ADDIS ABABA UNIVERSITY

ADDIS ABABA INSTITUTE OF TECHNOLOGY

SCHOOL OF ELECTRICAL AND COMPUTER ENGINEERING

Recurrent Neural Network-based Base Transceiver Station Power System Failure Prediction

By: Tewodros Kibatu

Adviser: Dr. -Ing. Dereje Hailemariam

A Thesis Submitted to the School of Graduate Studies of Addis Abeba University in Partial Fulfillment of the Requirements for the Degree of Master Science in Telecommunications Engineering.

Addis Ababa, Ethiopia

December 2019

ADDIS ABABA UNIVERSITY

ADDIS ABABA INSTITUTE OF TECHNOLOGY

SCHOOL OF ELECTRICAL AND COMPUTER ENGINEERING

Recurrent Neural Network-based Base Transceiver Station Power System Failure Prediction

Submitted by: Tewodros Kibatu ______________

Signature

Dr. -Ing. Dereje Hailemariam ______________

Advisor Signature

______________________________ ______________

Evaluator Signature

______________________________ ______________

Evaluator Signature

Declaration

I declare that this thesis is my original work and has not been presented for a degree in this or any other university, and all sources of materials used for the thesis have been fully acknowledged.

Tewodros Kibatu ______________

Name Signature

Place: Addis Ababa, Ethiopia

Date of Submission: ______________

This thesis has been submitted for examination with my approval as a university advisor.

Dr. -Ing. Dereje Hailemariam ______________

Advisor Name Signature

Acknowledgment

I would like to thank my advisor Dr. –Ing. Dereje Hailemariam for his devotion and continuous support during my work. This research would not have been realized without his time, constructive comments and follow up. I also thank Mr. Yonas Yehualashet for his skillful support and all others who helped me in completing this research.

Abstract

Global network infrastructures are increasing with the development of new technologies and growth in Internet traffic. As network infrastructures increases, maintaining and monitoring them will become very challenging since thousands of alarms are generated every day. Clearing those alarms by corrective maintenance activities require considerable effort and resources (car, labor, and budget).

In mobile networks, a Base Transceiver Station (BTS) is one key infrastructure element performing the task of connecting customer equipment with the cellular network. BTS services may be interrupted due to transmission, optical fiber cut, power system failure, natural disaster or many more. In the case of Ethio Telecom (ET), the sole telecom service provider in Ethiopia, power system failure takes the biggest share for interruption of BTS services. Minimizing power system failure will reduce downtime of the BTS thereby, guarantee customer satisfaction and maximize revenue. Recently, machine learning algorithms are used to predict failure in various areas like power distribution, hydropower generation plants, solar power generation plants, high voltage transmission grid and many more.

This thesis investigates predicting BTSs power system failure using a recurrent neural network (RNN) types namely, long short term memory (LSTM) and gated recurrent unit (GRU) with linear and sigmoid activation function applied for the output. In parallel, the prediction performance of LSTM and GRU has been compared. Data collected from five BTS sites for twenty weeks of observations are used to train and test the model. The data are prepared with two different data arrangements, which are a single site and multiple sites. The relevance of using different data size is, to check the impact of increasing data size with different arrangements on the prediction results. Mean squared error (MSE) and number of epoch are used to evaluate the performance of the models with different configurations. Based on the results found, GRU using sigmoid activation function with feature reduction achieves better performance than LSTM. In addition, both LSTM and GRU can be used for predicting BTS power system failure.

Keywords:- Base Transceiver Station, Gated Recurrent Unit, Long Short Term Memory, Recurrent Neural Network.

Table of Content

List of Table

List of Figure

Abbreviations

AC	Alternating Current
ACDC	Alternating Current/ Direct Current
AFR	After Feature Reduction
ANN	Artificial Neural Network
ATS	Automatic Transfer Switch
BBUV	Bus-bar Under Voltage
BFR	Before Feature Reduction
BLVD	Battery Low Voltage Disconnect
BTS	Base Transceiver Station
DC	Direct Current
DCDU	Direct Current Distribution Unit
DG	Diesel Generator
ET	Ethio Telecom
EMC	Electromagnetic Compatibility
GRU	Gated Recurrent Unit
GSM	Global System for Mobile Communication
HSPA	High Speed Packet Access
IDU	Indoor Unit
IP	Internet Protocol
LLVD	Load Low Voltage Disconnect
LSTM	Long Short Term Memory
LTE	Long Term Evolution
MW	Microwave
NAR	Network Available Rate
NE	Network Element
NMS	Network Monitoring System
NUR	Network Unavailable Rate
QoS	Quality of Services
RNN	Recurrent Neural Network
UNE	Unavailable Network Element

1. Introduction

Global network infrastructures are increasing with the development of new technologies and growth in Internet traffic; the latter is generated from cellular mobile users and different businesses like banks, manufacturing, and hospitals [1]. Mobile and fixed telecommunication infrastructures take the biggest portion from the global network infrastructure share [1].

Uninterrupted operation is the focus of any telecom operators to increase revenue, securing the quality of services (QoS) and customer satisfaction. As network infrastructure increases, maintenance and monitoring of the overall system will become very challenging. Since thousands of alarms are generated from mobile networks on a daily basis. So, fixing and clearing those alarms by corrective maintenance activities will diminish the performance of the overall system because corrective maintenance activities take place after the actual failure happen [2]. In addition, corrective maintenance requires considerable effort and facilities such as cars, labor, and budget.

In mobile networks, Base Transceiver Station (BTS) is one key infrastructure element that is performing the task of connecting the customer equipment with the cellular network. BTS services may be interrupted due to many reasons like transmission failure, optical fiber cut, power system failure, natural disaster or many more. In the case of Ethio Telecom (ET) power system failure takes the biggest portion. Because all the communication equipment is supplied from a common power source, their failure affects overall services provisioning. As power systems are critical elements in any communication system [3], their failure may cause an interruption to the complete services or failure of subsystems. Therefore, eliminating the failure causes of the BTS power system will minimize the downtime of the overall system and it will increase customer satisfaction thereby maximizing revenue. In addition to service interruption and customer dissatisfaction, such a failure may cause a loss of customers and the company data [4][5].

In the case of ET, because the power system failure takes the biggest share, minimizing these failures by following recommended maintenance trends would reduce downtime of the BTS and the overall services provisioning thereby, guarantee customer satisfaction and maximize revenue. Recently, machine learning algorithms are used to predict failure in various areas like power distribution systems, hydropower generation plants, solar power generation plants, high voltage transmission grid and many more [6].

In ET, there are around 7093 BTSs over the whole country and above 742 of them are found in Addis Ababa city, Ethiopia. NetEco real-time power monitoring system is installed to manage and check the events (operation conditions). This monitoring system collects real-time measurements, by using measuring tools and sensors which can generate alarms. Besides collecting real-time measurements, the monitoring system is capable of keeping history data and report these data for the central monitoring system [7].

Most of the BTS sites in Addis Ababa support cellular communication types Global System for Mobile Communication (GSM), High Speed Packet Access (HSPA), and Long Term Evolution (LTE). In addition to cellular communication, at the BTS sites looping and transmission services may take place. All of the services are supplied from a common power system, but with different priorities. Services with the highest priority are connected on separate distribution units (but from a common source) and services with the lowest priority are connected to another distribution unit. The main reason behind connecting the loads on different distribution units is, to easily disconnect them on different orders and conditions. Standard mobile services (GSM, HSPA, and LTE) are less prior than MW, transmission, and optical equipment [7]. However, both services get power from a common power source, but with different distribution systems and the availability or performance of the overall services significantly depend on the performance and availability of the power source. Giving priorities for services is not enough solution to guaranty the uptime of the services, but the maintenance trends of the power system may determine the performance of the overall services at BTS sites. Since both the equipment's supplied from a Therefore power source, following good maintenance strategies is mandatory in order to increase the performance of the overall system [2].

Nowadays, the maintenance trends shift from reactive maintenance to proactive maintenance, where reactive maintenance focuses on giving a solution for the problems after the actual failure happen and proactive maintenance focuses on preventing the failure by using preventive and predictive maintenance. Both of the maintenance approaches will have their own advantages and disadvantages. For instance, reactive maintenance will have its own advantage regarding consuming all the remaining component lifetime (until the actual failure happens), but service downtime will increase until the repairing or replacing action takes place[2][8]. Whereas, proactive maintenance has a greater advantage when the tasks or services are critical, which means the failure of the service may be prevented before it might cause damage to expensive equipment or loss of human life and incur a big loss.

Prediction of impending failure will increase equipment operation time and helps to realize the return on investment[5]. Moreover, predicting the future is used to implement better maintenance strategies, used to take correct decisions about the future, used to implement spare parts and resource assignment, better resource utilization, and many more. In order to engage in proactive maintenance activities, historical failure information, and different measurement data are required because the future is predicted based on the past [9].

1.1 Statement of the Problem

Nowadays, proactive maintenance activities are a must choice for operators in order to guarantee QoS and customer satisfaction [10][11]. Corrective maintenance will require a lot of effort and facility because it is engaged without a scheduled program. In addition, it will diminish the availability of a system since services are interrupted until the required repairing or replacing work takes place. Moreover, the cost of downtime is very high if failures are on critical components or critical sites, (e.g., a hub site failure may be the cause for other site service disruption), in off-hours, pick-hours, and remote locations. Therefore, predicting the upcoming failure would have a great advantage in solving unexpected service interruption, to guarantee QoS and thereby increase revenue.

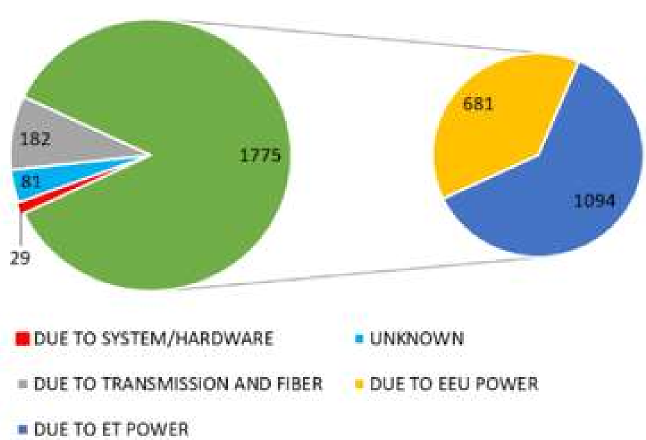

Figure 1.1 One-month failure root cause in April 2019.

Even failure prediction activities are taking place in different fields, but no prior works in BTS power system failure prediction. Therefore, this work might have a big contribution in enhancing the performance of the BTS power system. As recognized from one-month failure root causes for BTS report and tried to illustrate in Figure 1.1, a total of 2067 BTS failure registered for April 2019, of which 54 % of the failures are due to ET power system, 33 % due to commercial power outage (in this case the service by Ethiopian Electric Power utility (EEPU)), 9% is due to transmission and fiber cut, and the rest 4 % of the failure causes are unknown. Therefore, predicting BTS power system failure have a big impact to enhance the performance of the power system, thereby it will improve the provisioning of the overall service.

1.2 Objective

1.2.1 General Objective

The general aim of this research is to predict BTS power system failure by using 20 weeks time-series data which is collected from the NetEco power monitoring system in ET. To achieve these failure prediction activities, LSTM and GRU machine learning algorithms are used.

1.2.2 Specific Objectives

The specific objectives of this research work are:

- Study BTS power system architecture to understand the dependency and configuration of equipment in the BTS power system;
- Reviewing different papers on machine learning, failure prediction, BTS power system components, and configuration;
- Study RNN (GRU and LSTM) that are used to predicting the impending failure;
- By using BTS power system data, evaluate the predictive performance of LSTM and GRU using different activation function;
- Revise data preprocessing techniques especially on time-series data (normalization and outlier detection);
- Learn how to use python for prediction activity (specifically scikit-learn a library for machine learning in Python) and use for BTS power system failure prediction.

1.3 Methodology

To engage in this research work, there are some important conditions need to be considered during data collection, data preprocessing, reviewing BTS power system architecture, and studying different machine learning approaches. In order to formulate the problem correlation analysis between features and failure has been made. And bus-bar voltage, direct current (DC) load power and DC load current are found features that have high correlation results with the actual failure.

From the algorithm point of view, RNN (LSTM and GRU) is applied using Min-Max normalization for the input, linear and sigmoid activation function for the output. In addition, the model's results are checked with and without feature reduction techniques. Moreover, RNN hyper-parameters and recommended values are applied for the tests. MSE and number of epoch are used as performance measurements. MSE fundamentally measures an average squared error between the predicted and the actual values. The higher MSE value, the worse the model and the minimum MSE the better the model. In addition, the number of epochs is a parameter defines the times that the learning algorithm will work through the entire training dataset.

In addition, reviewing several related papers to learn the technics how others did the prediction activities is the primary operation. Data collection is conducted from ET that contains both good and bad measurements. The data is collected from a single server (Net-Eco power monitoring system sever) over 20 weeks period with a weekly collection program. The collected data has 11 features with a 5-minute sampling period. In both the tests (LSTM and GRU), the number of features and data size that has been are equal. 90% of the data is used for the training and 10% of the data is used to test the performance of the model. In addition, data merging and preprocessing have been engaged because the data is collected weekly-based and in a separate excel spreadsheet. So, it should have a sequential time-stamp. After all, an Excel spreadsheet for data preprocessing, TensorFlow which is an open source library used for machine learning is configured in Python.

1.4 Literature Review

Nowadays, a neural network is used for diverse applications including image processing, classification, and prediction. Neural networks can learn patterns and capture information automatically from the data. As research shows that, a neural network can be used to predict faults of individual components or failure of a subsystem. To the best of our knowledge, there is no prior work that applies machine learning techniques to predict BTS power system failure. The next section presents prediction works that are done in the area of component failure, transformer, power grid, hydraulic generator, and interconnected transmission system.

Chigurupati, R. Thibaux, and N. Lassar investigated the predictive abilities of a neural network technique by predicting individual component times until failure in advance of actual failure [12]. They employed a Support Vector Machine (SVM) for classifying the data. The data includes a 30-failed module having time information for the training with three classes. Data with three hours leading up to failure is assigned to class 2, data with less than eight hours and greater than three hours is assigned to class 1 and data with more than 8 hours are assigned to class 0. The developed neural network algorithm was able to monitor the health of 14 hardware samples and notify the coming failure well ahead of actual failure.

C. Haseltine et al used a neural network to determine performance and areas of concern (determine areas that require more attention by using the weight factor) of the power grid [13]. A single layer neural network with 10 years of data was applied. Through the investigation, they tried to consider all components of the power grid system (generation, load, transmission, and distribution). In addition, they included features containing information about the major failure causes of a power grid for the training. Those features are drought, hurricane, load growth, generation reserve, and proper control system operation rate. The limitation of this paper is, it is very difficult to set the proper control rate accurately which is used as a feature. Because it highly depends on the power grid technician and some unpredictable situation from the power generation through the distribution)

Most of the prediction activities take place by neural networks better than other data-driven approaches. K. Venugopal and et al recommend predicting transformer failure using Artificial Neural Network (ANN) [14]. They recommend that, in order to achieve better reliability, corrective maintenance activities are not sufficient since there will be a temporary interruption during maintenance. So, predicting the fault occurrence time has a greater advantage to develop reliable service provisioning and they configured a feed-forward backpropagation network with 100 neurons in the hidden layer that has been applied on MATLAB.

F. Wang, Z. Mi, S. Su, and C. Zhang develop an ANN based model to predict power generation of a grid which is connected to a photovoltaic plant [15]. They applied 16 inputs and 1 output model having two hidden layers with 15 neurons for the first layer and 7 neurons for the second layer. The input variables for training include solar radiation, ambient temperature, and power measurements were taken for 14 hours and they applied Mean Absolute Bias Error (MABE) and Root Mean Square Error (RMSE) as a performance measurement. They conclude power prediction generated power from the grid performs well and they found root mean squared error (RMSE) of 9.86% and the mean absolute error (MABE) of 7.16%.

N. Nadai et al, used a neural network for predicting the normal and abnormal operating conditions of hydraulic generator [16]. In order to do this work, one-month operational parameter measurements and sensorial inspection results collected from a monitoring system are used to train the neural network. Totally twenty-five measurement features are collected and out of these twenty-five features, two of the features which have much information about the performance of the hydraulics generator at the plant are selected as a target. These features are shaft vibration and temperature. In order to achieve better performance, they recommend using diverse data (data which contain both normal and abnormal operation measurements) which will give a better chance to the model to produce the best result.

L. T. Mar [17], uses an artificial neural network approach for predicting fault in a large interconnected transmission system and they conclude that ANN can predict the failure with better performance than the mean time between failure (MTBF) and mean time to failure (MTTF). In addition, they recommend careful selection of features that will have more information about the failure to be predicted which, will define the success of the prediction. So, they select the bus voltage and line current which has better information regarding the operating conditions.

Due to the nonlinear behavior of most of the time-series data and limitation of neural network with capturing long-term information, predicting the failure accurately is challenging. To deal with this problem RNN was introduced. J. Zheng, C. Xu, Z. Zhang, and X. Li [18] uses LSTM based RNN to show the capabilities to predict the electric loads. A total of 16 weeks of load measurement data with one -minute sampling period is used. Out of 16 weeks data, 14 weeks for training and 2 weeks for testing. Long term memory is problems that encounter in a standard feed-forward neural network. In addition, they show LSTM capability to predict both long term and short term faults.

Moreover, the papers discuss and compare multiple types of time series predicting method for future electric load forecasting. Finally, they develop an electrical load predicting scheme on GRU because they have got better performance than LSTM and other linear prediction techniques for complex nonlinear data. In addition, they recommend that long term forecasting is not feasible because there is an irregular increase in the demand for electric loads.

1.5 Scope and Limitation

The main scope of this research work is to predict BTS power system failure, by using RNN (GRU and LSTM) algorithms. However, the data collection process gets very difficult because the maximum data that can be collected from the monitoring system with 5 minutes sampling period is one week. Of course, it is possible to get one-month data with one hour sampling period, but the possibility of losing information will increase as the sampling period increase. Therefore, the data collection scheduled based on a weekly collection plan and the maximum data collected for this research is 20 weeks. As big data will give a better chance for RNN to model the data and capture more information. So, it could be better to collect more data and test the performance of the algorithm.

1.6 Contribution

This thesis work makes some key contributions to the development of BTS power system failure prediction activities, including the following:

- Organizing the data with two different sized arrangements. Single-site data starting from 12 weeks to 20 weeks and multiple sites data which is 20 weeks of observations for five sites;
- Compare the prediction performance of two RNN types which are LSTM and GRU;
- Motivate proactive maintenance activities for the company because reactive maintenance has its own limitations for assuring QoS;
- Motivate researchers to engage in a BTS power system area and can be used as supplements for the planning and optimization of BTS power systems in ET;
- Describes BTS power system components, architecture, and operating principles which may help staffs engaged in maintenance work;

1.7 Outline of thesis

The thesis is structured in seven chapters. Chapter one discusses the introduction, problem statement, literature review, and methodology used to handle the prediction activity. Chapter two presents BTS power system architecture, component configuration, operating principles of BTS power system, BTS power system failures type by categorizing them based on units. Chapter three describes RNN and its applications for prediction especially types called LSTM and GRU. Chapter four discusses the data collection, preprocessing and technics used like normalization, activation function, and correlation analysis. Chapter five presents the experiment setup which includes data set preparation, target selection, and the parameter setup. Chapter six discuss tests and results found using different data size. Finally, Chapter seven concludes the overall research and gives insight for future works.

2. Overview of BTS Power System

2.1 BTS Power System Equipment and Configurations

BTS is a transceiver that facilitates wireless communication between users' equipment and a network. The BTS is composed of different components like cellular network equipment, transmission equipment (which is used as a backbone or a link to other BTS), and power equipment. These equipment need electrical power to perform their operation properly and they have different priorities. Most of the time transmission equipment have higher priority than cellular network equipment (GSM, HSPA, and LTE). Transmission equipment is high prior because they serve as backbone links and they are used for linking remote BTS and exchanges [7]. Therefore, the failure of this high prior equipments will affect not only local cellular services but also other BTSs and exchanges.

BTS power system is responsible for supplying the required electrical power to the BTS subsystems and it is composed of different power sources, protection devices, switches, fuses, and circuit breakers. We may find the BTS power system with different configurations based on the location or available spaces at the site and site geographical conditions [19][20]. Most widely used configurations are:

- BTS with solar panel and battery bank;
- BTS with generator and battery bank;
- BTS with commercial power (mains) and battery bank;
- BTS with commercial power (mains), and battery bank;
- BTS with commercial power (mains), generator, solar panel, and battery bank.

The above power system configurations and requirements for BTS power system widely vary depending on a number of factors including whether the site is indoor or outdoor, estimated traffic (load), whether the site only gives mobile services or Hub sites (which is used as backbone link for other BTSs or exchange), service supported, site access (transport to and from the site), operation and maintenance constraints [20].

Figure 2.1 shows the power system configuration in most BTSs in Addis Ababa. Commercial power (mains), diesel generator and battery bank are used as power sources. These power sources are operating with different priorities. When the commercial power source fails to supply alternating current (AC) power, the generator automatically starts and the automatic transfer switch (ATS) transfers all the loads to the generators. The rectifiers get AC power from a commercial source or generator and it converts 220VAC to 48VDC to provide power to direct current (DC) loads and charge the battery banks [7].

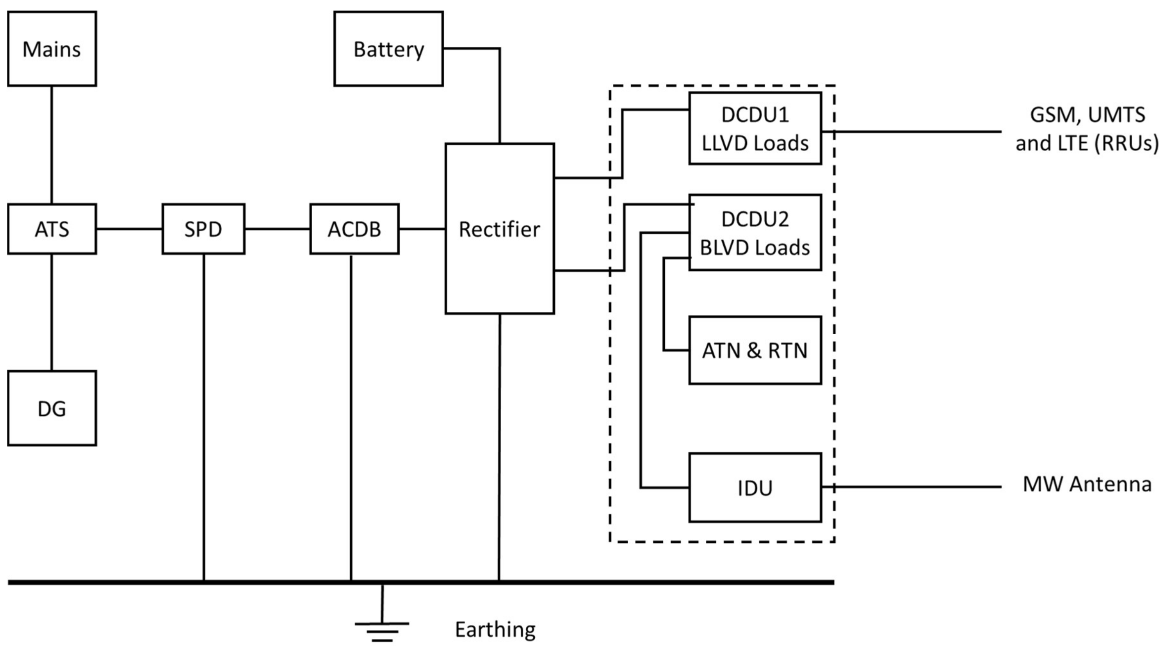

Figure 2.1 BTS power system configuration. *Source: Adapted from*[21].

2.2 Operation Principle of BTS Power System

The operation principle of a BTS power system varies based on the types of power equipment configuration used. Most of the time from the ET perspective, in cities like Addis Ababa the commercial power source is a primary power supplier while generator and battery bank are standby (secondary) power sources. For remote sites, solar power may be used to charge the battery bank and we may also find generator as a standby power source.

When the commercial source power is interrupted (i.e., totally turned off, phase lost or below the expected threshold), the ATS control module will send a start signal to the generator control module to start the generator and the generator start and supply AC power to the site. Then the rectifier received AC power from the commercial sources or the generator and convert it into DC 48VDC to supply the communication equipment and charge the battery banks.

Almost all the AC sources that we find in Addis Ababa BTSs are three-phase five-wire systems, The three phases are designated as R, S, T, neutral N and earthing E. The rated input voltage is 220VAC in single-phase or 380VAC line voltage in three-phase.

Moreover, the BTS power system encompasses different types of components, monitoring units, and protection devices including sensors, environmental equipment (air conditioner and fan), surge protection, AC or DC fuses, circuit breaker, relay and lightning arrester. The protection devices are used to protect the equipment from undesired events like a short circuit or overvoltage.

In general, a BTS power system structure and components configuration can be classified into four major parts namely:

- AC distribution unit;
- Rectifier;
- Direct Current Distribution Unit (DCDU);
- Monitoring Unit and each unit is explained in the next sections.

2.3 AC Distribution Unit

The AC distribution unit inputs its AC power from a commercial source or generator and distributes it to different loads. As shown in Figure 2.2, the AC part consists of commercial power (mains), stand by generator, ATS, main distribution board (MDB), surge protection devices including fuse, circuit breakers, and lightning arrester [7].

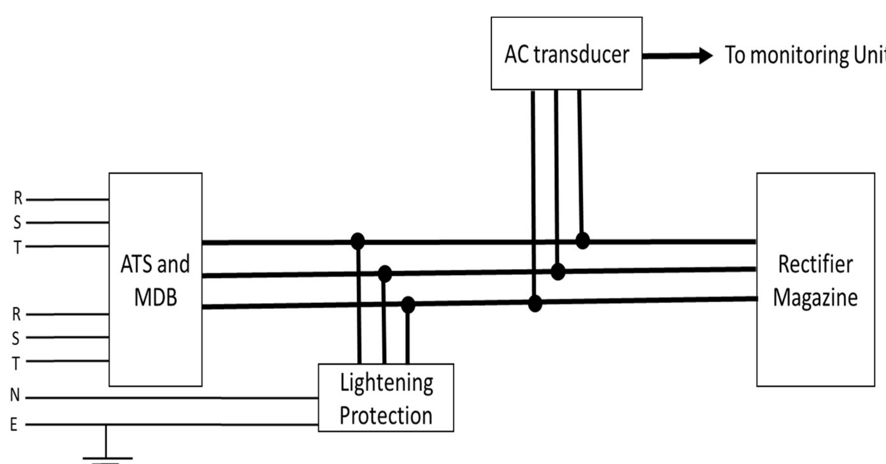

Figure 2.2 AC distribution unit. *Source: Adapted from* [7].

2.3.1 Automatic Transfer Switch (ATS)

ATS is an electromechanical switch, which is used to transfer the loads to commercial power or to the generator. When the commercial power fails to supply AC power, the ATS control module sends a signal to the generator control module to start the generator and the ATS transfer all the loads to the generator. When the commercial power restores, the ATS will transfer the loads to the commercial power source. The basic function of ATS is to harmonize and supply the AC voltages from different sources to the equipment.

2.3.2 Main Distribution Board

Main Distribution Board (MDB) is an electric distribution system board that divides an electrical power to secondary circuits based on the rating of the loads connected to it. Sometimes it uses a protective fuse or circuit breaker for each path in a common junction box. MDB gets its input power from a commercial source or a generator and supplies this power to the connected loads including the rectifiers.

2.3.3 Generator

A generator is composed of engine, alternator, and controller parts which convert mechanical energy into electrical energy. Most of ET generators that have been found at BTS sites use deasil for running the engine or to create mechanical energy. The rating of the generators depends on the loads which will be handled at the site. The generator is responsible to supply AC power when commercial power fails to supply the loads and when it the AC voltage became below the configured thresholds. When commercial power fails, the generator should have to start automatically and serves the site until commercial power recovered. If the generator fails to start while commercial power interrupted, the battery bank takes care of supplying the loads until it exhausts (minimum configured threshold) it's stored power.

2.3.4 Circuit Breakers, Surge Protection, and AC Transducers

A circuit breaker is a protective switch, which is used to safeguard equipment attached to it from undesired events (when the current becomes beyond safe level or the rating). Whenever a circuit breaker trips, it requires manual restoration.

Surge protection (surge suppressor) is a device designed to protect equipment from voltage spikes or lightning, mostly which occurs when the weather conditions become rainy or fogy. AC transducers (used as a sensor) is connected to the AC distribution system which is used sense the required AC input parameters (phase voltage, phase current, frequency, and phase angle) and communicate the readings to the monitoring system.

2.4 Rectifier

A rectifier is an essential element in the BTS power system, which is used to convert AC voltage into DC voltage because most of the BTS equipment operates on DC power. A rectifier is also used to charge the battery bank while the commercial power or the generator is available. In addition to the above-stated responsibilities, rectifiers also handle smoothing the voltage, protect sudden changes in voltage or current, DC to DC conversion, PFC and filtering harmonics.

Most of the communication equipment are very sensitive for interference so, rectifiers that will be used for BTS should be selected by considering the equipment's requirements. Rectifiers operate on a wide range of input sources from 80VAC to 300VAC and frequency range of 45Hz to 60 Hz. The relevance of operating on a wide range is to protect the sites from interruption on a poor power distribution environment [7].

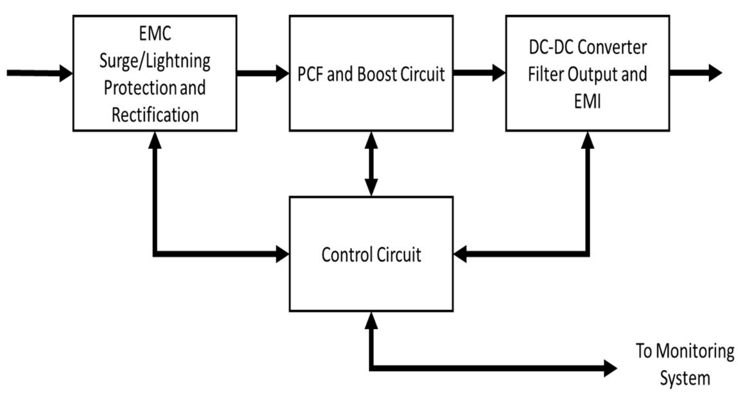

Figure 2.3 Rectifier internal components. *Source: Adapted from* [7].

As shown from Figure 2.3 above, rectifier internal structure, before rectified voltage enters to PFC and a boost circuits, electromagnetic compatibility (EMC), surge and lightning protection process have taken place. After PFC and boosting the voltage applied, the voltage is provided to DC-DC conversion. The PFC and DC-DC conversion enhance the reliability of the rectifier by increasing energy efficiency and eliminating different harmonics, respectively. After all this process, the rectified 48VDC will be supplied to the battery banks and DC loads. And all the status and the parameters inside the rectifier are communicated to the monitoring system by control interface.

2.5 DC Distribution Unit

The DC Distribution Unit (DCDU) gets its DC power from the rectifier or battery bank and distributes it to different DC loads that we found at BTS. As can be seen from the block diagram of DCDU in Figure 2.4 below, it includes battery banks, Load Low Voltage Disconnect (LLVD) contactor, Battery Low Voltage Disconnect (BLVD) contactor, and surge arresters. Rectified 48VDC is supplied to both DC loads and battery banks in parallel. There is a shunt in between the battery bank and the loads and the shunt is used to detect (sense) the total load current and battery current [22]. DCDU also has LLVD and BLVD function which is used to serve the loads with different priorities. When the battery voltage becomes below the configured threshold, the system will automatically disconnect less prior equipment by tripping off the LLVD contactor. The purpose of disconnecting less prior services is to keep active the critical loads until the battery bank exhaust its stored power (gets to the minimum configured voltage level) [7].

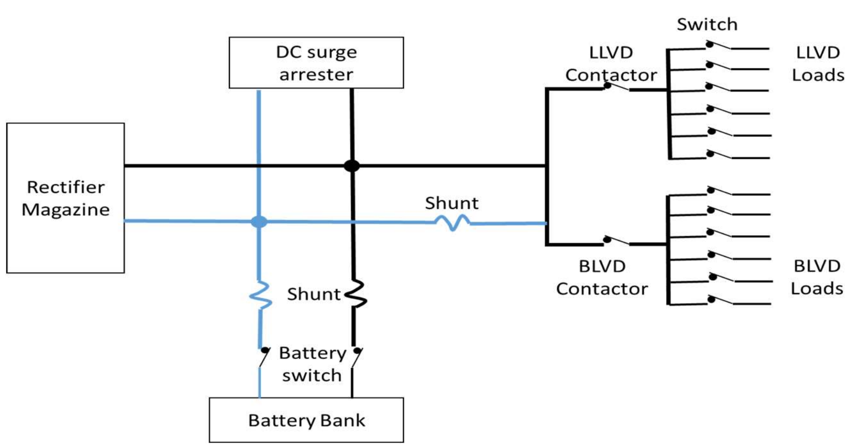

Figure 2.4 Direct Current Distribution Unit. *Source: Adapted from*[7].

2.5.1 Battery Bank

The battery bank is an essential element for the BTS power system, which is interconnected in a series-parallel manner. The series connection is used to increase the voltage level, whereas the parallel connection increases current generating capabilities (ampere-hour, AH) of the battery bank. Four 12VDC batteries are connected in series to achieve the 48VDC and the number of battery banks connected in parallel depends on the actual loads required at the specific site; as shown in Figure 2.7 most BTSs found in the ET Addis Ababa sites have two battery banks connected in parallel.

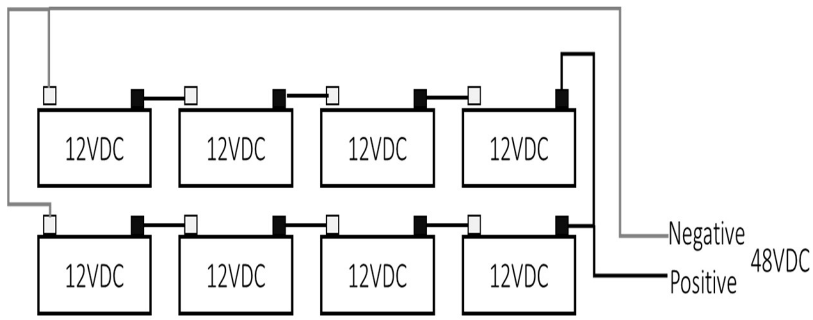

Figure 2.5 Battery bank configuration (two battery bank).

Generally, the battery bank is expected to supply the site for 8 hours and above when commercial power and generators fail. In addition, the battery bank handles the task of supplying DC power to the equipment until the generator starts in case the commercial power fail and while the ATS swaps the load from the generator to the commercial power or vice-versa.

BTS power system performance highly depends on the performance of the battery banks, especially for sites that exist in poor commercial power distribution lines (repeated commercial power failure encounter). The battery banks might lose their performance or the battery bank exhausts its stored power to quickly due to multiple reasons and there are many factors which may degrade the performance of the battery banks load handling capacity [23]:

- Corrosion of the positive grid structure due to oxidation of the grid;
- Higher working temperature environment;
- Discharge cycles (when charging and discharging cycle end the specification);
- Overcharging (causes excessive gassing);
- Undercharging (causes sulfating);
- Deep discharging (over-discharged may damage the battery elements).

The battery bank performance degradation can be detected by comparing the battery total discharge and total battery cycle with the manufacturer specification and age from the time of installation.

2.6 Monitoring Unit

The monitoring unit is a key element for the BTS power system, which highly supports the maintenance activities and helps telecom operators to guarantee QoS and used to minimize the restoration time when a failure occurs. The monitoring system detects the operation status of the system by using different kinds of sensors or measuring tools. Sensors are devices that are used to detect physical parameters and convert into a signal which can be measured. For instance: temperature, humidity, voltage, current, battery status, door contact, smoke detector, camera, and many more. Those sensors will sense the operation of the equipment's on a real-time manner and transmit the measured results to the local monitoring unit and as well as for remote monitoring systems [24]. In addition to data collection, monitoring unit has many functionalities including:

- Battery bank management;
- Battery protection;
- Data processing for the purpose of reporting;
- Gives an interactive interface for the technicians using liquid crystal display (LCD) and buttons that enable the engineers to configure and query necessary information from the monitoring unit.

Moreover, the monitoring system has the capability of protecting the equipment by shutting down in case of critical events encountered (flood, fire or smoke) and reports to the remote monitoring system.

ET uses the NetEco which is a short naming for Network Ecosystem, it provides management of site power system and environmental issues [24]. This monitoring system manages the activities in real-time and reports all the conditions to local and central monitoring system including failure information.

From the NetEco BTS power system monitoring point of view, alarms (failure or risk notifications) are categorized based on the damage they contribute. These alarms can be reconfigured based on the operation conditions and they are classified as critical, major, minor, and warning alarms. Critical alarm indicates a problem that may interrupt both the BTS operation or there may be a potential risk to interrupt the operation. Major alarms indicate the possibility of some service-related problems. Minor alarm indicates a problem of relatively low severity that may not obstruct the BTS operation but it may damage equipment if maintenance action is not taken. The warning alarms indicate a condition that can potentially cause a problem with the operation in the long run.

Table 2.1 BTS power system alarms category and descriptions.

Alarms Type	Category	Description
Abnormal battery current	Major	The current level of the battery current becomes beyond the expected range (+ve or –ve).
AC SPD fault	Major	Surge protection device malfunction or get damaged.
Air conditioner failure	Major	When the air conditioner refrigerant becomes low, fan failure which may cause stress on the system.
BLVD	Critical	Disconnecting all the loads that found at the BTS in order to protect the battery from getting a deep discharge.
Bus-bar under-voltage	Major	Pre-warning for LLVD to happen. If the battery performance is poor, there may be a possibility that LLVD to happen.
Communication between NMS & NE	Major	When the network monitoring system can not communicate with the network element or the site itself.
DG output under-voltage or over-voltage	Minor	When the generator voltage becomes under or overvoltage.
D.G running	Major	When the generator continues running while the commercial power is available.
D.G failure	Critical	When the generator fails to start or the generator encounter failure.
Door open alarm	Warning	The site door has opened notification.
Fan failure	Major	When fan failure encounter ate the BTS sites.
High ambient temperature	Minor	The temperature level becomes beyond the configured threshold.
LLVD	Major	Disconnecting less prior services when the battery voltage gets below the configured LLVD thresholds.
Low ambient humidity	Minor	The humidity of the site becomes lower than the recommended or configured thresholds.
Low fuel level	Major	When the generator fuel level gets the minimum level.
Mains failure	Major	When commercial power is gone.
Mains lose phase	Minor	When the commercial lines lost one or two phases.
Mains over-voltage	Minor	When the commercial power gets over the configured or recommended threshold.
Mains under-voltage	Major	When the commercial power voltage level gets below the recommended or configured thresholds.
Power system charge failure	Major	When the battery charging process interrupted or failed.
PSU fault	Major	When the power supply unit (rectifier) failed.

PSU lost	Minor	When the power supply unit disconnected from the rectifier rack.
High or low battery temperature	Major	When the battery temperature becomes below or higher than the recommended or configured thresholds.

2.7 BTS Power System Failure

BTS power system failure is an event when BTS power gets below the configured threshold or a power blackout (gone). For this research, failure is defined as partial BTS power system failure and total BTS power system failure in addition to a warning for partial failure. These failures can be detected from the bus-bar of the BTS power system. The bus-bar is a copper strip or bar, where DC loads, battery bank, and the rectifier output are connected in parallel. The measurements taken from the bus-bar of a power system are bus-bar voltage, DC load current, and DC load power. But, bus-bar voltage capture relatively higher information than that of DC load current and power.

As tried to illustrate by a diagram from Figure 2.6 when the voltage level gets 48.2VDC it is termed as BBUV and it is a pre-warning for the LLVD failure to occur. The duration until failure depends on the battery load carrying capabilities. When the voltage level becomes 46.2VDC, LLVD loads will be disconnected and the LLVD alarm generated. In addition, when the voltage level gets 45.2VDC, the BLVD loads will be disconnected and all the services at the BTS sites interrupted except the communication between the monitoring system and the site. Disconnecting the loads at the BLVD voltage level is used to protect the battery banks from getting deep discharging that may damage the battery elements permanently[23].

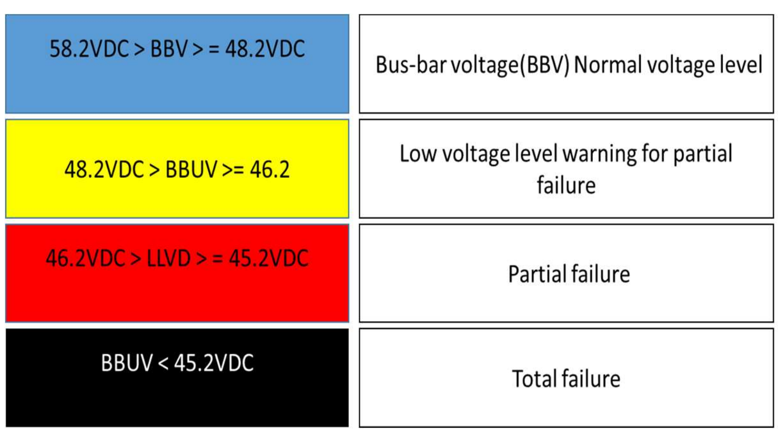

Figure 2.6 Failure representation on bus-bar voltage.

2.8 Factors for BTS Power System Failure

There are many factors that might cause BTS power system failure. Power system failure may occur because of commercial power (mains) outage, and the inability of subsequent backup system (generator, battery banks, and solar power system) or power-related devices (circuit breakers, fuse, and rectifiers) failure[25]. BTS power system failure may also arise due to cascading failure where the malfunction of one component could result failure of other parts of the subsystems[26]. Sometimes individual component failure may cause interruption on the power system thereby the whole BTS operation may be disrupted. For instance, if a major load fuse is broken, the load does not have any opportunity to get its operational power. So, the service interrupted even all the power sources and components are in good condition.

BTS may encompass transmission and cellular equipment (GSM, HSPA, and LTE) at the same compound or shared house. This equipment requires enough (based on the operating specification) power to operate properly. If the power to that service is interrupted, the equipment stops its operation. Therefore, it requires backup power to continue serving what they are expected. Therefore, the battery bank have a big impact on the BTS power performance under the interruption of commercial power and generator.

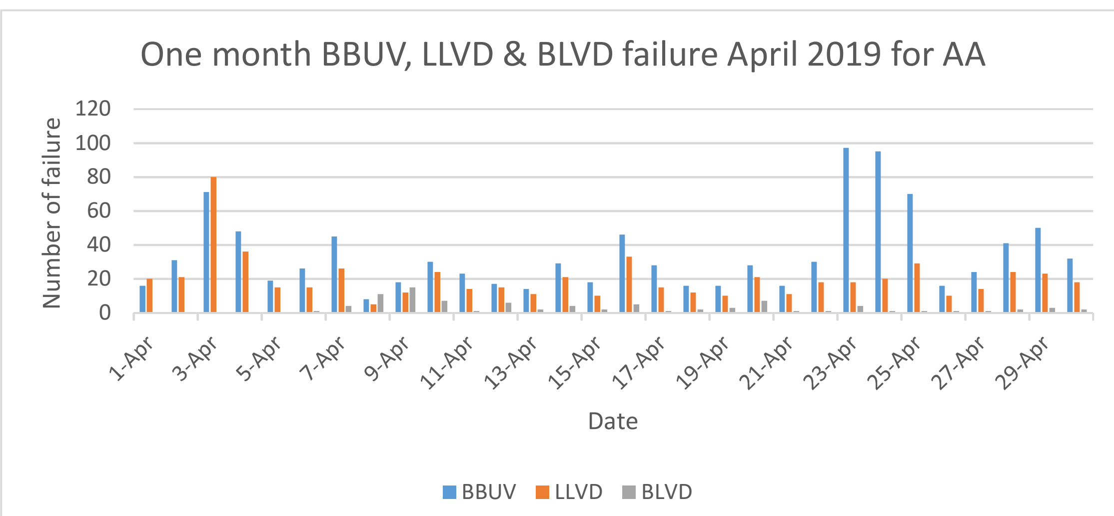

Figure 2.7 One-month BLVD, LLVD, and BBUV alarms.

Failures have a propagation behavior[26], as can be shown from Figure 2.7 above, the BBUV failure propagates and becomes LLVD failure which is a major alarm and the LLVD failure also propagates to BLVD failure which is a critical alarm. This propagation behavior may be deterred if the required maintenance action takes place and this can be shown from the figure below, in which most of the BBUV alarms have a higher number than LLVD and LLVD alarms also greater in number from BLVD. In addition, even they are different in number, almost for all the situation after BBUV happen, LLVD and BLVD alarms encounter. This is visibly demonstrated on one-month alarms in Figure 2.7 below.

2.9 Types of Alarms at BTS Power System

Besides classifying the alarms based on the damage they contribute, the NetEco monitoring system categorized them based on the unit or subsystem. Observing the alarms from the unit point of view have its own relevance to organize maintenance activities, resources (car, spare parts) allocation, and staff assignment. The classifications are[7]:

- Alarms of AC distribution unit;
- Alarms of DCDU;
- Environmental alarm.

2.9.1 Alarms of AC Distribution Unit

AC distribution failures are events, which occur in the AC parts of a BTS power system. These failures may lead to other types of failure. For instance, if commercial power fails to supply the site, the generator operates until it exhausts its fuel and then stops serving. Then, the responsibility of handling the site falls on the battery bank. Therefore, the impact of commercial power failure leads to the battery bank and lastly, the whole BTS operations may be interrupted. There are many types of failure which are categorized into AC distribution part [7]. Some of these are commercial power (mains) failure, commercial power (mains) phase loss, commercial power (mains) undervoltage, and commercial power (mains) overvoltage.

2.3.2 Alarms of DCDU

Failures of DCDU part are events, which occur on DC parts. There are many DC part failures which will cause service interruption fully or partially including DC undervoltage, DC overvoltage, major load fuse break, battery over-temperature, battery charging over current, which may arise due to the battery drain high current when it gets deep discharge and recharging, BBUV, BLVD, and LLVD.

Even most of DCDU failures affect the provisioning of the service, but LLVD and BLVD alarms could interrupt the whole operation of the BTS (fully or partially). LLVD alarm is generated when the loads are supplied from the battery bank and get below the configured threshold voltage level. LLVD event will interrupt all regular mobile services (GSM, UMTS and, LTE) in order to use remaining battery power for high prior equipment. BLVD alarm is also generated when the battery level reaches below the configured BLVD voltage level. BLVD refers to disconnecting all the loads at the BTS site for protecting the battery bank from getting a deep discharge. A deep discharge may damage the battery elements permanently, decrease battery life, load handling capabilities, it will cause an increase in battery temperature and battery current while charging to recover to the normal voltage level.

2.9.3 Environment Alarm

Environmental alarms are generated when temperature or humidity becomes beyond the recommended or configured level. Even though such alarms rarely happen, environmental alarms including smoke or fire and flood may encounter. For instance, high ambient humidity or low ambient humidity, low ambient temperature or high ambient temperature are environmental alarms. Most of the time the environmental problem comes from weather conditions, environmental equipment failure, site type (load), design problems and other failures which are not an environmental issue. Rarely, environmental alarms have cascading behaviors and affect BTS equipment operation. For instance, over-temperature will increase the burden on air conditioners and fans. This burden will cause an over temperature which may cause a fault on the equipment and may interrupt the whole operation at the BTS.

3. Machine Learning

3.1 Artificial Neural Network

ANN is a mathematical formation motivated by the study of interconnections between neurons found in the human brain [27]. It consists of an interconnected group of neurons, which processes information. A neural network can also be expressed as mapping an input(s) into the desired output(s). The input(s) for the neuron may be the output(s) of another neuron. The input and output neurons can be one-to-one, many-to-one or many-to-many. For example, as seen in Figure 3.1 two neurons feed a single neuron that has one output. Every neuron has different importance and this importance will be represented with different weights. In most cases, a neural network is an adaptive system that changes its structure and weights based on external or internal information that flows through the network during the learning phase. Nowadays, a neural network is used as modeling tools for non-linear statistical data and they are usually used to model complex relationships between inputs and outputs or to find patterns in data. In the artificial intelligence (AI) field, a neural network has been applied successfully to speech recognition, image analysis, and adaptive control, in order to construct software agents or autonomous robots [28].

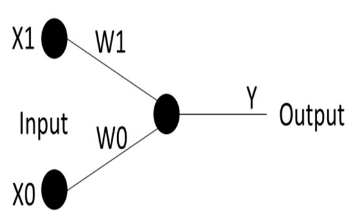

Figure 3.1 Neurons with waited input and output.

neural network models can be mathematically represented based on their structure. For instance, Figure 3.1 can be represented by using its inputs, output, and weights mathematically as in Equation (3.1) [28]. Where i represents the number of inputs, weight wi, input xi, and output y.

$$Y = \sum_{i=0}^{n}(w_i x_i) \tag{3.1}$$

3.2 Activation Functions

Activation functions are mathematical equations that are used to decide the output of machine learning algorithms [29]. Each neuron has its own activation function which decides the neurons to be activated or not based on the degree of information (importance) captured. It has a greater impact on the performance of neural networks by deciding the output of the neural network. In addition, it determines and improves the accuracy and the computational efficiency of the neural network [30].

An activation function has many variants, which are used to enhance the neural network reliability. The selection of an activation function is employed based on available data type, problem to mitigate and resource available. Binary step, linear, sigmoid and tangent hyperbolic (tanh) activation functions are some of the well-known and widely used activation functions. Each of them is explained in the below sections.

3.2.1 Binary Step Activation Function

A binary step activation function work based on a certain threshold to activated or not. When the input is above a certain threshold value, the neuron is activated and the output becomes one. Whereas, when the input becomes below the threshold value, the neurons will not be activated and the output becomes zero. Figure 3.2 [30] shows the activation function. The binary step activation function can be mathematically represented as the equation (3.2) [30].

$$f(x) = \begin{cases} 0 & for\, x < 0 \\ 1 & for\, x \geq 0 \end{cases} \quad (3.2)$$

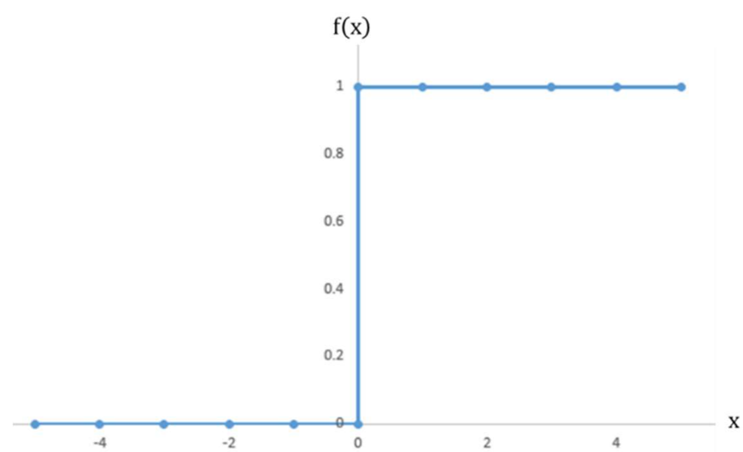

Figure 3.2 Binary step activation function.

One of the biggest problems with the binary step activation function is that it does not consider multiple level output classification [29], the result found for the output becomes 0 or 1.

3.2.2 Linear Activation Function

A linear activation function transforms the input by multiplying with a constant on each neuron and produce an output proportional to the input. If we compare a linear activation function with a step activation function, linear is better with respect to having multiple outputs classification, but it has a limitation for handling data with complex or non-linear relation for input and output [29]. Figure 3.3 [30] shows a linear activation function and can be mathematically represented as the equation (3.3) [30].

$$f(x) = x \tag{3.3}$$

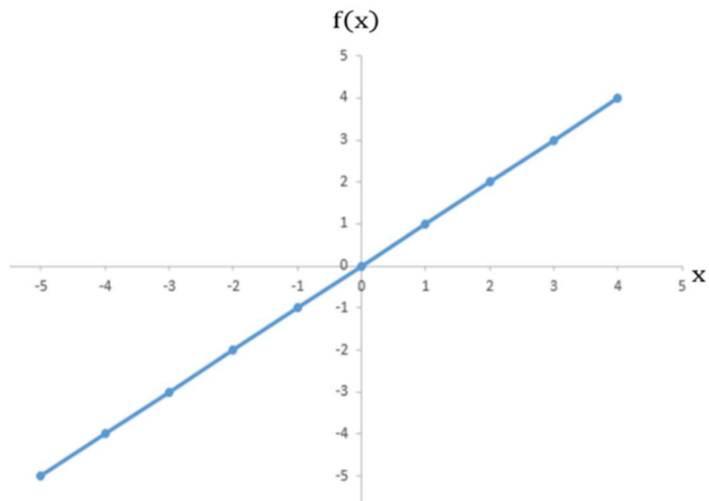

Figure 3.3 Linear activation function.

3.2.3 Sigmoid/Logistic Activation Function

A sigmoid activation function is a widely used activation function for neural networks, which maps the input data between zero and one. Since neural networks are used to learn and model complex data that experience input and output have a complex relationship, the sigmoid activation function becomes a choice for many neural network-based models. The main problem of using the sigmoid activation function is, the output values are not zero centered which means it has a limitation on modeling data with negative values. Since the sigmoid activation function squash the data in between 0 and 1, so both negative and positive data are forced to be mapped to positive value [29]. Figure 3.4 [30] represents the sigmoid activation function and can be mathematically represented as the equation (3.4) [30].

$$f(x) = \frac{1}{1 + e^{-x}} \tag{3.4}$$

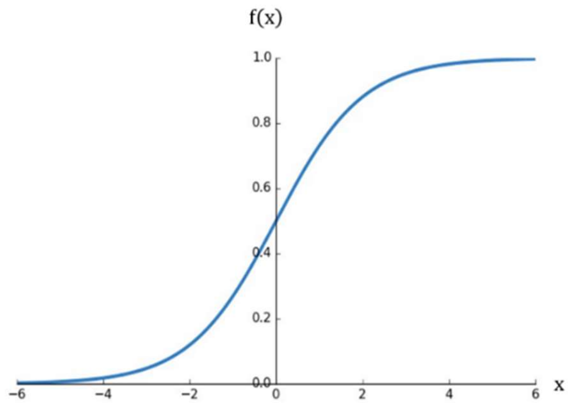

Figure 3.4 Sigmoid activation function.

3.2.4 Tangent Hyperbolic Activation Function

A sigmoid activation function, the tangent hyperbolic (tanh) activation function is also wildly applied for machine learning and it is very similar with a sigmoid activation function. However, the main difference is, tanh is zero centered, so it can easily represent both negative and positive values [29]. Figure 3.5 represents the tanh activation function and can be represented mathematically as equation (3.5) [30].

$$f(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \tag{3.5}$$

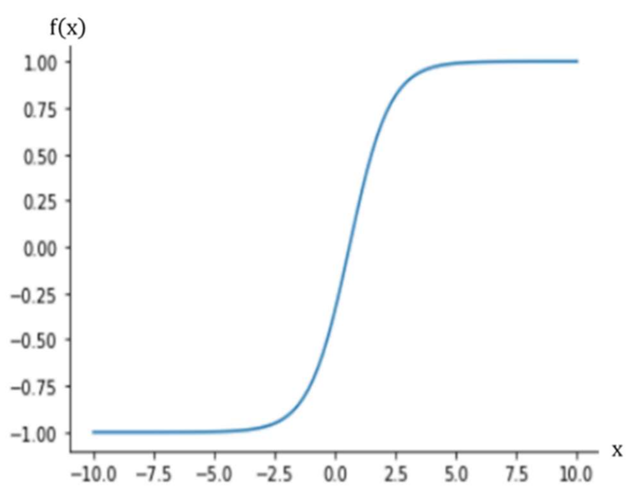

Figure 3.5 Hyperbolic tangent (tanh) activation function.

3.2.5 Parameters in Machine Learning

Parameter setup generally has a major impact on the achievement of machine learning algorithms. The process of assigning these parameters includes assigning of number of an epoch, batch size, number of neurons and many more. The machine learning algorithm operates on an iterative process to get the most optimal results by varying the weights at each epoch and in every batch of the process [31]. The iteration, the batch size and the number of epochs are used when the data is too big and it is not possible to pass all the data once. So, the data should be divided into smaller batches before gives it to the computer [31].

The number of epochs is a hyperparameter that is used to defines the number of times that the machine learning algorithm works through the whole training dataset [31]. One epoch means that the samples have passed forward and backward through the machine learning algorithms and get an opportunity to update the parameter at every epoch. The number of epochs and the error can be plotted to check whether the model has over-learned, under learned or is suitably fit to the training dataset [31]. As the number of epochs increases, the opportunity to update the weight and loss is also increased. Passing the dataset multiple times may increase the performance, but the possibility op the model to overfitting may be increased. Actually, the number of epochs and batch sizes is different for a different dataset and it is related to how diverse the data is.

3.3 Recurrent Neural Networks

Even activation functions have an advantage regarding speeding wall time (learning time) and bias between features, they have their own limitations. They introduce a vanishing and exploding gradient (exploding gradient problem is rare) problem for small value and large values, respectively, while backpropagation because it uses derivative and multiplication with respect to the parameters found in every layer in a neural network [32]. These problems prevent the algorithms to memorize and use long-term information, which diminishes the performance of the neural network.

Recurrent Neural Networks (RNN) is basically a neural network, which has a memory to store information at every recurrent unit, and it can remember things from the past. RNN feeds the outputs from neurons to other adjacent neurons, to themselves, or to neurons on preceding network layers and this capability makes RNN better to model complex works which is difficult to handle by a standard neural network [28]. RNN uses an interconnected individual recurrent cell. As shown from Figure 3.6 [28], a recurrent cell is an individual unit, with input, output, and, cell state which represents the contained information.

A traditional feed-forward neural network is not well suited to handle sequential data because it uses a fixed input sequence for learning the data. Because sequential data may have important information in the past sequencethe which may be used to represent the future and may be used to understand the entire data, storing this information and used when required determine the efficiency of the algorithm [28]. RNN solves the problems that arise to time-series and sequence data that suffer from capturing long-term information [28]. Sequence modeling mechanisms have their own criteria to be fulfilled. These are:

- It should support variable-length input;
- It should have the capabilities of tracking long term dependencies (information);
- It should have the capabilities to maintaining information order;
- It should have the capability to share parameters across the sequence.

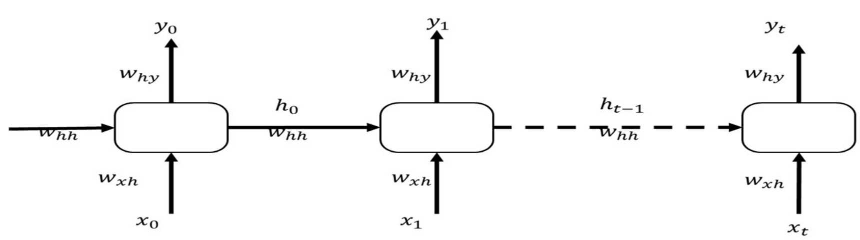

Figure 3.6 Connection of recurrent cells in RNN

RNN is widely used for time-series prediction which suffers from capturing long-term dependency. Actually, both long-term and short-term information are required to properly model the entire sequence. As discussed above, a standard neural network has a limitation to store earlier information from the data, because it uses fixed-length observations as input. This means it has a limitation to consider the entire sequence. In addition, while backpropagation, vanishing and exploding gradient problem encounter which may produce biasing between the features [28]. RNNs allow signals to travel both forward and backward and also it introduces loops in the network which allows internal connections among hidden units. With the help of such internal connections, RNNs are more suitable for consuming the information in the past data to forecast future data [18][27].

RNN employ interconnected recurrent cell to store information. As can be seen from Figure 3.6 above, each recurrent cell takes the previous output or hidden states as inputs. The composite input at time t has some historical information about the happenings at time T < t. RNN has diverse application including image recognition, classification, and prediction. Even RNN has multiple variants the most wildly used are Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) [28].

3.3.1 Long Short-Term Memory

LSTM is one variant of RNN that has feedback connections with itself and other neurons. It has the capability of considering the entire sequences of data. LSTM is consists of three essential parts which determine the output of every recurrent cell. These are forget, update and output gates. These gates decide what to store, erase, and output at every computation by open and close the information flow. To manage the information flow RNN uses a sigmoid gate and pointwise multiplication [33].

As explained in Figure 3.7 [33] below, the forget gate f_t remove irrelevant information by sigmoid function using previous cell output and x_t the new input. Equation (3.6) [33] shows the mathematical formulation for forget gates.

$$f_t = (W_f * \sigma [h_{t-1}, x_t] + b_f) \quad (3.6)$$

As stated from Equation (3.7) [33] the update gates decide what values to be updated using previous cell state C_{t-1} and the new candidate value f_t and add a new candidate value which is a function of (i_t , c_t).

$$C_t = f_t * C_{t-1} + i_t * c_t \quad (3.7)$$

Finally, the output gate uses a sigmoid(σ) function to decide what parts of the state to be output. The output O_t and the new version of cell state h_t is described mathematically by Equations (3.8) and (3.9) [33].

$$O_t = \sigma (W_o [h_{t-1}, x_t] + b_o) \quad (3.8)$$

$$h_t = O_t * \tanh(C_t) \quad (3.9)$$

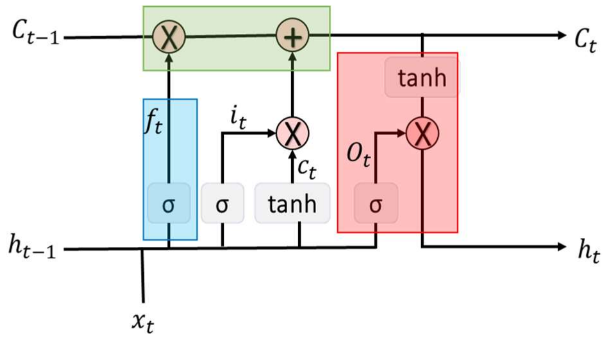

Figure 3.7 LSTM internal structure

3.3.2 Gated Recurrent Unit

GRU is another variant of RNN that is inspired by the LSTM unit but simpler than that of LSTM from the computation points of view because it is implemented with a simple internal structure. As can be noted from Figure 3.8 [33], GRU consists of an update & a reset gate.

The update gate defines how much previous memory to keep and the reset gate defines how to combine the new input with the previous memory and decide how much past information to forget and the update gate decide what information to add or remove. GRU can be mathematically expressed as r which is reset gate, h_t the new cell state and Z the update gate as equation (3.10), (3.11) and (3.12) [33].

$$r = \sigma(W_r\ h_{t-1} + U_r x_t) \quad (3.10)$$

$$h_t = (z * C) + ((1 - z) * h_{t-1}) \quad (3.11)$$

$$Z = \sigma(W_z\ h_{t-1} + U_z x_t) \quad (3.12)$$

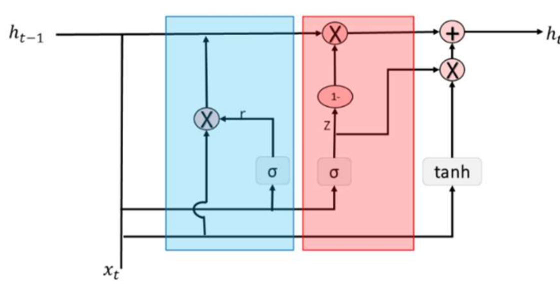

Figure 3.8 GRU internal structure

4. Data Collection and Preprocessing

There are multiple sensors installed to the BTS power system; these sensors take measurements continuously and communicate the data to the monitoring system. Most of the BTS power systems configured with different sensors like temperature, humidity, flood, smoke, voltage, current, and many more that measure and report events in real-time. The measurements represent unique observations and are termed as features in machine learning [28].

A neural network is a data-driven approach that is used to model complex and nonlinear relationships between inputs and outputs. To model this relationship, the neural network finds patterns (information) in the features. The accuracy of modeling this relationship depends on the degree of information that every feature hold and it is better for multiple inputs and diverse (good and bad) data. Even the complexity of the model increase with the increase in a number of features, but it will give a better chance for the neural network to model the data and give an opportunity to learn and capture rare events [34].

In addition to using multiple features, careful collection and preprocessing of the data determine the success and failure of using neural networks [28]. The data collected for this research work are only BTS power system related and collected based on a weekly data collection plan. A total of 20 weeks of data having 11 features with 5 min sampling period has been used. The features are working temperature, working humidity, battery current, battery voltage, battery temperature, battery total discharge power, battery total cycle times, AC/DC system output current, bus-bar voltage, DC load current, and DC load power. The data is collected from the ET NetEco power monitoring system. Actually, a large amount of observation will give the neural networks a better chance to represent the data and increase the ability to produce the desired output.

4.1 Features

A feature is an observation collectged using sensors and measuring instruments. The features may have information about the actual operation status of a system and sometimes the features may not have or capture small information. In the BTS power system, there are different measurements taken that capture operational information about good and bad events. Most of the time a feature is usually a numeric data, but one may find features in the form of string and graphs [34].

From the ET power system monitoring point of view, features can be categorized based on the unit (subsystem) that the measurement is taken from. These are:

- Environmental features;
- Battery system features;
- Load-related features.

4.1.1 Environmental Features

Ambient temperature and ambient humidity are categorized into an environmental feature. They contain basic information regarding the operation condition of the BTS power system. Maintaining the temperature and humidity to the recommended level has its own operational and energy cost but it has a direct effect on the performance of the whole system[35].

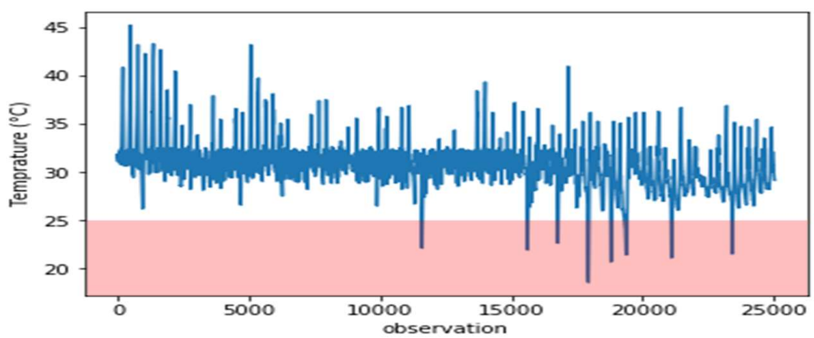

Figure 4.1 Working temperature

Humidity is a concentration of water vapor present in the air and expressed as in percentage (%). If humidity levels become too high, it introduces the accumulation of electrostatic charge on conductors and insulating materials [36] which may cause short-circuiting, corrosion, and rust on the equipment. The recommended humidity range for communication equipment is 40% to 55% [35]. As humidity, maintaining the temperature level to the recommended level has its own challenges and it is very critical for sustaining the health of the components in the BTS power system. The recommended operating temperature for communication equipment is from +15° C to +25° C [36][35].

Figure 4.2 indicates, humidity and temperature measurements taken from a specific ET BTS power system. We can see that the temperature and humidity are beyond the recommended values.

Exposing the equipment to operate beyond the recommended values for short periods of time may not be a problem, but running beyond the recommended range for a longer time could result in damage to equipment and degrade the performance of the system [36][35].

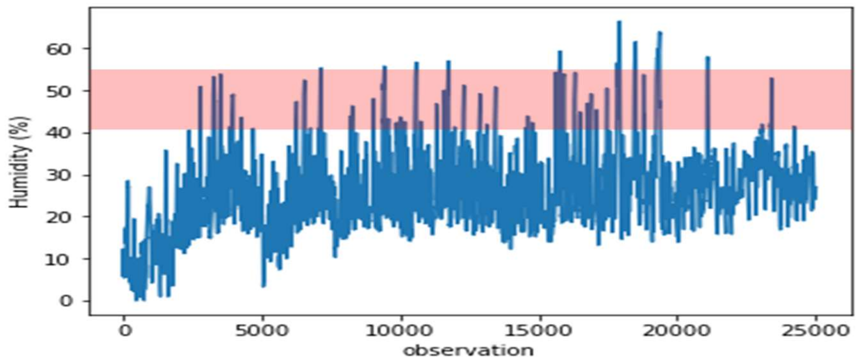

Figure 4.2 Working humidity

4.1.2 Battery System Features

The battery bank is the critical element of the BTS power system because it is used as a backup power source when the commercial power and generator fail to supply what they are expected. Battery voltage, battery current, battery temperature, battery cycle, and battery total discharges can be categorized into battery system measurements. Because of the battery and the loads are connected in parallel on the bus-bar, the battery voltage and the bus-bar voltage follow the same pattern.

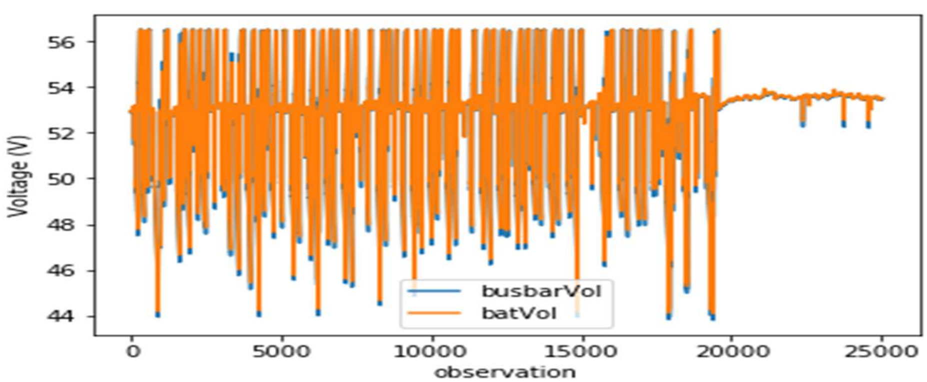

Figure 4.3 Battery voltage and bus-bar voltage

Two battery technologies that are widely used in power systems are flooded and lead-acid batteries. Flooded batteries require a separate room with their own air conditioning and require periodic follow-up [23]. The separate room and air conditioning are used to prevent the accumulation of flammable hydrogen gas, which is produced while the battery is in operation. Lead-acid batteries, on the other hand, develop very little hydrogen gas and it is free of moisture loss in the charging/discharging process. Therefore, it is possible to install lead-acid batteries in the same house with any communication equipment and devices.

All the battery banks that are found in ET BTS sites are lead-acid batteries. The communication equipment and the battery bank are contained in the same shelter and they do not require separate air conditioning for the battery bank. These batteries provided 48V DC power and as illustrated in Figure 2.5 in Chapter 2, the configuration uses 12v batteries in a series-parallel connection. The maximum DC voltage expected from the battery bank is 58.2VDC. When the voltage level gets 48.2VDC, it is termed as BBUV. When it becomes 46.2VDC, LLVD loads will be disconnected and when the voltage level gets 45.2VDC, the BLVD loads will be disconnected in order to protect the battery banks from deep discharge.

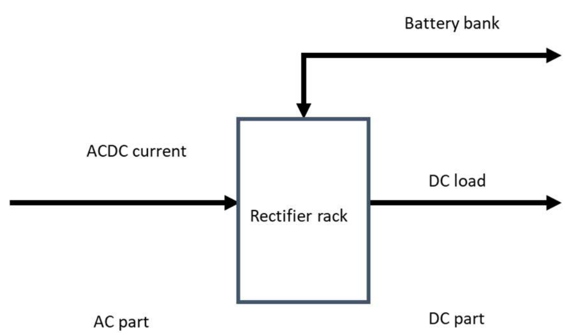

Figure 4.4 Battery bank, DC load, and AC part connection

Source: Adapted from [7]

The battery current is the current measurement taken from the battery output and as indicated in Figure 4.5 below, it may be positive (+ve) or negative (-ve). The +ve sign indicates floating current which means the battery is charging and the load is carried by the commercial power or a generator. When the sign is –ve, the battery handles the load and Alternating Current/Direct Current (AC/DC) becomes zero. AC/DC is an average current that is measured at the junction of the AC part and the DC part. As tried to demonstrate in Figure 4.4, it is an average AC current because it is measured on the three-phase input from the rectifiers. Because the rectifiers have load balancing and current equalization (the maximum current difference between the three phases is 1.5Ampper) capabilities

[7], the current that we find at the three phases are almost equal and represented by a single current measurement.

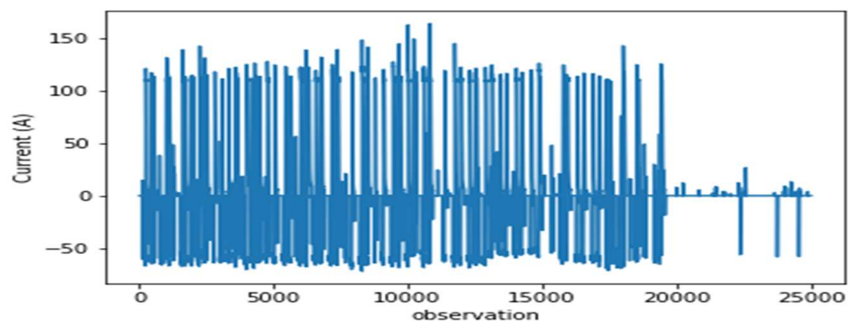

Figure 4.5 Battery current

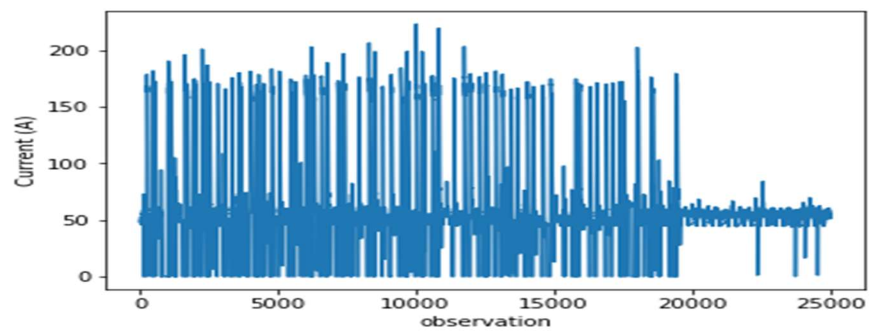

Figure 4.6 AC/DC current

In addition, we find battery temperature which tells about the internal temperature of the battery banks. The battery temperature is highly dependent on the battery current and operating environment [23]. Moreover, the total charge-discharge count and total power supplied by the battery bank starting from the first day of installation respectively may tell the battery current condition.

4.1.3 Load-related Features

Load related features are bus-bar voltage, DC load current, and DC load power which is measured from the bus-bar of the power system. The bus-bar is a copper strip or bar, where DC loads, battery bank, and the rectifier output are connected in parallel. BTS power system failures can be detected from the measurements taken from the bus-bar, especially from bus-bar voltage. As indicated in Figure 4.7 below, failures could be reflected in the bus-bar voltage including BBUV. As noted from Figure 4.8 and Figure 4.9, we cannot easily detect BBUV and LLVD from DC load current and DC load power. We can only detect BLVD from the DC load current and DC load power measurements.

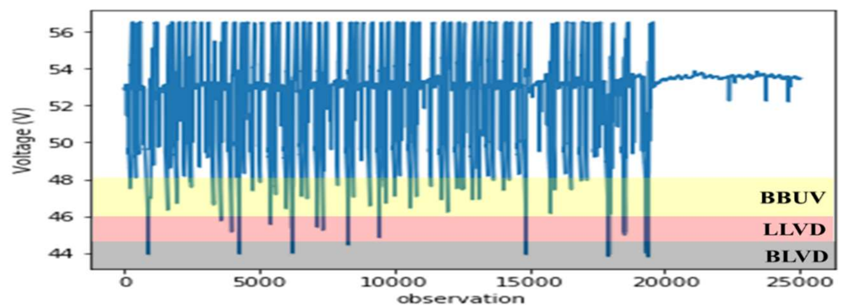

Figure 4.7 Bus-bar voltage

As discussed in Section 2.1, the loads that we find at the BTS have different priorities. For instance, mobile services (GSM, UMTS, and LTE) are termed as LLVD loads. They will be disconnected when the bus-bar voltage level gets around 46.2VDC (the threshold can be reconfigured). BLVD loads are the transmission, MW and optical equipment that serve not only the mobile services exist in the site but also serve other BTSs sites and exchanges.

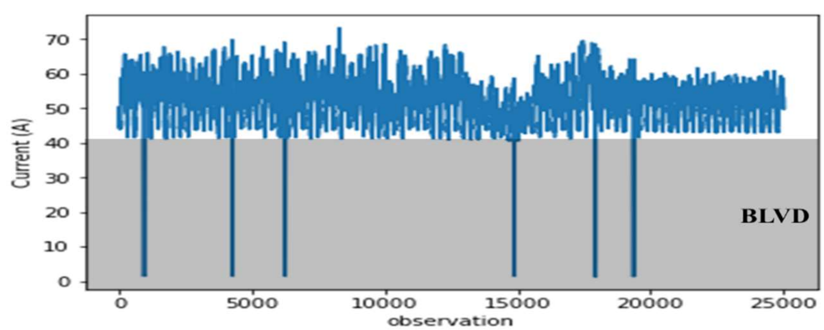

Figure 4.8 DC load current

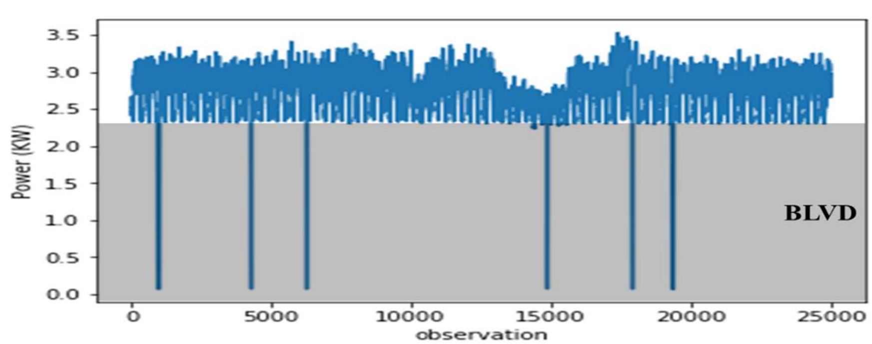

Figure 4.9, DC load power

To validate the above statements regarding load-related measurements, specific alarm and its voltage, current and power measurements are printed and shown in Table 4.1 below, DC load power and DC load current approach to zero and when the bus-bar voltage gets below the configured threshold and LLVD alarm is generated (the loads are disconnected). The alarm is recovered when the bus-bar voltage restored from 47.01VDC to 49.92VDC, the DC load power and DC load current start to raise from 0.07 to 2.81KW and 1.53 to 56.37A, respectively.

Table 4.1 Measurments and alarms.   Source: ET NetEco

Alarm Name	Occurrence Time	Cleared Time	Start Time	Bus Bar Voltage(V)	DC Load Current(A)	DC Load Power(kW)
LLVD	4/22/2019 19:55	4/22/2019 20:25	2019-05-12 10:50	47.01	1.57	0.07
LLVD	5/12/2019 3:41	5/12/2019 11:02	2019-05-12 10:55	47.01	1.53	0.07
LLVD	5/23/2019 21:03	5/23/2019 22:00	2019-05-12 11:00	48.30	1.53	0.07
LLVD	5/30/2019 17:53	5/30/2019 20:45	2019-05-12 11:05	51.09	15.72	0.80
LLVD	6/29/2019 18:14	6/29/2019 18:28	2019-05-12 11:10	49.92	56.37	2.81
			2019-05-12 11:15	47.09	60.51	2.85
			2019-05-12 11:20	48.12	55.75	2.68

4.2 Feature Selection and Extraction

Feature selection is a technique of identifying features that have better information regarding the works to be done. It has a big impact on the success of the neural network because it will minimize the computation time and required resources. If it is not selected wisely, it may diminish the performance of the neural network [28].

There are various methods for feature reduction technics used for neural network approaches. Most widely used are feature selection and feature extraction [28]. Feature selection is used to reject inputs, which is unrelated (having a small correlation) to the output. That means, eliminating the features that have small or no relation with the expected output. These uncorrelated features increase overhead on the neural network. In addition, feature extraction is the method of converting the raw data, to another form or generating new features that can be used for the neural network. Feature selection and extraction have many advantages regarding neural networks. Some of them are:

- Reduce redundant information;
- Reduce the required computational time (wall time);

- Reduce the resources required for the analysis (CPU, GPU, and RAM);
- Compress the amount of data to be used by reducing or extracting new features.

In addition to feature selection and extraction techniques, the amount of data has a big impact on the success of the neural network [28][37]. When the amount of data is small, the neural network may not capture the required information from the features that are used to model the data. Rarely, feature selection and extraction may reduce the performance of the neural network because the data must have diverse information in order to realize good performance and the neural network might learn rare events from the rejected features [28]. In addition to the data size and number futures, understanding the data have a big impact on the success of the neural network [28]. Therefore, all the features should be checked whether they contain outlier and wrong measurements. So, for this research, all the features are plotted and checked whether they are in between the expected range.

4.3 Data Normalization

Data normalization is a method of rescaling the values of the features from one range to a new range. This approach is used to protect one measurement from biasing other measurements while using them for a neural networks. Because both the features are transformed into the same range, the impact between features will be reduced. Most of the time, the normalized values are reduced to between 0 to 1 or -1 to 1. Normalization is also used to optimize the required resources (CPU, GPU, and RAM) for the computation, it reduces the effects of outliers in the data since the process starts from the same scale which has an impact on reducing training and testing time [28]. The performance of the neural network highly affected while training on different data representations (wide range). So normalization increases the performance of a neural network by giving a chance for the training and testing from the same data representation. Nowadays, there are many normalization techniques applied to machine learning. Statistical or Z-score normalization and Min-Max normalization are wildly used for a neural network.

4.3.1 Z-Score Normalization

Z-score normalization is a technic of normalization which transforms the inputs linearly to new output values having 0 mean and standard deviation of 1. In order to apply Z-score normalization, the mean (μ) and standard deviation (σ) for each input feature are required to produce the new data which will have a 0 mean and unit variance. The formula of Z-score for a sample X is expressed as equation (4.1) [37]:

$$Z_i = \frac{X_i - \bar{X}}{S} \tag{4.1}$$

Where $\overline{\mathrm{X}}$ (the sample mean), S (the sample standard deviation), X sample input and Z is Z score value.

4.3.2 Min-Max Normalization

Min-Max normalization is a normalization approach, which will change the input or output feature into a linearly transformed new range values. Most of the time, the rescaled values become in between 0 and 1 or -1 and 1. As explained in the min-max Equation (4.2), $min(x)$ and $max(x)$ are the minimum and maximum values respectively, where x is input or output samples and N is the new normalized values. Therefore, in order to calculate N, i.e. the normalized value of a member of the set of observed values of x, the formula of min-max normalization is expressed as equation (4.2) [37]:

$$N = \frac{x - min(x)}{max(x) - min(x)} \tag{4.2}$$

5. Experiments Setup

5.1 Data Set

The data set used for this research contains measurement information for the BTS power system which is taken from the ET NetEco power monitoring system. Totally, five (5) sites data are collected for twenty (20) weeks (from April to August 2019) by weekly data collection plan. The data set contains eleven (11) features with a five (5) minute sampling period. Every site data set have an equal number of observations with a similar time-stamp. The number of observations found in the dataset has thirty-four thousand one hundred fifty-six (34,156) rows with eleven (11) features. Table 5.1 illustrates the sample data structure of the collected observations for this research. The sample data for the test include measurements taken from both good and bad events.

Data preprocessing is an essential step while working on a data-driven algorithm, which improves the possibilities of achieving good results. Therefore, the data is preprocessed and the process includes removing biased (outliers) sample, normalization and merging the data based on time stamp (since the data is collected on a weekly base with a different spreadsheet).

Table 5.1 Collected data and features   Source: ET NetEco

Start Time	Working Temperature(degC)	Working Humidity (%)	Battery Current (A)	Battery Voltage (V)	Battery Temperature (degC)	Battery Total Discharge Power(Ah)	Battery Total Cycle Times	AC/DC System Output Current(A)	Bus Bar Voltage (V)	DC Load Current (A)	DC Load Power (kW)
2019-04-13 23:55	18.88	58.52	0.00	56.47	18.17	122818.61	426	46.37	56.47	44.86	2.53
2019-04-13 23:50	18.88	61.98	0.00	56.47	18.27	122818.61	426	47.50	56.47	46.15	2.61
2019-04-13 23:45	18.88	65.33	0.00	56.47	18.37	122818.61	426	48.49	56.47	45.84	2.59
2019-04-13 23:40	18.88	62.38	0.00	56.47	18.37	122818.61	426	48.76	56.47	46.14	2.61
2019-04-13 23:35	18.88	58.92	0.00	56.46	18.27	122818.61	426	50.35	56.46	47.53	2.68
2019-04-13 23:30	18.88	59.03	0.00	56.47	18.17	122818.61	426	49.45	56.47	47.58	2.69
2019-04-13 23:25	18.99	65.23	0.00	56.48	18.58	122818.61	426	49.45	56.48	46.82	2.64
2019-04-13 23:20	18.88	62.48	0.00	56.47	18.37	122818.61	426	49.43	56.47	48.07	2.71
2019-04-13 23:15	18.88	59.13	0.00	56.47	18.27	122818.61	426	50.77	56.47	48.21	2.72
2019-04-13 23:10	18.99	56.99	0.00	56.47	18.17	122818.61	426	51.39	56.47	49.93	2.82
2019-04-13 23:05	19.09	64.52	0.00	56.47	18.58	122818.61	426	50.22	56.47	48.52	2.74
2019-04-13 23:00	18.88	62.18	0.00	56.47	18.37	122818.61	426	48.11	56.47	47.88	2.70

5.2 Target Selection

Among those 11 features, target selection has been done by studying the relationship between the selected failures (BBUV, LLVD, and BLVD) and all the features. Therefore, bus-bar voltage, DC load current, and DC load power are found to have a high correlation with the actual failure. However, the bus-bar voltage has been selected as a target feature because of BBUV and LLVD failures are not well reflected on DC load current and DC load power. As indicated in Figure 4.8 and Figure 4.9 in Section 4, only BLVD failure is reflected in DC load current and DC load power. However, as can be noted in Figure 4.7, both BBUV, LLVD and BLVD failures information is reflected on bus-bar voltage.

5.3 Implementation

The implementation of the experiment engaged by applying two types of RNN (LSTM and GRU) machine learning platform, built-in python which has capabilities of learning and test the data automatically. LSTM and GRU are capable of remembering each information through the whole dataset. It is recommended in time-series prediction, because of its capabilities to predict the feature based on the information get from the very past. A sigmoid and linear activation function are selected for testing the results which are a linear and nonlinear activation function respectively. Because linear activation function has a limitation on handling nonlinear dataset so, the sigmoid activation function which is capable to represent nonlinear inputs and outputs is included for the test.

The tests also engaged with two different arrangements. Single site and multiple site tests. First by using a single site test but by varying the data size (starting from 12 weeks through 20 weeks) and by using 5 sites data starting from a single site then go through 5 sites data by using the whole 20 weeks observations. The main reason behind using different data sizes is to check the impact on the performance on LSTM and GRU which are used to predict the future BTS power system failure. As performance measurement criteria, MSE and number of epoch are used to measure the performance of the models.

5.3.1 Single Site Test

First, the proposed RNN (LSTM and GRU) based scheme tested using single BTS site data. Five types of data size are arranged, starting from 12 weeks to 20 weeks (12, 14, 16, 18, and 20). For bothe the tests, linear and sigmoid activation function used separatlly for the model output. Min-max normalization for the output.

5.3.2 Multiple Site Test

In addition to using single BTS site data, the analysis is conducted based on data collected from five (5) different BTS sites (starting from a single site through five BTS data). Again, the primary reason behind using multiple sites data for the training and predict the impending failures on a single site is to test whether using multiple sites for training will give a better chance for the RNNs to understand the target sites and predict coming failure for a single site.

5.4 Parameter Setup

Parameter setup is an important stage, which can make significant changes in machine learning performance. For this research purpose to get the correct parameter to be set, two weeks of data have been used initially to get better values. In addition, RNN hyper-parameters has been applied. The RNN hyper-parameters are suggested values by the developers of the RNN. As can be seen from Table 5.2, 256 recurrent units, 20 epoch, 100 steps per epoch, 144 samples as a sequence length and 288 samples as a batch size are applied.

The number of epoch used for the training is 20, where the maximum number of epoch achieved through the initial (two weeks) test is 15. In addition, the MSE does not improve while using the number of epochs above 20. Therefore, the number of patience has been set to 5 by using early stopping for preventing the model from overfitting. This means stop the training if the performance of the LSTM and GRU stope improving for 5 continuous epochs.

Table 5.2 Parameters setup.

Parameters	Values	Remarks
Number of recurrent unit	256 unit	Hyper parameters and by test
Number of epoch	20 epoch	By test
Step per epoch	100	Hyper parameters and by test
Sequence length	144 samples	12 hours
Bach size	288 samples	24 hours

In addition to the above parameters, a feature reduction technic is applied and compared the feature reduced data results with the results found without feature reduction. For the purpose of feature reduction, Pearson's correlation analysis between all features and the target features has been done.

The correlation results are presented in Table 5.3 and as Pearson's correlation criteria state that, features having results values between $\pm$ (0.5 to 1) have a strong correlation, $\pm$ (0.3 to 0.5) medium and $\pm$(0.1 to 0.3) small (weakly) correlated. The negative (-ve) sign indicates inverse correlation while positive (+ve) sign indicates a direct correlation. The feature reduced data contains 6 features including the target feature bus-bar voltage.

Table 5.3 Pearson's correlation results between features.

Features	Working Temperature	Working Humidity	Battery Current	Battery Voltage	Battery Temperature	Battery Total Discharge Power	Total Battery Cycle	ACDC Current	Bus-bar Voltage	DC Load Current	DC Load Power
Working Temperature	1	-.832**	-.043**	-0.01	.794**	-.098**	.093**	.171**	-0.354	.270**	.258**
Working Humidity	-.832**	1	0.01	-.026**	-.847**	.246**	-.027**	-.134**	.037**	-.179**	-.171**
Battery Current	-.043**	0.01	1	.485**	.040**	0.007	0.001	.697**	.529**	-.101**	-.018**
Battery Voltage	-0.01	-.026**	.485**	1	-.019**	-.050**	-.037**	.724**	.973**	.477**	.558**
Battery Temperature	.794**	-.847**	.040**	-.019**	1	-.333**	-.048**	.123**	-.089**	.110**	.103**
Battery Total Discharge Power	-.098**	.246**	0.007	-.050**	-.333**	1	.809**	-0.004	.044**	-0.008	-0.006
Total Battery Cycle	.093**	-.027**	0.001	-.037**	-.048**	.809**	1	.012*	.024**	.016**	.016**
ACDC Current	.171**	-.134**	.697**	.724**	.123**	-0.004	.012*	1	.777**	.634**	.694**
Bus-bar Voltage	-0.354	.037**	.529**	.973**	-.089**	.044**	.024**	.777**	1	.506**	.587**
DC Load Current	.270**	-.179**	-.101**	.477**	.110**	-0.008	.016**	.634**	.506**	1	.994**
DC Load Power	.258**	-.171**	-.018**	.558**	.103**	-0.006	.016**	.694**	.587**	.994**	1

As clearly seen from the correlation result in Table 5.3, working temperature, battery current, battery voltage, AC/DC current, DC load current, and DC load power have a better correlation result with the target feature bus-bar voltage. Since, battery current and AC/DC current has a strong correlation with each other, using two of them will become redundant for the RNN (LSTM and GRU) and will increase the overhead. Therefore, AC/DC current is selected to be part of the feature reduced data because it has a higher correlation with the target feature than battery current.

The proposed RNN (LSTM and GRU) based algorithms are compared with two different activation functions configured for the output. These activation functions are linear and sigmoid. Furthermore, after the minimum and maximum values checked, min-max normalization has been used for the input because the minimum and maximum values found are -91.16 and 125684.63 respectively. Where -91.16 is battery current in ampere and 125684.63 is the total power supplied by the battery (rated in AH).

The RNN will give output data, which is in the range of 0 to 1 because the normalization technique has been used for the inputs, and there is an activation function on every neuron, therefore, the output should be recovered (to the actual human interpretable values) by post-processing of the algorithm output, by using output coding on the target value. Generally, these are the types of configuration used to evaluate algorithms performance:

- Linear activation function with feature reduction using LSTM;
- Linear activation function without feature reduction using LSTM;
- Sigmoid activation function without feature reduction using LSTM;
- Sigmoid activation function with feature reduction using LSTM;
- Linear activation function with feature reduction using GRU;
- Linear activation function without feature reduction using GRU;
- Sigmoid activation function without feature reduction using GRU;
- Sigmoid activation function with feature reduction using GRU.

Out of 20 weeks of data, about 10% of the observation is used as the test sample and 90% of observation is used for the training (number of training and test observation are around 30,156 and 3,400 respectively).

6. Results and Analysis

6.1 Results with Single Site Test

During single site tests, the training and testing process engaged for forty (40) times using five (5) types of data arrangements. Using these data, linear and sigmoid activation functions configured separately for the output while min-max normalization used for the input. In addition, feature reduction techniques applied to test the model performance before feature reduction (BFR) and after feature reduction. The size of single site data is presented in Table 6.1.

Table 6.1 Dataset size for single site test.

	BFR		AFR	
	Number of features	Number of rows	Number of features	Number of rows
Weeks	25024	11	25024	6
14	27328	11	27328	6
16	29632	11	29632	6
18	31648	11	31648	6
20	34156	11	34156	6

Table 6.2 shows the results found for linear activation function at the output of both LSTM and GRU. As a comparison criterion, MSE and number of epoch is used. The test engaged by single site data starting from12 weeks off and goes through 20 weeks of observation. The major reason behind using different data sizes is to check the impact on the model performance and results found in addition to using different activation functions.

Table 6.2 Results using a linear activation function.

Linear Activation Function								
	GRU				**LSTM**			
Number of Week	**Number of Epoch BFR**	**MSE BFR**	**Number of Epoch AFR**	**MSE AFR**	**Number of Epoch BFR**	**MSE BFR**	**Number of Epoch AFR**	**MSE AFR**
12	9	0.0008605	6	0.0014	13	0.0093	11	0.0056
14	9	0.006	6	0.0045	15	0.00167	9	0.0034
16	14	0.0006919	7	0.0014	13	0.00198	12	0.0091
18	10	0.00043	7	0.000611	14	0.00123	12	0.0019
20	9	0.0011	7	0.000494	13	0.0098	12	0.0013

Table 6.3 shows the results found while the sigmoid activation function used at the output of LSTM and GRU in the test. Again, MSE and number of epoch is used for the performance measurements.

For the test, single site dataset has been applied but the data size starts from 12 weeks and go through 20 wiks.

Table 6.3 Results using a sigmoid activation function.

Sigmoid Activation Function								
	GRU				LSTM			
Number of Week	**Number of Epoch BFR**	**MSE BFR**	**Number of Epoch AFR**	**MSE AFR**	**Number of Epoch BFR**	**MSE BFR**	**Number of Epoch AFR**	**MSE AFR**
12	8	0.0012	6	0.00012	11	0.0094	10	0.0019
14	12	0.000171	6	0.000437	12	0.0065	11	0.00091
16	6	0.000127	7	0.0000763	10	0.0053	9	0.00021
18	8	0.000272	6	0.0000668	12	0.0071	9	0.00017
20	12	0.000414	7	0.0000632	11	0.0092	10	0.00011

6.2 Results with Multiple Sites Test

After setting all the parameters as previous but only by using GRU with sigmoid activation function and applying feature reduction, results have been checked for multiple sites dataset.

Table 6.4 Dataset size for multiple site test.

Number of sites	Number of rows	Number of features
1	34156	6
2	34156	12
3	34156	18
4	34156	24
5	34156	30

As can be seen from Table 6.4 above, the test has been done by using five (5) BTS power system data for training the GRU and predict a single BTS power system site failure. The number of rows used is the same for both the tests which have 34156 rows. But the number of features used are increasing as the number of sites increased starting from 6 features go through 30 features.

In multiple site tests, the reason behind using multiple sites data is to check whether using multiple sites for the training gives a better chance for the GRU to model the data and predict single BTS power system failure or does the data collected from multiple BTSs power systems have information about other BTS power system?.

Table 6.5 shows the MSE and the number of epoch results found for multiple site tests. For performance measurement criterion, MSE and number of epoch are used.

Table 6.5 MSE and number of epoch using multiple sites for training.

N0 of Site	Site 1		Site 2		Site 3		Site 4		Site 5	
	MSE	Epoch	MSE	Epoch	MSE	Epoch	MSE	Epoch	MSE	Epoch
1	0.000309	10	0.0076	6	0.000069	10	1.32E-05	10	0.0185	6
2	0.000828	8	0.008	9	7.04E-05	10	2.71E-05	7	0.0184	7
3	0.001	6	0.008	10	0.000069	10	1.74E-05	10	0.0191	6
4	0.002	7	0.008	10	7.29E-05	9	0.000016	10	0.0187	10
5	0.0016	8	0.0079	7	8.44E-05	9	0.000019	8	0.0195	8

6.3 Findings and Discussion

6.3.1 Single Site Findings

As can be observed from the MSE and the number of epoch results for single site tests, GRU archived better performance than LSTM specially AFR, even the activation is linear or sigmoid. From the activation points of view, sigmoid achieved better than linear. Even the LSTM gets good results AFR using sigmoid activation function, it swings and very difficult to generalized using a linear or sigmoid activation function and BFR or AFR for LSTM. As shown from the MSE plot above in Figure 6.1, MSE results after 16 weeks data seem to decrease especially for the GRU.

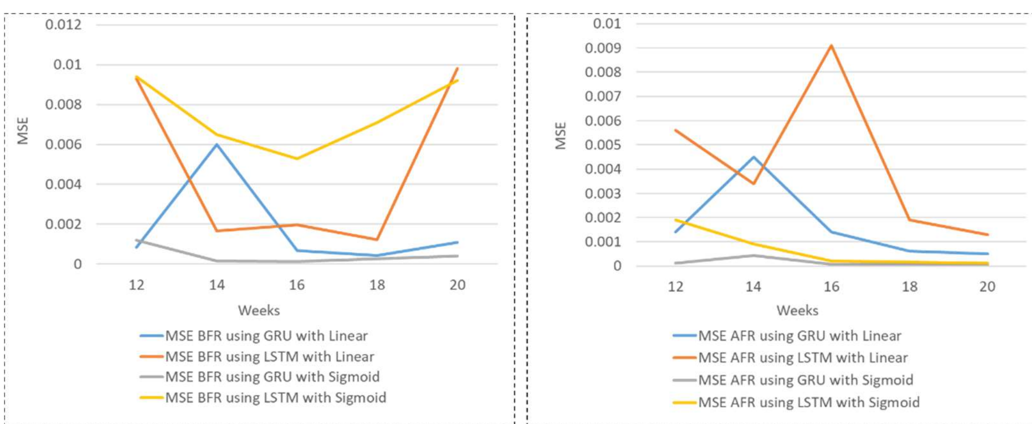

Figure 6.1 MSE using GRU or LSTM with sigmoid and linear activation function.

For all the tests, feature reduction achieved better results, especially for GRU. As can be seen clearly from all the results, using linear activation functions achieved worse MSE values than the sigmoid activation function. What is more, Figure 6.2 shows the MSE enhanced as the data size (the number of weeks) increases for a single site test. These MSE improvements realized whether the activation function is linear or sigmoid.

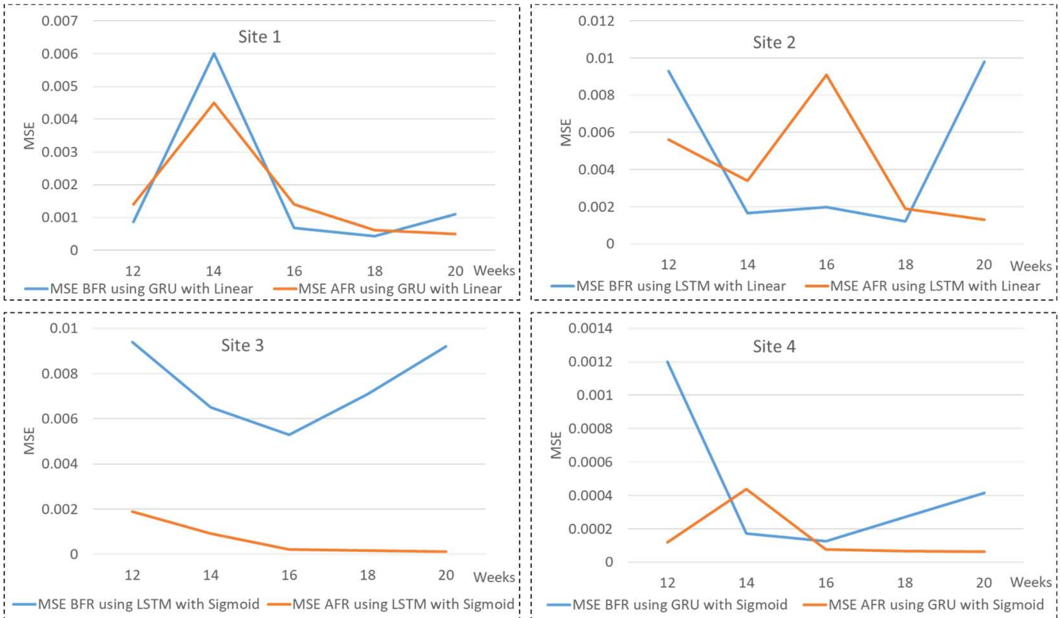

Figure 6.2 MSE of LSTM and GRU wit BFR or AFR.

As can be noted from Figure 6.2 above and detailed observation shows inTable 6.2 and Table 6.3, the minimum MSE achieved is $6.32e^{-5}$ using GRU with feature reduction by 20 weeks data with a sigmoid activation function. The maximum MSE achieved is $9.8e^{-3}$ using LSTM before feature reduction, using 20 weeks of data by a linear activation function.

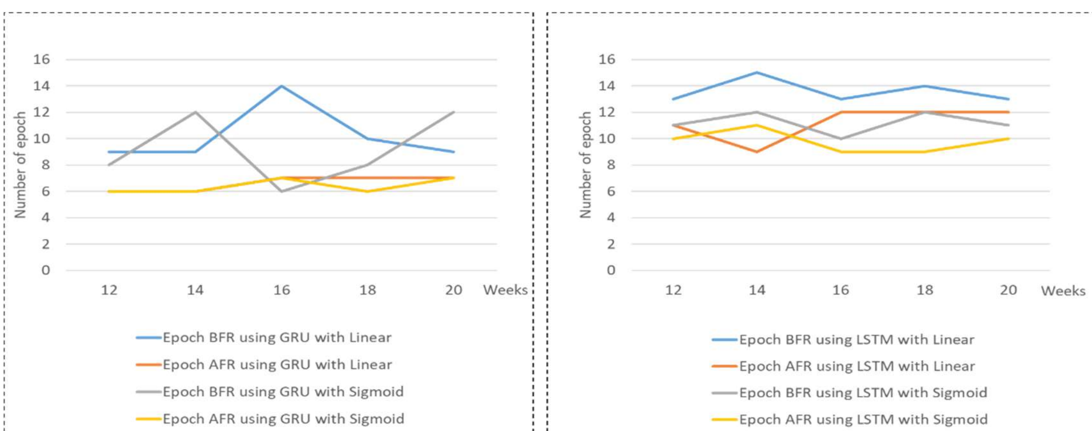

Figure 6.3 Number of epoch BFR or AFR with linear and sigmoid activation function.

The number of epoch found is very difficult to generalize for both GRU or LSTM, sigmoid or linear, and BFR or AFR. From the number of epoch sides, as can be noted from Figure 6.3, the minimum epoch achieved is 6 that is using LSTM by 12 and 14 weeks of data with linear activation function after feature reduction, using GRU by 12, 14, and 18 weeks of data after feature reduction, and using GRU by 16 weeks data before feature reduction. The maximum epoch achieved is 15 using LSTM before feature reduction by a linear activation function.

6.3.2 Multiple Sites Findings

MSE and number of epoch is used as comparison criterion for the multiple sites as a single site experiment. As can be observed from the MSE and the number of epoch results for multiple site tests from Table 6.5 the MSE is increasing as the number of BTS power system data increase to predict a single BTS power system failure. However, it is very difficult to generalize from the number of epochs points of view because it is around the same range.

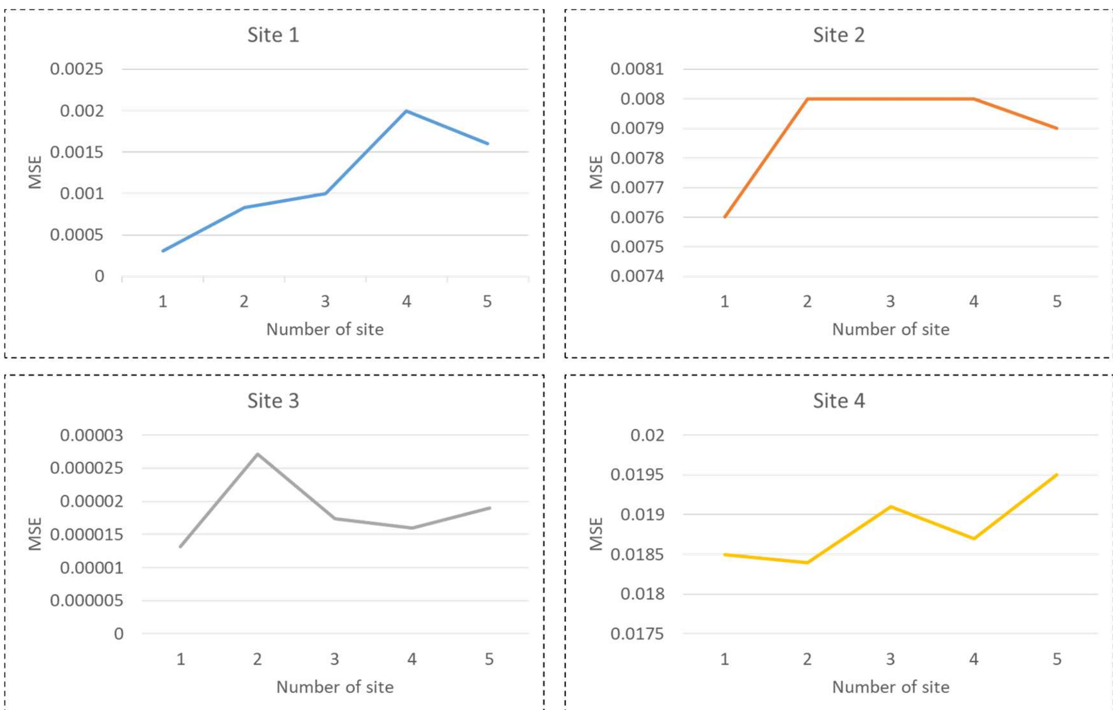

Figure 6.4 MSE using multiple site data for training and predicting single site failure.

As can be noted from Table 6.5 above, the minimum MSE achieved is $1.32e^{-5}$ using single site test. The maximum MSE achieved is $1.95e^{-2}$ using five sites for training and predict single site power system failure.

The major reason behind using multiple BTS power system site data for predicting single BTS power system failure is to check the impact on the model results. As understand from the result found, it seems very difficult to use different BTS power systems data to learn the model for the prediction. The possible cause of increasing MSE while using multiple BTSs power system data for the training to and predict single BTS power system failure, the BTS power system data of one site have different information with other BTS power system data which may impact the training and testing.

6.3.2 Discussion

After the GRU and LSTM is configured with similar parameters, MSE and the number of epoch have been used as comparison criteria. In order to enhance the model performance, data cleaning, activation function, normalization, early stopping, and output codding is applied. In addition, the model is tested by feature reduction and without feature reduction techniques. Single site MSE results are presented in Figure 6.5.

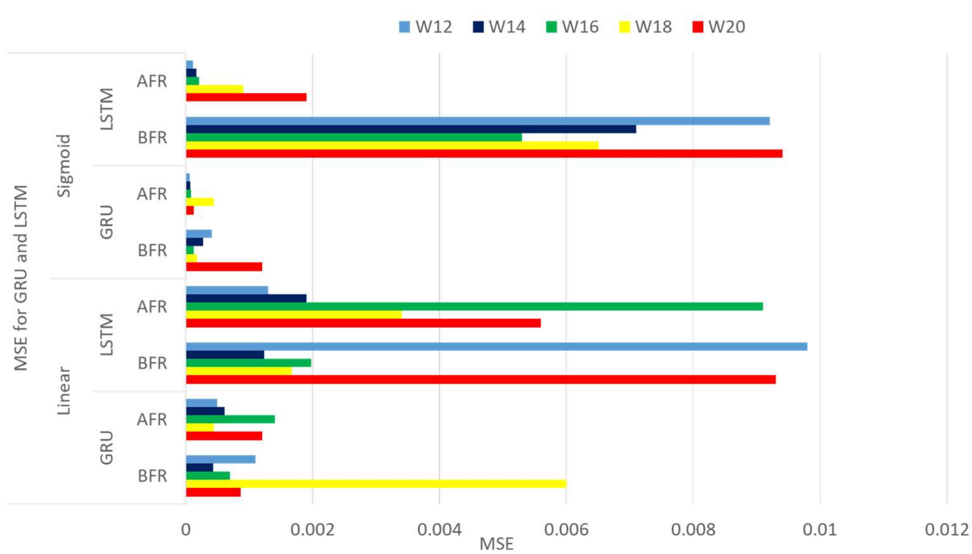

Figure 6.5 Single site test MSE results.

The tests have been also engaged by using different data sizes and configurations. As learned from the results, both LSTM and GRU can predict BTS power system failure using history time serious data even GRU achieved better results than LSTM. On the single site and multiple site tests, GRU shows good performance for all the prediction of BTS power system failure.

From the single site test points of view, the sigmoid activation function achieved better results compared with linear. This might show that the dataset used to have a nonlinear property which decreases the MSE results found for both the tests. Another observation is that the feature reduction technique reduces the number of epoch required to do BTS power system failure prediction work.

The possible reason behind the MSE increase for before feature reduction applied is, there are be features that may bias the BTS power system failure prediction work. An additional remark is that the feature reduction technique can improve the performance of the LSTM and GRU by reducing unrelated data that contain less information and redundant data, which may introduce a bias between the features. Again, for the number of epoch, feature reduction has an impact on achieving smaller number of epochs.

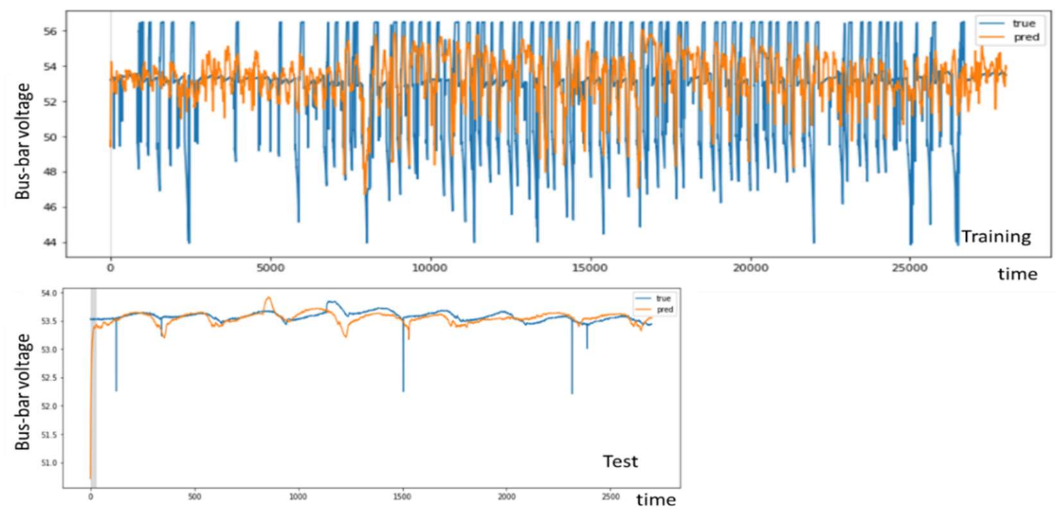

Figure 6.6 Comparison plot, training, and testing.

Moreover, the predicted result of bus-bar voltage values is plotted with the actual value for the purpose of compression. As can be seen from Figure 6.6, the blue line indicates the true (actual) values in the training and testing. The orange lines represent the predicted values in the training and testing. As demonstrated in Figure 6.6, model train well and the result coincides with the real data even they have a limitation on touching the upper and lower limits. This limitation also reflected in the test (validation) plot. Even it is difficult to check performance differences visually between GRU and LSTM from the plots, as the MSE and number of epoch tell the GRU performs better.

7. Conclusion and Future Works

7.1 Conclusion

Mobile and fixed telecommunication infrastructures are increasing with the development of new technology and increasing demand. Thereby the maintenance, monitoring, and resource assignment of the whole system become challenging. Especially taking care of the infrastructure with a traditional maintenance activity requires considerable effort and corrective maintenance will diminish the performance of the whole system and leads to poor resource utilization.

A power system is a backbone for the entire communication infrastructure and its failure may cause an interruption to the complete services. Therefore, following good maintenance procedures will have a big impact on the whole service provisioning. Nowadays, maintenance trends shift from reactive maintenance to proactive maintenance, which is taking action for the impending problems by using prediction techniques.

As illustrated in Section 1.1, the BTS power system failure takes the biggest share of the BTS service interruption. Therefore predicting these failures before they happen to some degree will have a big benefit for guarantying QoS and in that way enhance return on investments.

As learned from the results found, both LSTM and GRU can predict BTS power system failure using history time serious data. The minimum MSE achieved is 0.0000632 for GRU with feature reduction using 20 weeks data with 7-epoch. Even both LSTM and GRU can predict the impending failure in the BTS power system, GRU with sigmoid activation function with reduced features obtained minimum MSE and number of epoch.

7.2 Future Works

Many different data preprocessing, prediction methods and experiments left for the future because the data collection process has its own challenge. Therefore, future work requires a detailed analysis of different methods. This thesis work primarily motivated on the use of RNN types called LSTM and GRU for predicting the impending failure of the BTS power system but trying other methods is highly suggested because there may be a better approach regarding different performance measurement criteria’s. Thereby, it will supplement the research area (BTS power system) because there is a limitation on finding related studies. In addition, the following ideas may have a better contribution to improving the results found:

- It could be better to consider other BTS subsystems (transmission and MW) data for predicting the BTS failure and root causes;
- Obviously, using other types of prediction approaches has an advantage because it gives an opportunity to select for practical implementation;
- In addition, the preliminary results of these experiments may be improved if it is tested with different data preprocessing and parameter tuning.

8. Reference

[1] P. Release, "Network Infrastructure Market Size , Share," 2019. [Online]. Available: https://www.marketwatch.com/press-release/network-infrastructure-market-size-share-2019. [Accessed: 10-Dec-2019].

[2] E. U. Warriach, T. Ozcelebi, and J. J. Lukkien, "A Comparison of Predictive Algorithms for Failure Prevention in Smart Environment Applications," *Proc. - 2015 Int. Conf. Intell. Environ. IE 2015*, no. 1, pp. 33–40, 2015.

[3] S. Thanakodi, N. S. M. Nazar, N. F. Joini, H. D. M. Hidzir, and M. Z. K. Awira, "Power plant fault detection using artificial neural network," *AIP Conf. Proc.*, vol. 1930, no. February, 2018.

[4] T. Chalermarrewong, T. Achalakul, and S. C. W. See, "Failure prediction of data centers using time series and Fault Tree Analysis," *Proc. Int. Conf. Parallel Distrib. Syst. - ICPADS*, pp. 794–799, 2012.

[5] A. Ai, T. Data, S. Process, and M. Docs, "Azure AI guide for predictive maintenance solutions." [Online]. Available: https://docs.microsoft.com/en-us/azure/machine-learning/team-data-science-process/cortana-analytics-playbook-predictive-maintenance. [Accessed: 21-Apr-2019].

[6] N. Singh and G. Data, "Artificial Neural Networks Applications and Algorithms," 2019. [Online]. Available: https://www.xenonstack.com/blog/artificial-neural-network-applications/. [Accessed: 10-Oct-2019].

[7] ZTE Corporation, "ZXDU68 T601 DC Power System," Shenzhen, Version 4.0, 2008.

[8] A. S. Douglas Okafor Chukwuekwe, Per Schjølberg, Harald Rødseth, "Reliable , Robust and Resilient Systems : Towards Development of a Predictive Maintenance Concept within the Industry 4 . 0 Environment," Euromaintenance 2016 Conf., no. June, p. 10, 2016.

[9] H. Wang, "Prediction of Electrical Equipment Failure Rate Based on Improved Drosophila Optimization Algorithm," *2017 13th Int. Conf. Nat. Comput. Fuzzy Syst. Knowl. Discov.*,

no. 20160204022, pp. 1915–1921, 2017.

[10] A. Abraham, “Genetic Programming Approach for Fault Modeling of Electronic Hardware,” no. 1, pp. 1563–1569, 2005.

[11] J. Ma, W. Song, J. Kang, and X. Tian, “Equipment failure control model based on failure life cycle,” in *Proceedings of 2012 International Conference on Quality, Reliability, Risk, Maintenance, and Safety Engineering, ICQR2MSE 2012*, 2012.

[12] A. Chigurupati, R. Thibaux, and N. Lassar, “Predicting Hardware Failure Using Machine Learning,” *2016*.

[13] C. Haseltine and E. E. S. Eman, “Prediction of power grid failure using neural network learning,” *Proc. - 16th IEEE Int. Conf. Mach. Learn. Appl. ICMLA 2017*, vol. 2018-Janua, pp. 505–510, 2018.

[14] K. Venugopal, “Artificial Neural Network based Fault Prediction Framework for Transformers in Power Systems,” *Proc. IEEE 2017 Int. Conf. Comput. Methodol. Commun.*, no. Iccmc, pp. 520–523, 2017.

[15] F. Wang, Z. Mi, S. Su, and C. Zhang, “A practical model for single-step power prediction of grid-connected PV plant using artificial neural network,” *2011 IEEE PES Innov. Smart Grid Technol. ISGT Asia 2011 Conf. Smarter Grid Sustain. Afford. Energy Futur.*, pp. 0–3, 2011.

[16] N. Nadai, A. H. A. Melani, G. F. M. Souza, and S. I. Nabeta, “Equipment failure prediction based on neural network analysis incorporating maintainers inspection findings,” *Proc. - Annu. Reliab. Maintainab. Symp.*, 2017.

[17] L. T. Mar, “Application Of ANN To Power System Fault Analysis,” pp. 269–273, 2002.

[18] J. Zheng, C. Xu, Z. Zhang, and X. Li, “Electric load forecasting in smart grids using Long-Short-Term-Memory based Recurrent Neural Network,” *2017 51st Annu. Conf. Inf. Sci. Syst. CISS 2017*, pp. 1–6, 2017.

[19] N. K. Pal and B. J. Ifeanyi, “Technical overview of all sources of electrical power used in

BTSs in Nigeria," pp. 18–30, 2017.

[20] M. P. Mahbub, S. M. Abedin, S. R. Sabuj, and A. Chakrabarty, "A possible alternative power solution of base stations in Bangladesh," *Int. Conf. Energy Power Eng. Power Progress, ICEPE 2019*, pp. 1–5, 2019.

[21] "Huawei LLD for Addis Abeba, Ethiopia." Huawei, 2014.

[22] Resistorguide, "Shunt Resistor for Current Measuring," 2019. [Online]. Available: http://www.resistorguide.com/shunt-resistor/. [Accessed: 05-Oct-2019].

[23] A. Interface, *IEEE Recommended Practice for Installation Design and Installation of Vented Lead-Acid Batteries for Stationary Applications*, vol. 2002, no. August. 2003.

[24] Huawei, "P&E System Operation and Maintenance Training Manual," pp. 1–436, 2010.

[25] Techopedia, "Base Transceiver Station (BTS)," *Janalta Interact. Inc.*, pp. 1–12, 2010.

[26] L. Informatik and V. Technische, "Fault Analysis and Fault Tolerance of a Base Station System for Mobile Communications."

[27] C. Stergiou and D. Siganos, "Artificial Neural Networks," vol. 606, no. 66707180, pp. 602–606, 2018.

[28] K. L. Priddy *et al.*, *Artificial Neural Networks_ An Introduction*. Bellingham, Washington USA: SPIE, Bellingham, Washington, 2005.

[29] MissingLink, "7 Types of Neural Network Activation Functions: How to Choose? - :," *Missing Link*, 2018. [Online]. Available: https://missinglink.ai/guides/neural-network-concepts/7-types-neural-network-activation-functions-right/. [Accessed: 10-Jun-2019].

[30] N. N. Concepts, "7 Types of Neural Network Activation Functions : How to Choose ?," pp. 1–29, 2019.

[31] S. Sharma, "Epoch, Batch Size, Iterations and Wall Time," *Towards Data Science*, 2017. [Online]. Available: https://towardsdatascience.com/epoch-vs-iterations-vs-batch-size-4dfb9c7ce9c9. [Accessed: 02-Sep-2019].

[32] J. T. Connor, R. D. Martin, and L. E. Atlas, “Recurrent Neural Networks and Robust Time Series Prediction,” *IEEE Trans. Neural Networks*, vol. 5, no. 2, pp. 240–254, 1994.

[33] I. Goodfellow, B. Yoshua, and A. Courvill, *Deep Learning*. MIT press, 2016.

[34] A. Zheng and A. Casari, *Feature Engineering for Machine Learning and Data Analytics - Principles and Techniques for Data Scientists*. 2018.

[35] ASHRAE Technical Committee 9.9, “2008 ASHRAE Environmental Guidelines for Datacom Equipment: Expanding the Recommended Environmental Envelope,” 2008.

[36] F. Wan, D. Swenson, M. Hillstrom, D. Pommerenke, and C. Stayer, “The effect of humidity on static electricity induced reliability issues of ICT equipment in data centers - Motivation and setup of the study,” *ASHRAE Trans.*, vol. 119, no. PART 2, pp. 341–357, 2013.

[37] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An introduction to statistical learning with R 441pges*. .