she relies on [24].

Physical faults are anomalies that happen on the different components that constitutes the power grid. There are many types of physical faults that affect the system, including:

- Line outages
- Sudden load changes
- Transient events
- Voltage variations

All of which will be discussed in this section.

2.3.1 Bad Data

Working with PMU measurements is synonymous to working with a vast amount of data over a longer period. When dealing with this, the occurrence of bad data faults become inevitable [4]. Bad data can be organized into two categories:

1. Missing data
2. Incorrect data

These faults occur as the result of cyber attacks or temporal equipment failure on either measurement- or communication units or both.

2.3.1.1 Missing data Due to the high volume flow of data points that the PMUs deliver, even a short disruption of the data transmission can cause a large amount of missing data. Utilizing the PMU's time synchronization feature might prove useful when a case of data transmission error occurs. Since all data points are labeled with a time stamp, data points which aren't lost can be retroactively be integrated into the data set. However, when dealing with live online state estimation or anomaly detection, this might not be possible since the live data feed might already be analyzed before the temporarily lost data is restored [23].

2.3.1.2 Incorrect data The data received from the PMU might include incorrect values. These faults appear in different configurations, depicted in Figure 6, where four different types are presented [25].

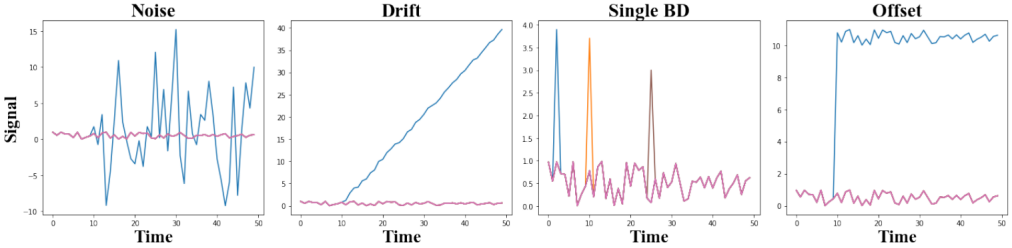


Figure 6: Different types of bad data examples [25].

The first type presented is noise. Noise can occur due to sensor or transmission malfunctions. These can cause large oscillations, or noise that can be recorded by the PMUs. Noise of low magnitude is normal for PMUs [14], and is of no real issue as built in noise reduction systems help calibrate and cancel this noise. However, the noise recorded by the PMUs at high magnitude, indicating a faulty system, can be troublesome for the components connected to the system. Power components subjected to noise and other power quality issues can lead to misoperation, damages and process disruption [26].

Synchronization errors might cause a drift over time, where the measurement errors add onto each other over time. It is of high importance that these types of anomalies and faults are detected [26]. If drift anomalies are not detected, system SE may be wrong, and as such production and loads may be utilized in a sub-optimal way. A drift may, in worst case scenario, cause production facilities to produce an abundance or a deficiency of power that the system requires. This can lead to a system collapse as the supply and demand of power is not in balance [27].

Sensor and transmission malfunctions can also cause spikes, or single bad data points. These spikes may not be of issue by themselves, but can warn TSOs of suspected damaged equipment or failures. If these are detected early on, faulty components can be taken out of service and replaced or fixed, reducing the chance for cascading faults that may be catastrophic for the system.

Sudden load changes and uncalibrated or malfunctioning measurement meter can cause an offset to appear, shifting the measurements higher or lower in magnitude. Offsets are not inherently dangerous for the power system if they are within the tolerable limits [28]. The problem occurs if the offsets breach the limits, which is why it is important to detect the offsets quickly. This allows for rapid handling like production increase or decrease and load shedding, to counteract the anomaly.

2.3.2 Topological Errors

When trying to estimate the operating state of a power system, having the correct topology of the system is vital [29]. The topology of the system means knowing the physical interconnections in all parts of the system, an example shown in Figure 7. All connected generators, transformers, busbars, loads and lines, as well as their connections, need to be known to make a correct estimate. Therefore, it is essential to know if the topology of the system is altered in any way without warning. In [29], the authors introduce a simple and effective approach to identify these topology errors. Extending a method used to identify multiple analog bad data, topology errors can be identified using tests to check the colinearity between the lagrange multiplier vector and the columns of the corresponding covariance matrix [30]. Tested on the Brazilian northern power system at the substation level, the method provides correct answers for different types of topology errors, including cases of low redundancy. The simplicity of the mentioned method makes it viable for real-time application, it also makes elaborated hypothesis testing unnecessary.

Topological errors can be split into three categories [31]:

- A line is in service but not included in the model.
- A line is not in service but included in the model.
- A line is open/not open ended in the model but not in the actual system.

As understood from these categories, topological errors are a mismatch between the model used for system analysis and SE, and has no real effect on the measurements done by PMUs. However, the events leading up to the topological errors can be both physical faults and bad data anomalies, explained further below in this thesis.

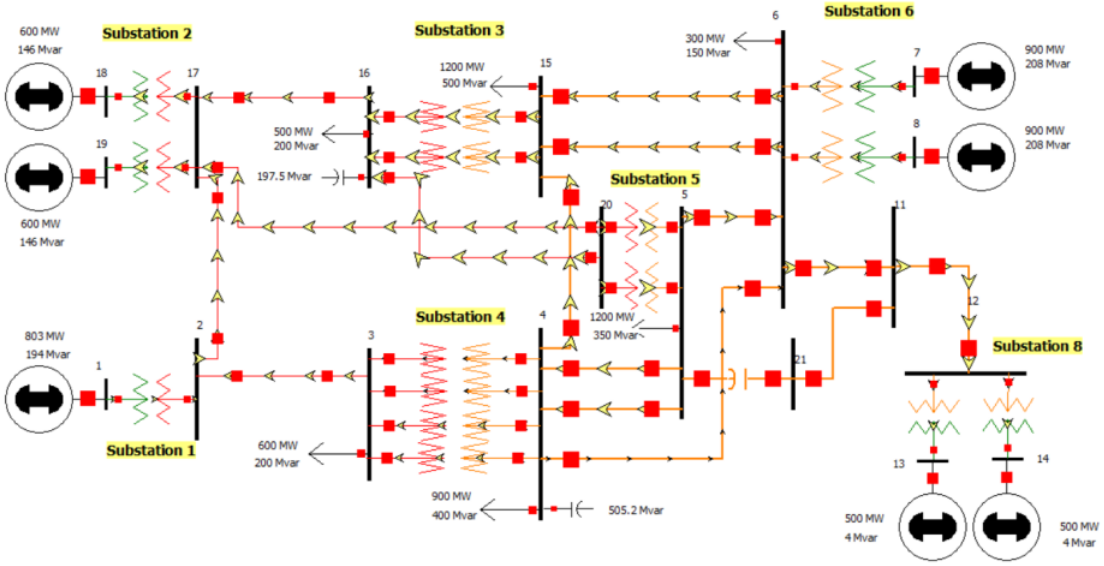


Figure 7: Single-line topology diagram of a 20-bus case study in [32].

2.3.3 Sudden Load Change

A Sudden Load Change (SLC) in the system can affect the quality of power delivered [27]. A sudden load change occurs when loads are suddenly added to or removed from the system. This is usually done in a normal and controlled way using load management techniques, like increasing generation while starting an industrial process [33]. Failing to properly handle a load change can lead to a SLC anomaly. A SLC anomaly affects the frequency and voltage levels of the system, which can lead to these levels exceeding the normal operating tolerances [27]. A sudden load change can also occur as a result of other faults. If a generator experiences a fault and needs to be taken off-line, spontaneous generation loss can occur, the system reacts to sudden generation loss in a similar fashion as sudden increase in load, shown in Figure 8. Similarly, the same can happen if a circuit breaker is used to clear a fault in the system, therefore topological errors and sudden load changes are closely linked.

Detecting sudden load changes and handling them effectively is important to preserve the Power Quality (PQ) of the system [35]. The impacts of PQ are further explained later in this thesis.

2.3.4 Transient Events

In [36], transient events are categorized and a method to process them is proposed. The paper explains transients in power systems as a description of short events on voltage and current signals. Furthermore, the paper classifies transient events into three:

- Events that adjust or change the voltage magnitude of the fundamental frequency over long

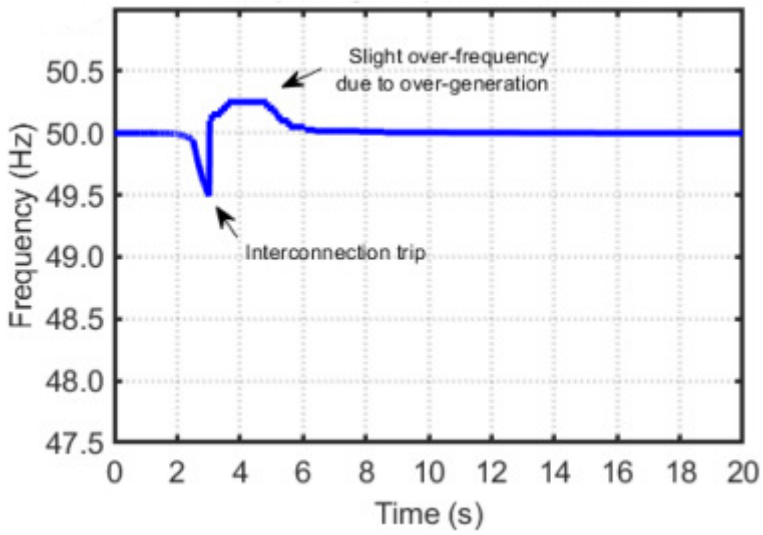


Figure 8: Frequency response after a short-term interconnection trip [34]

periods. These events can create voltage sag or swells for a period of time ranging from 50 ms to several seconds.

- Events that change the voltage magnitude for a short duration, typically fuse-cleared faults or self-extinguishing faults.
- Events where the fundamental voltage magnitude is of little importance. A lightning strike on the system goes under this category.

The IEEE has created a standard for categorizing transient phenomena which is illustrated in Table 1 [26]. This standard describes transients as events that cause undesired signals in voltages and currents. A synonym for transient in the context of the power system is *surge*. For simplicity, we will refrain from using that term in this thesis.

Table 1: Categorizing of transients [26].

Categories		Spectrum	Typical Duration	Typical Magnitude
Impulsive	Nanosecond	5 ns rise	<50 ns	
	Microsecond	1 μ s rise	50 ns - 1 ms	
	Millisecond	0.1 ms rise	>1 ms	
Oscillatory	Low frequency	<50 kHz	0.3 - 50 ms	0 - 4 pu
	Medium frequency	5-500 kHz	20 us	0 - 8 pu
	High frequency	0.5 - 50 Mhz	5 μ s	0 - 4 pu

2.3.4.1 Impulsive transient An impulsive transient is usually associated with lightning strikes since that is the most common cause of this phenomena. The impulsive transient is a sudden change in the steady state condition of voltage, current or both. Their categorization is given by their rise and decay time [26].

Impulsive transients are damped quickly due to their high frequency, and they therefore don't travel far from their source. However, they can resonance with the power system circuits in a way that creates another type of phenomena; oscillatory transients [26].

In Figure 9 an impulsive transient event is found in a 132 kV network. This type of signal is typical for many impulsive transients. One can clearly see that the networks response to the event induces low amplitude oscillations in the signal that lasts about 5 ms [36].

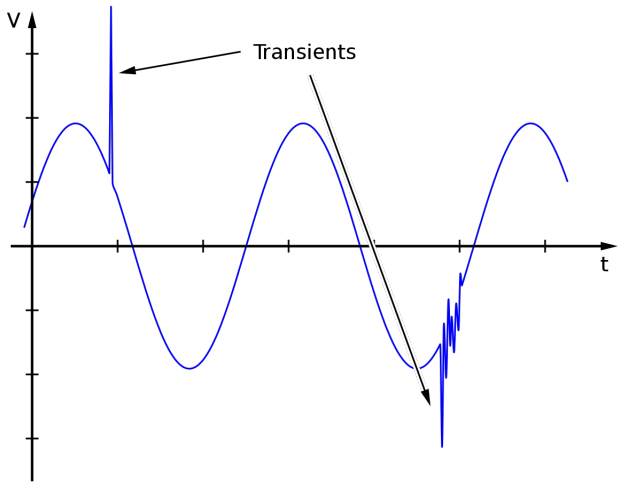


Figure 9: Impulsive transient illustration [37].

2.3.4.2 Oscillatory transient Oscillatory Transient events are characterized by the rapid changing polarity of current or voltage values. Each event is classified by their frequency rate in the categories low, medium or high as shown in Table 1 [26].

All high frequency oscillatory transients are usually caused by some sort of switching event. Usually this is caused by a local system response to an impulsive transient, that often comes from lightning strikes [26].

Medium frequency oscillatory transients currents can be caused by back-to-back capacitor energization, which happens when a capacitor bank is energized in the near presence to another capacitor bank already in use. Transient voltages in the same category occurs due to cable switching. It can also occur as a response to an impulsive transient [26].

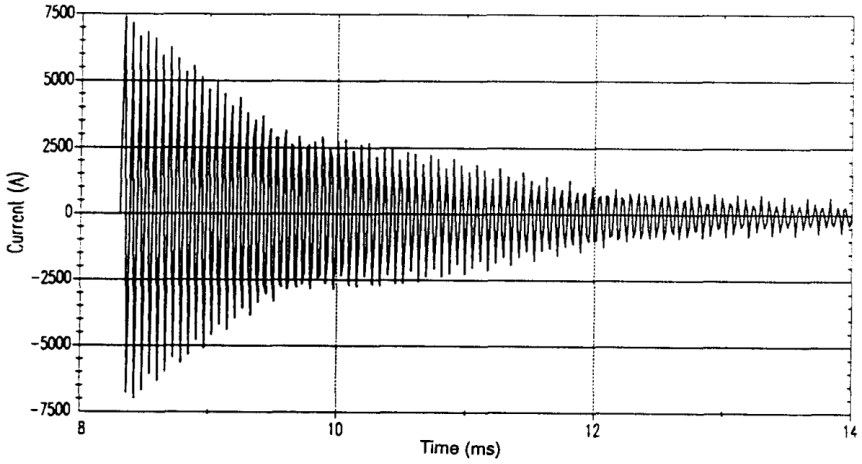


Figure 10: Medium frequency oscillatory transient current caused by back-to-back capacitor bank switching [26].

Low frequency oscillatory transients, visualized in Figure 10, has many different causes. Some examples are capacitor bank energization which induce transients between 300 and 900 HZ, shown in Figure 11, while ferroresonance and transformer energization can induce transients below 300 Hz. These events occur on subtransmission and distribution system levels.

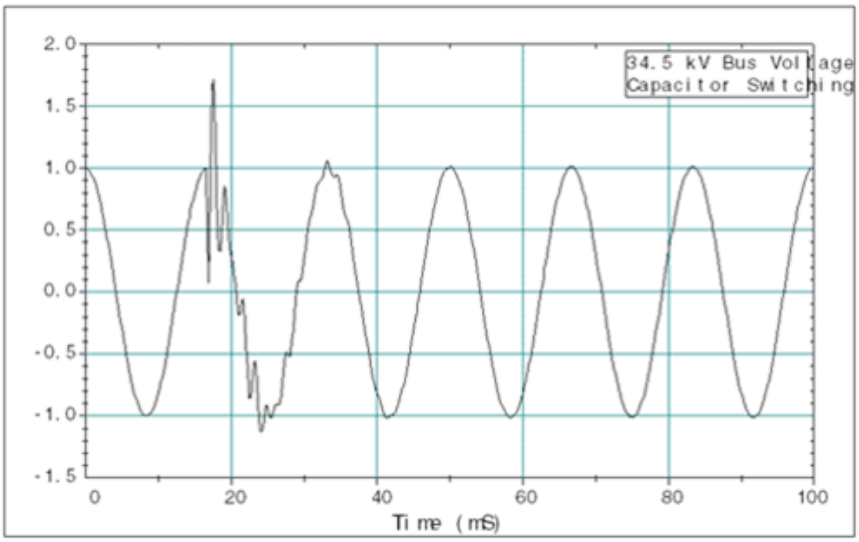


Figure 11: Capacitor-bank energization induced Low frequency oscillatory transient current [38].

2.3.5 Voltage Variations

Voltage variation anomalies are time periods where the voltage differs from the nominal level. Voltage fluctuations can cause damages to connected equipment, and is the most common type of PQ problems [39]. Voltage variations are categorized after the IEEE 1159 standard and is rendered in Table 2. Undesired change of voltage can either be voltage rises (swells/overvoltages), voltage drops (sags/undervoltages) or total loss of voltage (interruptions) [39].

Table 2: Categorizing of voltage variations [26].

Categories			Typical Duration	Typical Magnitude
Short duration variations	Insanteneous	Sag	0.5 - 30 cycles	0.1 - 0.9 pu
		Swell	0.5 - 30 cycles	1.1 - 1.8 pu
	Momentary	Interruption	0.5 - 3 s	<0.1 pu
		Sag	30 cycles - 3 s	0.1 - 0.9 pu
		Swell	30 cycles - 3 s	1.1 - 1.4 pu
	Temporary	Interruption	3 s - 1 min	<0.1 pu
		Sag	3 s - 1 min	0.1 - 0.9 pu
		Swell	3 s - 1 min	1.1 - 1.2 pu
Long duration variations	Interruption, sustained		>1 min	0.0 pu
	Undervoltages		>1 min	0.8 - 0.9 pu
	Overvoltages		>1 min	1.1 - 1.2 pu

2.3.5.1 Short duration variations Short duration variations are changes in voltage that are under a minute in length, where the shortest variations spans half a cycle of the voltages' frequency of typically 50 or 60 Hz. There are several causes of short duration voltage variations, the most prominent being equipment failures, system faults (for instance single line-to-ground), control system malfunction, damaged or loosely connected wires and SLCs [39].

In [40], several PQ surveys has been compared in relation to short interruptions and sags. The authors conclude that voltage sags are accountable for between 61% and 87% of all PQ incidents. According to [41] the same statistic for voltage sags is considered to be 70% to 90%. Furthermore the authors recognize short circuits and other faults, lightning strikes and inrush currents as the main causes of voltage sags.

2.3.5.2 Long duration variations Long duration voltage variation are defined when a breach of the ANSI limits [42] occur for more than 1 minute. One can see from Table 1 that this phenomena has three subcategories, namely sustained interruption, undervoltages and overvoltages. There are several causes for these types of faults. Some examples are load variations, erroneous tap setting on transformers, system switching operations, blown fuses and failed circuit components.

2.3.6 Drop in Frequency Events

In power systems, the frequency is a measure of the balance between load and generation. Generate too much and the frequency increases, generate too little and the frequency drops [43]. A drop in frequency impacts the power quality, as well as increasing the possibility of system failure, with eventual collapse of the grid. There are multiple reasons for a drop in frequency. All the anomalies explained above can be potential reasons, as they all have an impact on the balance between load and generation. In normal operation, the operating range for the frequency is commonly ± 0.2 Hz. It is therefore vital to detect the frequency changes rapidly so the threshold is met at all time.

2.3.7 Cyber Attacks

With the integration of Internet of Things (IoT) and digitization, the cyber-physical horizon of the power system is ever increasing [7]. With the need to control a power system that keeps increasing in complexity, the vast amounts of data collected needs to be recorded, centralized and handled quickly. This is done through the cyber-physical layer by communicating over the internet. This makes the system prone to cyber attacks, like any other internet-connected system. Cyber attacks can be categorized as follows: False Data Injection (FDI) attacks and Denial of Service (DoS) attacks.

2.3.7.1 False Data Injection Attacks The increasing number of PMUs in the power system make them more susceptible to attacks. FDI attacks revolve around getting access to the data flow and altering the measured data between the measuring point and the data collecting point in the system [44]. FDI attacks can prove effective if they are hard to detect and go unnoticed for a long period of time. Being able inject small amounts of false data while still being under normal operating conditions can have a major effect on the stability of the system. As the measured data is used for SE and to balance the production and loads in the system, the false data can destabilize the system and have catastrophic effects on the grid.

For microgrids, a common way to destabilize the system is by GPS spoofing and altering PMU data [44]. This is possible due to the time synchronization of PMU data. The spoofing is conducted by sending similar signals to the Phasor Data Concentrator (PDC), with the intention of injecting false data or altering the timestamp of the phasors, causing the magnitude and angles to be wrongly

synchronized [45].

2.3.7.2 Denial of Service attacks While injecting false data is not intended to be detected, Denial of Service (DoS) attacks will have an immediate effect on the system. The attacks are constructed to overload the network used to transmit data. This hijacking or overloading of the network has as intention to block out the normal data flow that is vital for the functioning of the system [46]. Detecting a DoS attack is not about subtle changes in measurement data, but rather understanding that an attack is ongoing instead of believing that a fault somewhere in the system is causing the disturbance of data flow.

2.4 The Impacts of Anomalies in Power Systems

As discussed, there are many types of anomalies that occur on the power system. All these anomalies have several consequences which will be explored in this section. The biggest threat when experiencing faults in the system is not being able to meet the power demand, causing severe economic consequences [47]. Electricity is an essential commodity for almost all human activity and a failure to supply also has large political and social. The severity of the supply failure ranges from load shedding of small areas or industrial plants, to large scale blackouts lasting over longer periods.

As our society and businesses continues to digitize and digitialize in an ever increasing pace, the economy is more reactive to anomalies in the power system [47]. These reactions are not only caused by the mentioned power outages, but also when there are issues with the PQ. These issues include everything that deviates from nominal power frequency and voltage levels; voltage sags, surges, transients and harmonics [48].

The economic loss from power anomalies are on an enormous scale. It is estimated, from a study in 2001, that the U.S. economy loses $104 to $164 billion each year due to power outages and major blackouts [48]. In addition another $15 to $25 billion a year is lost due to PQ-related losses.

In extreme cases consequences of power outages can be fatal. The 2021 Texas power crisis had an official death toll of 246 fatalities with direct or indirect connection to the power outages in Texas, February 2021 [49]. On top of the fatalities, it is estimated the damages due to the winter storm total to $195 billion [50]. The extent of the blackout can be seen in Figures 12 and 13.

2.4.1 Impacts of Transients

Transients, which are described in earlier subsections, can affect the PQ depending on the severity of the transient event. Both impulsive and oscillatory events effect equipment that rely on high PQ delivered by the TSO. This equipment includes motors, transformers, capacitors, cables, circuit breakers and other electronic components [39]. The high power transients, such as lightning strikes, has such a high energy concentration and voltage magnitude that it causes a degradation of insulation, dielectric breakdown (hazardous electric flow in the insulation) and thermal failure of components. The same degradation of insulation is also caused over time by repeated transients of lower voltage magnitude and power. Degraded insulation is a hazard, as it might not work as intended. Degraded insulation can cause physical failures on components which can lead to

Figure 12: Satellite imagery of Texas February 7th 2021, before extreme weather hit the State and caused a major blackout [51].

Figure 13: The same area exactly one week later. Texans experienced around 2,495,594 power outages during the extreme weather [51].

cascading faults on the power system. It is also a potential safety hazard to workers, as the risk of electrocution might be higher than advised and in unexpected locations. Heat losses can also increase which cause further degradation and reduces the efficiency of the components. Lastly, damaged insulation will experience a more rapid degradation during events which strengthens the threat of the aforementioned issues [39].

2.4.2 Impacts of Voltage Variations

All deviation from nominal voltage levels is undesired and can cause failures and damage in equipment [39].

Long duration voltage variations is a phenomena that occurs less than short duration voltage variations. However, the consequences of these sustained PQ issues is a serious economic and safety concern for utility companies and the industry. The consequence of sustained voltage interruptions is a complete shutdown of important equipment to our infrastructure. Overvoltage intensifies the aging of the power system. It can also trigger an immediate failure of electronic devices and disrupt operation. Undervoltage reduces the output of lighting appliances and causes overheating in motors due to current surges [39].

In [52], the authors has described concrete electricity failure events and their costs. One of them is a 15-minute voltage interruption in Vancouver 1995, corrupting data and back-up files at the city's stock exchange. The exchange didn't manage to fix the problem in time for opening the next day. They therefore lost revenues of around 30,000 CAD, and their members firms lost commissions such that the total losses were extended in to the millions. This is just one of many examples that the paper illustrates of how big of an economic impact a single event can have.

2.4.3 Social and Political Impacts

In [47], the effects of electricity outfalls in the UK is studied. From the survey that the study conducted, the authors report that people could, in general, cope with short disruptions. As long as the outage lasts less than 24 hours the social impact is minimal. However, our society is

becoming increasingly reliant on electricity due to electrification and digitalization. Hence, in the future the impacts of blackouts will only increase.

A modern example of electrical outage comes from present day Lebanon, where an economic crisis has left the country with a shortage of diesel which is used to produce electricity. The paper [53], describes how the country is embossed with social unrest due to this crisis. Although power shortage is the result of an economic crisis rather than a network anomaly, it still shows how vigorously the blackouts impact communities, and especially poor communities. This is due to these communities relying on power for basic health and income purposes. At the same time the more wealthy people of Lebanon continue living comfortably as they have enough money to buy diesel privately and run privately owned generators.

Electricity demand is extremely inelastic [47]. Hence, electricity shortfall will lead to rising power prices, which in turn increases fuel poverty. Wide-scale outages propels energy supply up in the political agenda, and can put serious pressure on a sitting government, having long lasting impacts. The three-day week policy in the UK in 1974 [54] is an example of this. The policy contributed to a change of government and to changes in the UK's energy policy that has lasted to this day [47].

2.4.4 Impacts of Cyber Attacks

In a system that is largely digital with a lot of data transmission, as a PMU integrated system is, the possibilities for cyber attacks are very real. Their impacts can have vast consequences, where having systems and personnel to deal with the attacks quickly is crucial for the systems reliability [7].

The impacts of cyber attacks are similar to other anomalies, as in they can cause blackouts and disruptions, however they have the potential to be much more complicated to fix. In 2015, Ukraine was a target for a cyber attack on its power system. The attack is believed to be successful through the use of phishing emails to inject malware into crucial systems to gain access and information, before unleashing a DoS attack. The attack resulted in 230,000 households losing power for up to 6 hours. This blackout could have lasted much longer, had the SOs not had some cyber security measures established. However, the attack could have been avoided altogether had better cyber security processes been in place [55].

The authors in [56] dig into the impact and control of load altering attacks on a 9-bus power grid. Distributed energy resources (DERs) are more and more common with the introduction of numerous wind parks and solar farms. The DERs need to communicate with the rest of the grid in order to provide information of how much they generate. The authors look into how cyber-physical attacks, namely dynamic-load altering attacks (DLAA), on this communication line can impact the system. By introducing a load variation in the system greater than 25% the system was destabilized with the collapse of frequency.

2.5 Methods for Anomaly Detection in Power Systems

Using PMUs in the data acquisition process provides a clearer and more precise picture of the state of the power system for the system operator [57]. There are however some challenges, as the vast amounts of data needs to be processed and analyzed, preferably in real time.

This section provides information of different types of anomaly detection methods have been used in the past. These methods focus on time-series data, as this resembles the data stream from PMUs. Here, ML algorithms of both supervised and unsupervised types are presented, as well as statistical methods.

For ML, we will focus on two different categories:

1. Supervised
2. Unsupervised

The difference correlates to if the outputs are known and labeled or not [7].

2.5.1 Supervised Machine Learning

In supervised ML, each set of input data contains a desired output value. In time series data this corresponds to the next real measurement. Knowing the next value allows for further corrections to tune the algorithms to better suit the desired output. The goal of the algorithm is learning to recognize labeled data, such that it can correctly analyze unlabeled data for identification and classification purposes.

An issue that can occur when training the algorithm is overfitting. This is when the algorithm is tuned or fit too carefully to the training data. This problem is most common when there is a lacking amount of training data [58]. This will give great performance factors for the data its training on, however, it will prove itself lacking or even useless on a similar, but different data set.

2.5.1.1 Artificial Neural Network Artificial neural networks (ANNs) are commonly used for analyzing and predicting time-series data [59]. ANNs are constructed by one or more hidden layers of artificial neurons as illustrated in Figure 14. These neurons are non-linear functions that transform the inputs. The connections between the different layers and neurons are called edges. Weights are added to the neurons and edges, with corresponding threshold values which decides the maximum accepted error between the forecast and real values. These start out as random values, which are then adjusted throughout training of the algorithm. The threshold values decide if the input will be transmitted through to the next neuron [7].

ANNs are under the category supervised learning, as the weights and parameters are changed to best predict a known target.

In [60], the authors propose a method to distinguish between topological and analog errors, as well as identifying where the error has occurred. This is done using an Artificial Neural Network (ANN) based on the Group Method of Data Handling (GMDH) [61]. The method is tested on a Brazilian 118-bus system. The tests show promising results in distinguishing topological and analog errors efficiently with real-time monitoring requirements. This is done while identifying branch or bus misconfigurations and inaccurate measurements. The computational time required to perform this analysis is also minimal, making it suitable for real-time applications for many different operating conditions.

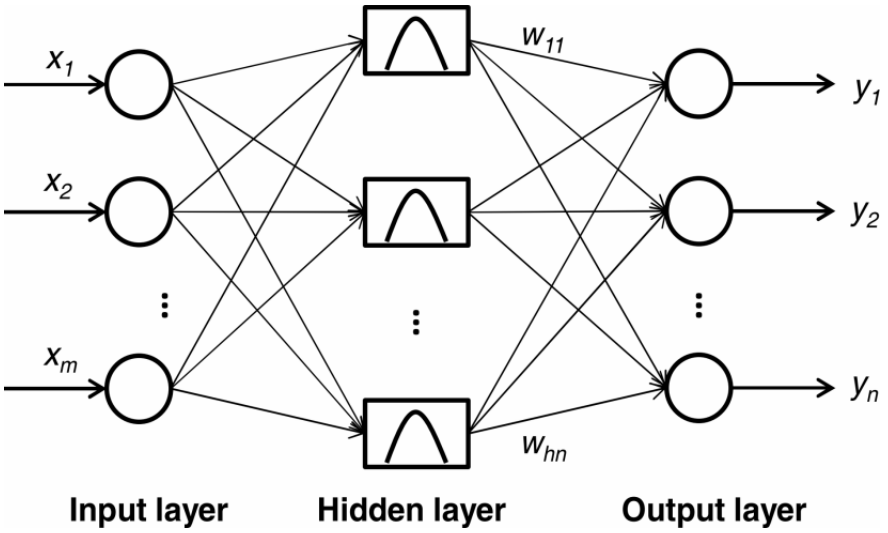


Figure 14: Artificial neural network structure [59].

2.5.1.2 Long Short-Term Memory Long Short-Term Memory (LSTM) is one type of deep neural network algorithms that is popular and well-suited for time series data processing [62]. The algorithm uses a feedback loop to reuse the outputs as inputs. These types of algorithms are called Feedback Neural Networks (FNN) which is a subcategory of Recurrent Neural Networks (RNN).

A well known issue for RNNs is their tendency to encorporate vanishing and exploding gradients [63]. This is one of the issues LSTM algorithms overcome [64]. The difference lies in the ability to store values in "memory cells". This allows for long term memory storage and therefore the ability to learn long term correlations.

In [6], the author uses two LSTM layers, shown in Figure 15, with a dropout layer after each one to regularize the data. The purpose of the dropout layer is to reduce overfitting issues. The method allows for tweaking of different parameters, such as the amount of hidden units in each LSTM layer and how many LSTM layers exist.

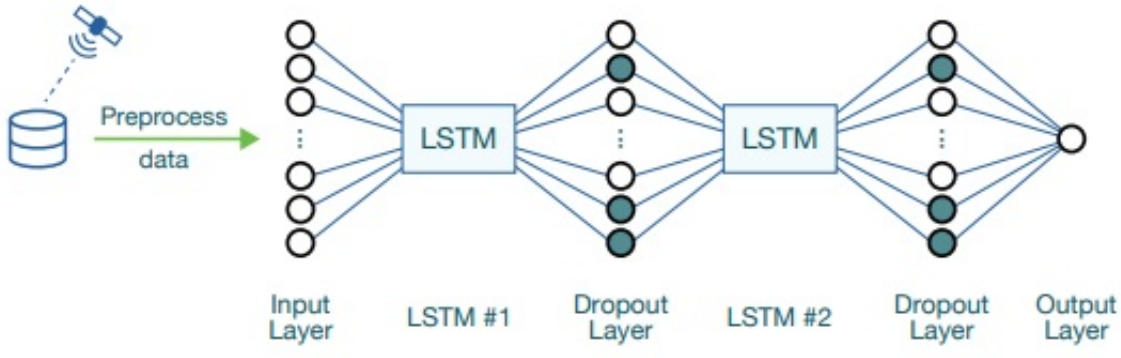


Figure 15: LSTM architecture [6].

In [6] the LSTM provided the best results, compared to other methods, for the purpose of detecting anomalies in time-series signals from a NASA spacecraft. Giving a $F_{0.5}$ score of 0.76 with 40 hidden units in the LSTM layer.

In [64], the authors propose to use stacked LSTM networks to learn higher level temporal patterns

using normal time series behaviour. These patterns are then used to predict the next time steps to compare the real values for anomaly detection. The method foregoes the need for a pre-specified time window, this allows for a flexible algorithm for anomaly detection. The authors focus on modelling and predicting normal behaviour of time series data, which leads to accurate detection of deviations or anomalies in the data.

The popularity of the LSTM neural network algorithms has produced a lot of various archetypes. Various approaches based on LSTM networks for anomaly detection in different technical systems using time series data are compared in [62]. These LSTM networks include regular LSTM, encoder-decoder based methods and hybrid approaches. The paper also presents recent trends in learning-based anomaly detection, where graph-based approaches and transfer learning approaches are presented.

2.5.1.3 Multilayer Perceptrons In [65], Multilayer Perceptrons (MLP) and cumulative sum technique are introduced to detect stealthy attacks on cyber-physical systems. Testing on data collected from a scaled down water treatment plant, the machine learning algorithm MLP and the statistical cumulative sum technique detected false data injection even when they were continuous and discrete. The authors mention that the model was tested on individual attacks, ie. individual components were injected with bad data one at a time. However, the model proved successful for its purpose in the paper.

2.5.1.4 Multi Class Support Vector Machines Multi Class Support Vector Machines (MSVM) is a method used with Support Vector Machines (SVM) for multi class classification. The method constructs decision hyperplanes with decision boundries in the high dimensional feature space. The mapping into high dimensional feature space allows for pattern recognition by linear classifiers, separating input data into different classes [66], shown in Figure 16.

As is common with other supervised ML methods, the data set should be split up into a training set and a validation set. SVMs use a different training approach which makes the data split obsolete [7].

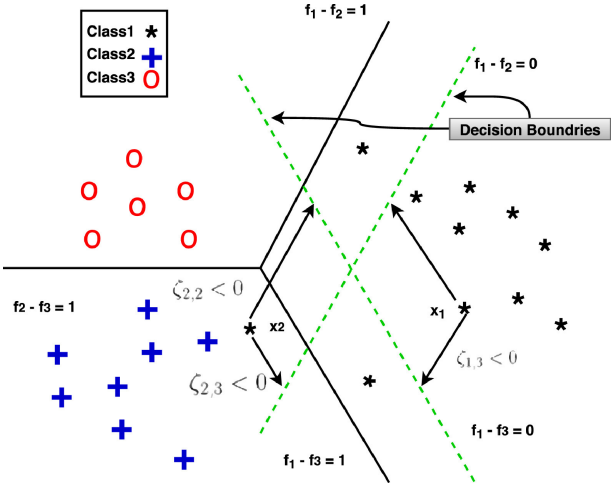


Figure 16: Three-class labels MSVM [66].

2.5.1.5 K-Nearest Neighbors For classification and regression, the k-Nearest Neighbors (kNN) technique is widely used [7]. The parameter, k, indicates how many of a given data points nearest neighbors are to be used in the process. Figure 17 shows an example of how clustering a given data set with k=3 looks like. The technique usually uses euclidean distance for classification or regression of data values. Where the euclidean distance between the points P_i and P_j in the xy-plane is given by [67]:

$$distance(P_i, P_j) = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2} \tag{4}$$

The method is non-parametric, meaning that the deciding factor depends on the k nearest neighbors. The technique can be used for classification or regression:

- For classification purposes, the data point is assigned to the class most common among the k nearest neighbors.
- For regression purposes, the value is determined by the average of the values of its k nearest neighbors.

For time series data application, [68] propose two novel approaches. These approaches looks into different ways of dividing the past data set for neighbor comparisons. The first approach compares the subsequence in question, with every possible subsequence of the same length from the past. The second approach reduced the amount of subsequence sets, therefore reducing the computational load. The results show that the first approach outperformed the second approach, however the computational load needs to be considered for the intended purpose and size of data set.

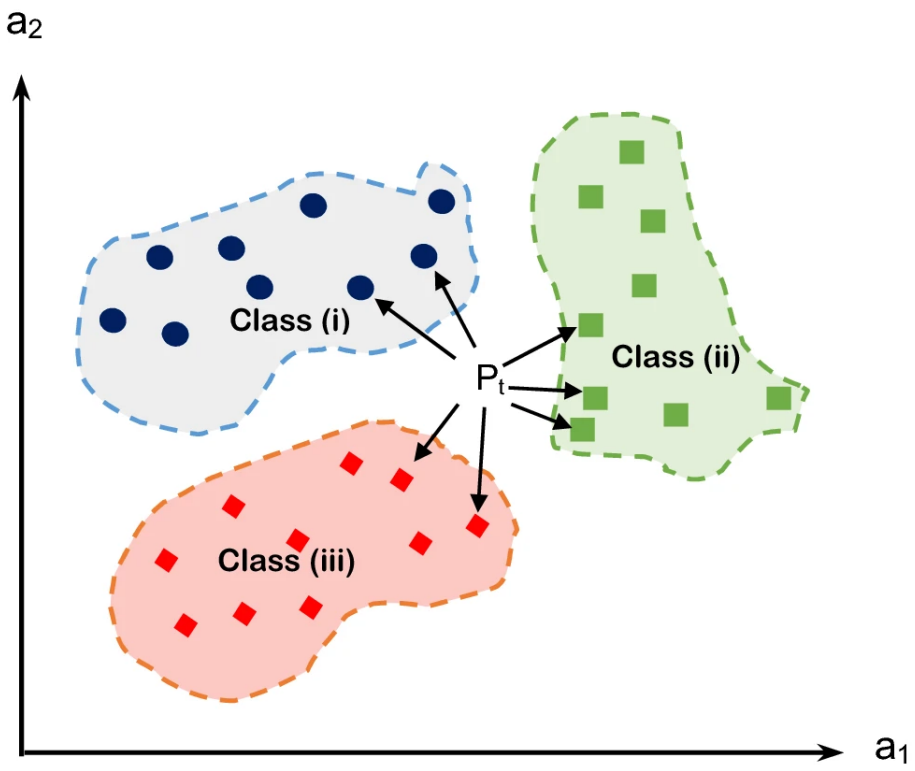


Figure 17: Example of k-nearest neighbors clustering with three clusters [69].

2.5.1.6 Correlation Based Neural Networks

For time series based data and systems, the time correlation between data points can be a good indicator for faults, bad data and cyber attacks. In [4], correlation analysis is combined with a neural network classifier. The correlation analysis is based on Pearson's correlation coefficient, shown in Figure 18. This suits the type of data that the PMUs generate, as the coefficient conveys how strong the linear relationship is between two data sets is.

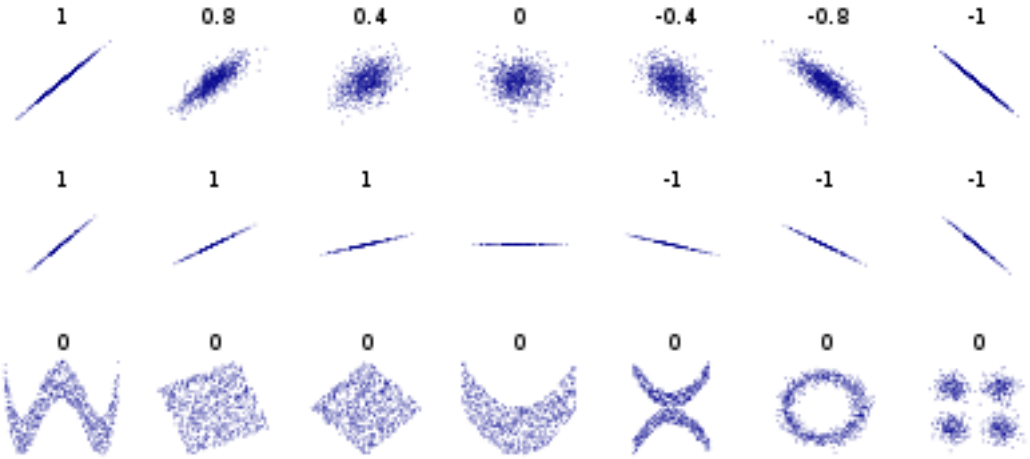


Figure 18: Examples of Pearsons correlation coefficients [70].

The authors propose an algorithm that detects and identifies bad data before SE is computed. This results in only one computation needed for SE, leading to quicker estimation as the bad data analysis is done pre-SE. The computational load for this method is lower than other typical methods used prior. The algorithm is also heavily data based, so network topology knowledge is not needed. Noise in the measurement data was one of the main performance factors that was tested. The results show that the amount of noise heavily affects the performance, due to higher noise being labeled as bad data. The authors conclude with the algorithm having significant potential in bad data detection. However, further work within real-time system analysis of real data needs to be conducted to assess how real system noise affects the algorithm.

2.5.2 Unsupervised Machine Learning

Unsupervised machine learning, as stated previously, does not have known outputs that correspond with the inputs. Instead the algorithms must discover patterns in the training data on its own. These types of algorithms require less work compared to supervised methods which need labeling and comparisons of the output. The unsupervised methods can also detect patterns that have not been known or visible to the engineers beforehand. This type of pattern recognition and categorization is called clustering, and is the primary use of unsupervised machine learning [7].

2.5.2.1 K-Means

Visualizing clustering in a 2-D plane can easily be shown with the k-means algorithm. Being the simplest unsupervised learning algorithm, the goal is to find the minimum sum of euclidean distances (Equation 4) between data points [7]. A 2-D visualization with 5 clusters is shown in Figure 19.

The algorithm is structured to cluster the data points to a predetermined, k, number of clusters.

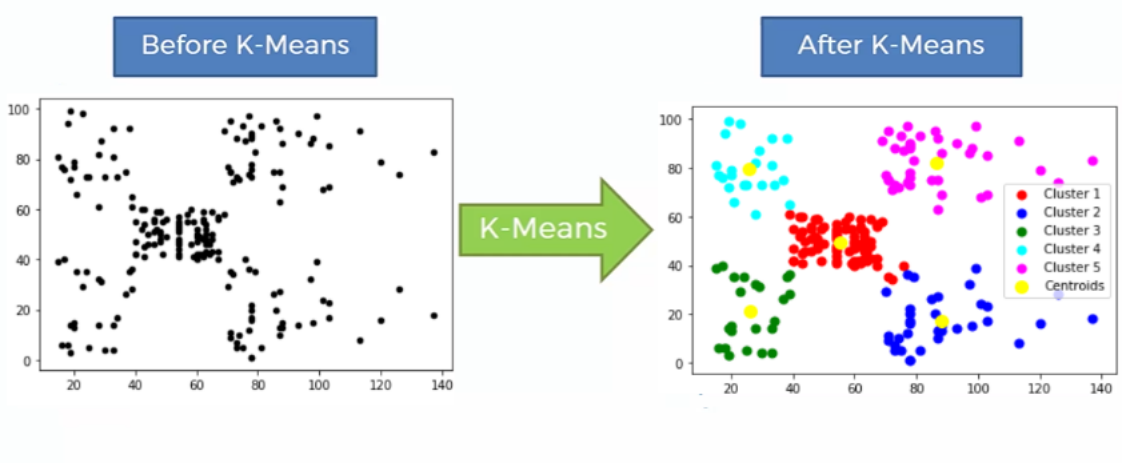


Figure 19: k-means clustering [71].

These clusters are centered around *centroids*. The centroids start of randomly placed in the data set, where each data point is assigned to its nearest centroid. Iteratively the centroids are moved to the mass center of its assigned data points, or where the sum of euclidean distances are the least. After which the data points are reassigned its closest centroid based on the squared euclidean distance [7]. This continues until no data points are assigned a different centroid after each successive iteration.

The optimal number of centroids will vary depending on the type of data and its intention. In [72], the optimal number of clusters are determined by the elbow method. This method plots the number of clusters, here from 1 to 10, and a cost function is computed for each number of clusters. The number of clusters that optimize the cost function is the optimal number of clusters to be used for its purpose.

2.5.2.2 Gaussian Mixture Model Similar to the k-means algorithm, the Gaussian Mixture Model (GMM) is used for clustering of unsupervised data [7]. This method, however, uses multiple Gaussian functions to cluster the data points. Each data point is assumed to be generated by different Gaussian distributions with a corresponding probability. The authors in [72], proposes that the maximum likelihood estimator of the GMM is found using two steps. The first step is the expectation step, the posterior probability is found using an expectation maximization algorithm and conditioning the weights, means and covariance using the k-means algorithm. These are then used in step two to optimize the probability of the parameters. This process is repeated until convergence.

In the case of anomaly detection, central in [72], the method proved in its ability to identify anomalies in streaming PMU data. The method was tested on different number of features in the data, spanning from 2 to 16. The F_1 score for the method in all cases was above 0.91, and 0.99 and above for 8-16 features when using historical data. The method was also tested on real time PMU streamed dataset, in which the F_1 score was even higher. The authors concluded with the method being able to adapt to the streaming PMU data, reducing the false positive rate.

3 Methodology

This section is mainly based on previous work done by the same authors [5]

High quality data establishes the foundation of anomaly detection models. This thesis will look at PMU data received from the Norwegian TSO, Statnett; as well as data obtained by the Texas Synchrophasor Network (TSN). The data sets consists of a total of over 50 million data points. The PMUs measure voltage magnitude, voltage angle, real power flow and frequency. This data needs to be processed so it is possible to use in further model training, prediction and anomaly detection. The data preprocessing consists of removing missing values, applying noise filtration and data normalization. The preprocessing enables the data to be used for a number of different ML models, which makes it practical to swap out and test different algorithms.

In this thesis different ML models for anomaly detection is trained, tested and compared on the given data. Different hybrids of CNN and LSTM algorithms are selected for this purpose, as it has in previous research shown promising results using time series data. Different ML models require different training and evaluation methods and therefore different approaches in splitting the available data. For different LSTM variants for instance, a train/test split of 80%/20% is preferred. Using the data, training and testing of the models is executed. When training the model the objective is to predict each next measurement with high precision. Therefore the model is tuned and altered so the predictions and measurements line up in the best possible fashion. After training the model, previously unseen test data is used to label any abnormalities as anomalies. Evaluating the algorithms performance is important to identify overfitting issues, and to analyze its potential. Comparing the models with an established third party model is part of the model analysis step, where the models are validated to indicate their efficacy. The actual anomaly detection tests is performed with false data injection before the models are deployed to detect anomalies on real anomalies. The flowchart of this process is presented in Figure 20.

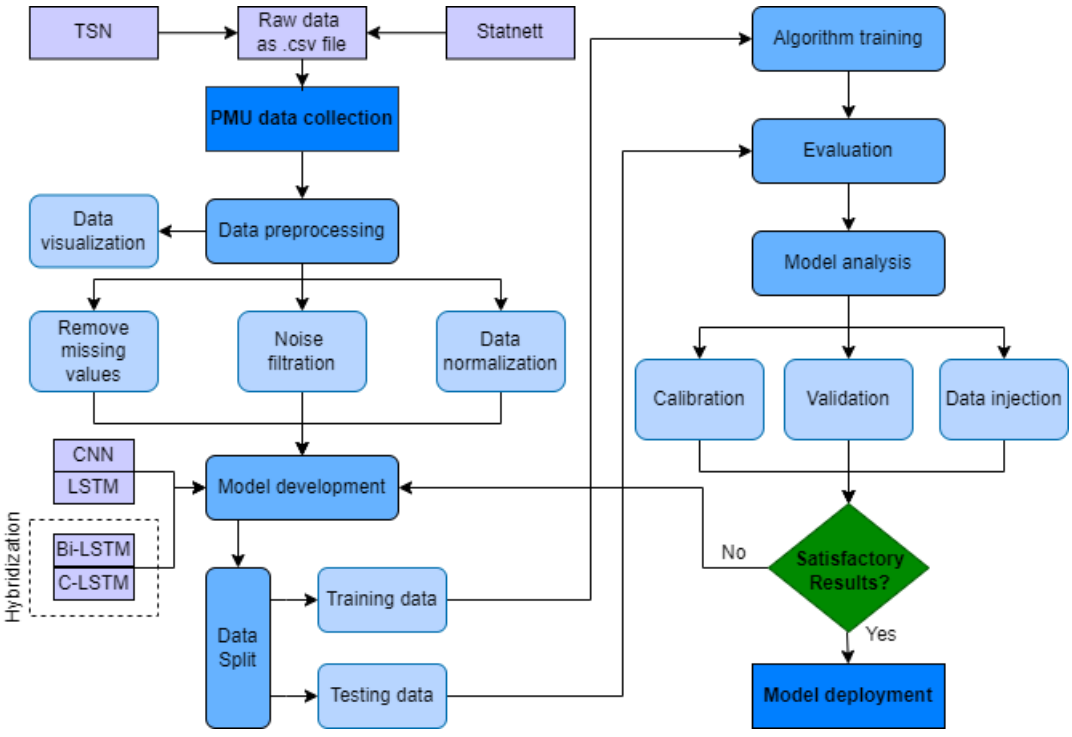


Figure 20: Methodology flowchart.

3.1 Data Gathering

A common issue when experimenting with data from power systems is that the data collected and used by the TSOs and power producers is considered critical information and is therefore confidential for the general public and hard to come by. This also applies for the PMU data that is used for this thesis.

At the same time, to work with ML and AI, having enough data is crucial for efficient and effective training and testing of the models. This thesis uses two different data sets. One of the data sets is obtained through the Norwegian TSO, Statnett. Obtaining this data was possible due to a close research collaboration between NTNU and Statnett, and important for this thesis' research. Statnett's PMUs measure data from power lines on the Norwegian transmission system, however, their exact location and the voltage levels are classified.

The second data set is obtained through the Texas Synchrophasor Network (TSN) by the University of Texas, using an independent system of six PMUs across Texas, USA. The data is described and analyzed briefly in [73]. This thesis will use five of the six PMUs in the system. The reason being one of the PMUs being on the costumer, 120 volt, side of the power system. The five stations are located within different zones of the Electric Reliability Council of Texas (ERCOT) and shown in Figure 21. The locations are:

- University of Texas - Austin
- University of Texas - McDonald Observatory in Fort Davis
- University of Texas - Pan American in Edinburg
- Houston in Harris County
- Waco in McLennan County

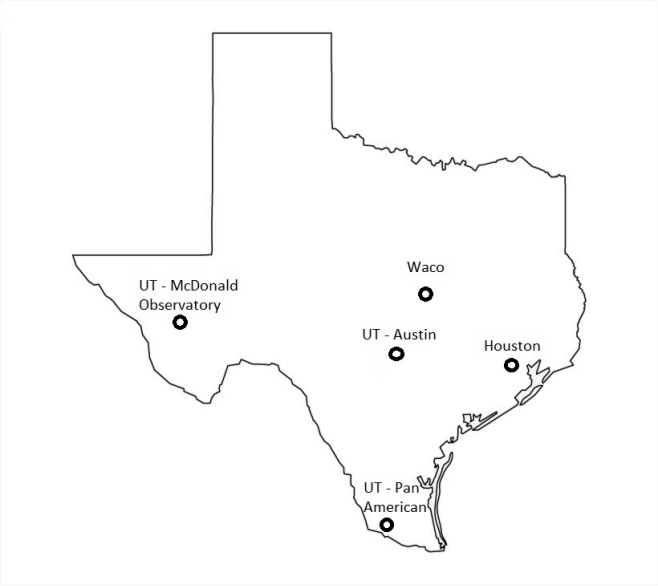


Figure 21: Positions of the PMU's considered in the study.

3.1.1 Data Format

The PMUs in the two different networks measure voltage magnitude, voltage phase angle and frequency, for each given timestamp. Statnett measures real power flow in addition to the features mentioned. Due to the nominal frequency difference of Norway and the U.S, the PMUs measure at different frequencies as well, corresponding to half the grid frequency. Statnett therefore measures at a rate of 25 Hz, while the TSN measures at 30 Hz. The units of the values given by the PMUs are also different, due to the data from Statnett is critical information, both the voltage magnitude and real power flow are given in per unit (pu) values. The TSN data does not have this constraint and the voltage magnitude is given in volts. The units of the values are therefore given as shown in Table 3:

Table 3: Units for measured values.

	Statnett	**TSN**
Voltage Magnitude:	[pu]	[V]
Voltage angle:	[°]	[°]
Frequency:	[Hz]	[Hz]
Real power flow:	[pu]	-

The Statnett data set contains measurements from three different PMUs. Each of the three PMUs covers coherent measurements spanning 48 hours at unknown locations at an unknown time in 2022. With a measuring frequency of 25 Hz, this totals to 4,320,000 different timestamps, each with the four measurements shown in Table 3. This corresponds to over 16 million data points for each PMU that can be analyzed and used to train and test the models.

The TSN data is more specific and covers measurements from one hour on January 3rd 2012 between 1 AM and 2 AM. With a measuring frequency of 30 Hz, this totals to 108,000 different samples for each station, each with the three measurements mentioned above. This corresponds to 1,620,000 data points for the given hour.

The data is in a time series format delivered by the PMU devices. This means that each measurement is synchronized to the exact same time by the GPS clock within the PMU device.

3.1.2 Computer Spesifications

All tests are performed on a Lambda Quad RTX 2080 Ti, which is a GPU computer that belongs to NTNU. The computer runs on an Intel Core i9-9920X @ 3.5 GHz, with four NVIDIA GV102 graphics cards. The tests are run from code written on personal computers through PuTTY which is an SSH and telnet client. Running tests on the Lambda computer greatly improves runtime compared to other alternatives available.

3.2 Preprocessing

The measurements done by the PMUs are not perfectly suited for anomaly detection right out the gate. The data needs to be processed before the models can be efficient and effective, this is called preprocessing. The preprocessing necessary for this thesis ensures that there are no missing values, data is normalized, noise filtered and voltage angles are unwrapped

3.2.1 Remove Missing Values

When a high measuring rate is used, any error or delay in the system, like faulty measurement devices, packet loss during transmission or software errors, can lead to data being lost or unavailable. These missing values are referred to as Not a Number (NaN) in the data sets. The action of removing missing data points is important for the function of the models used. Any missing value can disrupt the algorithms and generate a defective model.

For online applications, more sophisticated methods needs to be used to recover the missing values. In [74], four methods for missing value recovery is proposed. These are suitable for online applications, as the computational speed is sufficient to recover the missing data before the next measurement set is received. The authors have, however, set the time window to 20 seconds for each measurement set. This might be too slow for online anomaly detection purposes.

For this thesis, a simple linear regression technique is used. This method fills in the missing data points with a straight line between the last known data point in the series and the next known data point after the missing values. A pseudo code which explains the algorithm is given in Table 4.

Table 4: Remove missing values algorithm.

Algorithm for removing missing values (pseudocode)
Input: Dataset - Phase voltage, angle and frequency from a PMU device **Output:** Complete data set without missing values **Steps:** 1. **for** Feature in dataset: 2. **for** Value in range(length(Feature)): 3. **if** Value == NaN: 4. **while** next value == NaN 5. NumberOfNaN = NumberOfNaN + 1 6. Value = previous real value + (next real value - previous real value)/NumberOfNaN

The three PMU data sets from Statnett contains a different number of NaNs each. Data set from PMU 1, PMU 2 and PMU 3 contains 435, 1029 and 102787 missing values respectively. The data set from TSN does not contain any missing values.

3.2.2 Noise Filtration

The PMU measurements will contain a certain amount of noise. This noise is most prominent in the voltage magnitude and frequency measurements, where the noise is introduced by errors in transducers, quantization and signal processing [75]. Hence, the noise seen in the PMU data is introduced in the measurement or post-measurement stage. This means that the signal to noise ratios (SNR) are similar at different voltage levels.

Having the same SNR regardless of parameter or voltage level means that the same noise filtering techniques can be used for all measurements. Following the noise filtering technique presented in [75], a median filtering method is used.

The median filtering method is designed with a sliding window to calculate the median of the M closest values. Where M is the order of the median filter. Let $X(n)$ denote the parameter value at

time step n, where the parameter in this case can be voltage magnitude, voltage angle or frequency.

$$X_f(n) = median\{X(n - M/2), X(n - M/2 + 1), ..., X(n), ..., X(n + M/2)\} \tag{5}$$

Where $X_f(n)$ is the filtered parameter value.

The following algorithm explains the filtering process. Due to the high frequency of measurements, a simple filtering method, like the median filter, is preferred. With a time complexity of $\mathcal{O}(n)$, the median filtering method is quick and efficient for its intended purpose.

Table 5: Median filter algorithm.

Algorithm for noise filtering using median filter (pseudocode)
Input: Dataset - Phase voltage, angle and frequency from a PMU device **Output:** Filtered data **Steps:** 1. **for** X in dataset: 2. **for** n in range(length(X)): 3. **if** n < M/2: 4. X_f(n) = median(X[:n]) 5. **else**: 6. X_f(n) = median(X[(n-M/2):(n+M/2)])

3.2.3 Angle Unwrapping

When measuring the voltage phase angle, the reference point of the PMUs remain constant. For the phases, however, the angle shifts in tune with the speed of rotation compared to the nominal frequency of the grid. Due to the frequencies not being constant at nominal frequency, the difference will cause the voltage phase angle to shift over time compared to the constant reference of the PMU. As the phase angle does not increase or decrease to infinity, the angle jumps from +180 to -180 if it is steadily increasing, or from -180 to +180 if it is steadily decreasing. This jump also makes the visualization of the angles confusing and difficult to read. This can be seen in Figure 22, where unprocessed wrapped voltage angle is shown.

In anomaly detection each angle jump is an issue. Each jump of ±360 degrees does not correctly represent the actual change of angle between two measurements, and will give misleading information. To illustrate the problem, the angle data from Statnett PMU 1 is represented in Figure 22.

To solve this issue, an angle unwrapping technique is used. This technique will more accurately show the angle difference between two timesteps. However, this data is only useful for data processing, like anomaly detection, as the angles have no inherent real world value or meaning. To understand the technique, it is important to note that the change of angle between two measurements is:

$$\Delta\delta = \delta_2 - \delta_1 \tag{6}$$

Where δ_2 is the measurement which comes after δ_1. However, when the angle exceeds 180 degrees

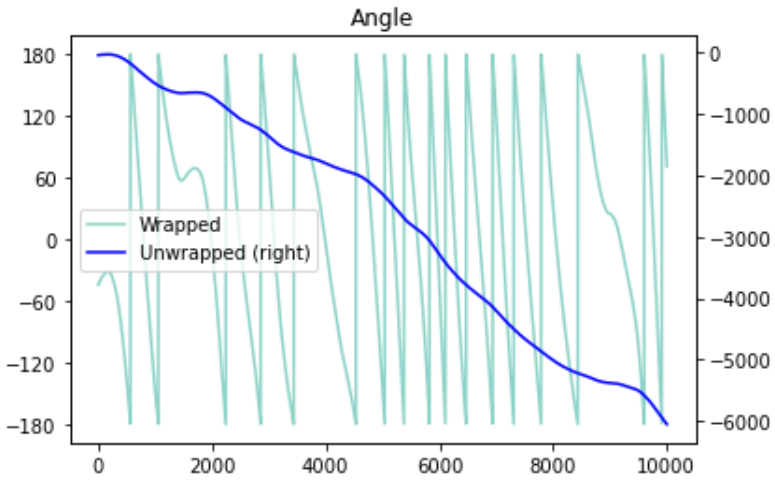


Figure 22: Wrapped and Unwrapped voltage angle measurements of the first 10000 values in the data set from Statnetts PMU 1.

the PMU device returns the new angle subtracted by 360 degrees. When the opposite happens the PMU returns the new angle with 360 degrees added to it. To get the actual change in these two special cases different equations are applied:

$$\Delta\delta_{angle>180} = \delta_2 - \delta_1 + 360 \tag{7}$$

$$\Delta\delta_{angle<-180} = \delta_2 - \delta_1 - 360 \tag{8}$$

The resulting algorithm for angle unwrapping is shown in Table 6.

Table 6: Unwrapping algorithm.

Algorithm for unwrapping phase angle data (pseudocode).
Input: angles - Phase angle data from a PMU device **Output:** Unwrapped angles **Steps:** 1. **for** i in range(length(angles)-1): 2. delta = angles[i+1] - angles[i] 3. **if** delta <-340: 4. angles[i+1:length(angles)] += 360 5. **if** delta >340: 6. angles[i+1:length(angles)] -= 360

Unwrapping the angles in such a way gives a more accurate representation of how the phase angle has shifted during the period of measurement.

3.2.4 Data Normalization

The data that is acquired from the PMUs are values that depict the actual voltage and power levels, frequencies and angles. These vary greatly in order of size, as the voltage level can both depict the high voltage level actually being measured, or their per unit values [76], together with

the frequency being around 50 og 60 Hz and voltage angles being between 0 and ± 180 degrees. Normalizing the data presents the models similar input values, meaning the features can be used interchangeably between models and analysis'. This is useful for ML algorithms, as the weights can be heavily influenced by the size of the input values. Normalizing the values removes this factor, and reduces the chance of avoidable errors.

The effectiveness of the algorithms are also dependent on which normalization method is used [77]. By using the correct normalization technique, it should be possible to implement and use the data in any ML algorithm without further processing. This makes it quick and easy to compare different methods.

In [78], a powerful and efficient normalization technique is proposed. The Tanh estimator, first introduced by Hample, is a technique considered to be robust against noise. This makes it a relevant candidate, considering the noise prone PMU data being used.

$$X_{norm}(n) = 0.5 \left[\tanh \left(\frac{0.01(x(n) - \mu)}{\sigma} \right) + 1 \right] \tag{9}$$

Equation 9 shows the fixed score normalization equation of the Tanh estimator. The technique takes the hyperbolic tangent of the given parameter value subtracted by the mean, μ, divided by the standard deviation, σ, of the respective parameter. The technique normalizes the data in the range $[0, 1]$ with a mean of 0.5 and standard deviation of 0.005.

3.2.5 Data Splitting

To train the algorithm, the data needs to be split into training and testing data. The different input types are as mentioned measurements of frequency, voltage magnitude, voltage angle and real power. The purpose of data splitting is so the algorithm can learn and train to predict based on the training data, while the testing data is used to see how well the model is able to do this. Another use for the testing data is to check for overfitting issues. If the model fares well on the training data, but not on the testing data, the model is overfit for the training data, and will not work well when implemented on new and unknown data inputs. There are several ways of performing the data split. The most natural way to do this with the time series data for anomaly detection is to train the model on the first 80% of all the measurements, and test it on the rest; corresponding to an 80%/20% split.

Furthermore, after a model is trained for a given set of parameters, it can be tested on other data. Since the model is already trained a different split is used, where the test data is the dominant part. This is because the training data only will be used to create an error threshold, which require much less data when training a model.

3.3 Model Development

Using different models in the analysis for anomaly detection is beneficial to find what works best for the given data, its structure and for the different anomalies present. The four models developed and analysed in this thesis are CNN, LSTM, C-LSTM and Bi-LSTM. These four models have different capabilities, with strengths and weaknesses. The following subsections will explain the models in depth.

3.3.1 LSTM

The use of LSTM-based neural network architecture as an anomaly detection algorithm is, as discussed in earlier sections, popular and well-suited for time series data processing. This type of architecture allows for dynamic and time-variant anomaly detection [62]. A common use for LSTM-based anomaly detection algorithms is to make a prediction model. This prediction model is used to predict future values. These values can then be compared with the measured values, where a difference error between the predicted and real value over a certain threshold is labeled as an anomaly.

The architecture consists of an input layer, one or more hidden layers, memory cells and an output layer. Using memory and forget cells, the algorithm is capable of preserving long term temporal relationships and patterns in the data.

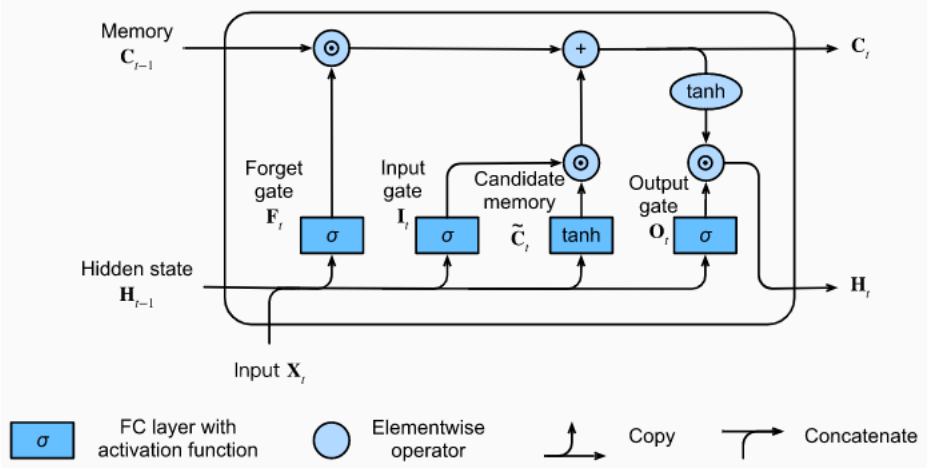


Figure 23: LSTM data flow [79].

Figure 23 shows the data flow in the LSTM model. Here we see the three different gates, and their outputs at step t, that control the memory cell; The forget gate, $\mathbf{F_t}$, input gate, $\mathbf{I_t}$ and output gate, $\mathbf{O_t}$. These gates are designed to better capture long range dependencies in the sequences.

$$\mathbf{I_t} = \sigma\left(\mathbf{X}_t\mathbf{W}_{xi} + \mathbf{H}_{t-1}\mathbf{W}_{hi} + \mathbf{b}_i\right) \tag{10}$$

$$\mathbf{F_t} = \sigma\left(\mathbf{X}_t\mathbf{W}_{xf} + \mathbf{H}_{t-1}\mathbf{W}_{hf} + \mathbf{b}_f\right) \tag{11}$$

$$\mathbf{O_t} = \sigma\left(\mathbf{X}_t\mathbf{W}_{xo} + \mathbf{H}_{t-1}\mathbf{W}_{ho} + \mathbf{b}_o\right) \tag{12}$$

Equations 10, 11 and 12 show how the output value of each gate is computed. Here σ represents the activation function, $\mathbf{W}$, the weight parameters and, $\mathbf{b}$, the bias parameters. $\mathbf{X_t}$ and $\mathbf{H}_{t-1}$ are the input value to the layer and the output of the hidden layer of the previous step, respectively.

Inside each memory cell, the activation function, σ, are all Sigmoid functions. This is a function that scales the values between 0 and 1. This solves the problem of vanishing gradients that can be seen in gradient recurrent networks. Tanh functions can also be seen in Figure 23. These create decision values, which decide if the transformed values from the input or stored memory is worth keeping or is to be used in the output of the cell.

The memory value C_{t-1} and C_t are shown at the top of Figure 23, these are the values that are kept from one step to the next. The purpose of the Equations 10, 11 and 12 is to choose both the output of the memory cell, and to choose what value is kept and memorized for future steps. Altering the parameters can alter how long term dependencies affect the end results, and a tuning process has to be completed for optimal results.

3.3.2 CNN

Convolutional Neural Networks (CNN) are a type of ANN designed for the purpose of managing variations in two dimensional data sets, like images. CNN models have earlier shown to be superior at these type of applications [80]. Nevertheless, the algorithms' capability to recognize spatial characteristics in imagery makes it an interesting potential candidate when picking algorithms for the purpose of anomaly detection on time series data [81].

When processing digital information, CNN has a significant advantage as it can extract important information with fewer weight coefficients [82]. In general, CNN is for the most part made up of four layer types:

1. Input layer
2. Convolutional layers
3. Pooling layers
4. Fully connected layers

The input layer is comprised of the input data.

In the convolutional layer convolutional kernels filter the input data as shown in Equation 13, where the number of convolutional kernels defines the weight coefficients. The function used in the convolutional layers utilizes the convolutional kernels to draw out information from the input data. Therefore, the size and step length of the kernels has a big impact on the performance of the convolutional layer. For two dimensional input we have:

$$x_j^l = f(\sum_{i \in M_j} x_i^{l-1} * W_{ij}^l + b_j^l) \tag{13}$$

Where f is activation function, W_{ij}^l is the weight coefficient corresponding to the convolution at the position (i,j) of the layer l. b_j^l represents the bias, x_i^{l-1} the feature mapping of the previous layer, M_j the set of feature mappings and x_j^l represents the feature map of the current layer.

The pooling layers often involve two popular strategies:

1. max pooling $P_{u,V}^{max}$
2. average pooling $P_{u,V}^{ave}$

The pooling layer filters out the main features of the input data, and reduces computational time significantly [82] [83]. Max pooling, as depicted in Figure 24, replaces the data in a certain area, called the pooling area, with the maximum value, shown in Equation 14. Average pooling does the same, but with the average value, shown in Equation 15. After pooling, the size of the data set and the weight coefficient are reduced, which in turn reduces computational time. Each dimension in the data set is reduced by a factor (stride value). Other pooling strategies include Mixed Max-Average Pooling, 'Gated' Max-Average Pooling, Tree Pooling, Spatial Pyramid Pooling, Stochastic Pooling, S3Pool, Rank-Based Pooling and Fractional Max-out Pooling [83].

$$P_{u,V}^{max} = \max_{i,j \in y_{u,v}} a_{i,j} \tag{14}$$

$$P_{u,V}^{ave} = \frac{1}{|y_{u,v}|} \sum_{i,j \in y_{u,v}} a_{i,j} \tag{15}$$

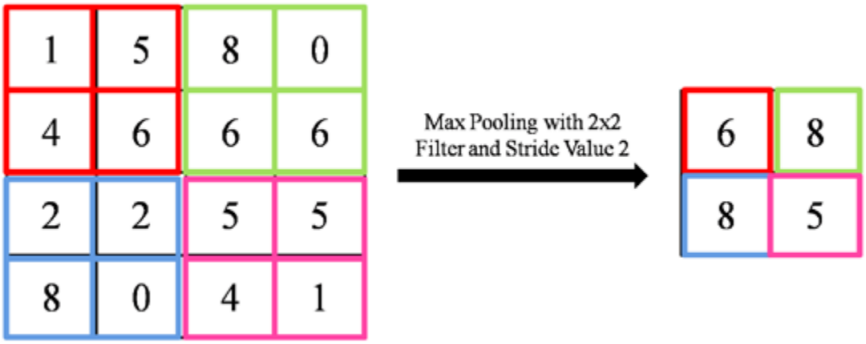


Figure 24: Example of max pooling with a pooling area of 2x2 and stride of 2 [83].

In the fully connected layer, the neurons of the previous layer is connected with all neurons of the current layer. The connection structure of the fully connected layer is identical to those of a traditional neural network. This is conventionally used as the last layer of the CNN [82]. An example of the CNN layers is presented in Figure 25

In Figure 26, the proposed architecture of the CNN model is presented. The model contains three one dimensional convolutional layers (Conv1D) which are optimized using the Talos optimizer package in Python (See section 3.4.1.1). This is followed by a Max-Pooling layer, whose task is to save the most important signals, before these signals are flattened to a single dimension. Lastly, the extracted features are interpreted in a dense fully connected layer.

3.3.3 Convolutional LSTM

Convolutional LSTM (C-LSTM) is a hybrid between CNN and LSTM, where both models are utilized. Combining the architectures will allow for both spatial and temporal features to be

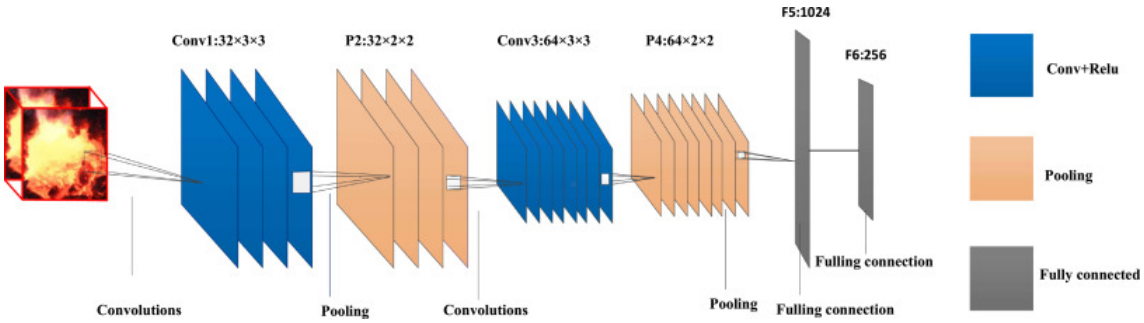


Figure 25: An illustration of layers and connections in a CNN model used to monitoring overheating [84].

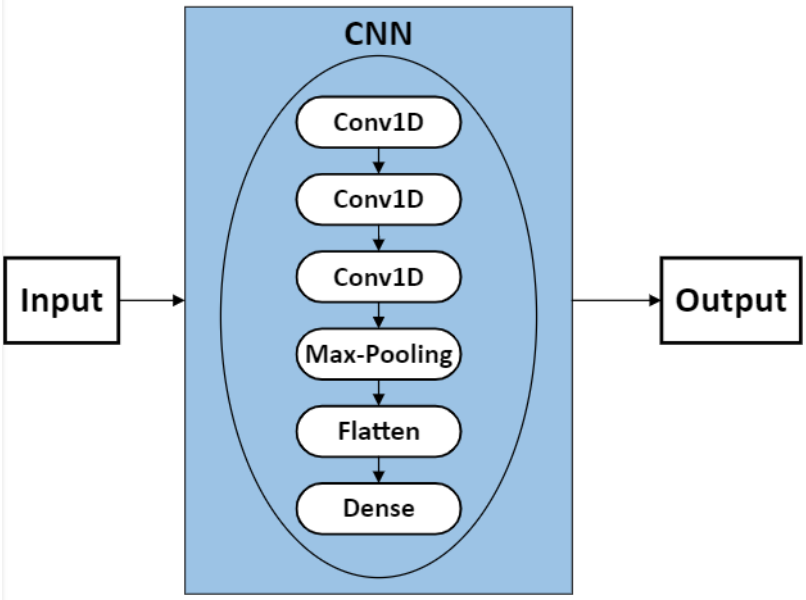


Figure 26: Proposed architecture of the CNN model.

captured. The input data is first fed into a CNN layer before the LSTM layers are used. Starting with a CNN layer will restructure and reformat the data in a way that will be different than the pure LSTM model. This restructuring and reformatting can prove beneficial in capturing anomalous points in the PMU data.

Previous uses of C-LSTM involve text recognition [85]. The combination of CNN and LSTM for text recognition allows for extracting higher-level sequences of words through the application of CNN, and the long-term dependencies through the application of LSTM. For time series data, having the ability to evaluate both higher-level sequences and long-term dependencies could prove important in the context of anomaly detection.

3.3.4 Bidirectional LSTM

Bidirectional LSTM (Bi-LSTM) is a sub-catagory of the LSTM neural network architecture. The Bi-LSTM model expands on the pure LSTM model by implementing and combining a forward propagating layer with a backward propagating layer in one single layer. The bidirectional model should allow for more accurate predictions, as it uses information both from the past and future [86]. This is possible due to the inputs being used in both the forward propagating neurons and

the backward propagating neurons before combining the outputs through the activation function. The architecture and data flow is shown in Figure 27. As seen in the figure, there is no interaction between the forward and backward propagating neurons. This non existent interaction is the difference between having a bidirectional layer and having two separate layers, one being forward and the other being backward propagating.

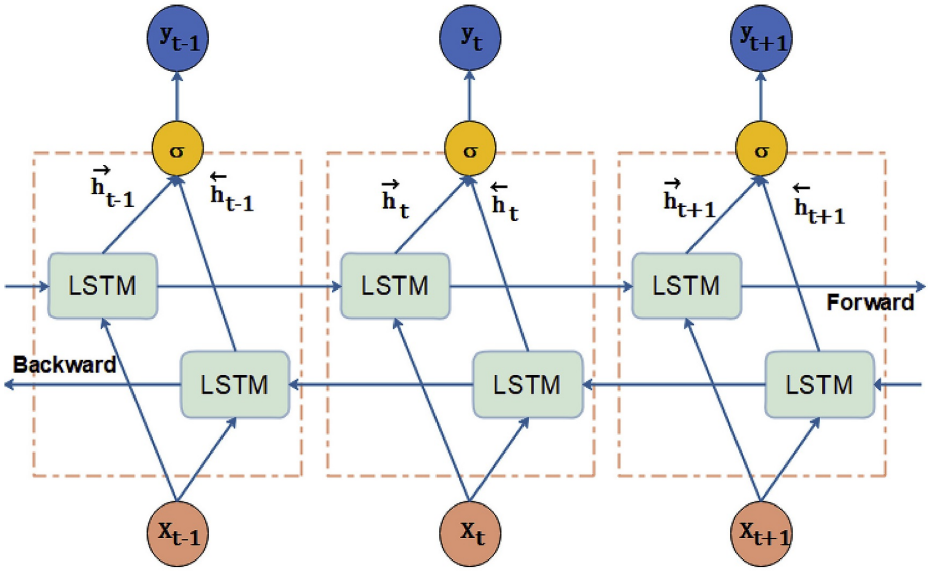


Figure 27: Bi-LSTM architecture [86].

3.3.5 Algorithm Training

The first part of the anomaly detection after completing the data preprocessing is the algorithm training. LSTM is trained under supervision using training data which is a part of the data where anomalies are being probed. The TSN data set is chosen to be used for training all the models. This is due to the model validation step performed later. In our case the training data will stand for 80% of the measurements in the data set, or the first 86400 measurements. While the model is trained the parameters are updated such that predictions for the next measurement, given an arbitrary timestamp, will be better. Anomalies are found by marking those measurements which are outside a given threshold of error from the predicted value.

For the LSTM based models previous signals aren't automatically applied to the ongoing step, but rather uses a set of cells to learn what previous measurements are beneficial for predicting the next one. The input, output and forget gates work together to predict the next step, by learning to open and close at the right moments [6].

When learning, the models are trained separately for each measurement type and for every PMU device. This way the model can learn to predict the values of each time series separately. The mean squared error (MSE) is used during the training as a metric for the performance of the model, as used in [6]. MSE is the average of the square of the error of the values that the model has predicted, shown in Equation 16:

$$MSE = \frac{1}{n}\sum_{t=1}^{n}(Y_t - \hat{Y}_t)^2 \tag{16}$$

Where Y_t is the measured value at the time t, and $\hat{Y}_t$ is the value predicted by the model for the same timestamp. MSE is a helpful tool in both the learning and evaluation process.

3.3.6 Activation Function

Due to the way the cells of the LSTM layers are constructed, non-linearity is ensured through the different gates. The sigmoid and Tanh functions are used in the forget, input and output gates, and as such every value is transformed in each layer. A further activation function is not necessarily needed, as the inputs are transformed within each LSTM layer. A further transformation might not increase the performance of the algorithm, but might instead increase the computational load.

3.3.7 Evaluation

Evaluation of the proposed method and its predictions can be excecuted in numerous ways. For anomaly detection purposes, the evaluation of the method is seen as its ability to detect the anomalies that are presented. As the LSTM method predicts future values, it is natural to use an error vector as a means to compare the predicted values and the real measurements done by the PMUs.

To evaluate the proposed method, the authors in [6] compare a dynamic error thresholding technique with a standard x-sigma thresholding technique. The dynamic error technique is first developed by NASA in [87]. The approach used in this paper will base itself on the work in [6], as it is interesting to investigate how their techniques fare on PMU data.

The first step for error thresholding techniques is to find the prediction error. This is very similar to MSE described above, however it looks at individual data points by themselves. The prediction error is given by:

$$e = |y_t - \hat{y}_t| \tag{17}$$

The dynamic approach, used in [6], requires a series of errors to determine the threshold. This is done by defining a window size, deciding the number of errors to combine into an error vector, $\mathbf{e} = [e_{t-n}, ..., e_{t-1}, e_t]$, where n is the window size.

LSTM based methods frequently experience error spikes [88], therefore, it is beneficial to smoothen the error vector, Exponentially Weighted Moving Average (EWMA) is used. The function is recursive, meaning that one observation is calculated using previous observations, the formula is shown below:

$$EWMA_t = \alpha * e_t + (1 - \alpha) * EWMA_{t-1} \tag{18}$$

The only parameter up for modification, is the parameter α. α is a number between 0 and 1, which defines how quickly the weights tend to zero, when propagating back in time.

The smoothed error vector, $\mathbf{e}_s$, is then to be used to calculate the error threshold. This can either be done supervised, comparing anomalies in the data previously, or unsupervised using the

dynamic technique proposed. The unsupervised method can be beneficial for online time series data, as a variable error threshold can reduce the number of false positives.

The threshold, ϵ, is selected to reduce the amount of anomalies that is marked. Reducing the amount has the benefit of negating false positives, however there needs to be a balance between the number of false positives and the number of false negatives. The threshold vector is found by:

$$\boldsymbol{\epsilon} = \mu(\mathbf{e}_s) + \mathbf{z}\sigma(\mathbf{e}_s) \tag{19}$$

Where μ and σ is the mean and standard deviation of the smoothed error vector, respectively. $\mathbf{z}$ is a positive number chosen to scale the threshold. A lower number means a lower threshold, and can lead to more false positives, in [88], a number between 2 and 10 is used.

Further, ϵ is chosen as:

$$\epsilon = argmax(\boldsymbol{\epsilon}) = \frac{\Delta\mu(\mathbf{e}_s)/\mu(\mathbf{e}_s) + \Delta\sigma(\mathbf{e}_s)/\sigma(\mathbf{e}_s)}{|\mathbf{e}_a| + |\mathbf{E}_{seq}|^2} \tag{20}$$

Where

$$\Delta\mu(\mathbf{e}_s) = \mu(\mathbf{e}_s) - \mu(\{e_s \in \mathbf{e}_s | e_s < \epsilon\})$$

$$\Delta\sigma(\mathbf{e}_s) = \sigma(\mathbf{e}_s) - \sigma(\{e_s \in \mathbf{e}_s | e_s < \epsilon\})$$

$$\mathbf{e}_a = \{e_s \in \mathbf{e}_s | e_s > \epsilon\}$$

and $\mathbf{E}_{seq}$ is a continuous sequence of $e_a \in \mathbf{e}_a$. The severity of the anomaly is also of importance, so a anomaly score is given to each detected anomaly by Equation 21:

$$s^{(i)} = \frac{max(\mathbf{e}_{seq}^{(i)} - argmax(\epsilon)}{\mu(\mathbf{e}_s) + \sigma(\mathbf{e}_s)} \tag{21}$$

3.4 Model Analysis

3.4.1 Model Calibration

The calibration process incorporates tuning the hyperparameters of the different models to perform in the best way possible. Using the Talos optimizer python package [89], easy and efficient parameter tuning is achievable. The optimizer package tests a given range of values for chosen hyperparameters. The outcome of the optimizer shows the value of evaluation metrics for each value of hyperparameters tested. As well as generating numerical values for metrics for the given values of hyperparameters, Talos also generates a correlation graph to show the correlation between different hyperparameters and evaluation metrics as shown in Figure 28. This correlation graph provides a visual overview of how the value of given hyperparameters affect the value of the evaluation metrics. In Figure 28, a positive value and red-gradient color indicates that the value of hyperparameter has a positive correlation to the corresponding evaluation metric. A negative and blue-gradient color indicates a negative correlation to the corresponding evaluation metric, meaning that a higher value of hyperparameter results in a lower value for the evaluation metric, which in this case is desired.

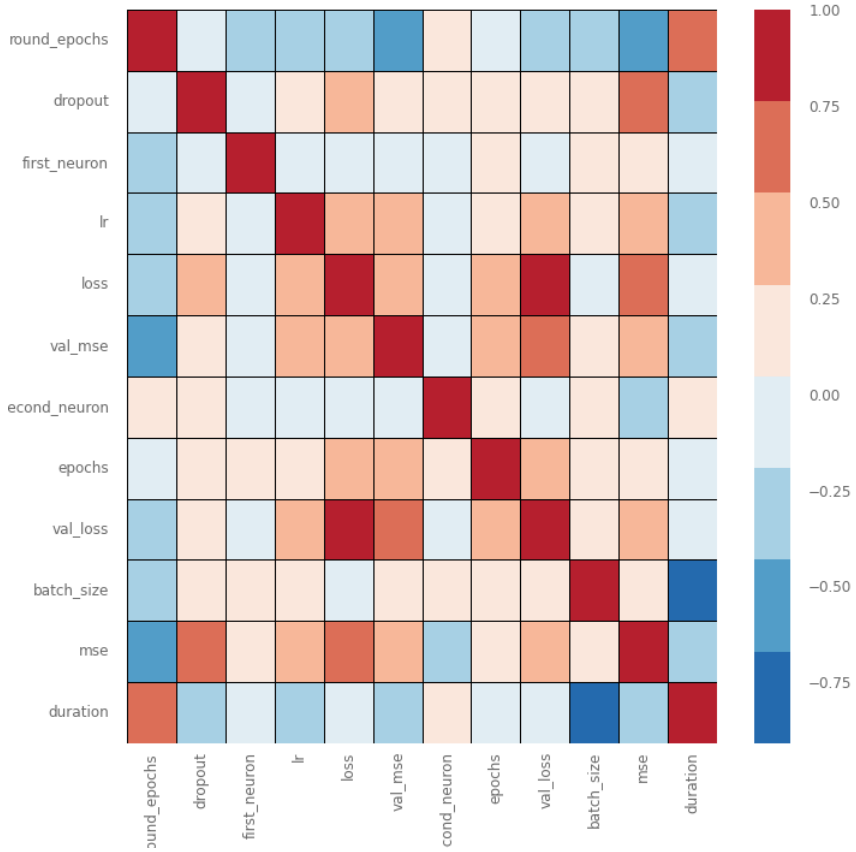


Figure 28: Talos optimizer correlation graph.

Table 7, 8 and 9 shows the parameters that generated the best outcome from the Talos optimization.

Table 7: CNN model parameters decided by the hyperparameter optimization.

CNN		
Conv1D	Filters	32
	Kernel size	3
	Activation	relu
Conv1D	Filters	16
	Kernel size	3
	Activation	relu
Conv1D	Filters	64
	Kernel size	3
	Activation	relu
MaxPooling	Pool size	2
Dense	Units	50
	Activation	relu
Dense	Units	1
Flatten	-	-

Table 8: LSTM and Bi-LSTM model parameters decided by the hyperparameter optimization.

LSTM/Bi-LSTM		
LSTM/Bi-LSTM	Hidden Units	64
Dropout	Dropout Value	0.12
LSTM	Hidden Units	32
Dense	Units	1

Table 9: C-LSTM model parameters decided by the hyperparameter optimization.

C-LSTM		
Conv1D	Filters	64
	Kernel Size	3
	Activation	relu
LSTM	Hidden Units	64
Dropout	Dropout Value	0.12
LSTM	Hidden Units	32
Dropout	Dropout value	0.12
Dense	Units	1

3.4.2 Model Validation

To ensure that the models proposed are calibrated correctly and function as intended, a model validation step is necessary. This validation step includes using the data obtained by the University of Texas and a statistical model proposed in [73]. The statistical method is developed by the National Renewable Energy Laboratory (NREL) and is specifically designed to detect anomalies

on the PMU data from Texas. This provides a great foundation to validate the models proposed in this thesis.

After the models are calibrated with tuning of the hyperparameters, the models are tested on the PMU data from Texas for anomaly detection purposes. The results are then compared to the anomalies found using the statistical method, to see if the ML models proposed here are able to capture the same anomalies.

The models that capture a satisfactory amount of the same anomalies is deemed well calibrated and functioning in the correct manner. This validation step gives credibility to the models used for real world application. It also gives the benefit of knowing that the anomaly detection of injected data, as seen later in this thesis, is not overfitted for the specific data injected, and instead accomplishes correct prediction of normal operating conditions.

As the models used for the Texas data and the data obtained from Statnett are the same, the validation step described above is valid for both instances. This provides a foundation for comparison between data sets, as well as testing how well the models that are calibrated for one PMU data set fares on another PMU data set with different input values and measurement frequencies. This comparison gives further insight into how well the models are able to adapt and find out if there are any strengths or weaknesses involved.

3.4.3 Data Injection

To test the models ability to detect bad data related to cyber security threats and anomalies, false data is injected into the data stream. As seen in Figure 6, there are four main types of bad data. The main focus of this thesis is injecting and detecting noise added to the system. Injecting noise is a way of mimicking a malfunction in a component or sensor, it is also similar to a Noise Injection Attack (NIA) [90].

To inject the false data, Gaussian noise is added to the data points over a time period. This time period is randomly chosen between 1 and 4 seconds in length at a random point during the whole length of data set. The amount of data affected is therefore in the range of 25-120 data points, depending on the measurement frequency of the PMUs. Gaussian noise is chosen due to its easy access through NumPy python package [91], and due to the noise generated having a probability density function equal to the normal distribution.

3.4.3.1 Evaluation Metrics To evaluate the models' performances a statistical analysis is deployed. The chosen metrics are popular in supervised ML:

- Recall
- Precision
- F1-Score

Recall describes how many of the true anomalies are found [92]. In other words, anomaly found by a model with a high recall score is not necessary truly an anomaly. However, an engineer can trust that few true anomalies are undiscovered.

$$Recall = \frac{Truepositives}{Truepositives + Falsenegatives} \tag{22}$$

Precision describes how many of the found anomalies are true [92]. Hence, an anomaly found by a model with a high precision is more likely to be true.

$$Precision = \frac{Truepositives}{Truepositives + Falsepositives} \tag{23}$$

The F1-score is the harmonic mean of precision and recall. It describes the overall performance of the algorithm [92].

$$F1 - Score = \frac{2 * Precision * Recall}{Precision + Recall} \tag{24}$$

When evaluating the results the F1-Score is the simplest way to decide which algorithm performs better. However, for different applications a higher Recall or Precision might be of a higher priority.

3.5 Model Deployment

After validating the models' abilities to detect both real and injected anomalies, the next step is to run the models on unlabeled data. That way, the models can be of help to SO's that would like to get quickly learn if there are any faults in their network. To simulate a real deployment of the models, the unlabeled 48 hours of data from Statnett is split into 20 equal parts; consisting of 2 hours and 24 minutes of measurements. For each part, the error threshold is updated using the first 5% of the data, and the whole set is checked for anomalies by the four different models. This type of test is similar to real world deployment, where the models are subject to new unseen data. Instead of splitting the data set into different equal parts, a sliding window technique is also possible to mimic a live data feed.

4 Results

4.1 Preprocessing Results

This subsection presents and visualizes the preprocessing results, showing its importance and how it's implemented.

4.1.1 Remove Missing Values

It was found that the only data sets with missing values were the data sets from Statnett. The data sets from PMU 1, 2 and 3 contained 435, 1029 and 102787 missing values respectively. Running the models on this data without removing the missing values resulted in errors, as the data being analyzed was not complete and always numerical. Figure 29 shows a snapshot of 100 data points from PMU 1, showing a section with missing values. Figure 30 shows how a linear regression is used to connect the two end points on each side of the missing value section.

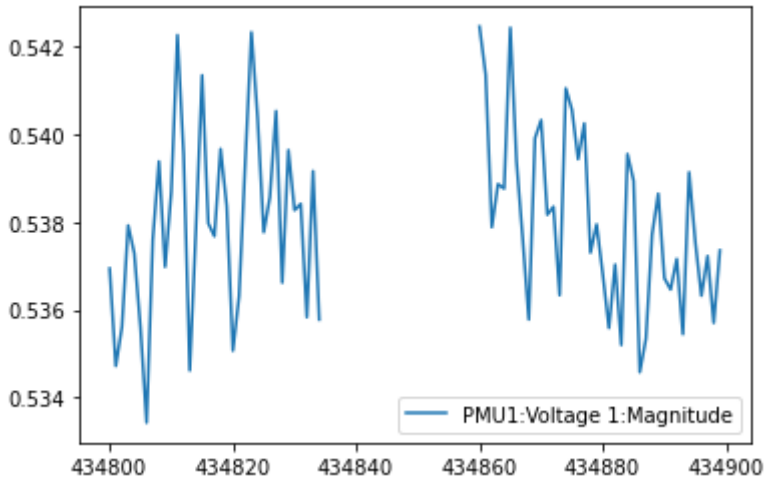


Figure 29: Missing values in Statnetts PMU 1 data set.

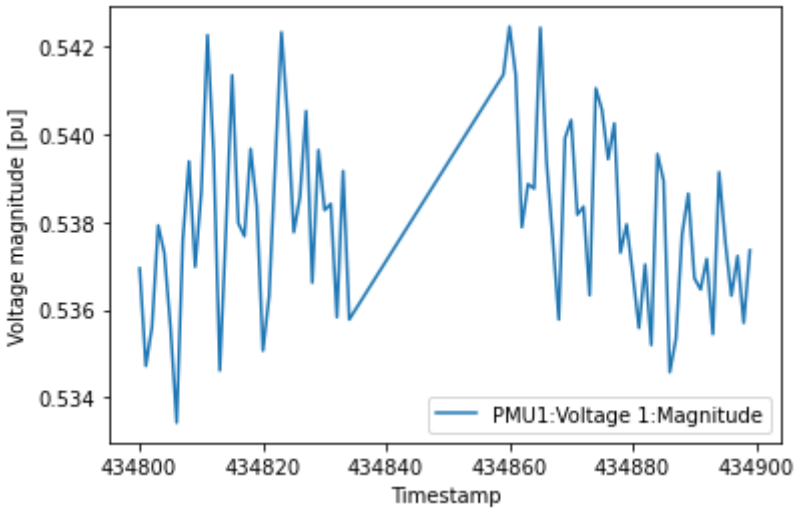


Figure 30: Missing values removed in Statnett's PMU 1 data set.

As seen in Figure 30, the straight line between the two points looks out of place and does not seem to replicate the real data in any meaningful way. However, this does not hinder or affect the models in anomaly detection context, as the amplitudes are within normal operating levels and the relative step change between data points are lower than what they normally are.

4.1.2 Noise Filtration Results

Testing the models on data with and without noise filtration shows the value of removing or dampening noise from the input data, shown in later subsections. As seen in Figures 31, 32 and 33, the noise filtration lowers the amplitude of various harmonics. This leads to a more uniform plot with less spikes and unwanted fluctuations.

In the noise filtration stage the order of the median filter M, is the only parameter that can be adjusted to balance filtering out the noise while not excluding any anomalies in the data. Figures 31, 32 and 33 show three different orders of filtering on a portion of the TSN data. Figure 31 shows the unfiltered phase voltage measurements, Figure 32 shows median filter phase voltage of order $M = 50$, and Figure 33 shows median filtered phase voltage of order $M = 150$.

The figures clearly show the advantages and disadvantages of filtering the data for anomaly detection purposes. The unfiltered phase voltage has a lot of minor spikes throughout, that may not be due to load variations on the grid. The filtered phase voltage of order $M = 50$ shows a much cleaner result, where load changes are more visible. A drawback is that some of the major spikes disappear, as they are too brief to have an impact in the filtering process. This is even more visible on the filtered phase voltage of order $M = 150$, here all major spikes disappear. This would be beneficial if the purpose was to use the data for SE or similar. However, as the purpose here is anomaly detection, it would be difficult to register the single point anomalies that are seen in the unfiltered phase voltage measurements.

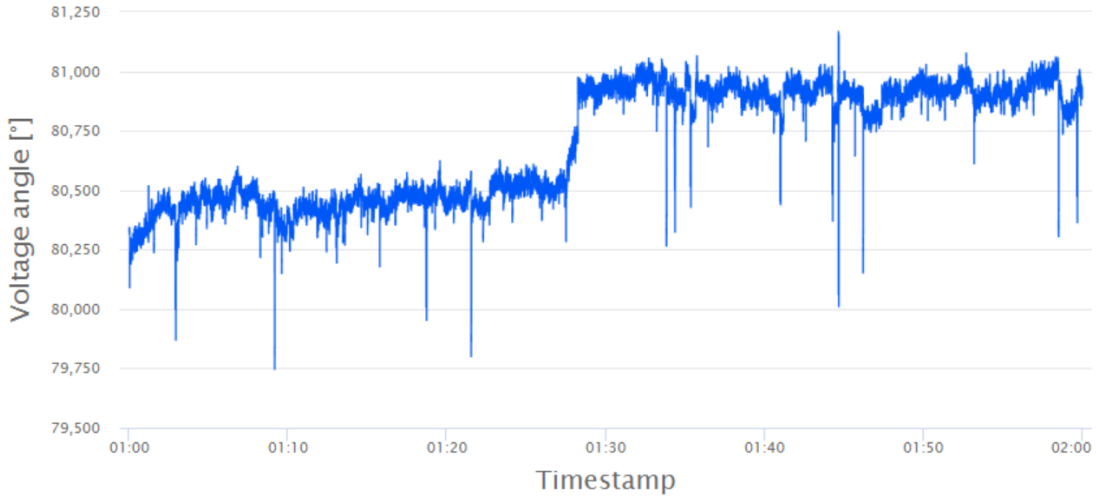


Figure 31: Unfiltered phase voltage.

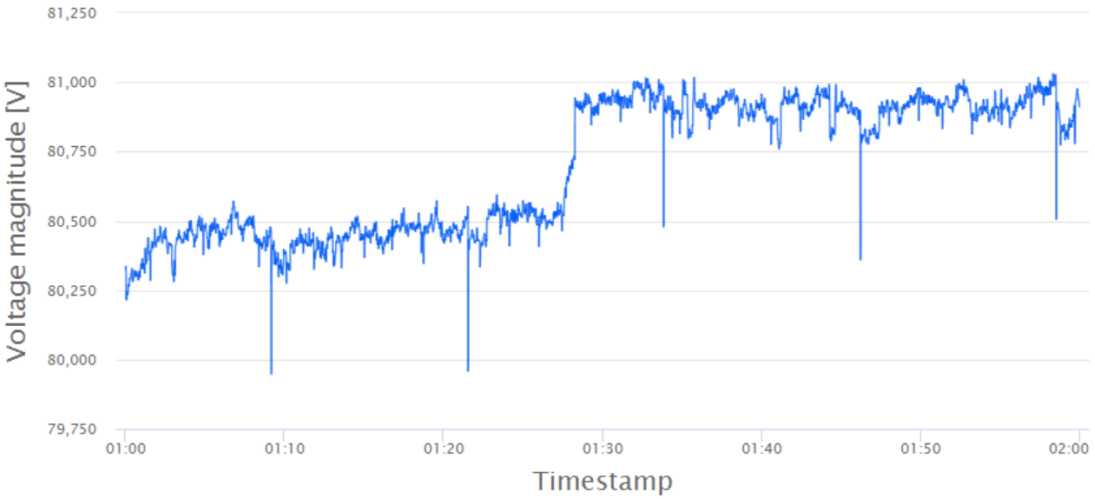


Figure 32: Median filtered phase voltage, M = 50.

Figure 33: Median filtered phase voltage, M = 150.

4.2 Model Validation Results

Using the statistical model developed by NREL on the TSN data, multiple anomalies are found. An example of an anomaly is shown in Figure 34. Using the machine learning models trained on the TSN data, the same anomalies were detected, as can be seen in Figure 35.

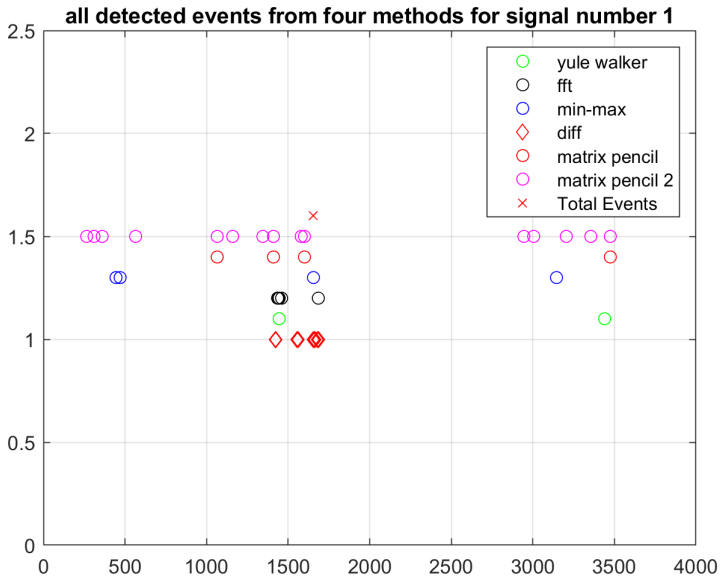


Figure 34: Anomaly labeled by the NREL model.

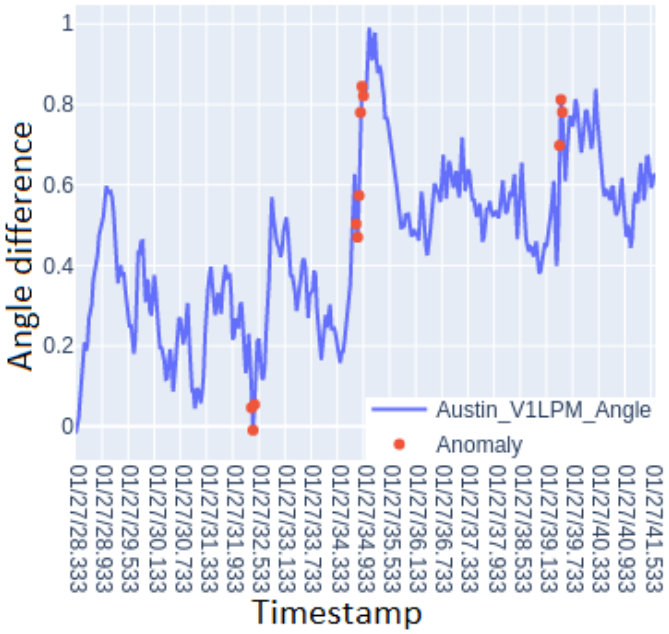


Figure 35: Anomaly detected using the C-LSTM model.

As seen in Figure 34, the anomalies found using the statistical method indicates an event at 1443 seconds from the beginning of the data set. This corresponds to a time of 24 minutes and 3 seconds. Figure 35 shows detected anomalies in this exact time frame, as expected. In this particular case, the C-LSTM model was used and a random event was chosen to compare the statistical method

developed by NREL with the models used in this thesis. Similar results were found for other events detected by the statistical method using all four ML models explained earlier.

4.3 Detecting FDI Attacks

Evaluation of the models efficacy is done using appropriate metrics within the field of ML. Due to the fact that the Gaussian noise is inserted in a small area relative to the whole data set, the accuracy metric is not a good indicator. Since accuracy examines the data set as a whole, differences between good and bad performance will make such a small difference that it will be difficult to evaluate. However, the metrics recall (Equation 22), precision (Equation 23) and F1-score (Equation 24) can examine any desired area of the data set. They are therefore fit to evaluate the different algorithms' performance when detecting FDIs.

In Table 10 and Table 11 the results of the algorithms performance on the TSN and Statnett data respectively, are presented. For each combination of model and noise filtration, 15 tests were run with the FDI being at different locations chosen at random for each test. The results shown in Table 10 and 11 show the average of the 15 tests for each metric, and each given combination. The models' performance vary on the different data sets within the different performance metrics. However, some patterns can be observed.

Table 10: Performance results for the different models using TSN data.

Model	**Noise filtration**	**Recall (%)**	**Precision (%)**	**F1-Score(%)**	**Computational time (ms/step)**
CNN	Yes	**97.42**	81.53	88.34	3
	No	95.25	84.50	89.21	3
LSTM	Yes	93.67	95.18	94.37	7
	No	82.67	**97.35**	89.30	7
Bi-LSTM	Yes	94.75	96.39	**95.47**	9
	No	87.08	96.20	91.35	10
C-LSTM	Yes	89.67	96.49	92.88	7
	No	80.25	96.11	87.36	7

Table 11: Performance results for the different models using Statnett data.

Model	**Noise filtration**	**Recall (%)**	**Precision (%)**	**F1-Score(%)**	**Computational time (ms/step)**
CNN	Yes	**96.53**	84.50	89.82	3
	No	95.13	84.10	89.19	3
LSTM	Yes	93.63	98.46	95.97	8
	No	83.47	99.10	90.59	8
Bi-LSTM	Yes	95.20	98.79	96.96	11
	No	84.77	**99.48**	91.50	11
C-LSTM	Yes	96.43	97.80	**97.10**	8
	No	87.13	98.43	92.41	8

The CNN model (Figures 36, 37, 38 and 39) struggles to correct itself after the added noise. This results in the model labeling to many anomalies, which in turn lowers the precision and F1-score. On the other hand, since it labels many anomalies, there are few false negatives which results in a high recall.

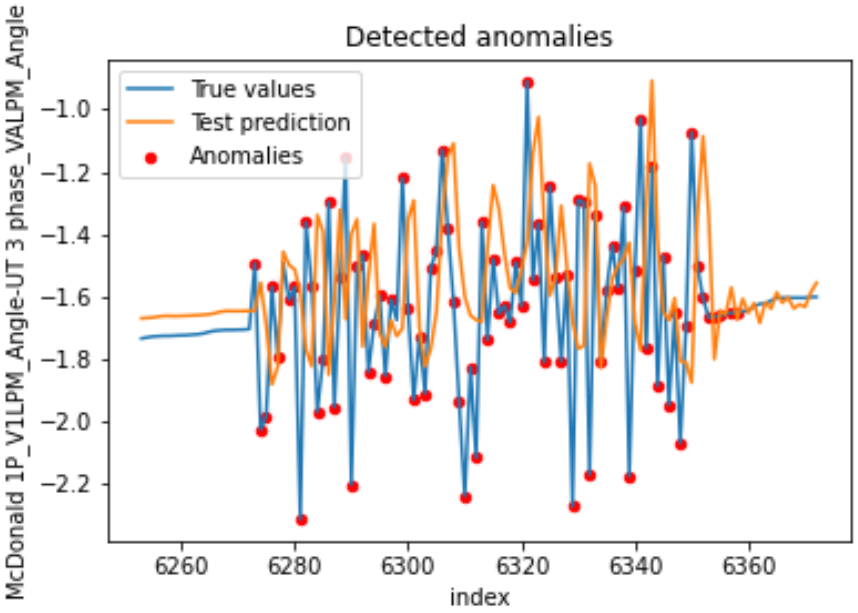


Figure 36: The CNN models' prediction of the noise filtered TSN data with added gaussian noise.

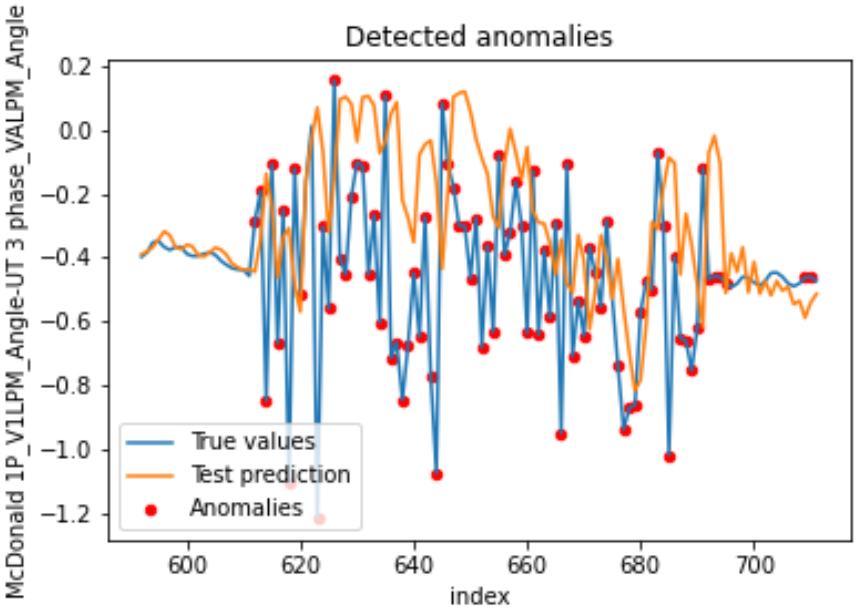


Figure 37: The CNN models' prediction of the unfiltered TSN data with added gaussian noise.

The LSTM model (Figures 40, 41, 42 and 43) generally follows the noise quite well, not labeling to many anomalies; giving the model a high precision. However, the models' recall score suffers a bit from the models' ability to follow the noise, resulting in false negatives. When the model is tested on a data set which has not gone through noise filtration, the recall score is punished even more, because of the higher error threshold that unfiltered data sets receive. This in turn results in fewer anomalies being labeled which increases the amount of false negatives.

The Bi-LSTM model (Figures 44, 45, 46 and 47) works very similar as the LSTM model, meaning that it follows the noise well. This results in few false positive, but some false negatives. However, a marginally better recall and precision score makes it outperform the LSTM models' F1-score in

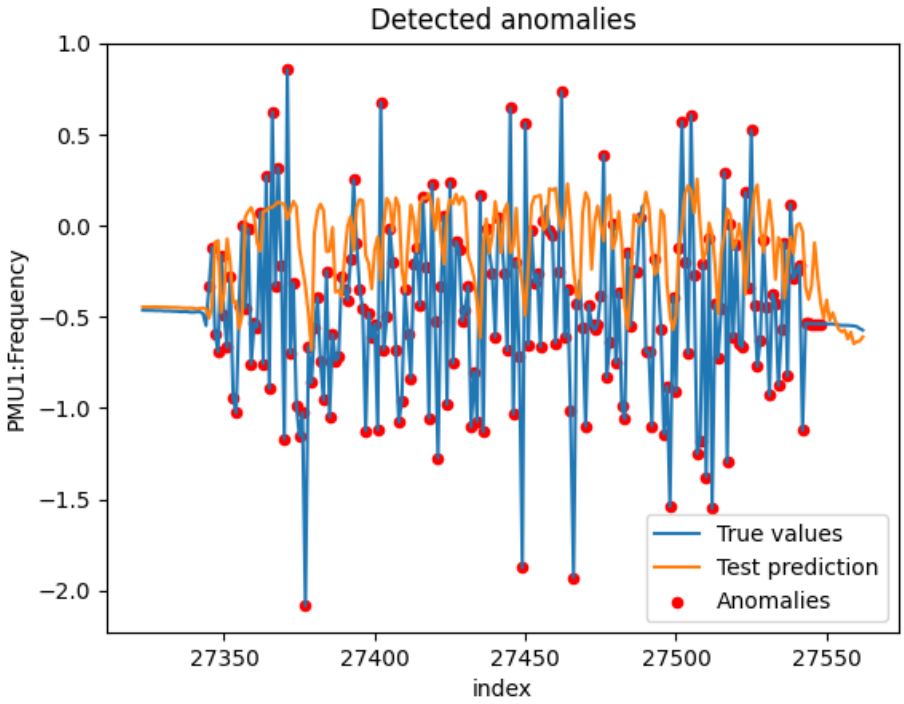


Figure 38: The CNN models' prediction of the noise filter Statnett data with added gaussian noise.

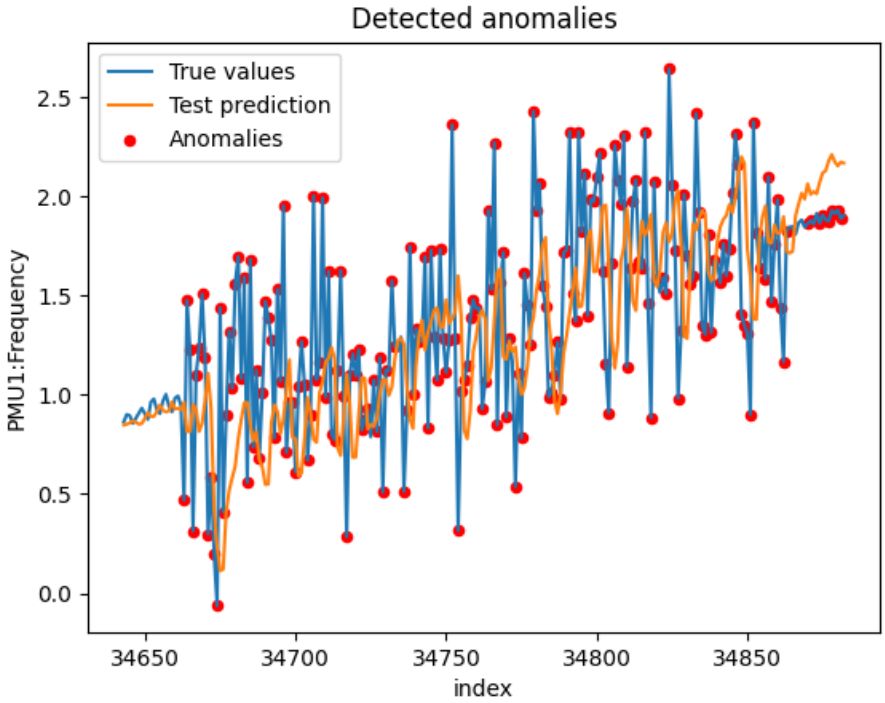


Figure 39: The CNN models' prediction of the unfiltered Statnett data with added gaussian noise.

all tests. Only in one of the tests (TSN data, without noise filtration) did the LSTM model get a slightly higher precision than the Bi-LSTM model. This might indicate that the Bi-LSTM model follows the noise a little less tightly then the LSTM model, without influencing the precision score.

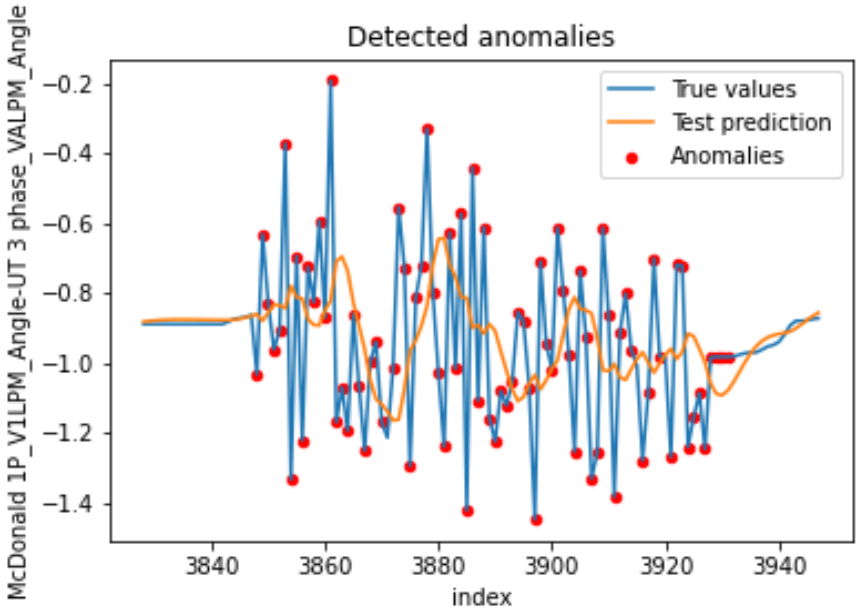


Figure 40: The LSTM models' prediction of the noise filtered TSN data with added gaussian noise.

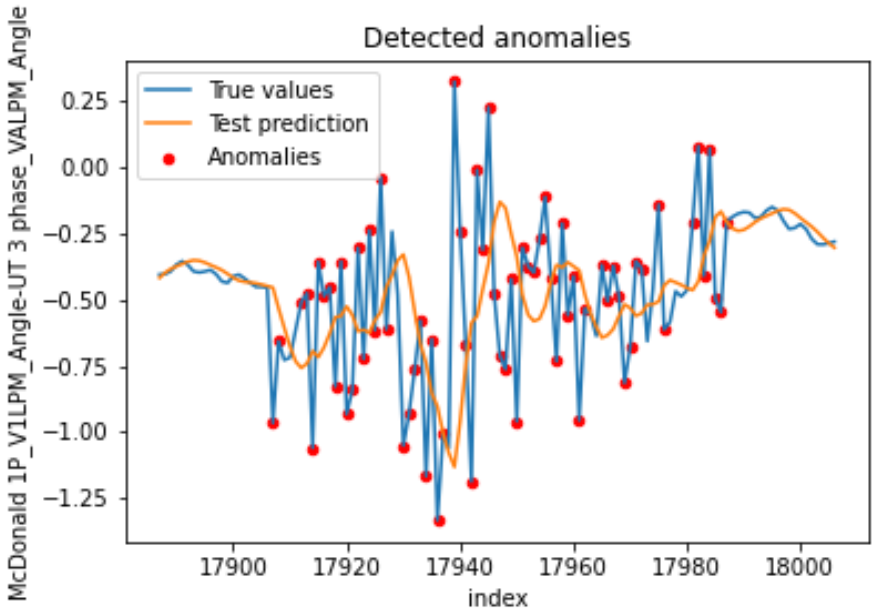


Figure 41: The LSTM models' prediction of the unfiltered TSN data with added gaussian noise.

The C-LSTM model (Figure 48, 49, 50 and 51) follows the test data too well. This means that it doesn't detect all anomalies which results in a lower recall score. This is the opposite of what happens with the CNN model. However, note there is an exception on the Statnett with noise filtration test. Here the C-LSTM has the highest recall score. It seems that the added noise is too difficult to follow, eliminating the punishment on the recall score, without harming the precision. On the other hand, the anomalies that are labeled are likely to be true since the precision score is high.

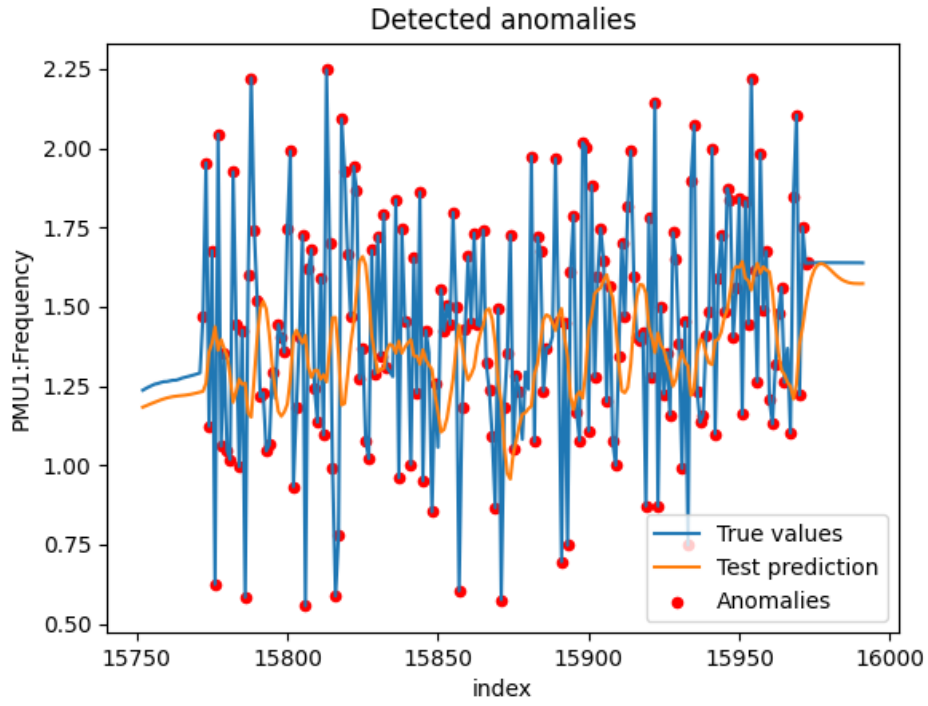


Figure 42: The LSTM models' prediction of the noise filtered Statnett data with added gaussian noise.

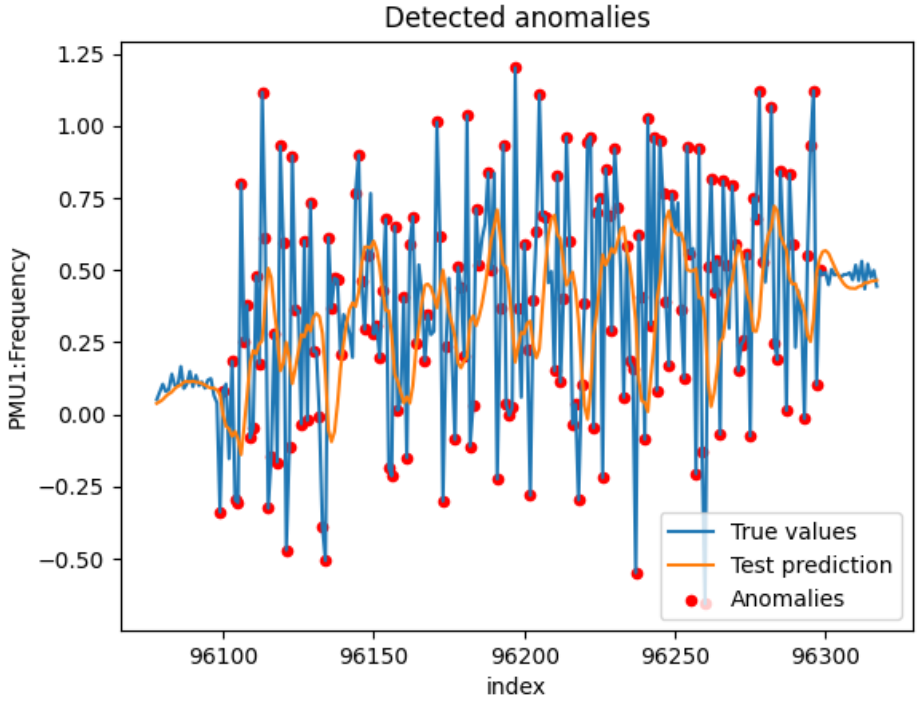


Figure 43: The LSTM models' prediction of the unfiltered Statnett data with added gaussian noise.

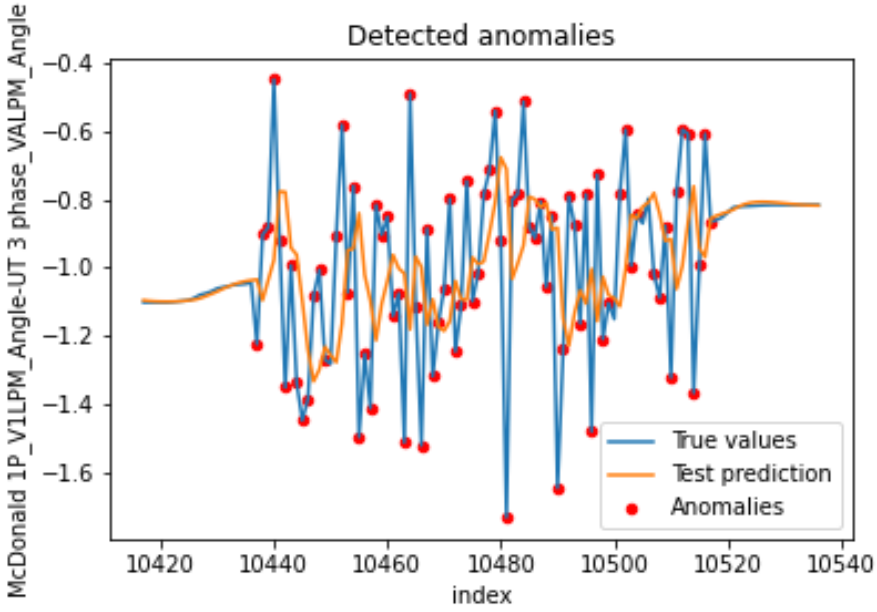


Figure 44: The Bi-LSTM models' prediction of the noise filtered TSN data with added gaussian noise.

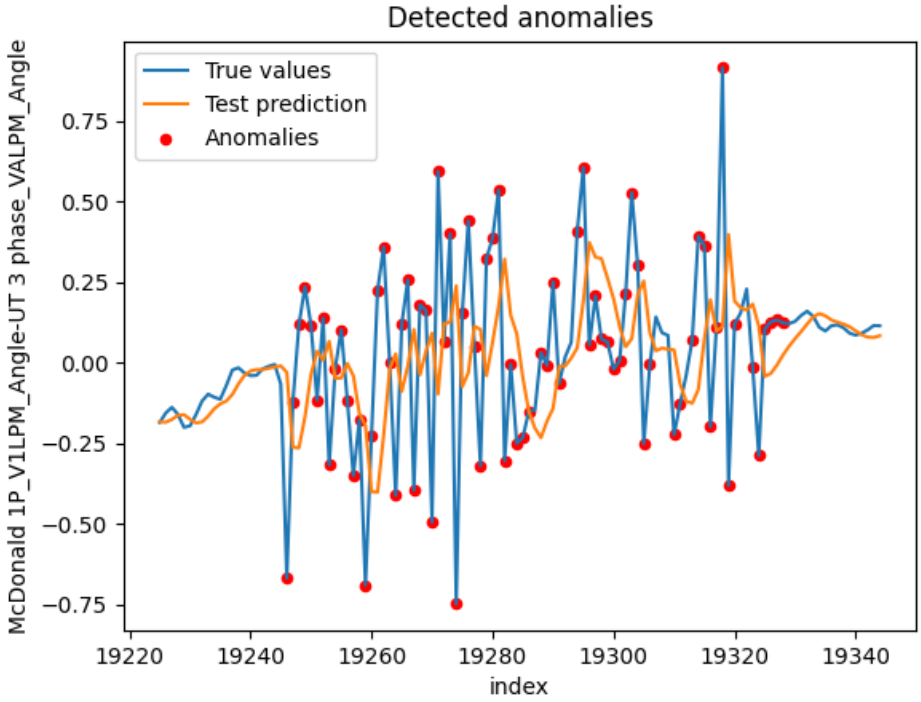


Figure 45: The Bi-LSTM models' prediction of the unfiltered TSN data with added gaussian noise.

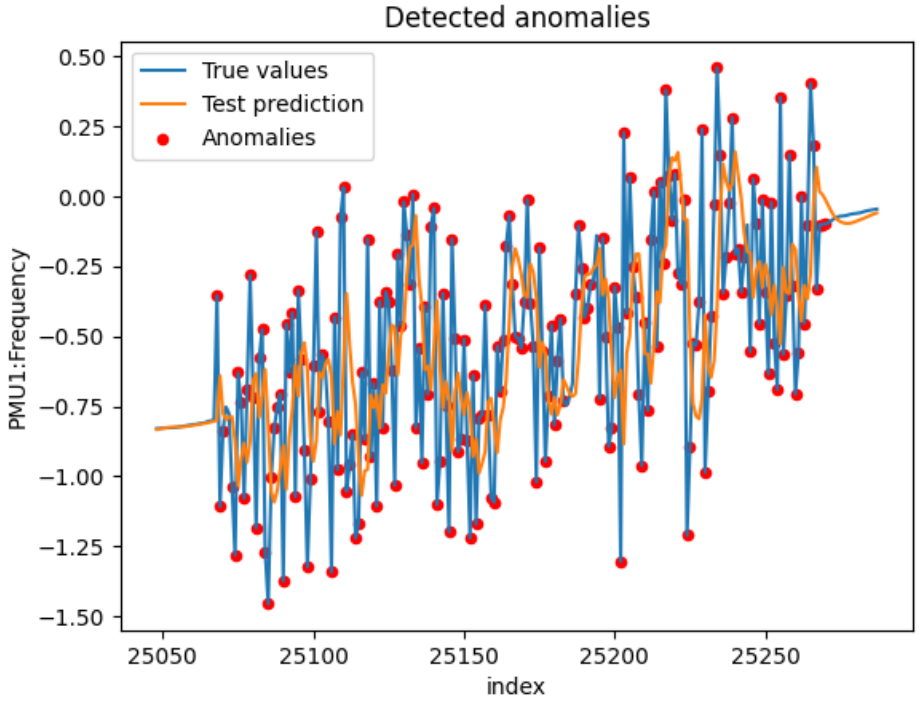


Figure 46: The Bi-LSTM models' prediction of the noise filtered Statnett data with added gaussian noise.

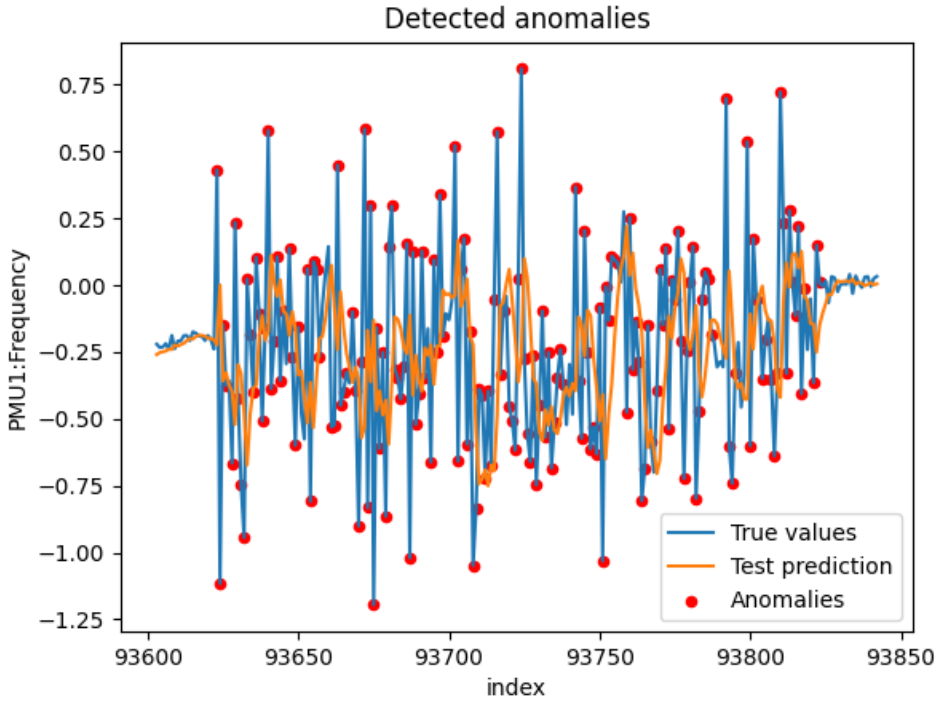


Figure 47: The Bi-LSTM models' prediction of the unfiltered Statnett data with added gaussian noise.

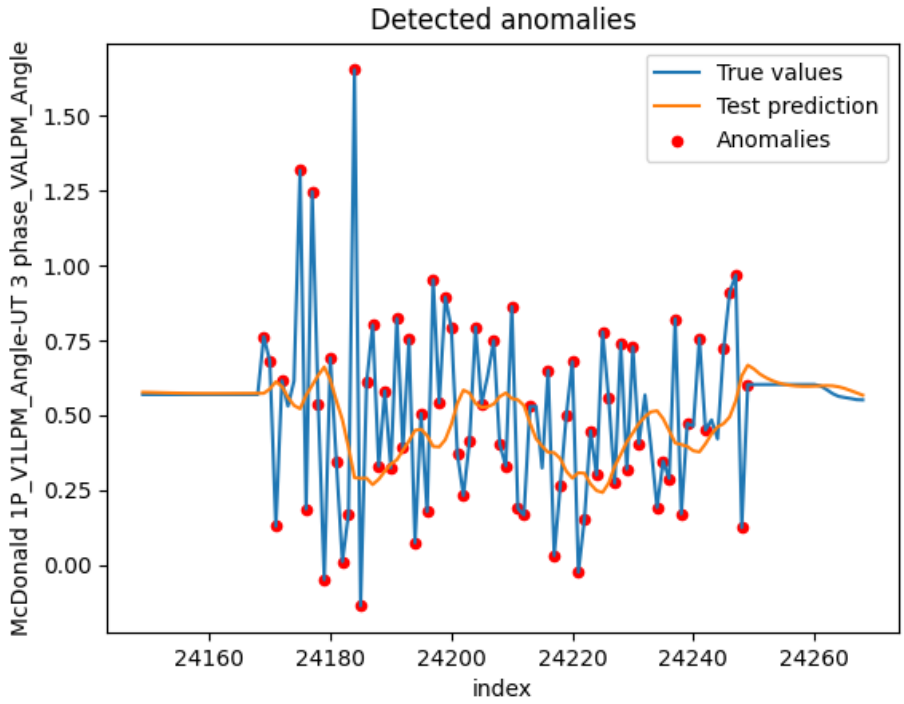


Figure 48: The C-LSTM models' prediction of the noise filtered TSN data with added gaussian noise.

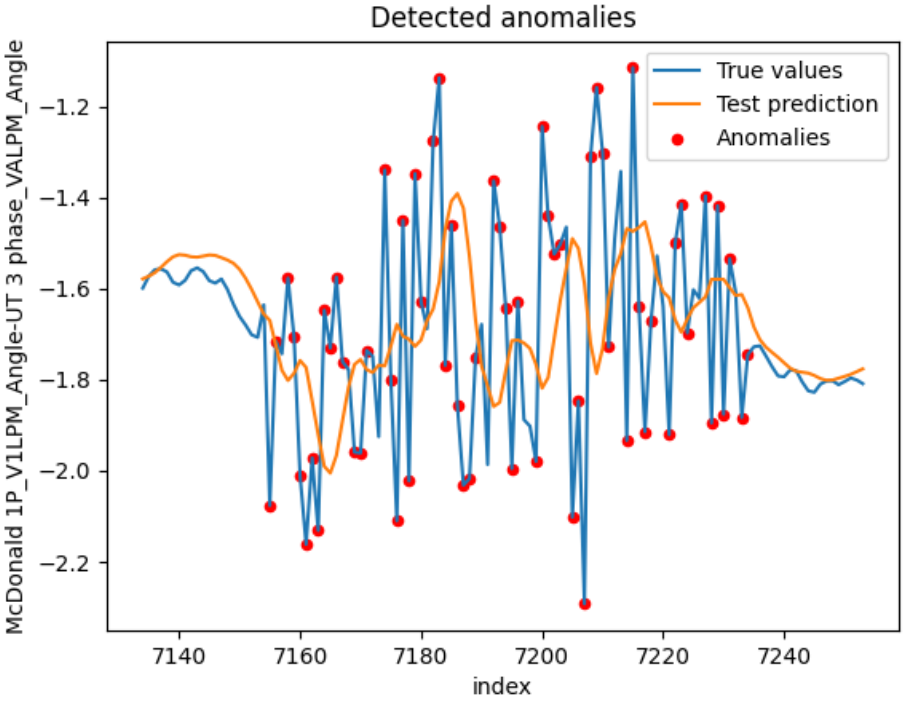


Figure 49: The C-LSTM models' prediction of the unfiltered TSN data with added gaussian noise.

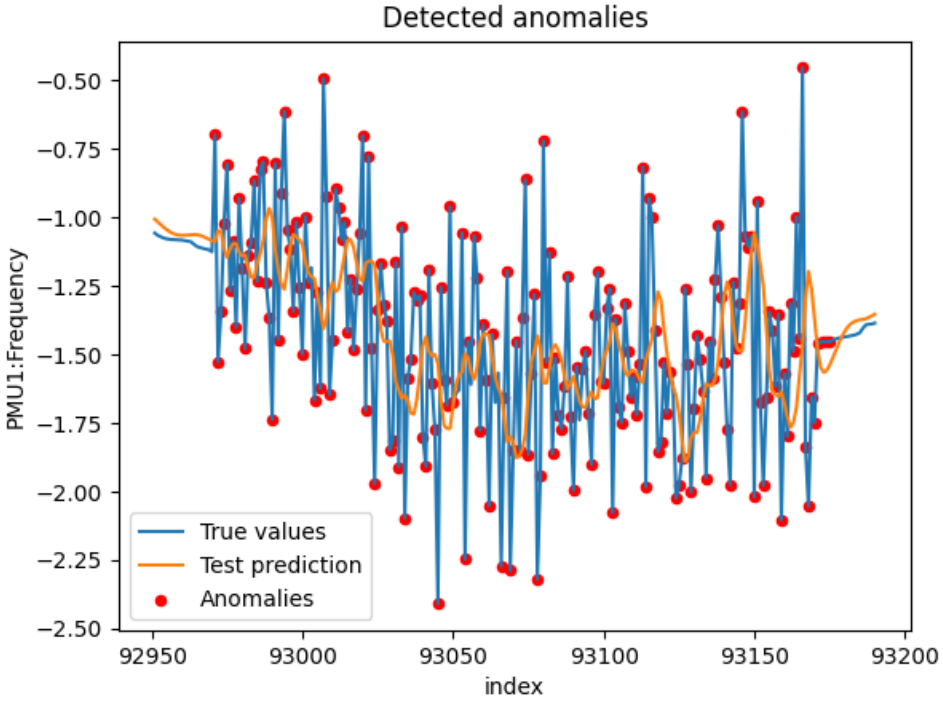


Figure 50: The C-LSTM models' prediction of the noise filtered Statnett data with added gaussian noise.

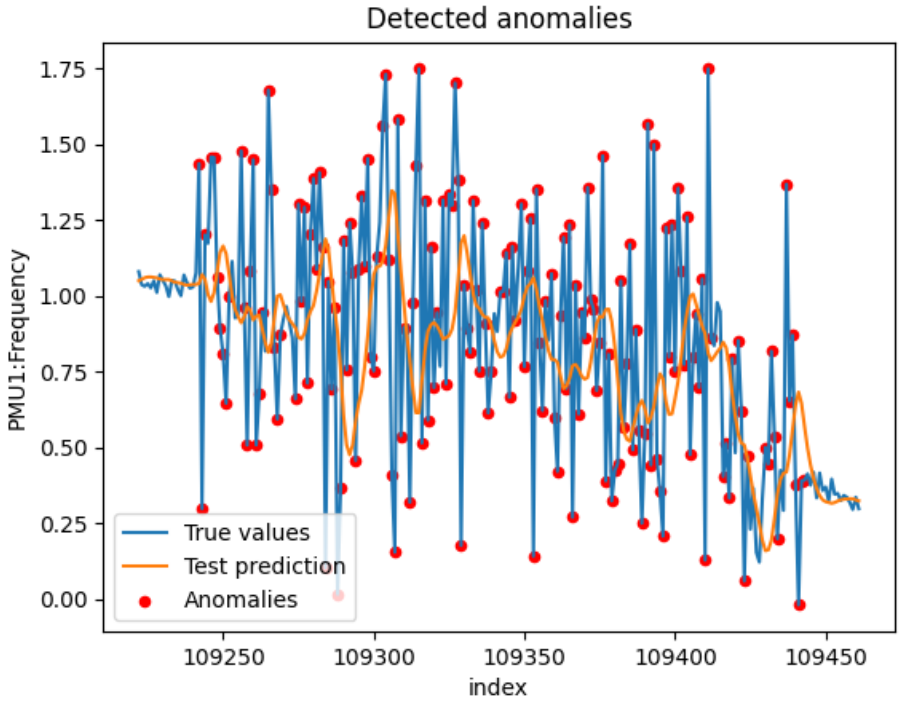


Figure 51: The C-LSTM models' prediction of the unfiltered Statnett data with added gaussian noise.

4.3.1 Drift Anomaly Detection

A linear drift was added to the Statnett data, at a random part of the test data. The drift continued from the start location and increased throughout the rest of the data set. Figure 52 clearly shows the drift anomaly. The relative change between single data points is minuscule, but throughout thousands of data points, the change adds up to alter the frequency. An increase in approximately 1 Hz in frequency over a time frame of 15 minutes can prove catastrophic for a power system.

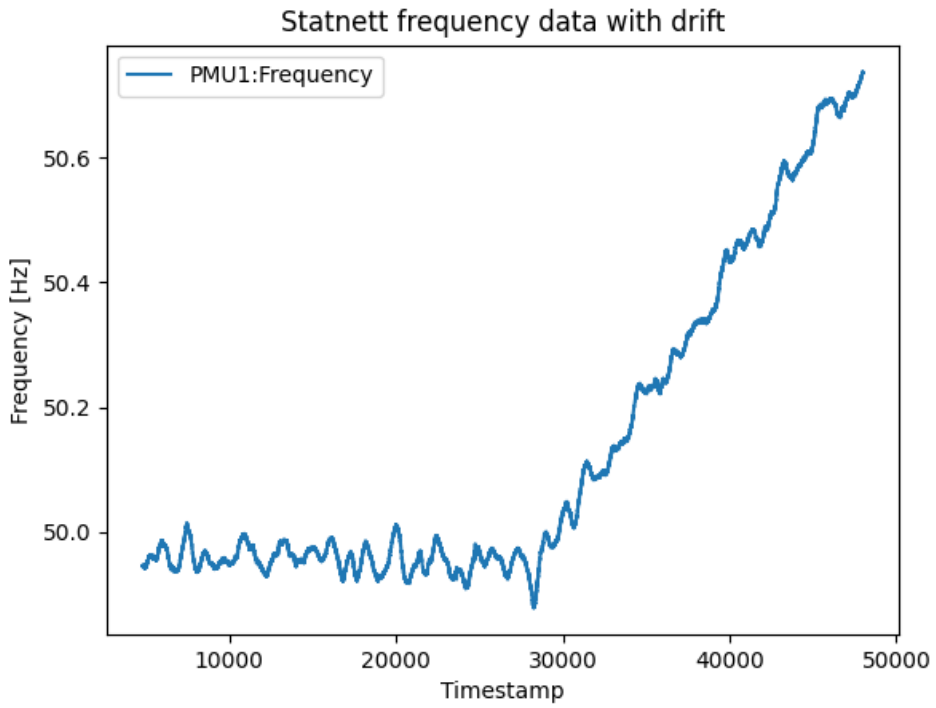
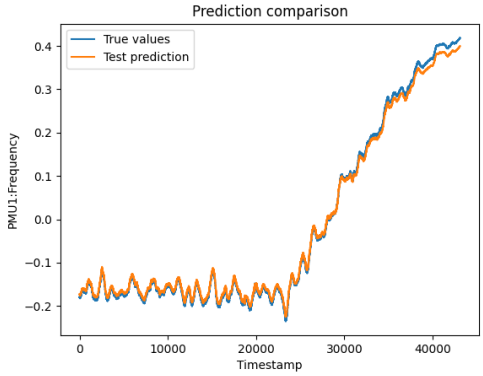
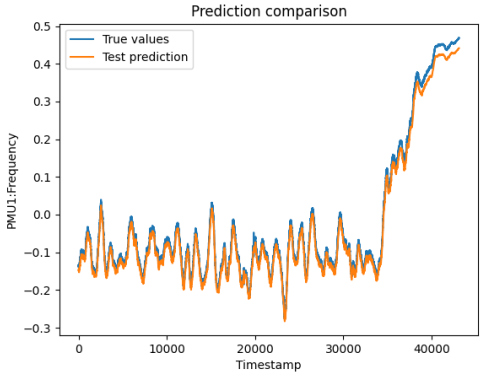


Figure 52: Statnett frequency data with injected drift anomaly.

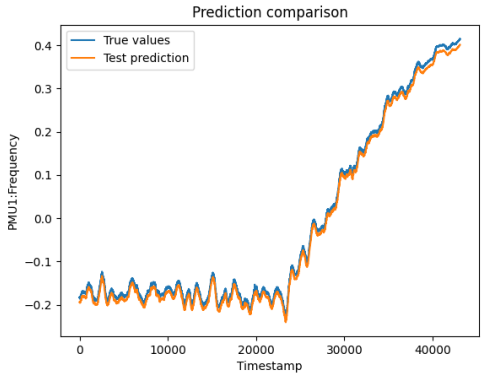
None of the models managed to detect the drift, as they adapted to the new trend quickly. Figure 53 shows the four models and their predictions on top of the data. The models performed similarly, as the drift does not disrupt the data values in big enough manner for the predictions to be off.



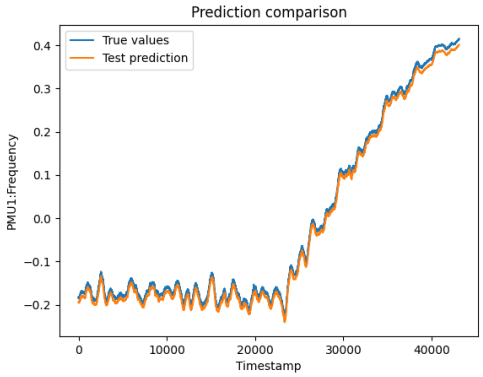
(a) CNN model.



(b) LSTM model.



(c) Bi-LSTM model.



(d) C-LSTM model.

Figure 53: Drift anomaly prediction and detection on Statnett data.

4.3.2 Spike Anomaly Detection

To test the models on spike anomalies, a random set of 5 spikes were injected in a snapshot of the Statnett data with a time length of 250 000 samples. The amplitude of the injected data was chosen to be small enough as to not be easily detected by a primitive visual inspection, and as such varies between 0.01 and 0.02 Hz. As seen in Figure 54, the injected data is hidden within the normal operating ranges and fluctuations. For this particular case the five spikes are located between timestamp 106400 and 106900. Analyzing Figure 54 shows no unusual behaviour at these timestamps.

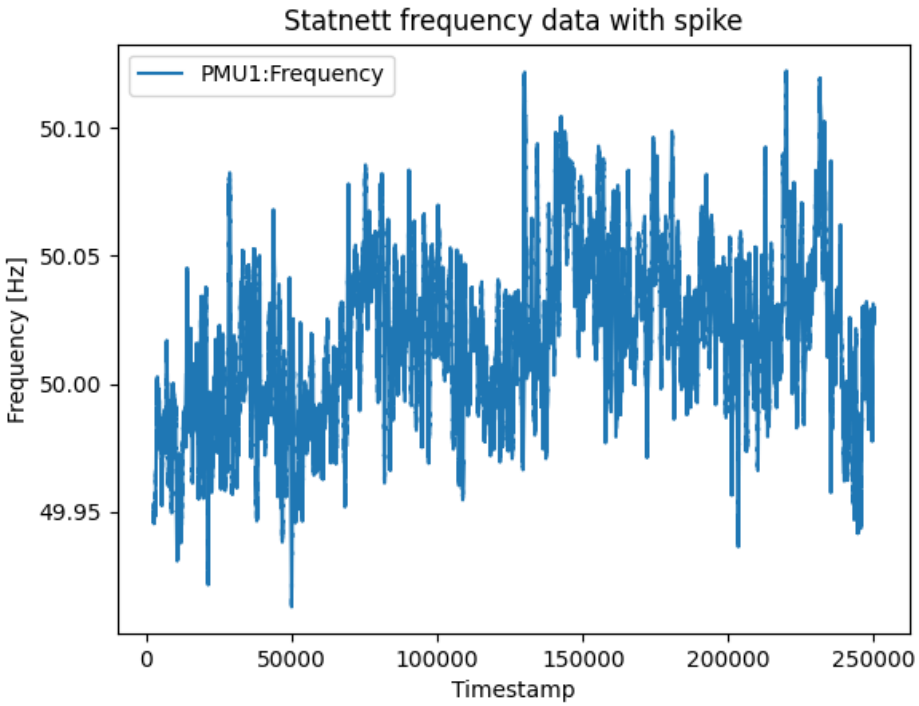


Figure 54: Statnett frequency data set with injected spike anomalies.

Figure 55 shows that the amplitudes of the spike anomalies vary between 0.01 and 0.02 Hz above or below the actual measured frequency. The short duration as well as the small amplitudes make these anomalies way within any operating limits.

The models fare similarly when detecting outliers, as the sudden change in amplitude is easily detected using prediction based models. Figure 56 shows the detection of all five spikes, however due to the model adapting to the increase or decrease in amplitude, some data points after the spikes are also labeled anomalies and as such increase the number of false positives.

Using the same test method as previous, Table 12 presents the recall, precision, F1-score and computational time accordingly. The numbers presented is the average of 15 tests run with 5 spikes injected into the data for each model with and without noise filtration beforehand.

Firsly, all models detected all anomalies throughout the 15 tests, as seen by the 100% recall. This proves that the models are excellent in detecting outliers and spike anomalies, as their values are far from the models predicted values. A large error in prediction to outlier results in a large error that is higher than the threshold value.

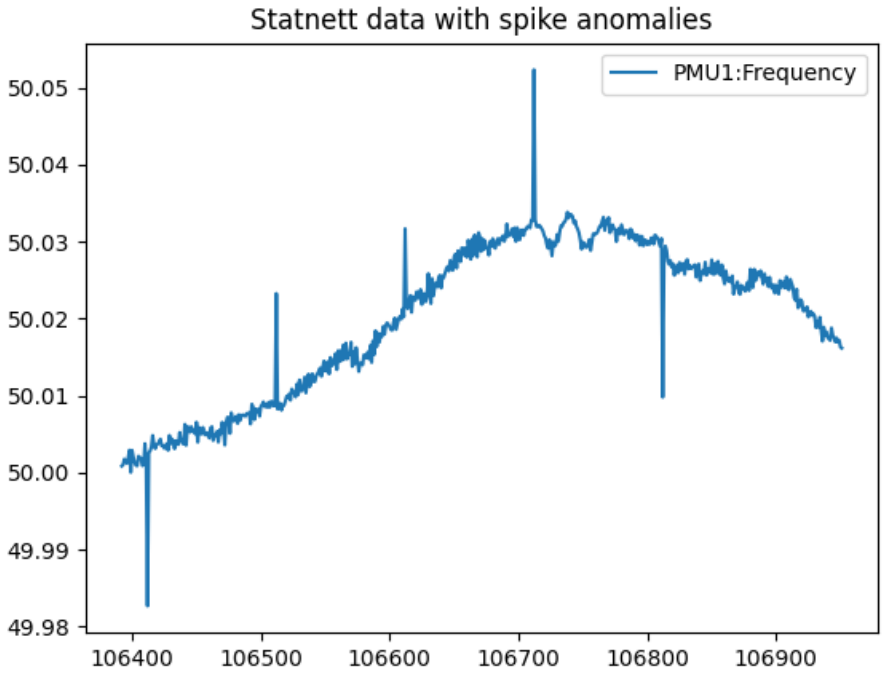


Figure 55: Enlarged data set with injected spike anomalies.

Table 12: Performance results for the different models using Statnett data on spike detection.

Model	**Noise filtration**	**Recall (%)**	**Precision (%)**	**F1-Score(%)**	**Computational time (ms/step)**
CNN	Yes	100	19.24	31.79	3
	No	100	28.04	42.91	3
LSTM	Yes	100	23.73	38.05	8
	No	100	34.67	50.81	8
Bi-LSTM	Yes	100	18.76	31.35	11
	No	100	24.2	38.82	11
C-LSTM	Yes	100	16.38	28.07	8
	No	100	**38.78**	**54.28**	8

In comparison to the noise tests, the precision and F1-score are lower in percentage by a great margin. The low amount of anomalies leads to any false positives having a greater impact on these scores. The false positives can be seen in Figure 56 and are located after the spikes occur. Because prediction models learn from the input data, the models tries to follow the spike anomaly value after it has occurred due to its large amplitude. This leads to a consequential error where the predictions for the subsequent data points are either below or above the normal operating values, depending on the anomaly amplitude. These consequential errors is what causes the false positives, where the error between predictions and real values are above the error threshold. The different models fare differently in how quickly they regain their normal prediction values corresponding to the real values.

Figure 57 shows test number 8 of 15 for the CNN model with noise filter active. The impulse response from the CNN model shows, as mentioned in the noise detection section, that it is highly affected by the spikes for a considerable amount of data points in some instances. Looking at

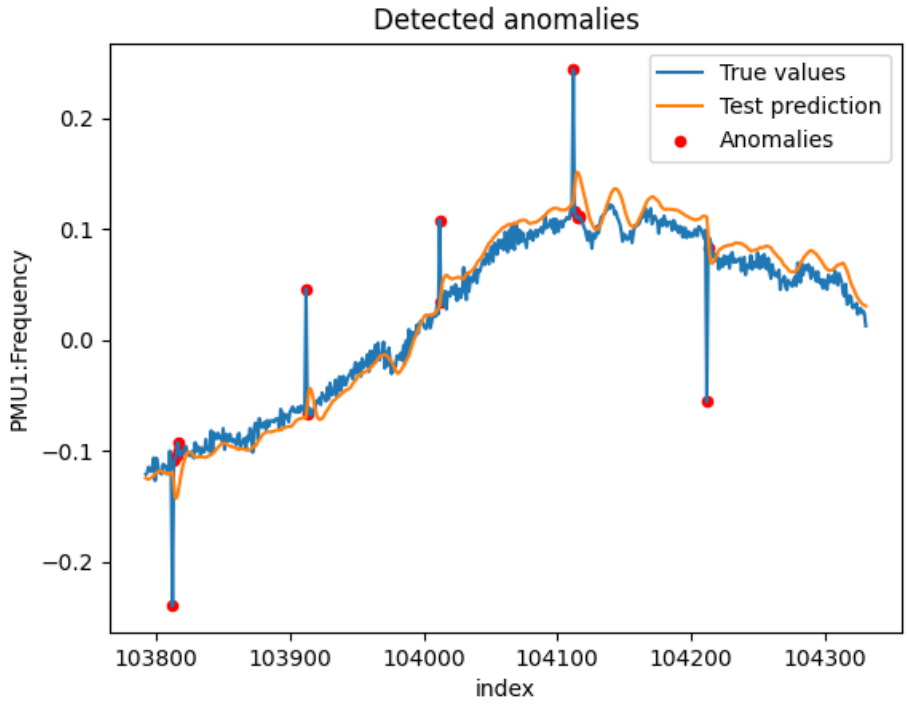


Figure 56: C-LSTM spike detection with noise filter.

the third spike from the left, the predictions are above the true values for a while after the spike, before it gradually converges on the true values. While the model gradually converges, oscillations are seen in the predictions. The fourth spike shows a different behaviour, here the predictions compensate for the spike anomaly in the opposite direction by predicting higher values than the true values, even though the anomaly has a value far lower than the true values.

Figure 58 shows test number 4 of 15 for the Bi-LSTM model with noise filter active. In comparison to Figure 57 using the CNN model, the Bi-LSTM model shows a different behaviour. The Bi-LSTM predictions rapidly detaches from the anomaly and predicts close to the true values again. The third spike from the left even shows the predictions over-shooting the true values as it tries to regain control. This shows tendencies for oscillations, just like the CNN model, but with a higher dampening.

For the purpose of anomaly detection, having a recall of 100% when detecting spike anomalies is more important than reducing false positives. The spike anomalies in themselves may be innocent on their own, but detecting and recording multiple spike anomalies over a time period can be beneficial. Multiple spike anomalies over a time period can mean faulty components in the measurement and data transmission systems. Detecting faulty components early can hinder cascading faults, and therefore hinder more expensive and system-affecting faults. The amount of false positives in close vicinity to the spike anomalies can also be favourable, as they assist in highlighting the data points, strengthening the anomalies' fault location and timestamp.

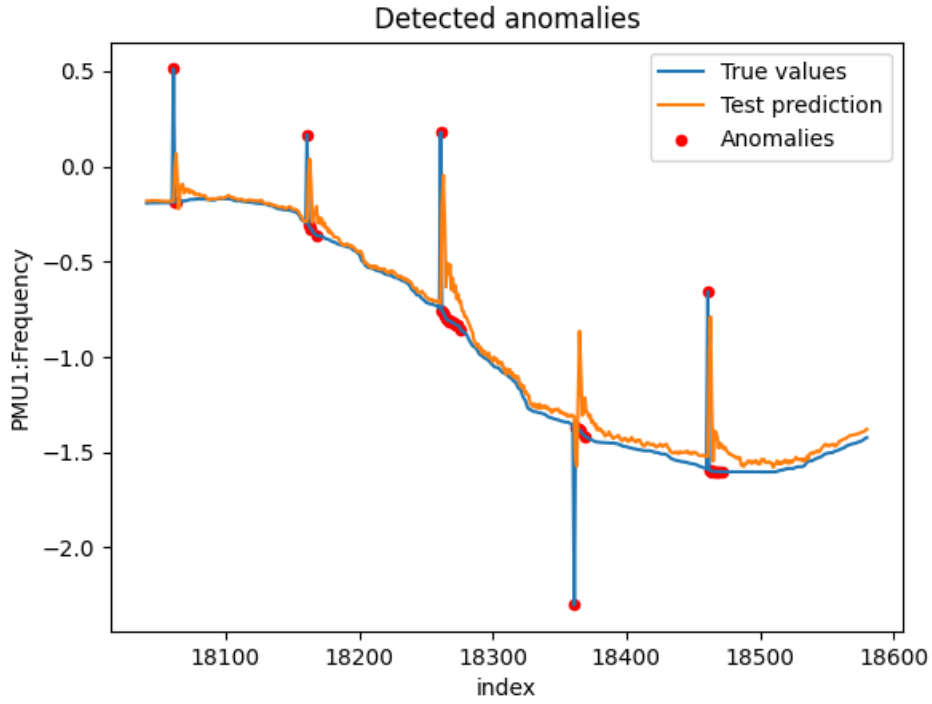


Figure 57: CNN spike detection with noise filter.

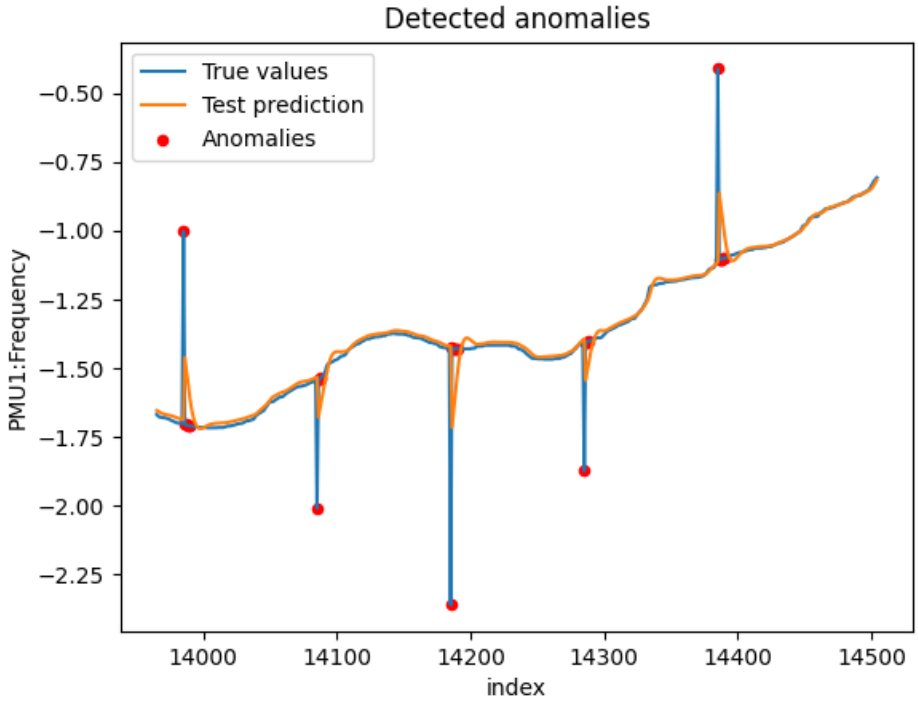


Figure 58: Bi-LSTM spike detection with noise filter.

4.3.3 Offset Anomaly Detection

Detecting offset anomalies is crucial for the operation of power systems, as mentioned earlier. Figure 60 shows the injected offset on a portion of the Statnett data. The offset was injected into the Statnett data to mimic a real offset. The comparison between the injected offset and a real offset can be seen by comparing Figure 59 and 60. The former being a real offset measured by the PMU located at the University of Texas-Pan American. The reason for the offset at UT-Pan Am is not known. A common reason for offsets similar to the one shown includes cyber attacks, SLCs and line faults. However, there is no reason to believe the offset shown in Figure 59 is due to a cyber attack, and more likely due to the other reasons mentioned.

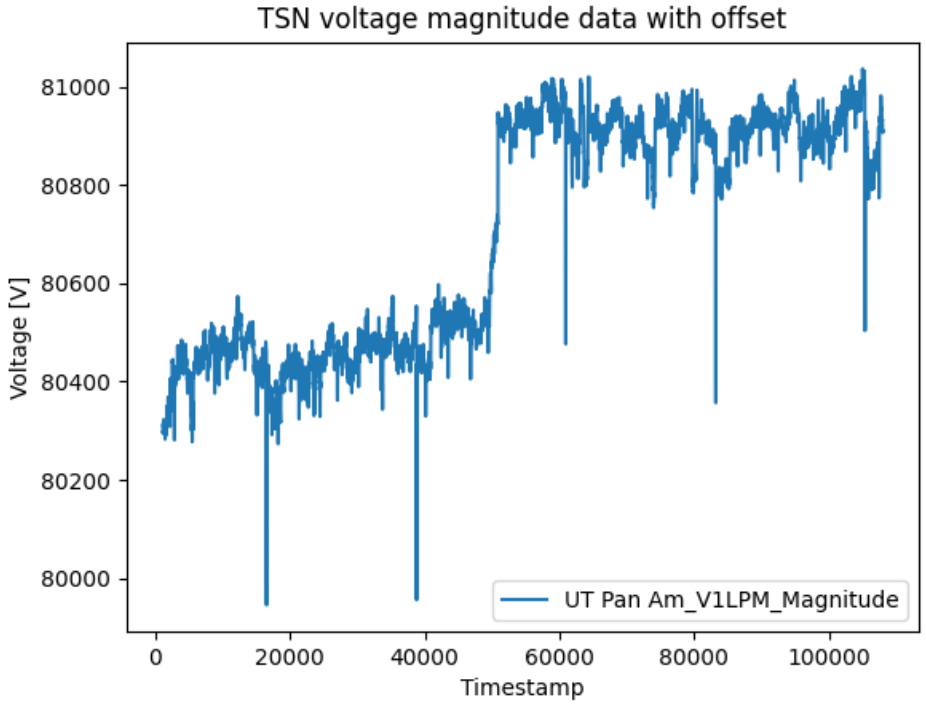


Figure 59: Real offset measured by the Pan-Am PMU in the TSN data set.

The models were tested on both the real offset data in Figure 59, and the injected offset in Figure 60. All models were able to detect both offsets, however in different fashions. Using metrics to display their efficacy is not beneficial in this case, as it is hard to define how many data points the offset consists of. Therefore using visual inspections of prediction and detection graphs is used. Figure 61 shows the Bi-LSTM model detecting the real offset in the TSN data, and Figure 62 displays all the models on injected offsets in the data from Statnett.

The real offset and the injected offset vary, in that the real offset has more of a curve after the spike, while the injected offsets aftermath is flatter. This leads to the prediction error being greater in the aftermath for the injected offset in Figure 62, which in turn results in more data points being labeled anomalies. This all stems from the oscillatory tendencies for the models after a larger shift in amplitude. The LSTM-based models' oscillations are similar to eachother, but with different dampening speeds. As seen in Figure 62d, the C-LSTM model dampens quicker than the pure LSTM or Bi-LSTM models. In contrast, the CNN model is much more chaotic and takes much longer before converging on the true values, this was also seen for the other anomaly types earlier.

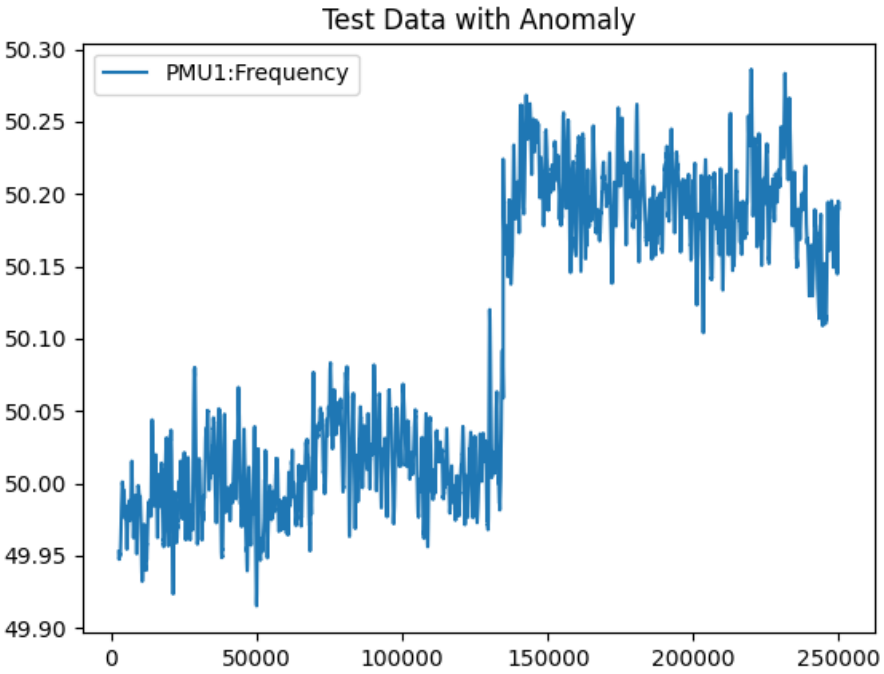


Figure 60: Statnett frequency data with offset injection.

The chaotic nature leads to an even greater amount of labeled anomalies after the offset occurred.

In the same way the spike anomalies led to some false positives in the aftermath of the anomaly, the same can be seen for the offset anomaly. The same argumentation can be used here, in that the amount of false positives, or a larger amount of labeled anomalies can lead to an offset anomaly being detected quicker, as it generates more emphasis on the given data.

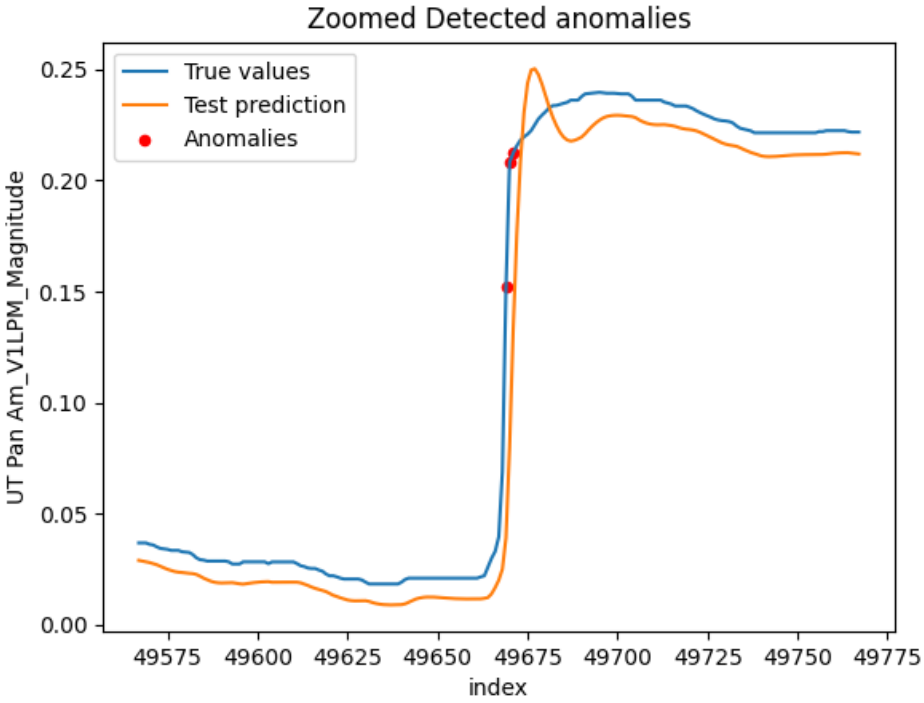
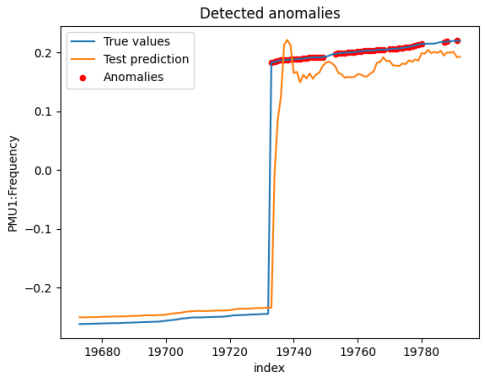
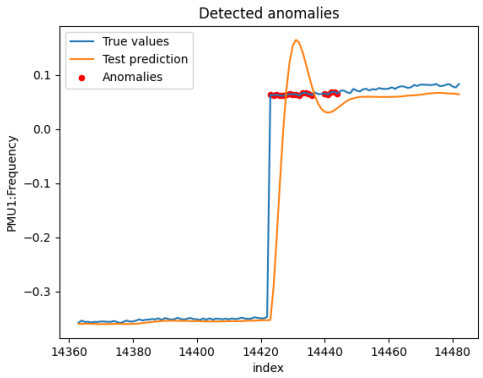


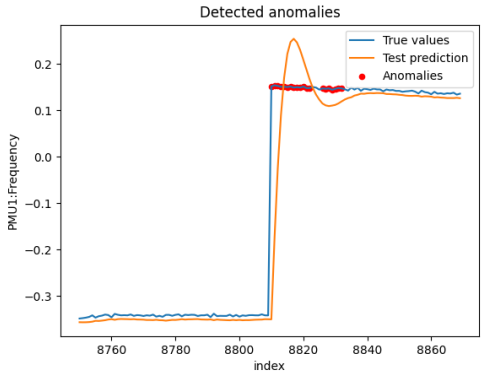
Figure 61: Bi-LSTM detection on the real offset in the Pan Am data.



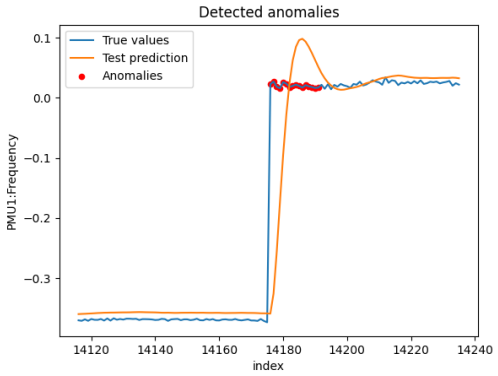
(a) CNN model.



(b) LSTM model.



(c) Bi-LSTM model.



(d) C-LSTM model.

Figure 62: Offset detection on Statnett data.

4.4 Detecting Real Faults

The model validation tests were run on the Statnett frequency data. Suprisingly, all of the models labeled the exact same faults. This might suggest that in real data application, without FDI and bad data, spending a lot of computer power on sophisticated and time consuming models might deem unnecessary. The models found four faults that are presented as an index number in Table 13. All faults lasted for 8 ms, or 2 measurements.

Table 13: The faults labeled by the four different models. The numbers represent the index of the faults.

	CNN	**LSTM**	**Bi-LSTM**	**C-LSTM**	Duration
First Fault	969238	969238	969238	969238	8ms
Second Fault	2648064	2648064	2648064	2648064	8ms
Third Fault	2648139	2648139	2648139	2648139	8ms
Fourth Fault	2962404	2962404	2962404	2962404	8ms

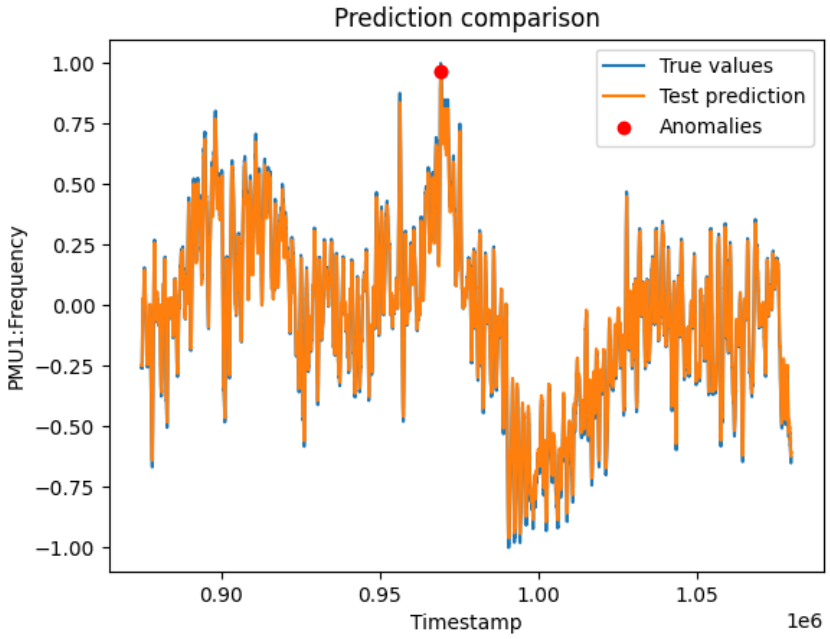


Figure 63: The first fault found by the models. This graph is the output from the Bi-LSTM models' test.

The first fault is presented in Figure 63. The graph is saved from the Bi-LSTM models prediction test. It contains a lot of data points and therefore it is difficult to see the differences between the models' prediction. Since they label the exact same faults it would be possible to conclude that all four models perform equally. However, as seen in other results the models do predict in different fashions, especially when dealing with FDI's.

The second and third fault are presented in Figure 64. There is a only a small time frame of 3 seconds between them and they represent noisy outliers in the data, before the data normalizes around 0 a short time after.

The fourth and last fault is presented in Figure 65. Similarly to the other faults it has a high value and appears in a noisy area before the data drifts back and oscillates around 0.

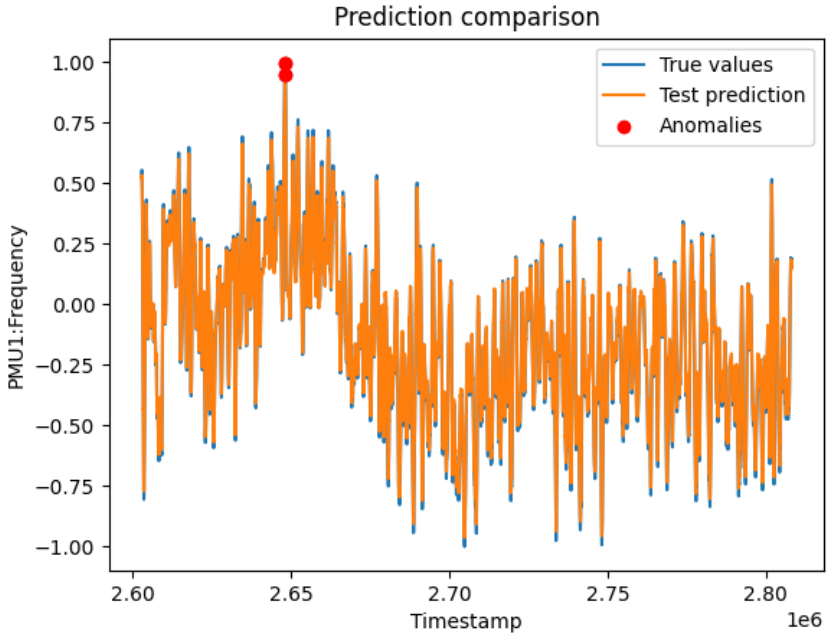


Figure 64: Fault number two and three found by the models. This graph is the output from the Bi-LSTM models' test.

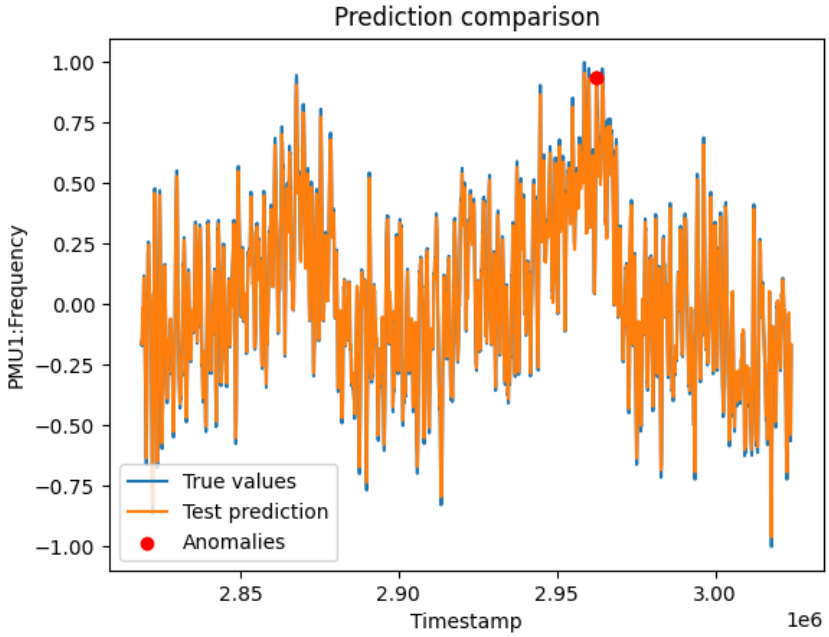


Figure 65: The fourth fault found by the models. This graph is the output from the Bi-LSTM models' test.

5 Conclusion and Discussion

Throughout the process of the model testing, two data sets with with seven different signals have been evaluated for anomaly detection. Four different ML models were trained on PMU data using optimal hyperparameters; namely CNN, LSTM, C-LSTM and Bi-LSTM. The trained models were saved for later testing on different PMU data from two sources. The research and data published by NREL and the University of Texas has been used to verify the validity of the models. Tests on the data provided by Statnett has shown that the models work on different data sets without further training, and that they therefore can be used to find both bad data, physical faults and FDI attacks.

When adding gaussian noise to the data sets we discovered that Bi-LSTM performed best on the TSN data with an F1-score of 95% when the data is subject to noise filtration. Meanwhile, the C-LSTM model performed best on the Statnett data with an F1-score of 97%, also when implementing noise filtration. Other type of bad data anomalies, such as single spikes and offsets, were also discovered by the models. All the injected spikes were found by all the models, while the C-LSTM model labeled the least amount of false positives, without noise filtration active. Both the real offset in the TSN data and the injected offset was discovered by all models, however the CNN model was struggling to converge on the data quickly after the offset. The drift was not discovered by any of the models in any of the data sets. In addition, some general characteristics of the models have been noted. Some examples are the CNN-models lack of ability to converge quickly after injected noise and spikes, or the C-LSTM and Bi-LSTM models ability to follow the test data tightly.

We have also learned the importance of run time and that there is a big difference between the models on this account. The CNN model performs worst on most parameters, but it is extremely quick with a step time of 3 ms. The C-LSTM and LSTM models, who both perform well on many tests and even best on some, has a step time of up to 8ms; which is an 167% increase in step time compared to the CNN model. The Bi-LSTM model is even slower with a step time of up to 11ms which corresponds to an 257% increase. Since data and computational power is a limited commodity, the run time is an important factor when deciding what models to use. Therefore, when working with implementing the models in industry applications, finding a model which performs well on a low run time is of a great importance.

When not considering the computational load, the hybrid models C-LSTM and Bi-LSTM, proved more effective than the pure models, CNN and LSTM, for detecting the FDI attacks tested in regards to the metrics used in this thesis. The value of detecting all of the anomalies injected has been an important factor for these tests. However, there has to be a balance between the amount of detected anomalies and amount of falsely labeled anomalies, the false positives. If the amount of false positives becomes too high, the true positives might be missed, as there will be too many other anomalies detected that need to be checked. We believe the models tested had a satisfactory balance between the amount of true and false positives.

While testing for noise all models proved effective in both high amounts of true positives, while keeping the false positives low. This can not be said for the spike and offset anomaly tests. Here the amount of false positives led to poor precision and F1-scores. However, in practical and real life anomaly detection use cases, these false positives might prove useful as a way to emphasize the locations for spikes and offsets. As seen in Figure 61, the real offset in the TSN data was detected

by only 3 data points. A greater number of labeled data points might prove more useful for emphasis' purposes. Therefore the number of false positives in spike and offset anomaly detection is deemed unproblematic.

All four models proved effective in detecting FDI attacks that mimic noise, spikes and offsets. Some improvements still needs to be considered before implementation. None of the models managed to detect the injected drift anomalies. Drift anomalies, as mentioned earlier, is a great potential threat to the stability of the power system if not detected and dealt with properly. Therefore different methods should also be implemented to detect these anomalies. Testing, improving and implementing the models that fit best to the SO's needs, would be a small step towards a system where computers help them to quickly detect anomalies when they occur.

The lack of detection for drift anomalies show some of the limitations for the models and their anomaly detection abilities. The models were only trained and tested on one feature at a time. An implementation that uses all measurements the PMUs offer might have improved the ML models, as there is more information to use. However, this was not possible with our time and knowledge limitations. At the same time, the increased computational load that follows the use of more features has to be taken into account.

The model validation step could also have been improved. While the validation using the NREL model together with the TSN data gives a clear indication that the models detect real anomalies using PMU data, no validation was done on the Statnett data. While it can be argued that the validation step using the TSN data is enough to assume the models work as intended, a thorough validation step using the Statnett data would be beneficial. The detected anomalies presented in Table 13 are not cross-checked with Statnett and there is no way for us to know if they are real anomalies or not, as this is classified information. Nonetheless, the results of this test proved useful as all four models labeled the exact same faults. Seeing that the CNN model found the same faults as the Bi-LSTM model, might indicate that running computationally heavy models to find areas of concern is unnecessary.

In spite of the future power systems's increasing cyber-physical threats, e.g. FDI's and volatile power generation, the results of this thesis show that there exist many opportunities to better the resilience of our future power supply. As discussed in section 2, the importance of this resilience is invaluable. The possible Economical, social and political consequences of an uncertain power supply are intolerable. Therefore, solving these challenges are crucial. Combining PMU data with unsupervised ML methods has the potential to alarm SO's of occuring faults as they happen, giving them the opportunity to act quickly. Therefore, we would lastly like to recommend readers, companies and governments to invest time and resources in the development of these and other types of solutions, that might help secure a stable power flow in the future.

6 Further Work

Although this thesis presents several results, research on this field is still in the early stage and there are still many things to investigate. The main goal with our research is to take a step towards a method that can be implemented in industry applications. Nonetheless, there is a lot of work between the tests that are presented in this thesis and real-time anomaly detection software for SOs.

Firstly, researchers interested in further experiments could test the data on models with different parameters, or even on new models altogether. There are many alternatives in the world of ML. Finding models that perform well on different data sets and with a low run time is a priority. As seen in the results, even though the CNN model is not the most precise, it does manage find most areas of concern on a very low run time, compared to the other models. Since it predicts the data less precise than the other models, it has a tendency to label more anomalies. Therefore, a possible implementation strategy might be to hybridize the models such that the CNN model rapidly finds the data that needs to be further inspected, while the heavier models conduct more precise anomaly detection schemes.

Secondly, the models work well when trying to detect bad data in certain type of FDIs where the PMUs are presumably measuring drastic spikes, noise or offsets. However, if a slow drift is injected the models predict the data too well - thus, not labelling the anomaly. Therefore, this is an area of improvement where a solution include experimenting with different thresholding techniques and statistical methods, like the x-sigma technique presented in [6].

Thirdly, for the real-time application the normalization technique presented in this thesis needs to be improved. Keeping in mind that a mixture of different input data needs to be considered equally by the models, and be kept within the range of [-1, 1].

Bibliography

[1] J. M. Barrett, 'Challenges And Requirements For Tomorrow's Electrical Power Grid', *Lexington Institute*, 2016.

[2] E. Commission, *2050 long-term strategy*. [Online]. Available: https://ec.europa.eu/clima/eu-action/climate-strategies-targets/2050-long-term-strategy_en.

[3] L. Zanni, 'Power-System State Estimation based on PMUs Static and Dynamic Approaches - from Theory to Real Implementation', p. 189, 2017. DOI: 10.5075/epfl-thesis-7665. [Online]. Available: http://infoscience.epfl.ch/record/228451.

[4] A. Karpilow, R. Cherkaoui, S. D'Arco and T. D. Duong, 'Detection of Bad PMU Data using Machine Learning Techniques', in *2020 IEEE Power Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, 2020, pp. 1–5. DOI: 10.1109/ISGT45199.2020.9087782.

[5] H. Vassdal and Y. Regev, 'Machine Learning Methods for Anomaly Detection on Phasor Measurement Unit Data', Trondheim, 2021.

[6] I. Tinawi, 'Machine Learning for Time Series Anomaly Detection', *Massachusetts Institute of Technology*, pp. 1–55, 2019. DOI: https://dspace.mit.edu/bitstream/handle/1721.1/123129/1128282917-MIT.pdf?sequence=1{\&}isAllowed=y.

[7] Umit Cali, Murat Kuzlu, Manisa Pipattanasomporn, James Kempf and Linquan Bai, *Digitalization of Power Markets and Systems Using Energy Informatics*, 1st ed. Springer International Publishing, 2021.

[8] Energifakta Norge, *STRØMNETTET ER VIKTIG INFRASTRUKTUR*, Apr. 2022. [Online]. Available: https://energifaktanorge.no/norsk-energiforsyning/kraftnett/.

[9] Norges vassdrags- og energidirektorat, *Nett*, Apr. 2022. [Online]. Available: https://www.nve.no/energi/energisystem/nett/.

[10] Circuit Globe, *Power system*, Apr. 2022. [Online]. Available: https://circuitglobe.com/power-system.html.

[11] A. Biancardi, M. Di Castelnuovo and I. Staffell, 'A framework to evaluate how European Transmission System Operators approach innovation', *Energy Policy*, vol. 158, p. 112 555, 2021, ISSN: 0301-4215. DOI: https://doi.org/10.1016/j.enpol.2021.112555. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0301421521004250.

[12] Statnett, *Eierskap og vedtekter*. [Online]. Available: https://www.statnett.no/om-statnett/eierskap-og-vedtekter/.

[13] T. B. I. Arun G. PHADKE, 'Phasor measurement units, WAMS, and their applications in protection and control of power systems', pp. 1–11, 2018.

[14] J. D. L. Ree, V. Centeno J. S. Thorp and A. G. Phadke, 'Synchronized phasor measurement applications in power systems', *IEEE Trans. Smart Grid*, no. 1, pp. 20–27, 2010. DOI: 10.1109/TSG.2010.2044815.

[15] K. U. Chandrarathna, *Historical phasor measurement unit (PMU) data analysis by using machine learning techniques for anomaly detection in power systems*, 2019. [Online]. Available: https://ttu-ir.tdl.org/bitstream/handle/2346/85938/CHANDRARATHNA-THESIS-2019.pdf?sequence=1&isAllowed=y.

[16] B. Vicol, 'MODERN TECHNOLOGIES FOR POWER SYSTEMS MONITORING', Jan. 2013.

[17] B. Pinte, M. Quinlan and K. Reinhard, 'Low voltage micro-phasor measurement unit (μPMU)', in *2015 IEEE Power and Energy Conference at Illinois (PECI)*, 2015, pp. 1–4. DOI: 10.1109/PECI.2015.7064888.

[18] J. Dhawal, 'Unsupervised learning for critical event detection using historical PMU data', 2020.

[19] A. G. Phadke and J. S. Thorp, *Synchronized phasor measurements and their applications*. Springer, 2011, pp. 3–79.

[20] M. A. Barry, D. Gault, G. Bolt, A. McEwan, M. D. Filipović and G. L. White, 'Verifying Timestamps of Occultation Observation Systems', *Publications of the Astronomical Society of Australia*, vol. 32, 2015. DOI: 10.1017/pasa.2015.15.

[21] T. Glla, *Selecting Phase-Locked Oscillators for Frequency Synthesis — 2019-09-05 — Microwave Journal*, 2021. [Online]. Available: https://www.microwavejournal.com/articles/32771-selecting-phase-locked-oscillators-for-frequency-synthesis.

[22] K. Zhu, S. Rahimi, L. Nordström and B. Zhang, 'Design phasor data concentrator as adaptive delay buffer for wide-area damping control', *Electric Power Systems Research*, vol. 127, pp. 22–31, Jun. 2015, ISSN: 03787796. DOI: 10.1016/j.epsr.2015.05.002.

[23] I. D. Melo and M. P. Antunes, 'Bad data correction in harmonic state estimation for power distribution systems: an approach based on generalised pattern search algorithm', *Electric Power Systems Research*, vol. 204, Mar. 2022, ISSN: 03787796. DOI: 10.1016/j.epsr.2021.107684.

[24] A. E. Elhabashy, L. J. Wells and J. A. Camelio, 'Cyber-physical security research efforts in manufacturing - A literature review', vol. 34, Elsevier B.V., 2019, pp. 921–931. DOI: 10.1016/j.promfg.2019.06.115.

[25] A. Karpilow, 'Detection of Bad PMU Data using Machine Learning Techniques', Tech. Rep., 2019.

[26] 'IEEE Recommended Practice for Monitoring Electric Power Quality', *IEEE Std 1159-1995*, pp. 1–80, 1995. DOI: 10.1109/IEEESTD.1995.79050.

[27] M. Mirošević and Z. Maljković, 'Effect of sudden change load on isolated electrical grid', in *Electrical Systems for Aircraft, Railway and Ship Propulsion, ESARS*, vol. 2015-May, IEEE Computer Society, May 2015, ISBN: 9781479974009. DOI: 10.1109/ESARS.2015.7101465.

[28] C. Collados-Rodriguez, M. Cheah-Mane, E. Prieto-Araujo and O. Gomis-Bellmunt, 'Stability and operation limits of power systems with high penetration of power electronics', *International Journal of Electrical Power and Energy Systems*, vol. 138, Jun. 2022, ISSN: 01420615. DOI: 10.1016/j.ijepes.2021.107728.

[29] E. M. Lourenco, A. J. A. Simoes Costa, K. A. Clements and R. A. Cernev, 'A Topology Error Identification Method Directly Based on Collinearity Tests', *IEEE Transactions on Power Systems*, vol. 21, no. 4, pp. 1920–1929, 2006. DOI: 10.1109/TPWRS.2006.881113.

[30] K. A. Clements and P. W. Davis, 'Multiple Bad Data Detectability and Identifiability, A Geometric Approach', *IEEE Power Engineering Review*, vol. PER-6, no. 7, p. 73, 1986. DOI: 10.1109/MPER.1986.5527890.

[31] R. L. Lugtu, D. F. Hackett, K. C. Liu and D. D. Might, 'Power system state estimation: Detection of topological errors', *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-99, no. 6, pp. 2406–2412, 1980, ISSN: 00189510. DOI: 10.1109/TPAS.1980.319807.

[32] C. Klauber and H. Zhu, 'Power network topology control for mitigating the effects of geomagnetically induced currents', in *Conference Record - Asilomar Conference on Signals, Systems and Computers*, IEEE Computer Society, Mar. 2017, pp. 313–317, ISBN: 9781538639542. DOI: 10.1109/ACSSC.2016.7869049.

[33] H. M. Long and J. C. Smith, 'Load management on the electric power system', in *1979 18th IEEE Conference on Decision and Control including the Symposium on Adaptive Processes*, vol. 2, 1979, p. 273. DOI: 10.1109/CDC.1979.270179.

[34] M. N. H. Shazon, Nahid-Al-Masood and H. M. Ahmed, 'Modelling and utilisation of frequency responsive TCSC for enhancing the frequency response of a low inertia grid', *Energy Reports*, vol. 8, pp. 6945–6959, Nov. 2022, ISSN: 23524847. DOI: 10.1016/j.egyr.2022.05.101. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S2352484722009507.

[35] J. Pardha Saradhi, R. Srinivasarao and V. Ganesh, 'Wavelet based multiresolution analysis of a 5-Bus system in the presence SVC controller under fault and sudden load conditions', *Materials Today: Proceedings*, 2020, ISSN: 2214-7853. DOI: https://doi.org/10.1016/j.matpr.2020.10.852. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2214785320384893.

[36] E. Styvaktakis, I. Gu and M. Bollen, 'Event-based transient categorization and analysis in electric power systems', vol. 5, Jan. 2003, 4176–4183 vol.5, ISBN: 0-7803-7952-7. DOI: 10.1109/ICSMC.2003.1245641.

[37] Wikipedia, *Voltage spike*, 2022. [Online]. Available: https://en.wikipedia.org/wiki/Voltage_spike.

[38] Andy Sagl, *Power quality: some fundamentals*, 2016. [Online]. Available: https://megger.com/electrical-tester/november-2016/power-quality-some-fundatmentals.

[39] S. Liu, C. H. Singer and R. A. Dougal, 'Power anomaly effects and costs in low-voltage mobile power systems', *IEEE Transactions on Aerospace and Electronic Systems*, vol. 42, no. 2, pp. 612–624, 2006. DOI: 10.1109/TAES.2006.1642576.

[40] J. C. Gomez, M. M. Morcos, C. A. Reineri and G. N. Campetelli, 'Behavior of induction motor due to voltage sags and short interruptions', *IEEE Transactions on Power Delivery*, vol. 17, no. 2, pp. 434–440, 2002. DOI: 10.1109/61.997914.

[41] J. C. Gomez and M. M. Morcos, 'Voltage sag and recovery time in repetitive events', *IEEE Transactions on Power Delivery*, vol. 17, no. 4, pp. 1037–1043, 2002. DOI: 10.1109/TPWRD.2002.803840.

[42] yepyep, *ANSI C84.1 ELECTRIC POWER SYSTEMS AND EQUIPMENT - VOLTAGE RANGES*. [Online]. Available: http://www.powerqualityworld.com/2011/04/ansi-c84-1-voltage-ratings-60-hertz.html.

[43] S. Homan and S. Brown, 'An analysis of frequency events in Great Britain', *Energy Reports*, vol. 6, Jan. 2020. DOI: 10.1016/j.egyr.2020.02.028.

[44] P. Risbud, N. Gatsis and A. Taha, 'Vulnerability Analysis of Smart Grids to GPS Spoofing', *IEEE Transactions on Smart Grid*, vol. 10, no. 4, pp. 3535–3548, 2019. DOI: 10.1109/TSG.2018.2830118.

[45] A. Jafarnia-Jahromi, A. Broumandan, J. Nielsen and G. Lachapelle, 'GPS Vulnerability to Spoofing Threats and a Review of Antispoofing Techniques', *International Journal of Navigation and Observation*, vol. 2012, Jan. 2012. DOI: 10.1155/2012/127072.

[46] B. Ramasubramanian, M. A. Rajan, M. Girish Chandra, R. Cleaveland and S. I. Marcus, 'Resilience to denial-of-service and integrity attacks: A structured systems approach', *European Journal of Control*, 2021, ISSN: 0947-3580. DOI: https://doi.org/10.1016/j.ejcon.2021.09.005. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0947358021001114.

[47] A. Walker, E. R. Cox, J. N. Loughhead and J. D. Roberts, 'Counting the cost: the economic and social costs of electricity shortfalls in the UK - A report for the Council for Science and Technology', 2014.

[48] D. C. Lineweber and S. McNulty, 'The Cost of Power Disturbances to Industrial & Digital Economy Companies', 2001, ES1–ES3.

[49] Smoc, 'February 2021 Winter Storm-Related Deaths-Texas', Tech. Rep., 2021.

[50] CITY OF AUSTIN and TRAVIS COUNTY, '2021 WINTER STORM URI AFTER-ACTION REVIEW FINDINGS REPORT', Tech. Rep., 2021.

[51] M. Impelli, *Satellite photos show extent of Texas power outages from space*, 2021. [Online]. Available: https://www.newsweek.com/satellite-photos-show-extent-texas-power-outages-space-1569942.

[52] J. H. Eto, J. G. Koomey, B. Lehman *et al.*, 'Scoping Study on Trends in the Economic Value of Electricity Reliability to the U.S. Economy', Tech. Rep., Jan. 2001, p. 148.

[53] K. Lien, *Mørke utsikter i Libanon*, 2021. [Online]. Available: https://www.vg.no/nyheter/utenriks/i/Jxv6o6/moerke-utsikter.

[54] Wikipedia contributors, *Three-Day Week — Wikipedia, The Free Encyclopedia*, 2021. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Three-Day_Week&oldid=1051180900.

[55] M. Mullane, *The five pillars of cyber security*, 2019. [Online]. Available: https://medium.com/swlh/the-five-pillars-of-cyber-security-d247cd2e49cb.

[56] E. Baron-Prada, E. Osorio and E. Mojica-Nava, 'Resilient transactive control in microgrids under dynamic load alteringphadke_thorp_2011 attacks', in *2017 IEEE 3rd Colombian Conference on Automatic Control (CCAC)*, 2017, pp. 1–5. DOI: 10.1109/CCAC.2017.8276400.

[57] Srikumar M.S., T. Ananthapadmanbha, F. Z. Khan and Girish V., 'Line Outage Detection Using Phasor Measurement Units', *Procedia Technology*, vol. 21, pp. 88–95, 2015, ISSN: 2212-0173. DOI: https://doi.org/10.1016/j.protcy.2015.10.014. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S221201731500242X.

[58] J. Chu, X. Liu, Z. Zhang, Y. Zhang and M. He, 'A novel method overcomeing overfitting of artificial neural network for accurate prediction: Application on thermophysical property of natural gas', *Case Studies in Thermal Engineering*, vol. 28, p. 101 406, 2021, ISSN: 2214-157X. DOI: https://doi.org/10.1016/j.csite.2021.101406. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2214157X21005694.

[59] L. Zhou, B. Wang, Z. Wang, F. Wang and M. Yang, 'Seasonal classification and RBF adaptive weight based parallel combined method for day-ahead electricity price forecasting', in *2018 IEEE Power Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, 2018, pp. 1–5. DOI: 10.1109/ISGT.2018.8403372.

[60] J. C. S. Souza, A. M. da Silva and A. P. de Silva, 'Online topology determination and bad data suppression in power system operation using artificial neural networks', *IEEE Transactions on Power Systems*, vol. 13, no. 3, pp. 796–803, 1998. DOI: 10.1109/59.708645.

[61] S. J. Farlow, *Self-Organizing Methods in Modeling: GMDH Type Algorithms*, ser. Statistics: A Series of Textbooks and Monographs. Taylor & Francis, 1984, ISBN: 9780824771614. [Online]. Available: https://books.google.no/books?id=G2_4Eu6hdQcC.

[62] B. Lindemann, B. Maschler, N. Sahlab and M. Weyrich, 'A survey on anomaly detection for technical systems using LSTM networks', *Computers in Industry*, vol. 131, p. 103 498, 2021, ISSN: 0166-3615. DOI: https://doi.org/10.1016/j.compind.2021.103498. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0166361521001056.

[63] N. Shreyas, M. Venkatraman, S. Malini and S. Chandrakala, 'Chapter 7 - Trends of Sound Event Recognition in Audio Surveillance: A Recent Review and Study', in *The Cognitive Approach in Cloud Computing and Internet of Things Technologies for Surveillance Tracking Systems*, ser. Intelligent Data-Centric Systems, D. Peter, A. H. Alavi, B. Javadi and S. L. Fernandes, Eds., Academic Press, 2020, pp. 95–106, ISBN: 978-0-12-816385-6. DOI: https://doi.org/10.1016/B978-0-12-816385-6.00007-6. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780128163856000076.

[64] P. Malhotra, L. Vig, G. Shroff and P. Agarwal, 'Long Short Term Memory Networks for Anomaly Detection in Time Series', Jan. 2015.

[65] G. Raman MR, N. Somu and A. P. Mathur, 'A multilayer perceptron model for anomaly detection in water treatment plants', *International Journal of Critical Infrastructure Protection*, vol. 31, p. 100 393, 2020, ISSN: 1874-5482. DOI: https://doi.org/10.1016/j.ijcip.2020.100393. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1874548220300573.

[66] A. A. Khan, O. A. Beg, M. Alamaniotis and S. Ahmed, 'Intelligent anomaly identification in cyber-physical inverter-based systems', *Electric Power Systems Research*, vol. 193, p. 107 024, 2021, ISSN: 0378-7796. DOI: https://doi.org/10.1016/j.epsr.2021.107024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0378779621000043.

[67] Y. Ma, R. Huang, M. Yan, G. Li and T. Wang, 'Attention-based Local Mean ¡math altimg="si25.svg" display="inline" id="d1e1490"¿ ¡mi¿K¡/mi¿ ¡/math¿ -Nearest Centroid Neighbor Classifier', *Expert Systems with Applications*, vol. 201, p. 117 159, Sep. 2022, ISSN: 09574174. DOI: 10.1016/j.eswa.2022.117159. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0957417422005504.

[68] S. Tajmouati, B. E. Wahbi, A. Bedoui, A. Abarda and M. Dakkoun, *Applying k-nearest neighbors to time series forecasting : two new approaches*, 2021.

[69] D. M. Atallah, M. Badawy and A. El-Sayed, 'Intelligent feature selection with modified K-nearest neighbor for kidney transplantation prediction', *SN Applied Sciences*, vol. 1, no. 10, Oct. 2019, ISSN: 25233971. DOI: 10.1007/s42452-019-1329-z.

[70] Wikipedia contributors, *Pearson correlation coefficient — Wikipedia, The Free Encyclopedia*, 2021. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Pearson_correlation_coefficient&oldid=1052186484.

[71] N. Singh, *MS Windows NT Kernel Description*, 2020. [Online]. Available: https://ai.plainenglish.io/what-is-k-means-clustering-3060791cb589.

[72] A. L. Amutha, R. Annie Uthra, J. Preetha Roselyn and R. Golda Brunet, 'Anomaly detection in multivariate streaming PMU data using density estimation technique in wide area monitoring system', *Expert Systems with Applications*, vol. 175, p. 114 865, 2021, ISSN: 0957-4174. DOI: https://doi.org/10.1016/j.eswa.2021.114865. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417421003067.

[73] A. Allen, M. Singh, E. Muljadi and S. Santoso, 'PMU Data Event Detection: A User Guide for Power Engineers', Jan. 2014. DOI: 10.2172/1160181. [Online]. Available: https://www.osti.gov/biblio/1160181.

[74] D. Osipov and J. H. Chow, 'PMU Missing Data Recovery Using Tensor Decomposition', *IEEE Transactions on Power Systems*, vol. 35, no. 6, pp. 4554–4563, 2020. DOI: 10.1109/TPWRS.2020.2991886.

[75] M. Brown, M. Biswal, S. Brahma, S. J. Ranade and H. Cao, 'Characterizing and quantifying noise in PMU data', in *2016 IEEE Power and Energy Society General Meeting (PESGM)*, 2016, pp. 1–5. DOI: 10.1109/PESGM.2016.7741972.

[76] Electrical Academia, *Per Unit Calculation — Per Unit System Examples*, 2022. [Online]. Available: https://electricalacademia.com/electric-power/per-unit-calculation-per-unit-system-examples/.

[77] S. Nayak, B. Misra and D. H. Behera, 'Impact of Data Normalization on Stock Index Forecasting', *International Journal of Computer Information Systems and Industrial Management Applications*, vol. 6, pp. 357–369, Jan. 2014.

[78] S. Bhanja and A. Das, 'Impact of Data Normalization on Deep Neural Network for Time Series Forecasting', Jan. 2018.

[79] D. into Deep Learning, *9.2. Long Short-Term Memory (LSTM)*, 2021. [Online]. Available: https://d2l.ai/chapter_recurrent-modern/lstm.html.

[80] Y. LeCun, L. Bottou, Y. Bengio and P. Haffner, 'Gradient-based learning applied to document recognition', *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2323, 1998, ISSN: 00189219. DOI: 10.1109/5.726791.

[81] W. H. Chung, Y. H. Gu and S. J. Yoo, 'District heater load forecasting based on machine learning and parallel CNN-LSTM attention', *Energy*, vol. 246, May 2022, ISSN: 03605442. DOI: 10.1016/j.energy.2022.123350.

[82] Y. Jing, L. Zhang, W. Hao and L. Huang, 'Numerical study of a CNN-based model for regional wave prediction', *Ocean Engineering*, vol. 255, p. 111 400, Jul. 2022, ISSN: 00298018. DOI: 10.1016/j.oceaneng.2022.111400. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0029801822007843.

[83] S. Sharma and R. Mehra, 'Implications of Pooling Strategies in convolutional neural networks: A Deep Insight', *Foundations of Computing and Decision Sciences*, vol. 44, no. 3, pp. 303–330, Sep. 2019, ISSN: 23003405. DOI: 10.2478/fcds-2019-0016.

[84] Y. Lei, X. Chen, M. Min and Y. Xie, 'A semi-supervised Laplacian extreme learning machine and feature fusion with CNN for industrial superheat identification', *Neurocomputing*, vol. 381, pp. 186–195, Mar. 2020, ISSN: 18728286. DOI: 10.1016/j.neucom.2019.11.012.

[85] C. Zhou, C. Sun, Z. Liu and F. C. M. Lau, 'A C-LSTM Neural Network for Text Classification', Nov. 2015. [Online]. Available: http://arxiv.org/abs/1511.08630.

[86] H. Sharadga, S. Hajimirza and R. S. Balog, 'Time series forecasting of solar power generation for large-scale photovoltaic plants', *Renewable Energy*, vol. 150, pp. 797–807, May 2020, ISSN: 18790682. DOI: 10.1016/j.renene.2019.12.131.

[87] F. Pukelsheim, 'The Three Sigma Rule', *The American Statistician*, vol. 48, no. 2, pp. 88–91, 1994, ISSN: 00031305. [Online]. Available: http://www.jstor.org/stable/2684253.

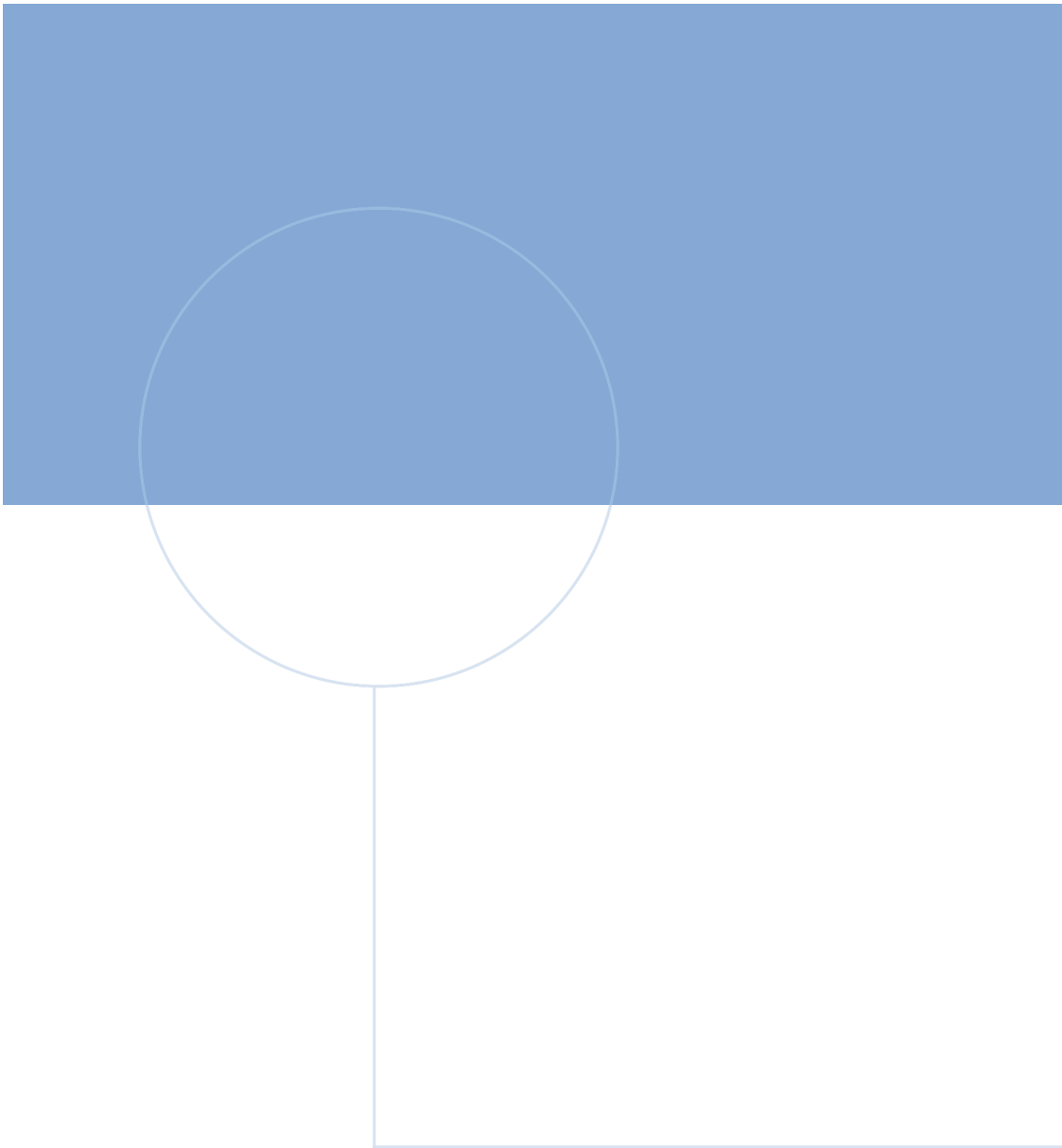
[88] K. Hundman, V. Constantinou, C. Laporte, I. Colwell and T. Soderstrom, 'Detecting Spacecraft Anomalies Using LSTMs and Nonparametric Dynamic Thresholding', in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, ser. KDD '18, New York, NY, USA: Association for Computing Machinery, 2018, pp. 387–395, ISBN: 9781450355520. DOI: 10.1145/3219819.3219845. [Online]. Available: https://doi.org/10.1145/3219819.3219845.

[89] Talos Contributors, *Talos*, 2022. [Online]. Available: https://autonomio.github.io/talos/#/.

[90] J. Yang, 'A Controllable False Data Injection Attack for a Cyber Physical System', *IEEE Access*, vol. 9, pp. 6721–6728, 2021, ISSN: 21693536. DOI: 10.1109/ACCESS.2021.3049228.

[91] NumPy, *NumPy*, 2022. [Online]. Available: https://numpy.org/.

[92] Z. DeVries, E. Locke, M. Hoda *et al.*, 'Using a national surgical database to predict complications following posterior lumbar surgery and comparing the area under the curve and F1-score for the assessment of prognostic capability', *Spine Journal*, vol. 21, no. 7, pp. 1135–1142, Jul. 2021, ISSN: 18781632. DOI: 10.1016/j.spinee.2021.02.007.



NTNU

Norwegian University of
Science and Technology