Addis Ababa University
Addis Ababa Institute of Technology
School of Electrical and Computer Engineering
Telecommunication Engineering Graduate Program

Machine Learning for Power Failure Prediction in Base Transceiver Stations: A Multivariate Approach

By: Sofia Ahmed

Advisor: Dr. -Ing. Dereje Hailemariam

A Thesis Submitted to the School of Electrical and Computer Engineering in Partial Fulfillment of the Requirements for the Degree of Master of Science in Telecommunication Engineering

October 4, 2023

Addis Ababa, Ethiopia

Addis Ababa University;
Addis Ababa Institute of Technology;
School of Electrical and Computer Engineering
Telecommunication Engineering Graduate Program

Machine Learning for Power Failure Prediction in Base Transceiver Stations: A Multivariate Approach

By: Sofia Ahmed

Advisor: Dr. -Ing. Dereje Hailemariam

Signed by Examining Committee:

Dr. -Ing. Dereje Hailemariam	___________	___________
Advisor	Signature	Date
Dr.Yalemzewud Negash	___________	___________
Examiner	Signature	Date
Dr. Bisrat Derebssa	___________	___________
Examiner	Signature	Date

____

Dean, School of Electrical and Computer Engineering

Declaration

I, the undersigned, declare that this thesis is my original work, 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.

__________-

Signature

Sofia Ahmed Mohammed

October 4, 2023

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

__________-

Signature

Dr. -Ing. Dereje Hailemariam

October 4, 2023

Abstract

The proliferation of mobile cellular networks has had a transformative impact on economic and social activities. Base transceiver stations (BTSs) play a critical role in delivering wireless services to mobile users. However, power failures in BTSs can pose significant challenges to maintaining uninterrupted mobile services, leading to inconveniences for users and financial losses for service providers.

This thesis introduces a novel approach to mitigating power system interruptions in BTSs using a machine learning-based power failure prediction framework. The framework leverages multivariate time-series data collected from the BTS power and environmental monitoring system. The methodology aims to preemptively predict power failures using three advanced machine learning techniques, specifically, Convolutional Neural Networks (CNNs), Long Short-term Memory (LSTM), and CNN-LSTM networks. These methods excel in capturing complex temporal relationships inherent in time-series data.

All the three algorithms reasonably capture the temporal patterns in the data. However, the LSTM model consistently outperforms the other two models having a MSE of 0.001 and 1.194 MAPE, albeit with longer training times which is more than three hours. On the other hand, the CNN-LSTM model stands out for its efficient training process, which takes notably less time than the LSTM model around two hours training time resulting 0.001 MSE and 2.528 MAPE. Furthermore, the CNN model takes notably less time to compute than the other two models with a prediction performance of 0.223 MSE and 2.843 MAPE.

essential to highlight that this study concentrates on the predictive aspect, which contributes significantly to the field by offering a robust and effective predictive model tailored specifically for BTS power systems. By enabling timely maintenance actions and minimizing downtime, our proposed methodology holds the possibility to significantly improve the reliability of telecommunications infrastructure, which will ultimately lead to better user experiences and streamlined service provider operations.

Keywords: BTS, Power failure, Deep learning, LSTM, CNN, hybrid CNN-LSTM Model,

Acknowledgements

To begin with, I express my gratitude to the Divine "Allah" for the excellent health and overall wellness that allowed me to complete this thesis.

Furthermore, I would like to convey my sincere gratitude to my mentor, Dr. -Ing. Dereje Hailemariam, for his consistent and invaluable assistance during my research expedition. His enduring guidance, remarkable patience, encouragement, and profound knowledge have been immeasurable. His mentorship has been instrumental in shaping this thesis, and I deeply appreciate his significant contributions, which have made this achievement possible.

Apart from my mentor, I extend my sincere gratitude to the individuals on my thesis committee: Dr. Yalemzewd Negash and Dr. Bisrat Derebssa. I deeply appreciate their invaluable perspectives, positive feedback, motivation, and support.

I also wish to express my gratitude to Dr. Tsegamlak Terefe for his valuable input and recommendations in this thesis. I would also like to acknowledge my colleagues and friends who offered valuable perspectives and moral backing during this endeavor. Their productive conversations and support have had a significant influence on the direction of my research.

I also want to convey my deep appreciation and admiration for my dear husband, Yesuf Ahmed, his constant encouragement and belief in my ability have been the driving force behind my achievements.

In conclusion, I want to express my gratitude to my parents, my young Adnu, my siblings, and my extended family for their steadfast support, patience, and empathy throughout the challenges and triumphs of this journey.

Contents

List of Figures

List of Tables

Acronyms

AC	Alternating Current
ACDB	Alternating Current Distribution Board
ANN	Artificial Neural Network
ARMA	Auto Regressive Moving Average mode
ATS	Automatic Transfer Switch
BLVD	Battery Low Voltage Disconnect
BTS	Base Transceiver Station
CN	Core Network
CNN	Convolutional Neural Networks
DC	Direct Current
DG	Diesel Generator
DL	Deep Learning
FFNN	Feed forward Neural Network
HLD	High Level Design
LLD	Low Level Design
LLVD	Load Low Voltage Disconnect
LSTM	Long Short Term Memory
MAPE	Mean Absolute Percentage Error
ML	Machine Learning
MSE	Mean Squared Error
NNOC	National Network Operation Center
OPEX	Operational Expenditure
O&M	Operation and Maintenance
PSU	Power Supply Unit
QoE	Quality of Experience
QoS	Quality of Service
RAN	Random Access Network
RCA	Root Cause Analysis
PSU	Power Supply Unit
ReLU	Rectified Linear Unit
RMSE	Root Mean Square Error
TT	Trouble Thicket
UNMS	Unified Network Management

0.1 Introduction

0.1.1 Background

In this digital era, networking is becoming essential in our daily lives. The vast majority of network functionality is carried out via cellular communication, which eliminates the need for wires or physical connections. The architecture of cellular systems was defined by the Third Generation Partnership Project (3GPP), and it consists of two components: the access network and the core network (CN). In the access network, there is a base transceiver station (BTS), which is responsible for interconnecting client devices to the CN through a radio interface, whereas the CN is responsible for connecting the access network to other networks. Hence, BTS can facilitate wireless communication between subscriber devices and telecom operators' networks [1].

As the number of mobile cellular network users increases, network operators densify BTSs to reach the service to their customers. BTSs consist of different components. These include an optical distribution frame (ODF) for a fiber-optic link to the central switching office, microwave devices for point-to-point communication, transmission routers to separate traffic to different services and RAN technologies, a baseband unit (BBU) for signal processing and radio resource management, a remote radio unit (RRU) for radio frequency (RF) signal transmission. Moreover, BTSs consist of air conditioning devices to maintain the room temperature.

All these components obtain power from a common source to accomplish their intended tasks. If power fails at a site, all the components' functionality is interrupted. Although BTSs service can be interrupted due to various reasons, like power system failure, microwave alignment issues, optical fiber cut, natural disasters and any other reasons, the root cause analysis report obtained from ethio telecom's Unified Network Management System (UNMS) indicates that power failure accounts for the major part [2]. A power failure may happen because of loss of commercial power or mains failure, failure of consequent backup power system, or the failure of power related equipment such as rectifiers, circuit breakers, or the fuses that are connected between the power input and the communications equipment. On the other hand, a power failure can be triggered by lightning or heavy rain, which can generate voltage fluctuations that result in a power surge and damage the rectifier or open circuit breakers [3].

Since BTS power failure is one of the critical issues for network operators, failing to handle this power failure results in degradation of quality of service (QoS), service availability, customer dissatisfaction, and high operational expenditure (OPEX). Finally, it results in a loss of revenue and economic loss to both mobile service users and network operators. This problem is worse if the site is a hub site because its failure results in the failure of other sites too. Also, in remote areas, the problem persists for a long time until technicians travels to reach the site for troubleshooting, resulting in a higher mean time to repair and maximum operational cost. So, to handle these challenges, failure prediction in advance is important for performing predictive maintenance, preventing the occurrence of failure and minimizing maintenance costs.

Failure prediction is the process of predicting whether a material system of interest will fail at a certain point in time in the future. Hence, failure prediction and detection in advance in any industrial system is an important task to assure reliable service by developing a machine learning model [4]. Nowadays, data-driven failure prediction is becoming popular and being implemented in different fields by using different types of machine learning algorithms. For example, [5] proposed machine learning models with Artificial Neural Networks, Logistic Regression and support vector machines to predict water pipe failures by utilizing the historical data of a large water supply network. On the other hand [6] implemented Multilayer Perceptron, Support Vector Regression and Linear Regression machine learning algorithms to develop a prediction model for aircraft system failure.

Predictive maintenance has recently gained popularity due to the seamless developments in contemporary technology [7]. But not much work has been done on the telecom power system. This research focused on proactive principles that can be implemented with a failure prediction technique using machine learning algorithms based on data extracted from Ethio telecom power and environment monitoring systems.

0.1.2 Statement of the Problem

The reliable operation of BTS sites is crucial for maintaining seamless communication networks. However, power failures can significantly interrupt network services and lead to user dissatisfaction [8]. Handling power failure in traditional way, which is solving the failure after its occurrence or by performing reactive maintenance would result in an extended period of unplanned down time, high emergency repair cost, poor customer satisfaction and loss of revenue. Therefore, there is a growing need for an advanced predictive system that leverages the potential of machine learning techniques and multivariate data to enhance the accuracy and timeliness of power failure predictions in BTS sites.

Predicting power failures in BTS sites offers several advantages that can significantly enhance the reliability and efficiency of communication networks. By accurately predicting the failure in advance, network operators can take proactive measures to mitigate the impact of failure which helps them to minimize service disruption and maintain seamless communication, mobile network users, whether individuals or business, highly demand uninterrupted service but power failures can lead to dropped calls, slow data speed and poor network connectivity.

So predictive power failure models enables operators to prevent such disruptions, leading to improved customer satisfaction and retention. The other advantage is, it enables operators to allocate resources more effectively by optimizing the usage of backup power sources and dispatching maintenance teams to sites that are at higher risk of experiencing outages. This not only reduces operational costs but also ensures that resources are directed where they are most needed. Predictive models provide network operators with data-driven insights into the factors contributing to power failure. This information can guide operational decisions, maintenance schedules and infrastructure upgrades based on empirical evidence rather than relying on reactive responses. This research aims to address the following key challenges:

- Complex Multivariate Relationships: BTS sites are influenced by a multitude of variables, including electrical load, generator status, and battery health, and other factors. Developing a predictive model that can effectively analyze the intricate relationships between these variables and their impact on power failures is a significant challenge.

- In advance prediction: Predicting power failures before the actual failure happen is crucial for proactively implementing preventive measures and minimizing service disruptions. Developing a model that can provide predictions based on historical data, allowing for early identification of potential failures with high accuracy, is a complex task.

In light of these challenges, the proposed research focused on developing a machine learning-based predictive model for power failure prediction in BTS sites. The model harnessed the capabilities of multivariate data analysis, feature engineering, and advanced machine learning algorithms to improve the accuracy and timeliness of predictions. Additionally, efforts had been directed towards enhancing the interpretability of the model's predictions, making it a valuable tool for network operators to proactively manage and mitigate power-related disruptions in BTS sites.

By addressing these challenges, this research will contribute to the advancement of communication network reliability and infrastructure management through the application of state-of-the-art machine learning techniques to power failure prediction in BTS sites.

0.1.3 Objective

General Objective

The general objective of this research is to develop a deep learning-based power failure prediction model for BTS sites using CNNs, LSTMs, and hybrid CNN-LSTM algorithms. The model will be evaluated to ensure reliable power systems for BTS sites.

Specific Objectives

In order to achieve the goal of the general objectives, the following specific objectives have been identified.

- Study BTS Power Sources and power system interconnection;
- Study different BTS power system and environmental alarms and identify their causes;
- Collect Data from Ethio telecom power and environment monitoring system;
- Prepare data set that serves as an input in the model building process;
- Perform the necessary data preprocessing techniques on the dataset;
- Build data driven Power failure prediction model using machine learning algorithm;
- Evaluate the prediction performance of the model; and
- Draw conclusions and recommendations based on the results and findings

0.1.4 Literature Review

Several studies have been conducted relation with data driven failure prediction in various fields. Some of them can be reviewed as follows. Author in [9] performs survey on fault prediction. The paper analyzes by dividing fault prediction into small sub-sample data and large sub-sample data based on the size of the data capacity. Auto Regressive Moving Average mode (ARMA) and Vector Auto Regressive model (VAR) are the most frequently used models for time series prediction using small sample data having a relatively small work load and predicting short term faults well. But in small size data since the prediction interval is short, it is difficult to obtain high precision fault prediction in medium and long term prediction. The author claimed that in large sample data, principal component analysis (PCA) is advantageous to convert high dimensional data information into low dimensional features through feature extraction. The use of neural networks in processing vast volumes of data and complex information is clearly advantageous as it does not require a lot of prerequisite knowledge and it possesses exceptional generalization, strong nonlinear neural network mapping, and high self-organization abilities.

The literature on fault prediction in industrial systems has witnessed a growing interest in recent years due to its potential for enhancing system reliability and reducing downtime. In [10], the authors present an innovative solution for predictive fault detection, utilizing data collected from Supervisory Control and Data Acquisition (SCADA) systems. Their approach focuses on generic fault and status prediction, employing a data-driven methodology centered on a self-organizing map (SOM) and a novel Key Performance Indicator (KPI). The model's evaluation, conducted on a group of three photovoltaic (PV) plants and over sixty inverter modules from different technology brands, demonstrates its effectiveness in predicting incipient generic faults with an impressive lead time of up to 7 days in advance, achieving a true positive rate of up to 9%. Remarkably, this model's adaptability and ease of deployment make it a valuable tool for online monitoring and anomaly detection in new PV installations, requiring only historical SCADA data, fault taxonomy, and inverter electrical datasheets, thus offering significant potential for improving system reliability and reducing downtime in industrial contexts.

On the other hand, [11] proposed a fault prediction method for line trip in power systems capturing electrical measurement, and multisource time series data like current, voltage, active and reactive power. The method was developed using LSTM to capture the temporal features of the data in the long time span and SVM for the classification of the faults. During model development, overfitting was avoided through the use of dropout and batch normalization. The model was trained using LSTM in 5-fold cross-validation and using RMS prop optimizer a good prediction performance was obtained and some improvement was observed after feature extraction. But in time series data, splitting using cross-validation is not recommended because test data may occur before training data, which biases the prediction. The model was also tested with single feature data and multiple-feature data and a better prediction was obtained with multi-feature data. In the implementation of SVM, the LSTM-trained temporal feature was put into the SVM classifier for fault classification, and a better and improved accuracy was obtained with a prediction accuracy of 97%. However the training time of SVM is much longer as it is much more computationally intensive.

Nowadays failure prediction is performed in different fields in order to sustain the service provided by the system. A failure prediction method also conducted in optical networks by [12] based on machine learning SVM and double exponential smoothing (DES). On the other hand [13], proposes a method for predictive maintenance that considers the loss of revenue if an equipment failure occurs. The method uses multivariate time series data to train a deep learning model that can predict the probability of an equipment failure. The model is then used to schedule preventive maintenance tasks to minimize the risk of failure and the associated loss of revenue. The paper was evaluated using data from a Microsoft case study. The data consisted of 24 hours of multivariate time series data for a specific piece of equipment. The model was able to predict the probability of failure with a root RMSE of 0.126. This suggests that the model could be used to effectively schedule preventive maintenance tasks and minimize the risk of equipment failure.

In many areas, including predicting machine failures, load forecasting, and smart manufacturing, which uses sensor data, machine learning approaches are being applied. Specifically, CNN-LSTM has been implemented in different fields such as, in mobile data traffic prediction, Other paper [14] examines limitations in traditional wind turbine fault prediction models concerning time series data handling and stresses the need to capture deep temporal connections for enhanced generalization. It presents an attention-based CNN-LSTM model addressing these issues by combining CNNs, LSTMs, and an attention mechanism. The model utilizes semantic sensor data annotated with SSN ontology, enhancing interpretability. CNN extracts features, LSTM captures temporal dependencies, and the attention mechanism improves fault-related insights. Employing the random forest algorithm for feature correlation analysis enhances model efficiency. Evaluations on wind turbine fault datasets demonstrate the model's superiority over RNN, LSTM, and XGBoost in efficiency, accuracy, and generalization. The integration of CNN, LSTM, and attention mechanisms, along with structured data and feature selection, offers significant improvements in fault prediction accuracy and efficiency. Overall, the paper advances wind turbine fault prediction models through this innovative approach.

The study in [15] analyzes the vital task of short-term solar irradiance forecasting by introducing a spatiotemporal correlation model grounded in deep learning. The model integrates a convolutional neural network (CNN) to capture spatial patterns from meteorological data across target and neighboring sites. Concurrently, a long short-term memory (LSTM) network extracts temporal features from historical solar irradiance data. These spatial and temporal facets are fused to predict global horizontal irradiance an hour ahead. Evaluation covers various seasons and sky conditions, using three datasets from 34 locations in Texas, USA. Comparative assessment with CNN, LSTM, and benchmark models demonstrates the proposed CNN-LSTM's superiority across five metrics. The results underscore its ability in short-term solar radiation prediction, indicating its potential as a robust alternative.

In the field of modern networks, which bring unparalleled convenience and efficiency to various aspects of life and work, the looming potential for losses due to network failures necessitates proactive fault prediction [16] This predictive capability not only readies personnel for predictive fault repairs but also minimizes repair durations and curtails associated losses. To address this critical need, this paper proposes a novel network log-based hybrid prediction model for wireless network faults, integrating CNN and LSTM networks. The model preprocesses network logs, extracting features via CNN, and inputting them to LSTM for prediction. Through comparative experimentation against CNN and Random Forest approaches, the study demonstrates the superior predictive performance of the CNN-LSTM model, offering promise in enhancing fault prediction accuracy for wireless networks.

0.1.5 Methodology

In this research, the methodology employed to develop an effective machine learning-based power failure prediction model for BTS sites using multivariate data involves a comprehensive framework that integrates:

- Literature review: Reading various literature, journals, articles, books and reviewing Ethio telecom Low Level Design (LLD), High Level Design (HLD), studying different manuals and modules and analyzing different BTS sites equipment and their interconnection.

- Data collection: multivariate time series data was collected from Ethio telecom Power and environment monitoring system, which was used for developing a machine learning model.

- Data preprocessing: data cleaning, feature engineering, model selection, training, and validation.

- Prediction model building: the preprocessed data was split into training, validation, and test set. CNN, LSTM and hybrid CNN-LSTM were used to develop the BTS power failure prediction model.

Although conventional ML models are implemented in failure prediction in various fields, they can only extract shallow features and usually require complex feature engineering operations that will be conducted in human effort [17]. This issue associated with conventional ML models has been addressed in recent years by the development of deep learning, such as convolutional neural network (CNN) and long short-term memory (LSTM) network, which rely on the representation learning of data [13]. Deep learning models improve the accuracy of classification

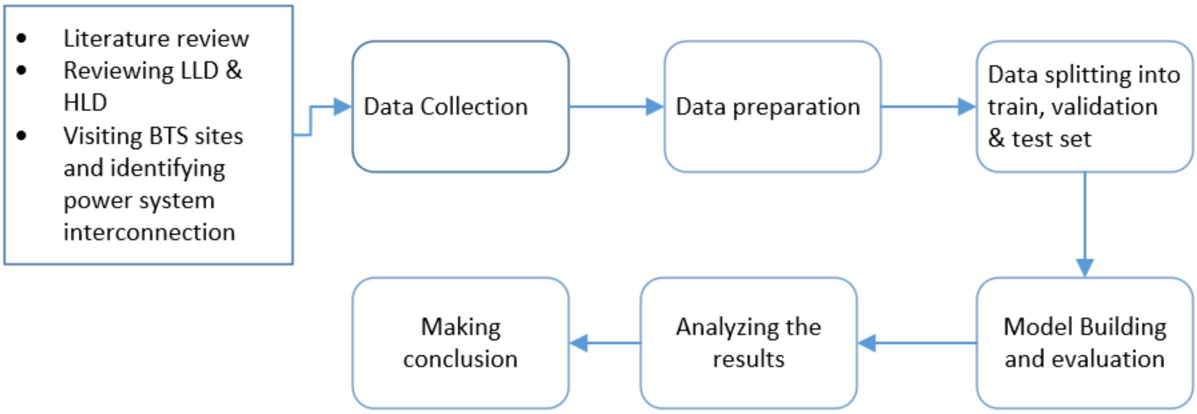

Figure 0.1.1: Flow diagram of general methodology.

or prediction by using deep nonlinear network structures to realize complex function approximation, as well as by the use of deep features from massive samples. These models can extract features spontaneously through multiple stacked hidden layers, and therefore alleviate the need for complex feature engineering operations [17]. Finally the prediction performance of the model was evaluated using performance evaluation metrics. For the simulation purpose, Microsoft Excel and python with a deep learning library Keras and Tensorflow was used. The general steps are indicated as follows.

0.1.6 Scope and Limitations

Scope of the Thesis

The scope of this study encompasses the development and implementation of a machine learning-based power failure prediction model specifically tailored for BTS sites. The study involve the collection and analysis of multivariate data, including variables such as weather conditions, electrical load, and battery related features, and potentially other relevant factors. The predictive model was designed to anticipate power failures in advance, enabling proactive measures to mitigate disruptions and enhance network reliability.

The study also explores various deep learning algorithms and techniques suitable for handling complex multivariate relationships and real-time prediction. The scope extends to the incorporation of interpretability methods to enhance the transparency of the model's predictions, providing valuable insights for network operators.

Limitations of the Thesis

The primary objective of this study is to develop a resilient model for predicting power failures in BTS sites, employing machine learning techniques and leveraging multivariate datasets. While power system failures encompass a variety of telecom networks, this investigation exclusively concentrates on BTS power systems. Specifically, data derived from chosen BTS power and environment monitoring systems serve as the foundation for building the predictive model. It's important to note that this study is delimited to BTS systems, and as such, the investigation excludes core networks and data centers. The study's focus is directed towards forecasting the occurrence of future ten days failures rather than delving into the causal factors that initiate these failures. This limited exploration arises from the unavailability of maintenance records from operation and maintenance teams, which constrains a comprehensive root cause analysis within the scope of this research.

0.1.7 Contributions

The proposed study offers several significant contributions to both the field of telecommunication and predictive analytics:

- The development of a machine learning-based predictive model for BTS sites fills a critical gap in existing power failure prediction approaches. The models ability to analyze complex multivariate relationships and provide accurate predictions could revolutionize how network operators anticipate and manage power-related disruptions.
- By accurately predicting power failures, the study contributes to enhancing the overall reliability of communication networks which has a direct positive impact on user experience, customer satisfaction and retention rate this in turn increases revenue for telecom companies.
- The predictive model empowers network operators with the ability to proactively manage and allocate resources by identifying sites at higher risk of power failures, prioritize maintenance efforts, optimize backup power sources, and ensure uninterrupted services

0.1.8 Thesis Organization

The rest of the paper is organized as follows. Chapter one gives a brief overview of the thesis by providing statement of the problem, thesis objective, scope and limitation and main contribution of the study. Chapter two gives a brief overview of BTS power system and alarms related to power failure. Chapter three describes about machine learning and deep learning algorithm along data preprocessing techniques, various model evaluation metrics and model tuning hyperparameters and different data preprocessing techniques like Min-Max normalization and Z-score normalizations are discussed. In Chapter four result and discussion of the thesis is presented. Finally, conclusion of the study and future work is presented in chapter five.

0.2 Overview of Power System

0.2.1 Power System in Base Transceiver Stations

Telecom power systems, specifically BTS, are designed to provide reliable and uninterrupted power supply to the telecommunication infrastructure. These power systems typically consist of multiple components such as mains (utility) power, diesel generators, backup battery systems, Rectifiers, different distribution boards like AC and DC distribution boards, circuit breakers, and fuses [3].

Most BTS have utility power sources as their initial power systems. Diesel generators and batteries are used as backup power sources. In the event of a utility power outage, a controller activates the backup battery supply and initiates the generator. An automatic transfer switch guides the generator's power to a controller, which subsequently transitions from utilizing backup battery power to employing emergency generator power. When the standard utility power is restored,

0.2.2 BTS Power System Architecture and its Components

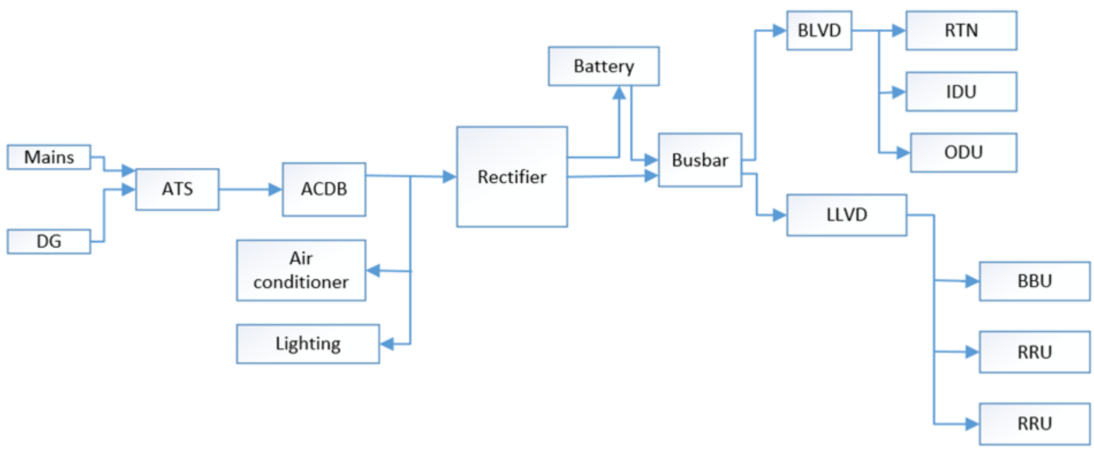

Figure 0.2.1: BTS Power System Architecture.

Mains

It is an alternating current (AC) electric power obtained from a utility or grid power supply. Which is the primary source of power for a BTS. Mains power is supplied through electrical lines and is the most reliable and cost-effective source of electricity. Hence, they are called primary power source. From the utility side three phase power source each having a voltage of 220 VAC phase to neutral or 380VAC phase to phase was used to supply power.

The telecom power system connects to the mains power supply, ensuring a continuous power under normal operating conditions. But the main power supply is prone to failure and interruption.

There are many possible reasons for mains failure. Some of the most common reasons include:

- Electrical equipment failures (e.g., damage to service transformers, distribution lines, or substations);
- Natural disasters: Natural disasters such as storms, floods, and earthquakes can damage power lines and transformers, leading to mains failures;
- Weather (storm, lightning, wind);
- Trees falling into lines, vehicle collusion with poles;
- Human error: Human error, such as accidents or vandalism, can also cause mains failures;
- High power demands may result in overburdened electric cables, transformers, and other electrical equipment can melt and fail;
- Overload: Overloading of the power grid can cause mains failures. This can happen when there is too much demand for power, or when there is a problem with the power grid itself;
- Installation and Maintenance issues, such as poor wiring or faulty equipment, can also lead to mains failures;
- Sabotage: Sabotage, such as intentional damage to power lines or transformers, can also cause mains failures. In some cases, the cause of a mains failure may be unknown.

Diesel Generators

Generator is a device that converts mechanical energy into electrical energy. They are installed in a fixed location and used as a standby power source to provide power to communication devices during mains failure or interruption. Standby power sources are typically far more expensive than prime/commercial power sources, and they are only used when commercial power fails. The generator supplies power to the site until utility grid power is restored or until its fuel is depleted or until it fails. Although they can supply a 3-phase AC power like a utility power system, they are designed to supply backup power for much shorter time periods until mains or utility power is restored [18]. But generators can face different failures or they may fail to start because of different reasons:

- Battery failure (due to loose connection, sulphate buildup on the plate or dead battery);
- Low coolant level (Without radiator coolant, an engine would soon overheat, leading to mechanical breakdown and engine failure);
- Low fuel level in the tank;
- Low Oil Levels in the Engine.

Automatic Transfer Switch

Automatic Transfer Switch (ATS) is a self-operating, intelligent power switching device with specific control logic. The control logic or automatic controller is typically microprocessor-based and continuously tracks the electrical characteristics (voltage, frequency) of the primary and backup power sources. The ATS will automatically switch the load circuit to the backup power source in the event that the connected power source fails. Mains is the primary source to supply electric power to the load (network elements), when there is a mains outage or when the voltage drops below a predetermined threshold level, ATS will switch the load circuit from the primary source to the secondary which is the backup generator. This enables critical loads to obtain power to accomplish their tasks. Upon sensing mains failure, 3-4 seconds are required for the generator to start and for the AC voltage to stabilize. During this phase, the battery backup system provides power to the telecommunication equipment [19].

ACDB (AC Distribution board)

ACDB stands for "Alternating Current Distribution Box" commonly known as an AC Distribution Panel or AC Power Distribution Unit (AC PDU). It is an essential component used to distribute AC power from the common AC source to various loads. It acts as a centralized distribution point, receiving incoming AC power from the utility grid or backup generator and route power to different devices, equipment and systems that require AC power to operate. This can include rectifier, air conditioner, lighting and other auxiliary equipment.

Rectifier System

Rectifier system is a critical component that converts the source AC power from the utility grid or backup generator into direct DC power and supplies DC power to the telecommunication load equipment as all telecom equipment are working with DC power. Also, the rectifiers are responsible for charging the backup battery system, which covers the site in the case of electricity cut-off.

Rectifier Modules – PSU (Power Supply Unit): The core component within the rectifier is the module, tasked with the conversion of AC voltage to DC voltage and the charging of batteries. Enhanced rectifier performance and accelerated battery charging rates are achievable by increasing the number of modules, a factor contingent upon site-specific loads and the quantity of battery strings. Consequently, the modules are consistently organized in an N+1 configuration, ensuring a minimum of two rectifiers at each site, and adhere to a load-sharing approach in their operation.

Rectifiers operate either in a float mode or in equalize mode. During float mode, it supplies power to load equipment and float charges the batteries during normal (when there is an AC power) operation to avoid battery internal loss. During equalize mode, it operates load equipment and recharges the batteries after an AC power failure.

The rectifier system is interconnected to the load equipment through LLVD (Load Low Voltage Disconnect) and BLVD (Battery Low Voltage Disconnect) contactors as it can be shown in the below diagram. In BTS sites, the loads are classified as critical and non-critical loads [20]. Less priority loads are connected to the LLVD contactor and higher priority loads like transmission equipment, Microwave equipment, ATN, and RTN are connected to BLVD contactor.

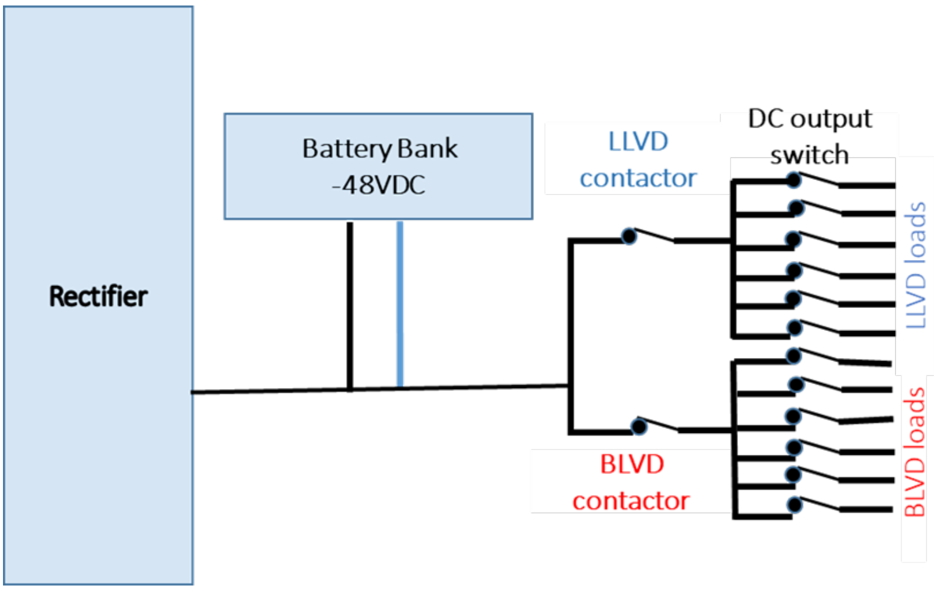

Figure 0.2.2: Rectifier System interconnection.

Normally Rectifier system voltage is -48Vdc and it reach to its maximum up to 54.5V dc.

But in case of power cutoff, this voltage reads the voltage of the battery strings which will decreased every time till the batteries become empty of charge. To protect the battery from deep discharge and to protect the site from total service interruption, a threshold value is sated as 46.5V for LLVD and 44.5V for BLVD. when the battery voltage reaches 46.5V, LLVD contactor open and less priority loads are disconnected. If the mains or ac power either from grid supply or DG does not restore, the battery continues to discharge and when it reaches to its BLVD threshold level, the BLVD contactor disconnect all critical or high priority loads. This time the site will experience total service interruption.

Backup Battery System

Batteries are energy storage devices in telecom power systems that can store electrical energy for later use and supply to the load on demand. Hence, each site is configured with a backup battery system to supply the site with DC voltage in the case of AC power cutoff. Since, telecom devices operate with -48V DC, each string has to be configured to provide this voltage. For example, if each battery has 2V DC, so the string will have 24 batteries (24 batteries * 2V DC = 48V DC).

During normal operation of AC power, the rectifier supplies the required voltage level to the load and float charges the battery to overcome losses due to self-discharge of the cells in the battery. If the rectifier fails to supply the required power due to AC power failure or due to some other reasons like rectifier module problem, the battery string provides DC power to the load until its discharge rate reaches its setted threshold level. The discharge phase of the cycle could run anywhere from a few minutes to 8 hours, depending on how long rectifiers are unavailable and also it depends on the battery age. The reserve time and discharge current dictate how quickly the battery discharges. A totally depleted battery recharges in (usually) 8 to 24 hour when the rectifier's ac power source is restored, and it then floats at a constant voltage until the next power supply failure.

0.2.3 Alarm Management System

A BTS consists of a couple of interconnected components like cooling elements, modem, switch, transmission equipment, microwave equipment, and power devices. All these equipment obtain power from a common source. These components can produce an alarm to indicate sort of abnormal situation. An alarm management is one of a network management system that helps to collect, analyze and manage alarms happening on the site.

In ethio telecom, the power and environment management or monitoring system is called Network ecosystem (Net-Eco) to manage the site operating status and sends real time data to the central server. A fault detection is performed through standardized performance and equipment functionality monitoring by means of threshold management.

Since all BTS components obtain power from a common source and power failure at the site results total service interruption, the power and environment are monitored regularly through connected sensors to each devices. These sensors reads their real time value and send the data to central server. Each device normal operating status or threshold value is sated and if the reading of these sensors are out of the sated threshold value, an alarm is generated and the network operation teams at National Network Operation Center (NNOC) are notified of the alarms. NNOC teams generate Trouble Thicket (TT) for each alarm and dispatch to technicians for maintenance. The site may be in sever condition until the technicians reach for repair.

Actually the occurrence of one alarm may result to the occurrence of another alarm. For example, lack of commercial power results battery under voltage alarm and if it is not restored before the battery voltage depleted, LLVD and BLVD alarms will be happen. Or if the air conditioner is not function properly, over temperature alarm may happen and if it is not fixed on time, the room temperature may rises and the cables may melt or it may cause power interruption.

0.2.4 Power System Failure in BTS Site

Power failure form a special class of telecommunication system failure. In analyzing a power failure, there are often several reasons to be considered. A power failure might happen because of loss of utility/grid power or malfunctioning of backup power systems. It could also be caused by failure of other power related devices like rectifiers, circuit breakers, or fuses that are linked between the power input and the communication equipment. These devices are connected to different sensors which can read and detect their measurements and the sensors are in turn connected to the monitoring system. A monitoring system is system that monitors the status or functionality of the system based on readings obtained from the sensors. If the reading of the sensor is debated from the actual value an alarm is generated to indicate abnormalities on the site.

A power system is said to be fail if there is no enough power to operate the load or if the power is totally not available on the site. If the power system is failed, the communication service both voice and data will be interrupted. Hence it is important to have a backup power system like generator and Battery to supply power to load equipment. If the utility power is lost, the generator has to be started automatically and supply to the load. If the generator is started successfully and fails for some reasons like piston seizure, depletion of fuel or startup battery failure, the backup battery system starts to operate the load equipment. With all this redundant power system, majority of telecom service disruption is occurred due to power and power related issues.

Actually, failures within the BTS power system can be identified through monitoring the busbar voltage, which provides more comprehensive information compared to DC load current and power. Under normal operating conditions, the busbar DC voltage should remain above 48.2 Volts. If it drops below this threshold, it's considered busbar under voltage (BBUV), serving as an early warning sign for LLVD issues. The duration until a full LLVD failure occurs depends on the battery's load-carrying capacity. When the voltage level falls below 46.5VDC, LLVD loads are disconnected, and an LLVD alarm is triggered. Furthermore, if the voltage level dips below 44.5VDC, BLVD loads are disconnected, resulting in the interruption of all BTS site services except communication with the monitoring system. This BLVD voltage-level disconnection is a protective measure to prevent deep discharging that could potentially harm the battery components irreversibly. These voltage levels are typically set as default values by operators but can be adjusted based on specific site conditions and operational requirements.

0.2.5 Power Alarms Triggering BTS Power Failure

Before the actual power failure happens, there have been notification or alarms that can indicate abnormal function of the system. Those alarms are categorized based on their severity and classified as critical, major, minor, and warning alarms. The severity level indicates the seriousness of a fault. Critical alarms indicates that the fault affects normal operation of the system, effective measures should be taken immediately otherwise it leads to BTS service interruption. Major alarms indicates that, the fault decreases the system performance and it results to a possibility of some service related problems and performance of a device or resources decreases significantly. Minor alarms are a low severity alarm and it indicates that a fault occur but does not interrupt BTS service but if it is not solved on time it decreases equipment and system performance. Warning alarm indicate that an error may occur and affect the system performance in a long term.

There are different types of power and environmental alarms that can lead to BTS service interruption. From the mains or AC power side; there are a number of AC power system alarms that can trigger BTS service interruption. These alarms include: Mains failure alarm which is caused by lack of utility power, mains phase loss alarm, phase under voltage alarm, and phase overvoltage alarms which can occur due to factors such as fluctuating grid power, load imbalances, or faulty equipment. Both overvoltage and under voltage scenarios can severely impact the functionality of sensitive electronic components within the BTS. Overvoltage can lead to equipment damage or failure, while under voltage can result in reduced equipment performance or complete shutdown. Mains alarms or unavailability of utility power leads to other types of alarms like battery under voltage and if the alarm is not handled on time it result to total BTS service interruption.

BTS sites often rely on backup battery systems to ensure continuous operations during power outages. Alarms associated with the battery system include low battery voltage, battery disconnection, battery over temperature, battery discharge alarm. A low battery voltage alarm can signify that the backup power source is nearing depletion, potentially leading to service disruption if not addressed promptly. Other Alarms from the DC side that can cause BTS service interruption includes, DC under voltage, load fuse break alarm, LLVD low voltage disconnect alarm, and BLVD low voltage disconnect alarm.

Environmental alarms are initiated in situations where the ambient temperature or humidity surpasses prescribed or configured thresholds. While such alarms occur infrequently, they encompass an array of environmental triggers, including instances of smoke, fire, and flooding. To elaborate, these alarms can manifest as scenarios involving exceptionally elevated or diminished ambient humidity levels, as well as instances of markedly low or high ambient temperatures.

This issue generally emanates from various external factors, such as adverse weather conditions, equipment malfunctions, the specific operational load of the site, inherent design deficiencies, and other technical failures that aren't inherently tied to the immediate environment. It is noteworthy that on rare occasions, these environmental alarms can precipitate a cascade effect that reverberates through the operational efficiency of BTS equipment. A related illustration lies in higher temperatures worsening the workload imposed on cooling systems and fans, thereby potentially inducing a feedback loop of further temperature escalation. This augmented thermal load may then serve as a catalyst for equipment malfunctions, culminating in disruptions to the comprehensive functionality of the BTS system.

Hence, thorough comprehension of these environmental alerts assumes paramount importance, necessitating a proactive approach to effectively anticipate any negative impact on the uninterrupted and dependable functionality of the BTS Services.

0.3 Basics of Deep Learning Algorithms

0.3.1 Machine Learning

Machine learning is a subset of artificial intelligence (AI) that empowers computers to learn from data and improve their performance over time without being explicitly programmed. By enabling machines to distinguish patterns, make predictions, and uncover hidden insights, machine learning has become an indispensable tool across diverse domains [21]. Furthermore, machine learning has revolutionized the way we approach complex problems and make informed decisions in various fields. At the heart of this revolution lies the process of building machine learning models, which involves transforming raw data into predictive insights. This transformation occurs through a sequence of steps where data is fed into an algorithm, and the algorithm adjusts its internal parameters to uncover underlying patterns. These patterns can manifest as correlations, trends, or complex relationships within the data. The algorithms then leverages these patterns to make predictions, decisions, or classifications when presented with new, previously unseen data [22].

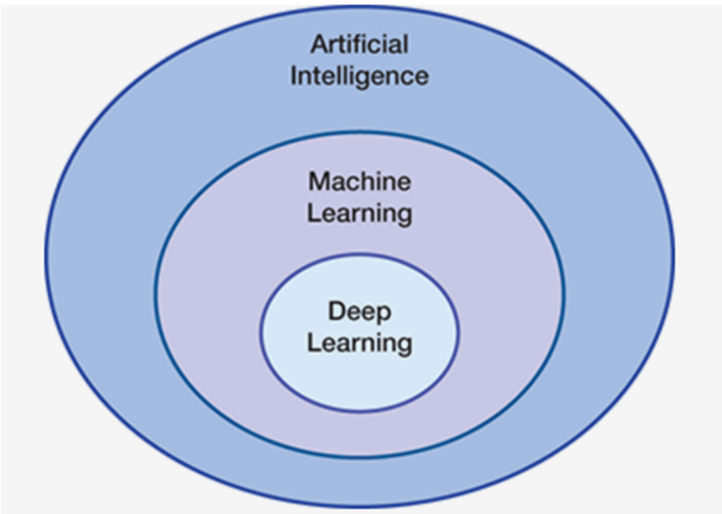

Figure 0.3.1: Artificial Intelligence, Machine Learning and Deep Learning [23].

0.3.2 Deep Neural Network Basics

AI is the capability of a machine to imitate intelligent human behavior. Machine Learning is a subset of artificial intelligence that makes computers to learn from past or historical data, improve performance from experiences, and predict things without being explicitly programmed. Deep Learning is a subfield of ML that uses algorithms called Artificial Neural Network (ANN), which are inspired by the structure and function of biological neural networks in the human brain and are capable of self-learning. ANNs are trained to learn models and patterns rather than explicitly programmed [24].

Components of Neural Network

The fundamental processing units in a neural network are called neurons. Each neuron receives input data, processes it, and produces an output signal. Neurons are organized into layers, including an input layer, one or more hidden layers, and an output layer.

- Input layer: receives raw data, which is then passed through the network's connections to the neurons in the subsequent layers.
- Hidden layers: process the input data by performing mathematical operations on the weighted sum of inputs and applying the activation function. These layers capture and learn hierarchical representations of the data.
- Output layer: the output layer takes input from the neighboring hidden layer and uses an activation function to compute and produce an output which can be a prediction, classification, or any other relevant result[25].

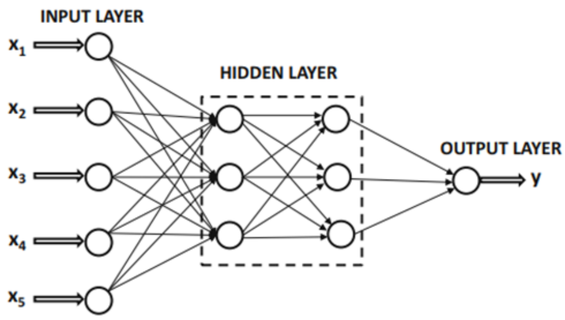

Figure 0.3.2: Artificial Neural Network Architecture [25].

0.3.3 Deep Learning Models for Power Failure Prediction

There are different types of deep learning algorithms for time-series data analysis such as Long Short Term Memory (LSTM), Gated Recurrent Unit (GRU), and Convolutional Neural Network (CNN). The deep learning algorithms used for this research are CNN, LSTM and hybrid CNN-LSTM. The proposed model are discussed in the subsequent section.

Long Short-Term Memory

LSTM is a specialized type of recurrent neural network (RNN) architecture designed to overcome the challenges of handling long-range dependencies and vanishing gradient problems in sequential data. It has been reported that RNN networks are incapable of handling long-term dependencies in sequential data due to vanishing and explosion gradient. LSTMs were introduced to address limitations in traditional RNNs, which has a limitation to capture meaningful patterns in sequences that have time lags or dependencies over extended periods [26].
LSTMs achieve this by incorporating memory cells and gates, which allow the network to learn and store information for long durations [27]. An LSTM network contains an input gate, out-

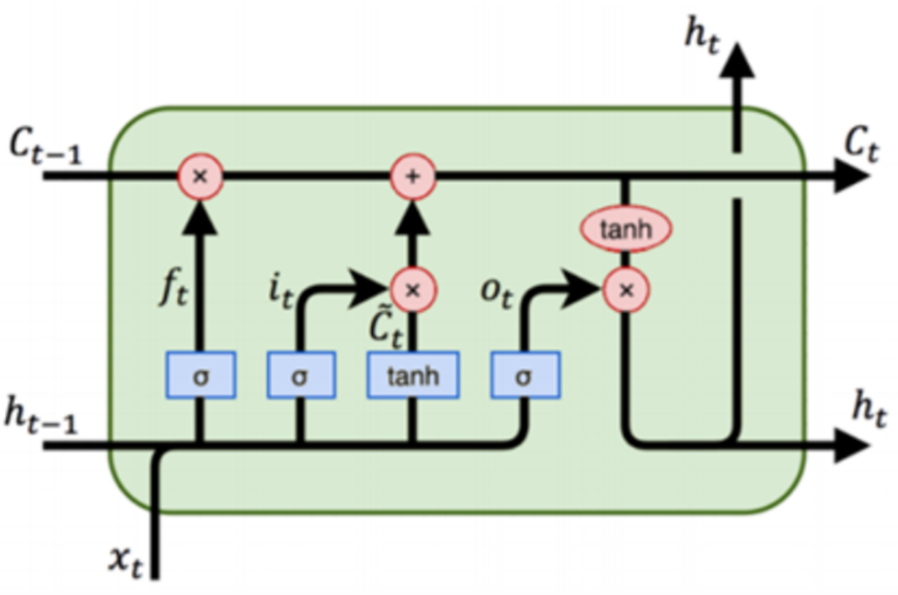

Figure 0.3.3: Basic Architecture of LSTM Model [26].

put gate and forget gate [28]. The function of the gates and the memory cell state within the LSTM network is explained as follows.

- **Forget gate:** The forget gate is the main element of the LSTM architecture. It plays a crucial role in controlling the information flow within the LSTM cell over sequential time steps. It decide which information from the previous time step should be retained or forgotten in the current time step. It evaluates the combination of the current input and the output from the previous time step, producing a value between 0 and 1 for each element in the memory cell [28]. This gate applies a sigmoid function on the output of the last hidden state $h_{(t-1)}$ and input data x_t:

$$f_t = \sigma(\mathrm{w}_f.[h_t \;, x_t]) + b_f \tag{3.1}$$

- **Input gate:** The input gate operates by combining the current input layer with the output from the previous time step or hidden state. It employs a sigmoid activation function to generate a value between 0 and 1 for each element in the memory cell, this determines the extent to which values should be updated. The input gate equation is described as follows.

$$i_t = (w_i.[h_t \;, x_t]) + b_i \tag{3.2}$$

- **Cell state:** The cell state is designed to store long-term dependencies and patterns present in the input data.

$$c_t = (f_t * \mathrm{c}_{t-1} + i_t(\tanh[wc.[h_{t-1} \;, x_t)]]) \tag{3.3}$$

- **Output gate:** The output gate plays a pivotal role in regulating the output of the LSTM cell, determining which information is disseminated to subsequent time steps and the final prediction. This gate is combined with a cell state, which is activated by tanh function to get the final output,h_t as given below.

$$o_t = \sigma(\mathrm{w}_o.[h_{t-1} \;, x_t \;]) + b_o \tag{3.4}$$

$$h_t = o_t * [tanh(\mathrm{c}_t \;) \tag{3.5}$$

Where:

- f_t, i_t and o_t are values of forget gate, input gate and output gate at time t
- b_f , b_i andb_o are bias vectors for forget gate, input gate and output gate respectively
- X_t is the input vector to the memory cell at time t
- W_f, w_i,w_c and w_o are weight matrices for gates and cell state
- σ and $tanh$ are activation functions

Convolutional Neural Network

CNN is a specialized type of artificial neural network. Even though, it was initially designed to process and analyze visual data, particularly images, its applicability has expanded to data organized in matrix-like forms. Notably, time-series and textual data can be represented as 1D vectors, while images are represented by 2D matrices of pixels. The main difference between a CNN and a regular neural network lies in the utilization of convolutions to manage the underlying mathematical operations. The convolution operation entails performing linear transformations, often via matrix multiplication, on at least one layer of the neural network. This methodology facilitates the network's adeptness at capturing complex patterns and relationships within grid-like data structures, thus making CNNs highly versatile across various domains [29].

The distinctive feature of CNNs lies in their ability to automatically learn and extract hierarchical features from data. This is achieved through a series of layers, including convolutional layers, pooling layers, and fully connected layers. Here's a brief overview of key components [30].

- **Convolutional Layers:** In CNN architecture, the basic important building block is the convolution layer. These layers perform convolutions, a mathematical operation that involves multiple sliding filters (also known as kernels). These are used to convolve the input feature map to generate the output feature map. A kernel is a grid of discrete numerical values, with each value referred to as a kernel weight. At the beginning of CNN training, random numbers are assigned as the initial weights for the kernel. Subsequently, during each training iteration, these weights are iteratively adjusted. Hence, through this process, the kernel progressively learns to extract important features [31].

- **Pooling Layer:** The main function of the pooling layer is to perform sub-sampling of the feature maps, which are obtained through the application of convolutional operations. In essence, this process involves reducing the dimensions of large feature maps to create smaller ones. However, this reduction retains the key information or features, ensuring that the dominant aspects are preserved throughout each stage of the pooling process. Common pooling operations include max pooling, which selects the maximum value from a group of neighboring pixels, and average pooling, which calculates the average value.

- **Fully connected Layer:**This layer is typically positioned towards at the end of each CNN architecture. In this layer, every neuron is interconnected with all neurons from the preceding layer, a methodology termed the Fully Connected approach. They aggregate the information extracted by convolutional and pooling layers and materializes in the form of a vector generated through the process of flattening feature maps and make final predictions as illustrated in the following Figure 0.3.4.

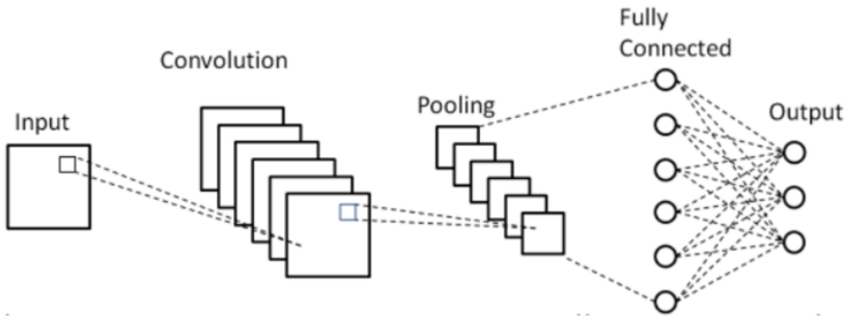

Figure 0.3.4: Schematic diagram of CNN Architecture [26].

There are different variants of CNN such as 1D, 2D and 3D tailored for processing data with distinct dimensional structure. While a 1D CNN (One-Dimensional CNN) designed to analyze sequential data, such as time-series data or sequences of text, where convolutions are performed along one dimension, capturing patterns and relationships within the sequential data [32]. Whereas a 2D CNN (two-Dimensional CNN) primarily employed for image data, where each pixel is represented as a two dimensional grid. The convolutions are applied across both width and height dimensions, enabling the network to detect visual features in images. But, a 3D (Three-Dimensional CNN) is applied to volumetric data, like medical imaging or video frames. In a 3D CNN, convolutions are extended to three dimensions, encompassing width, height, and depth. This depth dimension corresponds to time or third dimension of volumetric data [33]. So, each variant is optimized for specific types of data, utilizing convolutions along the respective dimensions to capture patterns and features. Hence, in this research, a 1D CNN was implemented because of the data types, as it is a time series sequential data.

0.4 System Model and Machine Learning for BTS Power Failure Prediction

0.4.1 System Model

The model building process aims to comprehensively capture the distinct characteristics and constituents of the dataset. Notably, the power and environmental system data is conceptualized as time series data, reflecting its temporal nature. This model development unfolds through the creation of three distinct algorithmic variants: CNN, LSTM, and the combined CNN-LSTM architecture.

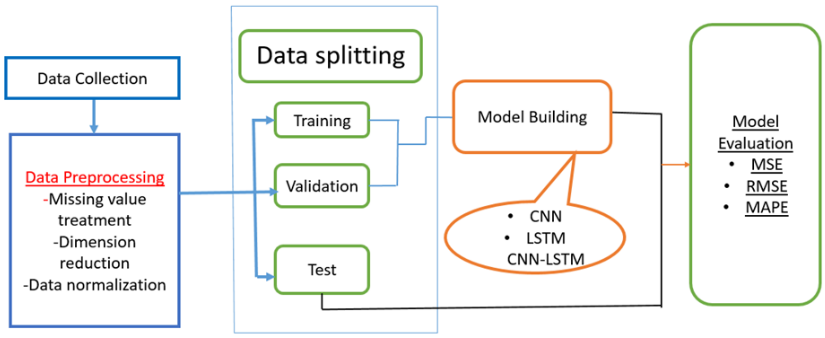

Figure 0.4.1: System Model.

0.4.2 Machine Learning for BTS Power Failure Prediction

In the realm of modern telecommunications, BTS sites serve as the backbone of mobile networks, ensuring seamless communication for millions of users. The uninterrupted operation of these sites is crucial for maintaining network reliability and providing a quality user experience (QoE). However, power failures can pose a significant challenge, leading to service disruptions and inconveniences. Machine learning has emerged as a potent tool for predicting failures in various fields, offering the potential to enhance system stability, reduce downtime, and improve overall operational efficiency [4]. Machine learning leverages historical data and complex algorithms to identify patterns and correlations that might be imperceptible to human observation [34]. When applied to power failure prediction in BTS sites, machine learning models analyze a variety of factors, such as weather conditions, voltage fluctuations, and traffic loads. By learning from these patterns, the models can make accurate predictions about the likelihood of power failures, allowing network operators to take proactive measures.

Advantages of Power Failure Prediction

Power failure prediction can give several advantages. Some of them are listed below:

- **Early Warning System:** Machine learning models act as early warning systems, enabling network operators to anticipate power failures before they occur. This proactive approach empowers technicians to intervene promptly, minimizing the impact of downtime;
- **Reduced Downtime:** By accurately predicting power failures, machine learning models contribute to reduce downtime and service disruptions. This leads to enhanced user satisfaction, as users experience fewer instances of dropped calls or data outages.
- **Operational Efficiency:** Predicting power failures allows network operators to plan maintenance activities strategically. By scheduling maintenance during periods of lower network activity, operators can minimize the impact on users and streamline operational workflows.
- **Resource Allocation:** With predictive insights, resources such as backup power systems and field technician schedules can be optimized. This ensures that resources are deployed effectively, reducing unnecessary costs and enhancing overall efficiency.
- **Data-Driven Insights:** Machine learning models can uncover hidden correlations between seemingly unrelated variables and power failures. This data-driven insight can lead to more informed decision-making regarding infrastructure upgrades, preventive maintenance, and network optimization [35][36] [37].

Challenges of Power Failure Prediction

Although machine learning holds significant promise for power failure prediction in BTS sites, there are some challenges to be considered. This include:

- **High-Quality, Relevant Training Data:** For a machine learning model to effectively predict power failures, it needs to be trained on a dataset that accurately represents the real-world scenarios and conditions that lead to power failures. This means having access to historical data that includes various factors like weather conditions, voltage fluctuations, traffic load, and more. Collecting such high-quality and relevant data can be challenging, as it requires ensuring the data is accurate, complete, and representative of the situations the model will encounter in the field.

- **Potential for False Alarms:** False alarms occur when the machine learning model predicts a power failure that doesn't actually happen. These false positives can lead to unnecessary disruptions, as resources might be deployed when they aren't needed, causing inconvenience and potentially wasting resources. Avoiding false alarms is crucial to maintain the credibility and trustworthiness of the predictive model.

  So, while machine learning offers substantial potential for predicting power failures in BTS sites, addressing challenges related to data quality and false alarms is essential. By carefully collecting data, refining the model, incorporating domain expertise, and fine-tuning decision thresholds, network operators can harness the power of machine learning to create a reliable and effective predictive system that enhances network stability and user satisfaction.

0.4.3 Data Selection

For this study the dataset is collected from Ethio telecom power and environment monitoring, Net-Eco system for a duration of five months with a five minute sampling period. The dataset contains multivariate features such as environment related feature like ambient temperature and humidity, AC related feature like the three phase voltage, battery related feature like battery temperature, battery remaining capacity, battery charge discharge rate, battery voltage and load related features like DC load power, DC output voltage, total DC load current. There was a total of fourteen features extracted.

Input Parameters	
battery voltage	total DC load power
battery current	total DC output current
battery remaining capacity	**DC output voltage**
battery temperature	room Temprature
battery charge-discharge rate	indoor Humidity
DC output current	phase voltages (L1, L2, L3)

Table 0.4.1: Data features

During the model build up 80% of the data, that is four month data (March 1-June 30) is used to train the model. From the next month (July), ten days of data is used for validation in order to predict the next ten days of power failure.

0.4.4 Multi Variate Time Series Data Preparation

Time series data is a type of data where observations are collected and recorded at specific time intervals. In other words, it's a sequence of data points ordered chronologically over time. Data preprocessing frequently plays a substantial role in influencing the generalization capabilities of a supervised machine learning algorithm [38]. Time series data preprocessing involves specific techniques and considerations to handle the temporal nature of the data. But it is challenging to prevent issues like data redundancy, missing values, errors, and unforeseen anomalies during collection and transmission. As a result, the process of data preprocessing becomes imperative and essential for effective analysis of time series data. In this research different data processing techniques are followed. Like data selection, missing value treatment, data normalization and dimension reduction [39].

Data Cleaning

Once the data has been collected in a comprehensive manner, the next step involves its exploration and assessment to identify significant patterns and irregularities. The primary objectives of data quality assessment are to identify missing values, outliers and noisy data. Missing data commonly occur in the majority of real-world datasets [40] [41]. Dealing with missing values is a crucial step in data preprocessing, as these gaps in data can stem from a multitude of factors, ranging from technical problems in sensors to human oversights during data collection. To ensure the integrity of the subsequent machine learning model, addressing missing values is imperative due to their potential to introduce biases that subsequently distort the model's predictions. These can have significant implications for the accuracy and reliability of prediction results in machine learning and data analysis. When missing values are not handled properly, they can introduce bias into the model, reduce model performance, unrepresentative results, misleading conclusions and finally, inaccurate business decisions [42].

Several strategies exist for handling missing values, each catering to distinct scenarios. The initial approach involves eliminating data instances that contain missing values, which is suitable when the prevalence of such values is minimal. Otherwise, it has the potential to disturb genuine data patterns and can even trigger further significant information or data loss when entire rows or columns are eliminated due to a few missing values within the dataset [40].

The alternative method entails employing missing value imputation techniques to populate the missing values, which is a process that involves substituting these gaps with estimated values derived from the available data. This approach mitigates the impact of missing values on the model's performance. Imputation draws on various statistical techniques, such as calculating the mean or median of the existing data and assigning those values to the missing entries. But such method imposes high bias because the newly imputed data are the same as the mean of the observed data [42].

Imputing missing values with zero is a common strategy when the missing values are indicative of a true absence of the variable being measured. This approach is appropriate when the missing values are not due to random or technical errors, but rather have a specific meaning in the context of the data. There are some scenarios where imputing missing values with zero might be appropriate such as; if the missing value represents a measurement that is expected to be zero due to the nature of the measurement or the context, then imputing with zero is logical [42]. On the other hand, if domain knowledge or the specific circumstances surrounding the data collection suggest that missing values should correspond to zero, imputing with zero aligns with the understanding of the data. It's important to note that imputing missing values with zero is not always the best approach.

It's crucial to consider the domain context, the nature of the data, and the reasons for the missing values. In some cases, other imputation methods or data handling strategies might be more appropriate. For example, if missing values are due to measurement errors or other technical issues, imputing with zero might not accurately reflect the true data distribution and more advanced imputation methods could be considered [43]. Ultimately, the decision to impute missing values with zero should be guided by a clear understanding of the data and the implications of using this approach on subsequent analyses or modeling tasks.

In our specific context, examining the dataset reveals instances of missing values occurring in the Battery temperature. This is due to the site's battery temperature sensor problem. Since all column values reading is NA for battery temperature, it is better to remove that column. The other missing value is observed in mains phase voltages which is occurred due to the mains failure alarm. Upon closer inspection, it's evident that these gaps coincide with periods when utility power is absent at the site. In response, a strategic decision has been made to impute these missing values with zeroes. This course of action aligns with the domain knowledge that, in the absence of utility power, phase voltages should indeed read as zero.

Target Variable Selection

In building a successful supervised learning model, selecting the appropriate target variable which is known as a dependent variable or outcome variable is a critical step. The target variable is a variable that our model is going to predict based on the input or independent features [44]. The choice of the target variable can influence the selection of relevant input features. Features that have a strong relationship with the target variable are likely to be more informative for the model. This dictates how the problem is shaped, how the model performs, and how effectively the analysis as a whole. Hence, Careful consideration of the target variable is essential for creating a valuable and accurate predictive model [45].

Hence, to select the target variable it is essential to analyze which feature is highly correlated with power failure from the features described in Section 0.4.3. In BTS site, power failure refers to the situation where the electrical power supply to the load equipment is interrupted or fails resulting the site unable to provide the required communication services. This may happen because of various reasons as stated in Section 0.2.1 of this study. As it can be shown in the following diagram in Figure 0.4.2, whenever there is mains power from the utility side the rectifier converts 220V-AC power to 48V-DC and supplies power to the load equipment at the same time it charges the battery bank. If mains power interrupted, the backup generate takes over. At a time when both mains and backup generator is not supplying power, the battery provides the required power to the load and it becomes discharged.

Under normal conditions, the busbar DC voltage should stay above 48 Volts. When it drops below this threshold, it's known as busbar or DC under voltage (BBUV), serving as a preliminary warning of LLVD issues. The duration until this issue becomes a failure depends on the battery's capacity. If the voltage falls below 46.5VDC, LLVD loads are disconnected, triggering an LLVD alarm. Furthermore, when the voltage dips below 44.5VDC, BLVD loads are disconnected, resulting in the interruption of all BTS site services except for communication with the monitoring system. This step safeguards the battery banks from deep discharging, which could otherwise permanently damage the battery system. While these voltage levels are set as defaults by operators, they can be adjusted to suit different site conditions and operational needs. In Net-Eco system before the power failure happens different alarms occur sequentially like mains failure alarm, battery under voltage alarm, high temperature alarm, if AC power is not restored, the final stage is LLVD low voltage alarm and BLVD low voltage alarm. So, the power failure is highly related with LLVD and BLVD alarms. These alarms are generated from the bus-bar Voltage (DC output voltage). Hence bus-bar voltage is selected as a target or dependent variable. Hence, a power failure means when LLVD and BLVD threshold level is reached and the load equipment does not obtain power.

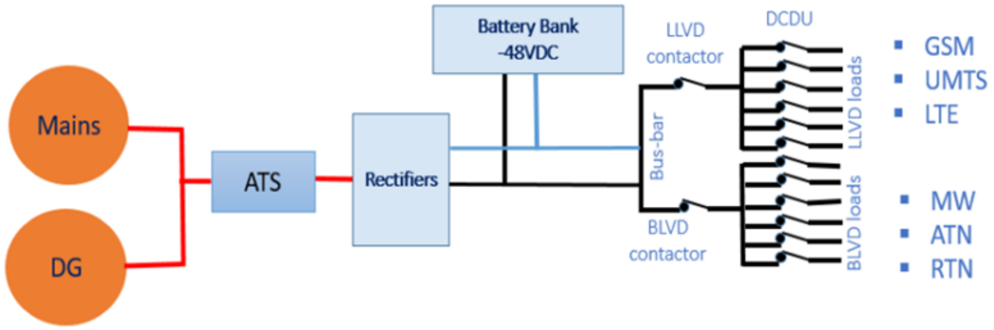

Figure 0.4.2: high level architecture for BTS power System.

Dimension Reduction

Rael world data representation involves multiple attributes, but only few of these attributes might be genuinely correlated to the target. Redundancy might exist, where certain attribute show correlation, making it unnecessary to include all of them in modeling. Feature subset selection entails identifying and eliminating irrelevant and redundant details, aiming to reduce data dimensionality. This process can enhance the efficiency of learning algorithms, enabling them to operate more swiftly and effectively [40]. Among the various dimensionality reduction techniques, correlation analysis is implemented to eliminate redundant features. Therefore, from the load side, total DC load power and total DC load current can have a linear relationship from Ohm's law of power ($P = I * V$), hence total DC load current is selected.

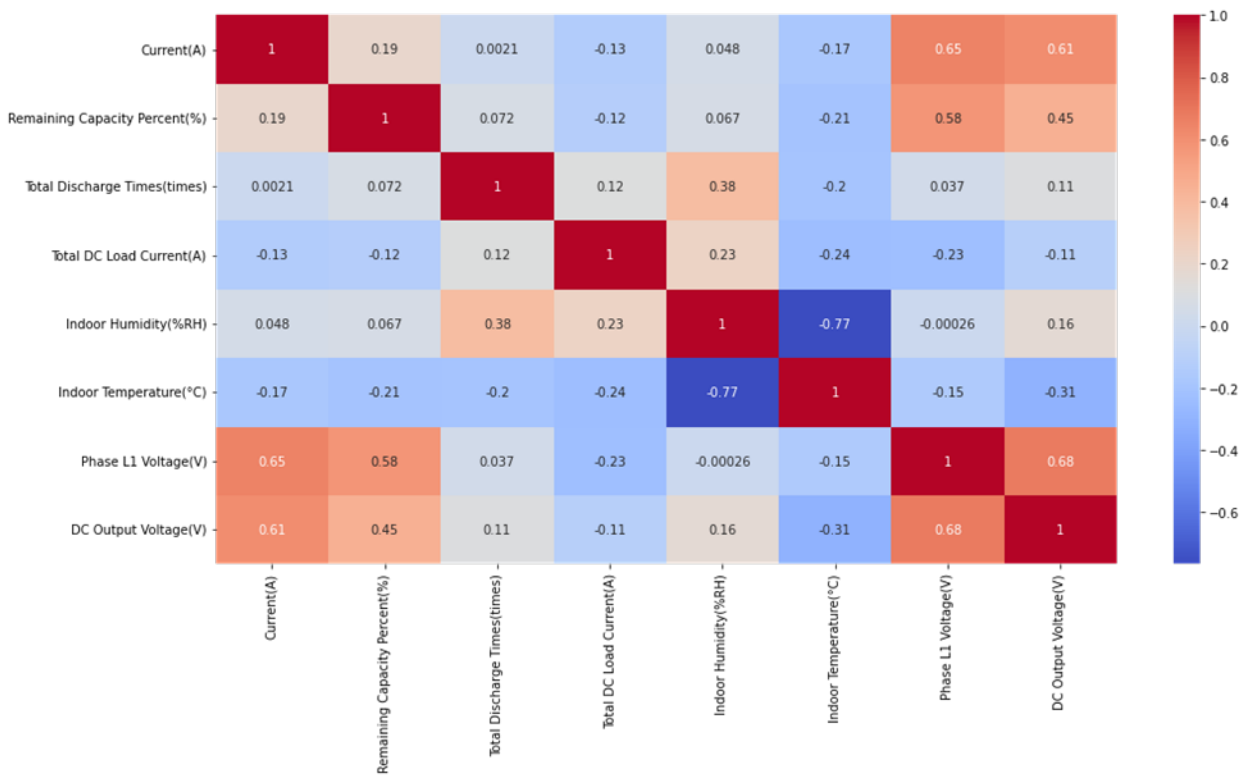

Figure 0.4.3: Feature correlation analysis.

Among the various fourteen features, eight are selected: Battery current, Battery remaining capacity, total charge-discharge times, total DC load current, Indoor humidity, phase voltage and bus-bar or DC output voltage and their correlation analysis is after feature reduction is plotted in figure 0.4.3.

Data Normalization

Normalization, refers to the process of adjusting the values of features in a dataset to ensure they are on a similar scale. The goal of normalization is to bring all features to a common scale without distorting the differences in the ranges of values. The process of normalization involves "scaling down" the features. The maximum and minimum values inside a feature are frequently very different, for example, 0.01 and 1000. The magnitudes of the values are scaled to noticeably low values after normalization. Normalization is crucial for many neural network because the algorithm perform better when features are normalized. It also prevents features with larger magnitudes from dominating the learning process. Z-score normalization and Min-Max normalization are the two most commonly used normalization techniques.

- **Min-Max Normalization:**
  Min-Max normalization scales features to a specific range, often between 0 and 1, which means the min value is 0 and the max value is 1. It's calculated as:

$$x - scale = \frac{x - min(x)}{max(x) - min(x)} \tag{0.4.4.1}$$

  Where x is the original value, min the minimum value of the feature, and max is the maximum value of the feature.

- **Z-score Normalization:** This method transforms features to have a mean of 0 and a standard deviation of 1. It's calculated as:

$$x(normalized) = \frac{(x - mean(x))}{(standarddeviation)} \tag{0.4.4.2}$$

  Where, x is ts he original value, $mean$ is the mean of the feature, and standard deviation is the standard deviation of the feature.

0.4.5 Model Performance Evaluation Metrics

Model evaluation metrics are quantitative measures used to assess the performance and effectiveness of machine learning and statistical models. These metrics provide a standardized way to compare different models and determine how well they are performing on a specific task. Model evaluation metrics help to understand the strengths and weaknesses of the model, make informed decisions about model selection and parameter tuning. When a model is built, we want to ensure that it generalizes well to unseen data and produces accurate predictions or classifications. Model evaluation metrics provide a way to quantify this performance. By comparing the model's predictions to the actual outcomes, these metrics allow us to test how closely the model's predictions match reality [46].

Different types of machine learning tasks require different evaluation metrics. For example, in classification where predicting categorical labels (classes), metrics like accuracy, precision, recall, F1-score, and ROC-AUC are commonly used to evaluate the model's performance. On the other hand, in regression algorithm, where predicting continuous numerical values, metrics like Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE) are often used to assess the model's accuracy [13]. Hence, it's important to choose the right evaluation metric based on the nature of the task and the goals that are aimed to be achieved. Since to develop our model a regression type algorithms are selected, MSE, RMSE, and MPAE are used to evaluate the performance of the model.

- **MSE:** calculates average squared difference between the predicted values and the actual values.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \quad (0.4.5.1)$$

- **RMSE:** represents the difference between the actual and the predicted value. If the difference between the errors is large, hyperparameters has to be tuned to minimize the error.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \quad (0.4.5.2)$$

- **MAPE:** it calculates the percentage of the absolute value of the error.

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}|\frac{(y_i - \hat{y}_i)}{(y_i}| * 100 \quad (0.4.5.3)$$

Where y_i is actual value and y_i predicted value The lower value of the error indicates better prediction accuracy.

0.4.6 Hyper Parameters

Training any machine learning algorithm requires selecting optimal parameters to give best result. hyperparameter are defined as the parameters that are explicitly defined by the user to control the learning process. They are set by the user before training begins. These settings influence how the model is trained, how it learns from the data, and how it makes predictions. Hyperparameters play a crucial role in determining the performance, convergence speed, and generalization ability of a model. They are often tuned and adjusted to find the best combination for a given problem [47].

Common Hyperparameters to train a deep learning algorithms include:

- **Learning Rate:** Determines the step size taken during each iteration to update model parameters. A higher learning rate may lead to faster convergence but can overshoot the optimal solution, while a lower learning rate may converge more slowly but with greater stability.

- **Number of Hidden Units or Layers:** The number of hidden layers and the number of neurons (units) in each layer influence the complexity and capacity of the model. Too few units or layers may result in under fitting; too many may lead to overfitting [48].

- **Batch Size:** Batch size defines the number of training examples used in each iteration. It can influence the speed of convergence and the memory requirements of training.

- **Epochs:** The number of training epochs is a hyperparameter that defines how many times the entire training dataset is processed by the model. Setting the right number of epochs is crucial to prevent overfitting or underfitting.

- **Activation Functions:** common activation functions include Rectified Linear Unit (ReLU), sigmoid, and tanh. The choice of activation function can impact the network's ability to model nonlinear relationships.

- **Dropout Rate**: Dropout is a regularization technique where a certain percentage of neurons are randomly deactivated during training to prevent overfitting. The dropout rate is a hyperparameter that determines the fraction of neurons dropped out.

- **Optimizer**: The choice of optimization algorithm (e.g., Adam, RMSprop, stochastic gradient decent (SGD)) is a hyperparameter that impacts the update rules for the model's weights during training.
- **Number of Convolutional Layers:** Determines the depth and complexity of the model.
- **Filter Size and Number of Filters:** The size and quantity of filters in each convolutional layer. Larger filters capture more complex features, but also increase computational cost.
- **Pooling Size and Type:** Determines the size of pooling regions and the type of pooling (e.g., max-pooling).

Tuning these hyperparameters often involves experimentation and can be done using techniques like grid search or random search. It's also important to perform proper validation and testing to ensure the chosen architecture and hyperparameters generalize well to unseen data [21].

0.5 Result and Discussion

This chapter presents an overview of the functionalities of experimentation's model and tools. Additionally, it provides a clear explanation of the dataset employed for model development. The metrics chosen for evaluating model performance are also outlined. The study's findings and their significance are explored, presenting insights gained from the analysis. The positive outcomes and challenges are examined through a close review of the results. Context is provided by comparing and interpreting the results, aiding in drawing conclusions, identifying future research areas, and contributing to the broader topic's understanding.

0.5.1 Dataset Description

The input to the system is multivariate time series data, which means that it is a collection of data points that are collected for five months (March to July, 2023) and that have multiple variables. The data set is collected from ethio telecom power and environment monitoring, called Net-eco system. The data has fourteen features and forty-four thousand and fifty-two samples which is collected in five-minute intervals, that means each data point represents the measurement of the fourteen variables at a specific five-minute time window. This allows the system to capture the dynamics of the power system and to identify patterns that may be indicative of a power failure. The multivariable data set contains battery related features, rectifier related feature, environment related features, mains supply related feature and load related features. The multivariate data is a valuable source of information for predicting power failures. By analyzing the data, the system can identify patterns that are associated with power failures. These patterns can then be used to train a machine learning model to predict power failures.

Once the input data is collected, it is pre-processed to remove noise and outliers, treat missing values, normalize the data, and remove correlated features. This is important for the machine learning model to have good generalization, which means that it can accurately predict power failures on new data that it has not seen before.

Missing values can occur in data for a variety of reasons, such as sensor malfunctions or human errors. It is important to treat missing values before training the machine learning model, as they can bias the model and lead to inaccurate predictions. There are a number of different ways to treat missing values. One common approach is to impute the missing values, which means to replace them with estimated values. The estimated values can be generated using statistical methods, such as the mean or median of the data. Another approach to treating missing values is to delete the data points that contain missing values. This can be done if the number of missing values is small, or if the data points are not essential for the analysis. In our data set missing value record is obtained in phase voltages when there is no utility power on the site. Hence it is imputed with zero.

Data normalization is the process of scaling the values of the features in the data so that they have a common range. This is important for machine learning models, as it helps to ensure that the different features have a comparable influence on the model. There are a number of different ways to normalize data. One common approach is to scale the values of each feature to the range [0, 1]. This can be done by using min-max normalization technique.

Feature correlation is the degree to which two features are related to each other. Features that are highly correlated can introduce redundancy into the machine learning model, which can lead to overfitting. Overfitting occurs when the machine learning model learns the patterns in the training data too well, and is unable to generalize to new data. This can lead to inaccurate predictions [49]. To avoid overfitting, it is important to remove features that are highly correlated. This is done by using a correlation matrix to identify the features that are most correlated, and then removing the least important features. After removing least important features, the dimension of the feature is reduced to eight.

The target variable is the variable that the machine learning model is trying to predict. In this case, the target variable indicates whether or not a power failure will occur. It is important to select the target variable carefully, as the performance of the machine learning model will depend on the quality of the target variable. Hence the target variable is selected based on its relation on the site power failure that leads the site service to disrupt. In our case DC output voltage is selected.

These pre-processing steps are essential for ensuring the accuracy and performance of the machine learning model. By removing noise and outliers, treating missing values, normalizing the data, removing correlated features, and selecting the target variable carefully, the machine learning model is less likely to be biased and is more likely to learn the underlying patterns in the data.

0.5.2 CNN-based Power Failure Prediction Model Building

The use of multivariate data for predicting power failures is a promising approach. This approach has been shown to be effective in a variety of studies. However, there are still some challenges that need to be addressed. One challenge is the need for large volumes of data.

The data for power failure prediction is collected in a monthly basis, Net-eco system does not provide more than one month dat. Another challenge is the need to develop machine learning models that are able to learn complex patterns in the data. Despite these challenges, the use of multivariate data for predicting power failures is a promising approach that has the potential to significantly improve the reliability of power systems.

The model development process encompasses the utilization of three distinct algorithms: CNN, LSTM, and a hybrid architecture known as CNN-LSMT a family of deep neural network. Rooted in the domain of deep learning, a subset of machine learning methodologies, these algorithms exhibit enhanced potency and versatility compared to traditional machine learning techniques. Notably, the chief advantage of deep learning algorithms lies in their capacity to automatically extract features directly from the data, avoiding the necessity for manual feature extraction, a task that conventionally requires human intervention [50].

The model development within this thesis adopts a stratified approach to dataset division, allocating segments for training, validation, and testing purposes, pertinent to the CNN, LSTM, and hybrid CNN-LSTM models. The data extraction is from March-1 to July-31 2023, 80% of the dataset serves to establish the model's parameters and specifications during the training phase, with an additional 10% reserved for the validation subset. Given the potential impact of hyperparameters on the ultimate model's speed and accuracy, a preliminary validation data assessment is conducted before the model's final deployment. The remaining 10% is dedicated to serving as the test dataset, enabling an empirical performance evaluation among the various prediction models.

CNN-Based Prediction Model

CNNs have emerged as a powerful tool in the field of image and sequence analysis. Originally designed for image recognition tasks, CNNs have been adapted to various applications including time series data analysis, such as equipment failure prediction, network failure prediction, stock market forecasting and so on.

In the context of power failure prediction, CNNs excel at capturing temporal patterns within the data [51]. By utilizing convolutional layers, these algorithms automatically learn hierarchical features from the input sequences, effectively recognizing relevant patterns, such as battery voltage, battery charge discharge cycle, battery remaining capacity, room temperature, rectifier output current, DC load current, DC output voltage, that might indicate an imminent power failure event. CNNs' ability to discern complex relationships within temporal data makes them a promising candidate for accurate and robust failure prediction models [52].

The model is built by training dataset (80% of the data) and by tuning different hyperparameters. Hyperparameters play a vital role in shaping the learning process of a neural network and ultimately influence the model's performance and ability to generalize to new data. The specific values chosen for hyperparameters can significantly impact the training process, convergence speed, and the final performance of the model. It's crucial to experiment with various combinations of hyperparameters to find the optimal configuration that yields the best results for a given problem. The hyperparameters used to develop CNN model are listed in the following table. The process of training a model involves finding the right balance of hyperparameters

CNN Model Hyperparameters	
Hyperparameters	Values
Number of Filters	512
Kernel Size	5
Pool Size	2
Learning Rate	0.001
Number of Epochs	200
Activation Function	ReLu
Optimizer	Adam

Table 0.5.1: CNN Model Hyperparameters.

to ensure it neither over-fits nor under-fits the data. Over-fitting occurs when the model grasps training data details and noise to a degree that harms performance on new data, while under-fitting results in a model that struggles to represent both the training data and unseen data. Optimal hyperparameters are determined through parameter tuning, using a training dataset. Model performance is then assessed using a validation dataset. Prior to final evaluation on a test set, the models fit, whether overfitting, under fitting, or appropriate is carefully examined. This approach ensures the model attains its highest potential performance. Hence, training and validation loss is curve is plotted as in Figure 0.5.1, to see overfitting or underfitting of the model.

As it can be seen from the plot, the model has a good fit and it is now ready to evaluate its performance with a test data set.

When assessing the developed model with the test data, the prediction output is graphically represented in Figure 0.5.2 below, showing a comparison between the ten days predicted DC output voltage and the actual DC output voltage. Notably, the two plots exhibit a strikingly similar pattern, indicating a meaningful alignment between the predicted values and the actual value. This alignment suggests that the model's predictive capabilities accurately capture

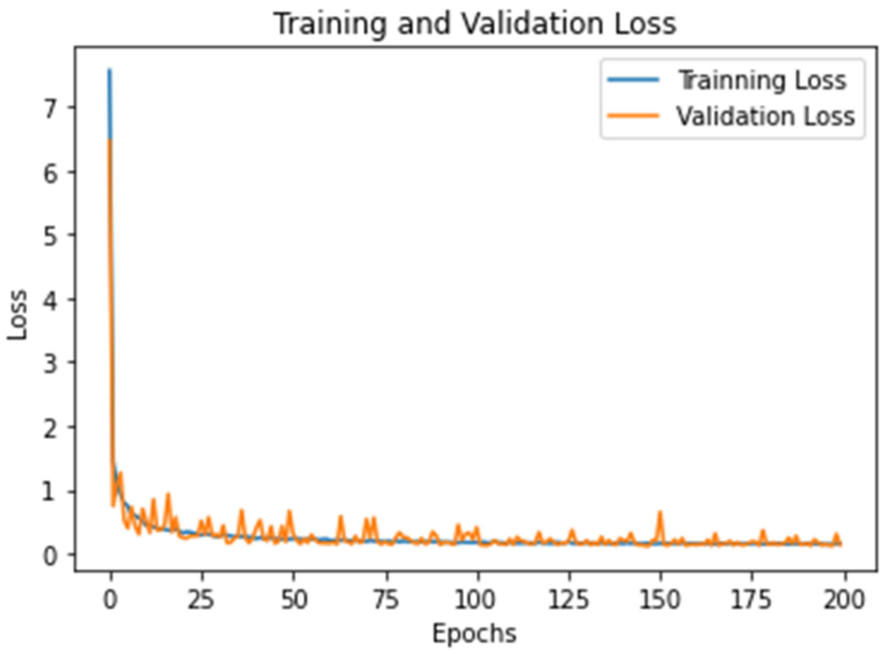

Figure 0.5.1: Training vs. Validation Loss for CNNN Model.

the underlying trends and variations in the DC output voltage. The proposed CNN model

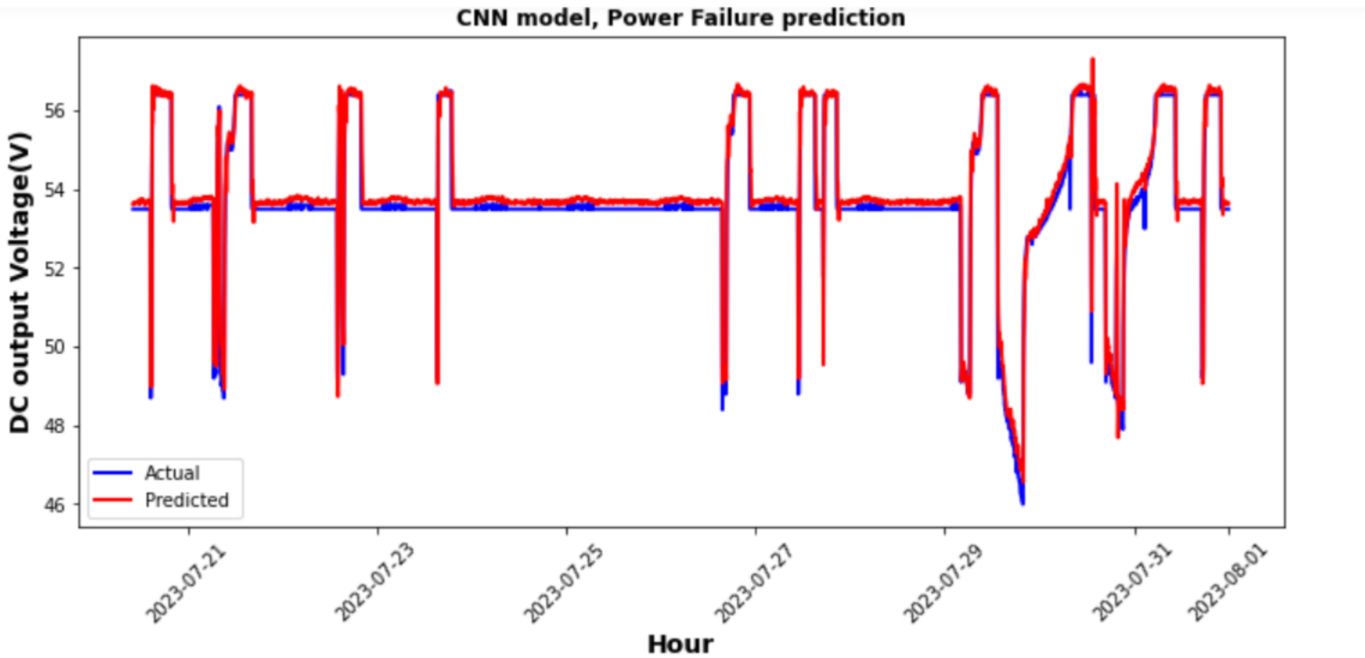

Figure 0.5.2: Actual verses predicted Plot for CNN Model for power failure prediction.

performance is evaluated with three regression problem evaluation metrics, and the result of evaluation are provided in in Table 4.2 below

CNN Model Evaluation Result	
Evaluation Metrics	Evaluation Result
MSE	0.0223
RMSE	0.472
MAPE	2.643

Table 0.5.2: CNN Model Evaluation Result.

0.5.3 LSTM-Based Prediction Model

In the realm of artificial intelligence, LSTM models have emerged as a cornerstone technology for processing sequential data. As a specialized type of RNN, LSTMs excel in capturing intricate temporal dependencies, making them exceptionally well-suited for a wide array of applications, ranging from natural language processing and speech recognition to time series forecasting and beyond [39].

Training an LSTM network using unscaled data that spans a wide range of values, such as quantities ranging from the tens to the hundreds, can lead to potential challenges. Specifically, when the inputs to the network include large values, they have the capability to impede the learning process and hinder the network's convergence. In certain scenarios, this situation can even inhibit the network from effectively grasping the intricacies of the problem at hand. As such, data scaling becomes pivotal as it allows the network to handle inputs more uniformly and facilitates a smoother learning trajectory, ultimately enhancing the network's ability to effectively learn and generalize patterns within the data. Hence, the data is normalized using min-max normalization technique. Finally, the LSTM model is built using the hyperparameters listed in Table (4.2). The model is built using a training samples (80% of the total data) to train power failure prediction model [53].

LSTM Model Hyperparameters	
Hyperparameters	Values
Number of Neurons	64
Hidden layer	3
Learning Rate	0.001
Number of Epochs	500
Activation Function	ReLu
Optimizer	Adam

Table 0.5.3: LSTM Model Hyperparameters.

LSTM model is built by tuning different hyperparameters until best fit is obtained. Figure 0.5.3 below illustrates the training vs validation loss. The best fit of the model is attained by using hyperparameters listed in Table 0.5.3 above.

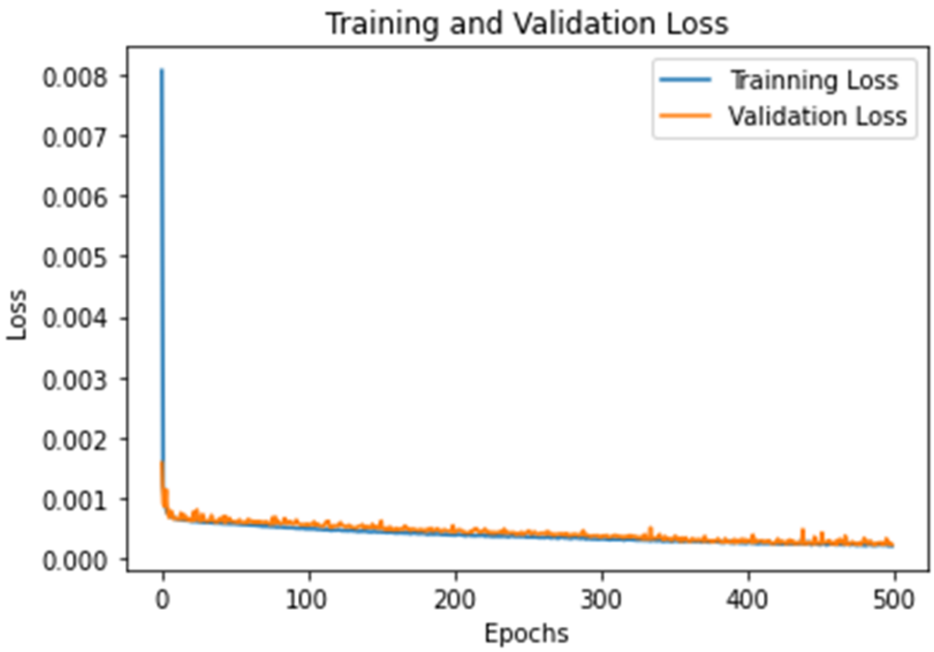

Figure 0.5.3: Training vs. Validation Loss for LSTM Model.

Once the best fit of the model is obtained, the actual and predicted result of the model for power failure prediction is demonstrated as in Figure 0.5.4. AS it can be seen clearly from the plot, the actual DC output voltage and the predicted DC output voltage almost have similar pattern. The actual and predicted plot for power failure model shows how the model learns

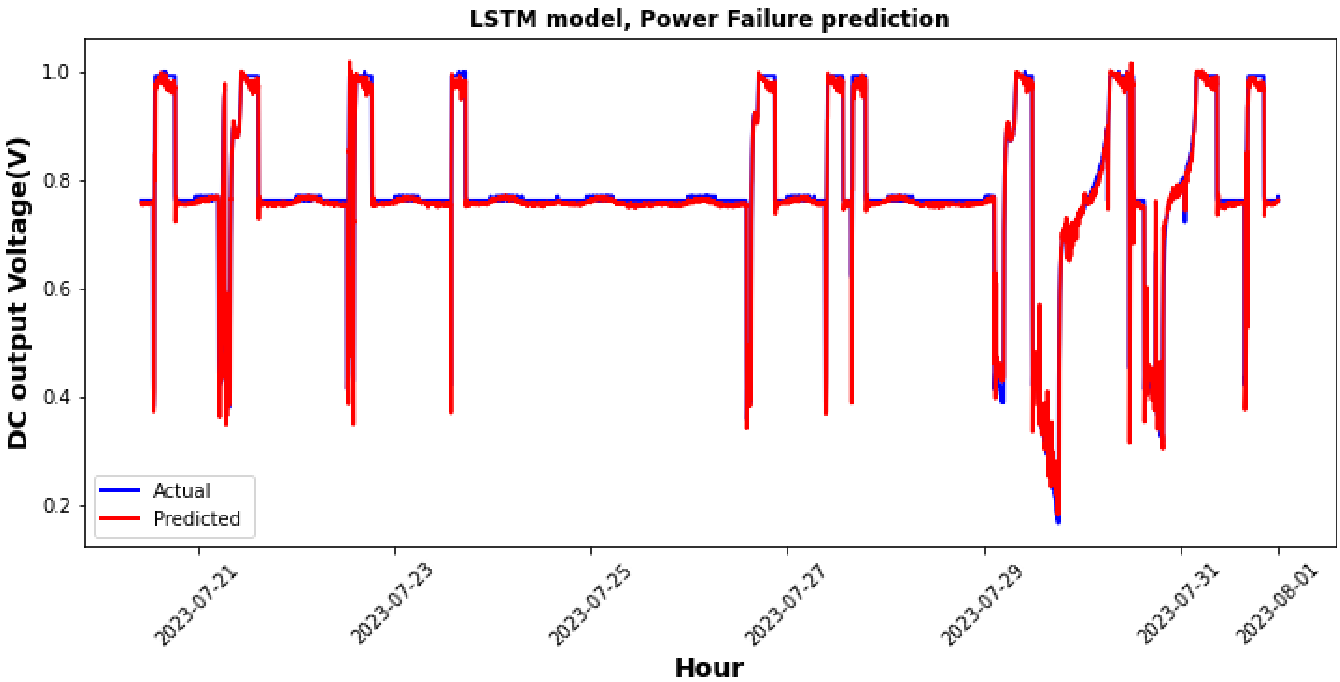

Figure 0.5.4: Actual verses predicted Plot for LSTM Model for power failure prediction.

the pattern very well and when the performance of the proposed model is evaluated, it has a minimal error as indicated in Table 0.5.4 below.

LSTM Model Evaluation Result	
Evaluation Metrics	Evaluation Result
MSE	0.001
RMSE	0.037
MAPE	1.194

Table 0.5.4: LSTM Model Evaluation Result.

0.5.4 Hybrid CNN-LSTM Prediction Model

The other machine learning algorithms that can be used for this purpose is a hybrid CNN-LSTM model. CNNs are well-suited for extracting features from time series data, while LSTMs are well-suited for capturing long-term dependencies in the data. The hybrid CNN-LSTM model combines the strengths of both CNNs and LSTMs, making it a powerful tool for predicting power failures [34].

The CNN-LSTM model receives time-series input in the same way as the previous two models. The CNN-LSTM model used in this study is shown in Figure 0.5.5. The convolution layer and pooling layer of CNN extract the features of the power data by convolution operation and pooling operation respectively, and the features containing the most important information of the input sequence can be obtained. The feature matrix extracted from CNN model is input into LSTM neural network structure for power failure prediction. The 1D convolution layer has

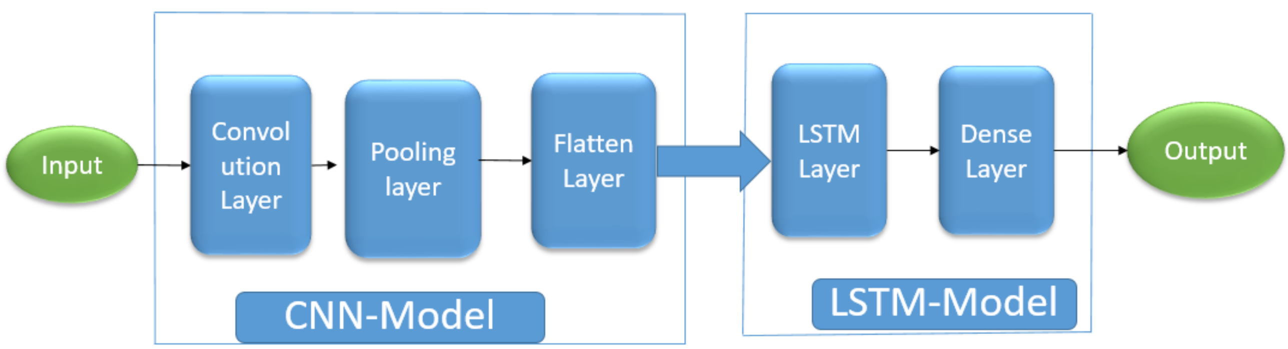

Figure 0.5.5: Actual Vs Predicted plot for CNN-LSTM Sytem Model.

256 filters with a kernel size of five, and uses ReLU activation. The pooling layer has a size of two. LSTM uses 64 neurons and ReLu activations, Adam is as optimizer and all parameters are regularized with a learning rate of 0.001 as stated in Table 0.5.5 below.

Training verses validation loss of the model is plotted indicating the normal fit of the model by employing the selected hyperparameters. The output of the proposed model for the power

CNN-LSTM Model Hyperparameters	
Hyperparameters	Values
Number of Filters	256
Kernel Size	5
Pool Size	2
Number of Neurons	64
Hidden layer	3
Learning Rate	0.001
Number of Epochs	500
Activation Function	ReLu
Optimizer	Adam

Table 0.5.5: CNN-LSTM Model Hyperparameters.

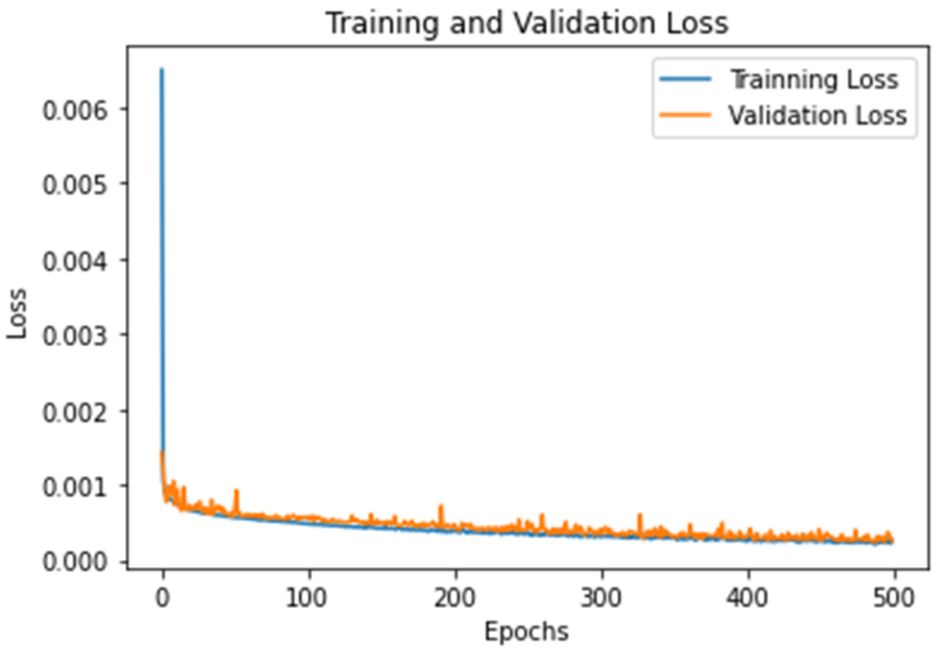

Figure 0.5.6: Train vs Val loss for CNN-LSTM.

failure prediction is illustrated in Figure 4.5 below, which aids in comparing and visualizing the actual and prediction result. As it is clearly shown the actual and predicted result for power failure prediction model almost have similar pattern which enables it to capture the pattern in the data set.

The proposed model performance is assessed with thre evaluation metrics and the result of the evaluation metrics are provided in Table 4.4 below. The performance of the three models was assessed based on their prediction accuracy and the time it took to train them. All three models performed well on the power failure problem, but there were some differences in their performance.

The CNN model had a slightly higher error rate than the other two models, but it took the least amount of time to train, at one hour. The LSTM model had the best accuracy, with a

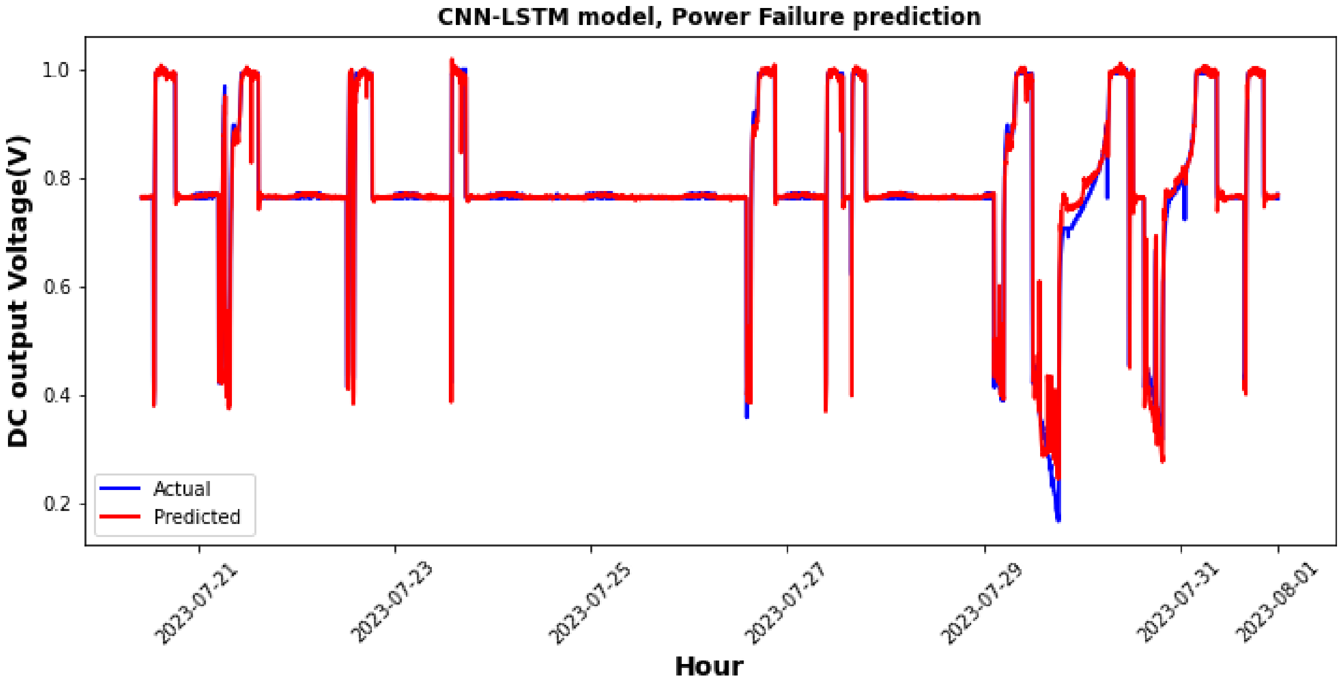

Figure 0.5.7: Actual verses predicted Plot for CNN-LSTM Model for power failure prediction.

CNN-LSTM Model Evaluation Result	
Evaluation Metrics	Evaluation Result
MSE	0.001
RMSE	0.038
MAPE	2.528

Table 0.5.6: CNN-LSTM Model Evaluation Result.

MSE of 0.001, but it took the longest time to train, at three and half hours.

The hybrid CNN-LSTM model had a computation time that was less than that of the LSTM model, but its MSE was almost as good as the LSTM model. This is because the hybrid model takes advantage of the strengths of both the CNN and LSTM algorithms

0.6 Conclusion and Recommendation

0.6.1 Conclusion

This study investigated the intricate realm of Machine Learning for Power Failure Prediction in BTS: A Multivariate Approach. Aiming to enhance the reliability of communication service by minimizing service disruptions whose major cause was power failure on the site. This can be realized by developing a machine learning model. Findings underscored the efficacy of employing a multivariate approach to predict power failures, demonstrating its potential to significantly mitigate service disruptions and boost operational efficiency.

In this study, three distinct deep learning variants CNN, LSTM, and the hybrid CNN-LSTM were used to develop a robust predictive model. Throughout the study, multi variate data was collected for five months and by employing precise feature engineering, the predictive method exhibited notable performance across diverse evaluation metrics.

Among the three algorithms, the CNN model emerged as the swiftest in training, demanding an hour to converge. On the other end of the spectrum, the LSTM model exhibited the most extended training time, spanning three and a half hours, yet compensating with superior prediction accuracy. The hybrid CNN-LSTM model, capitalizing on the strengths of its parent algorithms, wielded the power of CNN for extracting salient features from complex, nonlinear datasets. Concurrently, it harnessed LSTM's adeptness in capturing long short time-series dependencies. This dual advantage culminated in a predictive mechanism that not only reduced training time when compared to LSTM now standing at two hours but also maintained prediction accuracy comparable to that of LSTM. The synthesis of these methodologies has yielded a predictive framework that empowers stakeholders with advanced insights into potential power failures, facilitating proactive measures to avert service interruptions.

As it is reflected on this research journey, it becomes evident that the symbiotic relationship between machine learning and power failure prediction holds immense promise for the telecommunications sector. The successful integration of predictive analytics into BTS management not only enhances service reliability but also minimizes downtime and maintenance costs.

0.6.2 Recommendation

Future work could delve into the integration of real-time data streams from diverse sources within BTS, such as network traffic data and O&M records will expand the dataset which could provide a more comprehensive understanding of power failure predictions and enhance the predictive capabilities of the models.

Additionally, investigating the adaptability of the developed models to different BTS sites and geographical locations could yield insights into the transferability and robustness of the predictive framework.

Finally, an in-depth economic analysis of the implemented predictive system's cost-effectiveness in terms of minimizing service disruptions and operational downtime would provide valuable insights for industry adoption.

References

1. Mishra, A. R. *Advanced cellular network planning and optimisation: 2G/2.5 G/3G... evolution to 4G* (John Wiley & Sons, 2007).

2. Tefera, Y. Y., Kibatu, T., Shawel, B. S. & Woldegebreal, D. H. *Recurrent Neural Network-based Base Transceiver Station Power Supply System Failure Prediction* in *2020 International Joint Conference on Neural Networks (IJCNN)* (2020), 1–7.

3. Chayanam, K. *Analysis of Telecommunications Outages due to power loss* PhD thesis (Ohio University, 2005).

4. Wahid, A., Breslin, J. G. & Intizar, M. A. Prediction of Machine Failure in Industry 4.0: A Hybrid CNN-LSTM Framework. *Applied Sciences* **12,** 4221 (2022).

5. Fan, X., Wang, X., Zhang, X. & Yu, P. A. X. B. Machine learning based water pipe failure prediction: the effects of engineering, geology, climate and socio-economic factors. *Reliability Engineering & System Safety* **219,** 108185 (2022).

6. Celikmih, K., Inan, O. & Uguz, H. Failure prediction of aircraft equipment using machine learning with a hybrid data preparation method. *Scientific Programming* **2020** (2020).

7. Nguyen, K.-A., Do, P. & Grall, A. Multi-level predictive maintenance for multi-component systems. *Reliability engineering & system safety* **144,** 83–94 (2015).

8. Rauf, H., Khan, H. A. & Arshad, N. *Optimized power system planning for base transceiver station (BTS) based on Minimized power consumption and cost* in *2019 8th International Conference on Renewable Energy Research and Applications (ICRERA)* (2019), 773–779.

9. Yuan, L., Zhuojian, W., Zhe, L. & Zihan, J. *Research on fault prognosis methods based on data-driven: a survey* in *IOP Conference Series: Materials Science and Engineering* **1043** (2021), 042008.

10. Betti, A. *et al.* Fault prediction and early-detection in large pv power plants based on self-organizing maps. *Sensors* **21,** 1687 (2021).

11. Zhang, S., Wang, Y., Liu, M. & Bao, Z. Data-based line trip fault prediction in power systems using LSTM networks and SVM. *Ieee Access* **6,** 7675–7686 (2017).

12. Wang, Z. *et al.* Failure prediction using machine learning and time series in optical network. *Optics Express* **25,** 18553–18565 (2017).

13. Zheng, L. *et al. A fault prediction of equipment based on CNN-LSTM network* in *2019 IEEE international conference on energy internet (ICEI)* (2019), 537–541.

14. Xie, Y. *et al.* Attention mechanism-based CNN-LSTM model for wind turbine fault prediction using SSN ontology annotation. *Wireless Communications and Mobile Computing* **2021,** 1–12 (2021).

15. Zang, H. *et al.* Short-term global horizontal irradiance forecasting based on a hybrid CNN-LSTM model with spatiotemporal correlations. *Renewable Energy* **160,** 26–41 (2020).

16. Tan, Z. & Pan, P. *Network fault prediction based on CNN-LSTM hybrid neural network* in *2019 International Conference on Communications, Information System and Computer Engineering (CISCE)* (2019), 486–490.

17. Lee, J., Choi, W. & Kim, J. A cost-effective CNN-LSTM-based solution for predicting faulty remote water meter reading devices in AMI systems. *Sensors* **21,** 6229 (2021).

18. Reeve, W. D. *DC power system design for telecommunications* (John Wiley & Sons, 2006).

19. Rao, S. S. & Amer, G. *Power supply architecture for telecom application: A review* in *2010 international conference on communication control and computing technologies* (2010), 262–264.

20. Deevela, N. R., Kandpal, T. C. & Singh, B. A review of renewable energy based power supply options for telecom towers. *Environment, Development and Sustainability,* 1–68 (2023).

21. Probst, P., Boulesteix, A.-L. & Bischl, B. Tunability: Importance of hyperparameters of machine learning algorithms. *The Journal of Machine Learning Research* **20,** 1934–1965 (2019).

22. Ongsulee, P. *Artificial intelligence, machine learning and deep learning* in *2017 15th international conference on ICT and knowledge engineering (ICT&KE)* (2017), 1–6.

23. Gupta, A. Introduction to deep learning: part 1. *Chemical Engineering Progress* **114,** 22–29 (2018).

24. Ishaq, O. *Image Analysis and Deep Learning for Applications in Microscopy* PhD thesis (Acta Universitatis Upsaliensis, 2016).

25. Aggarwal, C. C. & Aggarwal, C. C. Teaching deep learners to generalize. *Neural Networks and Deep Learning: A Textbook,* 169–216 (2018).

26. Ghimire, S., Deo, R. C., Raj, N. & Mi, J. Deep solar radiation forecasting with convolutional neural network and long short-term memory network algorithms. *Applied Energy* **253,** 113541 (2019).

27. Rajagukguk, R. A., Ramadhan, R. A. & Lee, H.-J. A review on deep learning models for forecasting time series data of solar irradiance and photovoltaic power. *Energies* **13,** 6623 (2020).

28. Kumari, P. & Toshniwal, D. Deep learning models for solar irradiance forecasting: A comprehensive review. *Journal of Cleaner Production* **318,** 128566 (2021).

29. Ju, Y. *et al.* A model combining convolutional neural network and LightGBM algorithm for ultra-short-term wind power forecasting. *Ieee Access* **7,** 28309–28318 (2019).

30. Alzubaidi, L. *et al.* Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *Journal of big Data* **8,** 1–74 (2021).

31. Yang, Y., Guo, X.-M., Wang, H. & Zheng, Y.-N. Deep learning-based heart sound analysis for left ventricular diastolic dysfunction diagnosis. *Diagnostics* **11,** 2349 (2021).

32. Shawel, B. S., Mare, E., Debella, T. T., Pollin, S. & Woldegebreal, D. H. A Multivariate Approach for Spatiotemporal Mobile Data Traffic Prediction. *Engineering Proceedings* **18,** 10 (2022).

33. Wu, Y., Yang, F., Liu, Y., Zha, X. & Yuan, S. A comparison of 1-D and 2-D deep convolutional neural networks in ECG classification. *arXiv preprint arXiv:1810.07088* (2018).

34. Moradzadeh, A., Teimourzadeh, H., Mohammadi-Ivatloo, B. & Pourhossein, K. Hybrid CNN-LSTM approaches for identification of type and locations of transmission line faults. *International Journal of Electrical Power & Energy Systems* **135,** 107563 (2022).

35. Wang, R., Vanga, V. L., Zaiken, Z. B. & Bennett, J. Analysis of the electric power outage data and prediction of electric power outage for major metropolitan areas in Texas using Machine Learning and Time Series Methods. *SMU Data Science Review* **6,** 5 (2022).

36. Jaech, A., Zhang, B., Ostendorf, M. & Kirschen, D. S. Real-time prediction of the duration of distribution system outages. *IEEE Transactions on Power Systems* **34,** 773–781 (2018).

37. Dashti, R., Daisy, M., Mirshekali, H., Shaker, H. R. & Aliabadi, M. H. A survey of fault prediction and location methods in electrical energy distribution networks. *Measurement* **184,** 109947 (2021).

38. Lim, B. & Zohren, S. Time-series forecasting with deep learning: a survey. *Philosophical Transactions of the Royal Society A* **379,** 20200209 (2021).

39. Sagheer, A. & Kotb, M. Time series forecasting of petroleum production using deep LSTM recurrent networks. *Neurocomputing* **323,** 203–213 (2019).

40. Fan, C., Chen, M., Wang, X., Wang, J. & Huang, B. A review on data preprocessing techniques toward efficient and reliable knowledge discovery from building operational data. *Frontiers in Energy Research* **9,** 652801 (2021).

41. Alasadi, S. A. & Bhaya, W. S. Review of data preprocessing techniques in data mining. *Journal of Engineering and Applied Sciences* **12,** 4102–4107 (2017).

42. Pratama, I., Permanasari, A. E., Ardiyanto, I. & Indrayani, R. *A review of missing values handling methods on time-series data* in *2016 international conference on information technology systems and innovation (ICITSI)* (2016), 1–6.

43. Kotsiantis, S. B., Kanellopoulos, D. & Pintelas, P. E. Data preprocessing for supervised leaning. *International journal of computer science* **1,** 111–117 (2006).

44. Han, J., Zheng, H., Xing, Y., Chen, D. Z. & Wang, C. V2V: A deep learning approach to variable-to-variable selection and translation for multivariate time-varying data. *IEEE Transactions on Visualization and Computer Graphics* **27,** 1290–1300 (2020).

45. Sparks, R., Zucchini, W. & Coutsourides, D. On variable selection in multivariate regression. *Communications in Statistics-Theory and Methods* **14,** 1569–1587 (1985).

46. Kwon, J. E., Shifat, T. A., Kareem, A. B. & Hur, J.-W. RUL prediction of switched mode power supply using a Kalman filter assisted deep neural network. *Processes* **10,** 55 (2021).

47. Bergstra, J., Yamins, D., Cox, D. D., *et al. Hyperopt: A python library for optimizing the hyperparameters of machine learning algorithms* in *Proceedings of the 12th Python in science conference* **13** (2013), 20.

48. Panchal, G., Ganatra, A., Kosta, Y. & Panchal, D. Behaviour analysis of multilayer perceptrons with multiple hidden neurons and hidden layers. *International Journal of Computer Theory and Engineering* **3,** 332–337 (2011).

49. Miao, J. & Niu, L. A survey on feature selection. *Procedia computer science* **91,** 919–926 (2016).

50. Graham, S. A. *et al.* Artificial intelligence approaches to predicting and detecting cognitive decline in older adults: A conceptual review. *Psychiatry research* **284,** 112732 (2020).

51. Hoseinzade, E. & Haratizadeh, S. CNNpred: CNN-based stock market prediction using a diverse set of variables. *Expert Systems with Applications* **129,** 273–285 (2019).

52. Ji, W., Duan, S., Chen, R., Wang, S. & Ling, Q. *A CNN-based network failure prediction method with logs* in *2018 Chinese Control And Decision Conference (CCDC)* (2018), 4087–4090.

53. Brownlee, J. *Deep learning for time series forecasting: predict the future with MLPs, CNNs and LSTMs in Python* (Machine Learning Mastery, 2018).

Appendix

Machine Learning for Power Failure Prediction in Base Transceiver Stations: A Multivariate Approach

1st Sofia Ahmed
Department of Electrical and Computer Engineering
Addis Ababa Institute of Technology
Addis Ababa, Ethiopia
ahmed.sofia0@gmail.com

2nd Dereje Hailemariam
Department of Electrical and Computer Engineering
Addis Ababa Institute of Technology
Addis Ababa, Ethiopia
dereje.hailemariam@aait.edu.et

***Abstract*—The proliferation of mobile cellular networks has revolutionized economic and social activities, with Base Transceiver Stations (BTSs) serving as crucial components in delivering wireless services. However, BTS power failures disrupt mobile services, inconveniencing users and causing financial losses for providers. This paper presents a novel approach to preemptively address BTS power interruptions through a machine learning-based framework. Leveraging multivariate time-series data from BTS power and environmental monitoring systems, we employ Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM), and CNN-LSTM networks to predict power failures.**

All three algorithms effectively capture temporal patterns, with LSTM consistently outperforming the others, albeit with longer training times (over three hours). The CNN-LSTM model offers efficient training (around two hours) with competitive performance (0.001 MSE and 2.528 MAPE), while the CNN model provides swift computations (0.223 MSE and 2.843 MAPE). This research emphasizes the predictive aspect, offering a robust model tailored for BTS power systems. Enabling timely maintenance and reducing downtime, our approach enhances telecommunications infrastructure reliability, leading to improved user experiences and streamlined service provider operations.

***Index Terms*—BTS, Power failure, Deep learning, LSTM, CNN, hybrid CNN-LSTM Model,**

I. Introduction

Base transceiver station (BTS) plays a crucial role in interconnecting client devices to the Core Network through a radio interface, so, BTS can facilitate wireless communication between subscriber devices and telecom operators' networks [1].

There are different network components in BTS site which can obtain power from a common source to accomplish their intended tasks. If power fails at a site, all the components' functionality is interrupted. A power failure may happen because of loss of commercial power or mains failure, failure of consequent backup power system, or the failure of power-related equipment such as rectifiers, circuit breakers, or fuses that are connected between the power input and the communications equipment. On the other hand, a power failure can be triggered by lightning or heavy rain, which can generate voltage fluctuations that result in a power surge and damage the rectifier or open circuit breakers [2].

Since BTS power failure is one of the critical issues for network operators, failing to handle this power failure results in degradation of quality of service (QoS), service availability, customer dissatisfaction, and high operational expenditure (OPEX). Finally, it results in a loss of revenue and economic loss to both mobile service users and network operators. This problem is worse if the site is a hub site because its failure results in the failure of other sites too. Also, in remote areas, the problem persists for a long time until technicians travels to reach the site for troubleshooting, resulting in a higher mean time to repair and maximum operational cost. So, to handle these challenges, failure prediction in advance is important for performing predictive maintenance, preventing the occurrence of failure and minimizing maintenance costs.

Failure prediction is the process of predicting whether a material system of interest will fail at a certain point in time in the future. Hence, failure prediction and detection in advance in any industrial system is an important task to assure reliable service by developing a machine learning model [3]. Nowadays, data-driven failure prediction is becoming popular and being implemented in different fields by using different types of machine learning algorithms [4] [5].

Predicting power failures in BTS sites offers several advantages that can significantly enhance the reliability and efficiency of communication networks. By accurately predicting the failure in advance, network operators can take proactive measures to mitigate the impact of failure which helps them to minimize service disruption and maintain seamless communication, mobile network users, whether individuals or business, highly demand uninterrupted service but power failures can lead to dropped calls, slow data speed and poor network connectivity. So predictive power failure models enables operators to prevent such disruptions, leading to improved customer satisfaction and retention. The other advantage is, it enables operators to allocate resources more effectively by optimizing the usage of backup power sources and dispatching maintenance teams to sites that are at higher risk of experiencing outages. This not only reduces operational costs but also ensures that resources are directed where they are most needed. Predictive models provide network operators with data-driven insights into the factors contributing to power failure. This information can guide operational decisions, maintenance schedules and infrastructure upgrades based on empirical evidence rather than relying on reactive responses.

A. *Literature Review*

The literature on fault prediction in industrial systems has witnessed a growing interest in recent years due to its potential for enhancing system reliability and reducing downtime. In [6], the authors present an innovative solution for fault prediction method using SCADA data, focusing on generic fault detection with a self-organizing map (SOM) and a unique Key Performance Indicator (KPI). Tested on three PV plants and over sixty inverter modules, the model demonstrated the ability to predict generic faults up to 7 days in advance with a remarkable 95% true positive rate. Its adaptability and simplicity, requiring only historical SCADA data, fault taxonomy, and inverter datasheets, make it a valuable tool for enhancing industrial system reliability and reducing downtime.

In [7] proposed a fault prediction method for line trip in power systems capturing electrical measurement, and multi-source time series data like current, voltage, active and reactive power. The method was developed using LSTM to capture the temporal features of the data in the long time span and SVM for the classification of the faults. During model development, overfitting was avoided through the use of dropout and batch normalization. The model was trained using LSTM in 5-fold cross-validation and using RMS prop optimizer a good prediction performance was obtained and some improvement was observed after feature extraction. But in time series data, splitting using cross-validation is not recommended because test data may occur before training data, which biases the prediction. The model was also tested with single feature data and multiple-feature data and a better prediction was obtained with multi-feature data. In the implementation of SVM, the LSTM-trained temporal feature was put into the SVM classifier for fault classification, and a better and improved accuracy was obtained with a prediction accuracy of 97%. But the training time of SVM is much longer as it is much more computationally intensive.

On the other hand [8], proposes a method for predictive maintenance that considers the loss of revenue if an equipment failure occurs. The method uses multivariate time series data to train a deep learning model that can predict the probability of an equipment failure. The model is then used to schedule preventive maintenance tasks to minimize the risk of failure and the associated loss of revenue. The paper was evaluated using data from a Microsoft case study. The data consisted of 24 hours of multivariate time series data for a specific piece of equipment. The model was able to predict the probability of failure with a root RMSE of 0.126. This suggests that the model could be used to effectively schedule preventive maintenance tasks and minimize the risk of equipment failure.

In the field of modern networks, which bring unparalleled convenience and efficiency to various aspects of life and work, the looming potential for losses due to network failures necessitates proactive fault prediction [9] This predictive capability not only readies personnel for predictive fault repairs but also minimizes repair durations and curtails associated losses. To address this critical need, this paper proposes a novel network log-based hybrid prediction model for wireless network faults, integrating CNN and LSTM networks. The model prepro- cesses network logs, extracting features via CNN, and inputting them to LSTM for prediction. Through comparative experimentation against CNN and Random Forest approaches, the study demonstrates the superior predictive performance of the CNN-LSTM model, offering promise in enhancing fault prediction accuracy for wireless networks.

The main objective of this paper is is to develop a deep learning-based power failure prediction model for BTS sites using CNNs, LSTMs, and hybrid CNN-LSTM algorithms. The model will be evaluated to ensure reliable power systems for BTS sites

II. Power System in Base Transceiver Stations

BTS power systems typically consist of multiple components such as mains (utility) power, diesel generators, backup battery systems, Rectifiers, different distribution boards like AC and DC distribution boards, circuit breakers, and fuses [2] Most BTS have utility power sources as their initial power systems. Diesel generators and batteries are used as backup power sources. In the event of a utility power outage, a controller activates the backup battery supply and initiates the generator. An automatic transfer switch guides the generator's power to a controller, which subsequently transitions from utiliz- ing backup battery power to employing emergency generator power. When the standard utility power is restored.

A. *Mains*

Mains power, sourced from the utility grid, serves as the primary and cost-effective electricity supply for BTS. Operating at voltages of 220 VAC phase to neutral or 380VAC phase to phase, this three-phase power supply is critical for uninterrupted BTS functionality. However, mains power is susceptible to various failure factors, including electrical equipment breakdowns like damaged service transformers, distribution lines, or substations. Additionally, natural disasters such as storms, floods, and earthquakes can lead to mains outages. Weather-related issues, such as lightning strikes and strong winds, contribute to the vulnerability of this primary power source. Overloads due to excessive power demands and installation or maintenance issues like faulty equipment or wiring deficiencies can strain the power grid, resulting in disruptions.

B. *Diesel Generators*

Generators, which convert mechanical energy into electrical energy, serve as essential standby power sources during mains failures for communication devices. Despite their crucial role, generators are typically more costly than primary power sources and are activated solely when commercial power fails. They sustain the site with electricity until utility grid power is restored, their fuel depletes, or they encounter operational issues. While capable of supplying 3-phase AC power like the utility grid, generators are designed for shorter backup durations. However, they are prone to various failures, such as battery issues arising from loose connections, sulfation buildup, or dead batteries. Additionally, low coolant levels can lead to engine overheating and mechanical breakdowns, while low fuel levels and insufficient engine oil levels are also common reasons for generator startup failures [10].

C. Rectifier System

The rectifier system plays a pivotal role in telecommunications infrastructure, converting alternating current (AC) power from the utility grid or backup generators into essential direct current (DC) power for telecom equipment, as they primarily operate on DC. Additionally, rectifiers are responsible for charging backup battery systems, which serve as a crucial lifeline in the event of power disruptions. At the heart of the rectifier system are modules that facilitate the conversion of AC voltage to DC voltage and manage battery charging. Typically, rectifier modules are organized in an N+1 configuration, guaranteeing a minimum of two rectifiers at each site and implementing a load-sharing approach for their operation. The rectifier system is interconnected with load equipment through Load Low Voltage Disconnect (LLVD) and Battery Low Voltage Disconnect (BLVD) contactors, with critical and non-critical loads at BTS sites allocated to different priorities, with less critical loads linked to LLVD contactors and higher-priority equipment, such as transmission and microwave are connected to BLVD contactors as depicted in the below diagram.

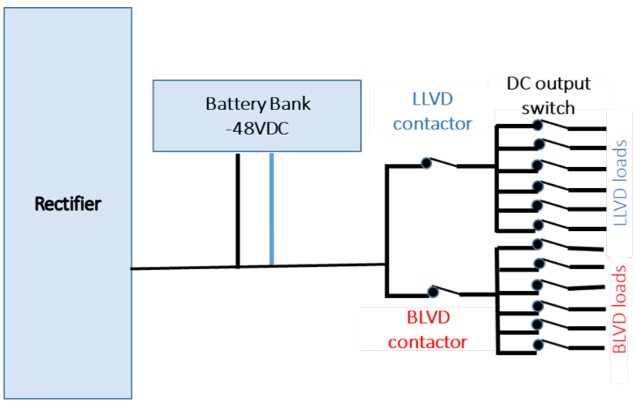

Fig. 1. Rectifier System interconnection.

Normally Rectifier system voltage is -48Vdc and it reach to its maximum up to 54.5V dc. But in case of power cutoff, this voltage reads the voltage of the battery strings which gradually decreases as the batteries deplete. To safeguard against deep discharge and maintain site functionality, predefined thresholds are set at 46.5V for LLVD and 44.5V for BLVD. When battery voltage reaches 46.5V, LLVD contactor opens, disconnecting less critical loads. If AC power isn't restored, battery discharge continues, and when it hits the BLVD threshold at 44.5V, the BLVD contactor disconnects all critical loads, resulting in a complete service interruption at the site.

D. Backup Battery System

The backup battery system in telecom power setups serves as a critical energy reservoir, storing electrical power for on-demand supply to the load in case of AC power interruptions. Telecom equipment typically operates on -48V DC, necessitating the configuration of battery strings to meet this voltage requirement. For instance, if each battery provides 2V DC, a string is formed by connecting 24 batteries in series to yield the required 48V DC output. During normal AC power operation, the rectifier efficiently delivers the necessary voltage to power the load while simultaneously float charging the batteries to counteract self-discharge losses within the cells. However, in the event of a rectifier failure, whether due to AC power disruptions or module malfunctions, the battery string seamlessly takes over, supplying DC power to the load until its discharge rate reaches a predefined threshold level. The duration of this discharge phase varies, spanning from a few minutes to up to 8 hours, contingent upon factors like the rectifier's unavailability and the age of the battery.

E. Alarm Management System

all components on BTS site obtain power from a common source. The site power and environment are monitored regularly through connected sensors to each devices. these sensors include: voltage, current, power, battery status, temperature and humidity. Hence, the sensors reads real time values and send to the central server for monitoring the site.

III. Power System Failure in BTS Site

Power failures in telecommunication systems are a distinct class of failures, often stemming from various factors. These failures can result from the loss of utility or grid power, malfunctioning backup power systems, or issues with power-related components like rectifiers, circuit breakers, and fuses, which are interlinked between power sources and communication equipment. These components are equipped with sensors that monitor and detect their measurements, with the sensor data feeding into a monitoring system. The monitoring system continuously assesses the system's status based on sensor readings, generating alarms if deviations from expected values are detected, thereby signaling abnormalities at the site. A power system is considered failed when there's insufficient power to operate the load or when power is entirely unavailable, leading to interruptions in both voice and data communication services. Actually, failures within the BTS power system can be identified through monitoring the busbar voltage, which provides more comprehensive information compared to DC load current and power. Under normal operating conditions, the busbar DC voltage should remain above 48.2 Volts. If it drops below this threshold, it's considered busbar under voltage (BBUV), serving as an early warning sign for LLVD issues. The duration until a full LLVD failure occurs depends on the battery's load-carrying capacity. When the voltage level falls below 46.5VDC, LLVD loads are disconnected, and an LLVD alarm is triggered. Furthermore, if the voltage level dips below 44.5VDC, BLVD loads are disconnected, resulting in the interruption of all BTS site services except communication with the monitoring system. This BLVD voltage-level disconnection is a protective measure to prevent deep discharging that could potentially harm the battery components irreversibly. These voltage levels are typically set as default values by operators but can be adjusted based on specific site conditions and operational requirements.

IV. Machine Learning for BTS Power Failure Prediction

In the world of modern telecommunications, BTS sites are the backbone of mobile networks, ensuring uninterrupted communication for millions of users. However, power failures can disrupt these critical operations, impacting network reliability and user experience. Machine learning, with its ability to discern hidden patterns from historical data [11], has emerged as a valuable tool for predicting such failures. By analyzing factors like weather conditions, voltage fluctuations, and traffic loads, machine learning models can make accurate predictions about the likelihood of power failures in BTS sites. This predictive capability empowers network operators to proactively address potential issues, reducing downtime, and ultimately enhancing the overall efficiency and reliability of their systems.

A. Long Short-Term Memory

LSTM, represents a specialized variant of recurrent neural networks (RNNs) designed to tackle the challenges associated with managing long-range dependencies and mitigating the vanishing gradient issue in sequential data analysis. Traditional RNNs struggle with handling long-term dependencies due to gradient-related problems, making them less effective in capturing meaningful patterns in sequences with extended time lags. LSTMs were introduced as a solution to these limitations by introducing memory cells and gating mechanisms that enable the network to acquire, retain, and utilize information over prolonged sequences, making them particularly well-suited for tasks involving extended temporal dependencies in data [12]. LSTM model contains an input gate, output gate and forget gate. The forget gate determines whether information from the previous timestamp should be retained or forgotten based on its relevance. The input gate assesses the significance of new input data. Meanwhile, the output gate decides which updated information from the current timestamp should be passed on to the next timestamp.

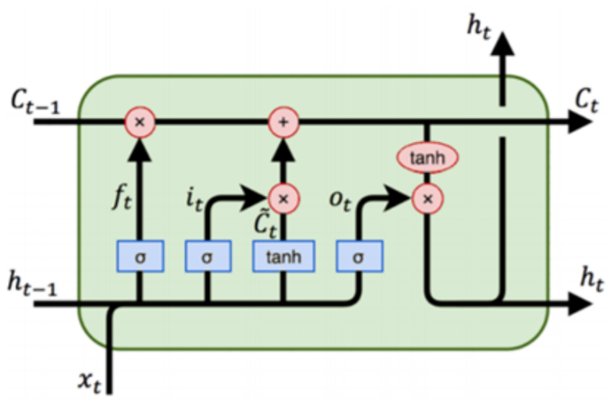

Fig. 2. LSTM Architecture.

B. Convolutional Neural Network

CNN is a specialized type of artificial neural network, which was initially designed to process and analyze visual data, particularly images, its applicability has expanded to data organized in matrix-like forms. Notably, time-series and textual data can be represented as 1D vectors. It utilizes convolution operation that entails performing linear transformations, often via matrix multiplication.

The distinctive feature of CNNs lies in their ability to automatically learn and extract hierarchical features from data [13]. This is achieved through a series of layers, including convolutional layers, pooling layers, and fully connected layers. Convolutional Layers detect local patterns whereas pooling layers reduce computational complexity and fully connected layers make predictions.

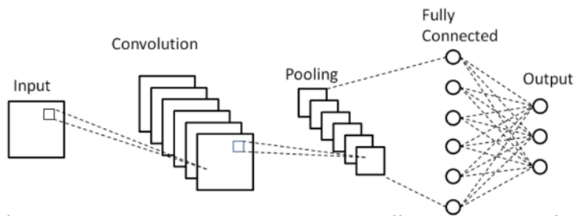

Fig. 3. CNN Architecture.

V. Data preparation

In this section, the methods followed to prepare the dataset is discussed.

A. Data Selection

For this study, the data is collected from power and environment monitoring, Net-eco system for five months with five minutes sampling interval. The data set contains fourteen features as indicated in the following table.

Input Parameters	
battery voltage	total DC load power
battery current	total DC output current
battery remaining capacity	**DC output/ busbar voltage**
battery temperature	room Temprature
battery charge-discharge rate	indoor Humidity
DC output current	phase voltages (L1, L2, L3)

TABLE I
Data features

B. Data Cleaning

the data cleaning process includes treating outliers and noisy data and handling missing values. addressing missing values is important due to their potential to introduce biases that subsequently distort the model's predictions, and reduce model performance finally resulting to inaccurate decision. Hence it is important to handle missing values appropriately.

C. Target Variable Selection

In building a successful supervised learning model, selecting the appropriate target variable which is known as a dependent variable or outcome variable is a critical step. The target variable is a variable that our model is going to predict based on the input or independent features [14]. The choice of the target variable can influence the selection of relevant input features. Features that have a strong relationship with the target variable are likely to be more informative for the model.

Under normal conditions, the busbar DC voltage should stay above 48 Volts. When it drops below this threshold, it's known as busbar or DC under voltage (BBUV), serving as a preliminary warning of LLVD issues. The duration until this issue becomes a failure depends on the battery's capacity. If the voltage falls below 46.5VDC, LLVD loads are disconnected, triggering an LLVD alarm. Furthermore, when the voltage dips below 44.5VDC, BLVD loads are disconnected, resulting in the interruption of all BTS site services except for communication with the monitoring system. This step safeguards the battery banks from deep discharging, which could otherwise permanently damage the battery system.

D. Hyper Parameter Tuning

In machine learning, hyperparameters are user-defined settings crucial for model training. They influence learning, prediction, and overall performance. Hyperparameter tuning, a vital step, optimizes these settings for improved model convergence and generalization to new data, enhancing overall effectiveness [15]. The development of the CNN model requires rigorous tuning, of hyperparameters, including kernel size, filter size, hidden layers, optimizer, activation functions, and epochs. Table 2 provides a comprehensive overview of the hyperparameters employed during the construction of the CNN model. The selection of these precise hyperparameter combinations represents a pivotal aspect in the pursuit of constructing a high-accuracy model.

On the other hand, the LSTM model takes critical considerations encompassing the determination of essential hyperparameters, including the optimal number of hidden layers, the appropriate count of LSTM cells within each layer, learning rate and activation function. The hyper parameters are outlined the table 2.

similarly the hybrid model development also considers tuning the combination of the two hyper parameters, the CNN and LSTM. So, it demands to consider appropriate kernel size, filter size, hidden layer, number of neurons, learning rate and number of epochs as indicated in table 3 below.

VI. Result and Discussion

During the model build up 80% of the data, that is four month data (March 1-June 30) is used to train the model. From the next month (July), ten days of data is used for validation in order to predict the next ten days of power failure. The results of the three model is presented in this section

A. CNN-Based Prediction Model

The CNN model is developed using the following hyperparameters.

CNN Model Hyperparameters	
Hyperparameters	Values
Number of Filters	512
Kernel Size	5
Pool Size	2
Learning Rate	0.001
Number of Epochs	200
Activation Function	ReLu
Optimizer	Adam

TABLE II
CNN Model Hyperparameters.

When the predicted failure is compared with actual data

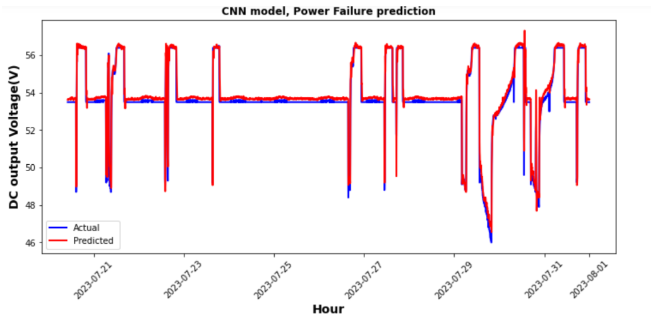

Fig. 4. Actual verses predicted Plot for CNN Model.

B. LSTM-Based Prediction Model

LSTM Model Hyperparameters	
Hyperparameters	Values
Number of Neurons	64
Hidden layer	3
Learning Rate	0.001
Number of Epochs	500
Activation Function	ReLu
Optimizer	Adam

TABLE III
LSTM Model Hyperparameters.

C. Hybrid CNN-LSTM Prediction Model

A power failure prediction model is developed using three variants of deep learning algorithms, CNN, LSTM and Hybrid CNN-LSTM model. The performance of the model is compared using three distinct evaluation metrics such as mean squared error (MSE), root mean squared error (RMSE) and mean absolute percentage error (MAPE). as indicated in Table V below.

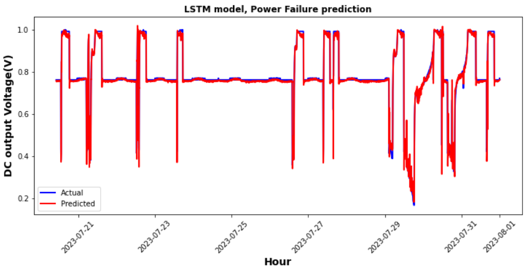

Fig. 5. Actual verses predicted Plot for LSTM Model.

CNN-LSTM Model Hyperparameters	
Hyperparameters	Values
Number of Filters	256
Kernel Size	5
Pool Size	2
Number of Neurons	64
Hidden layer	3
Learning Rate	0.001
Number of Epochs	500
Activation Function	ReLu
Optimizer	Adam

TABLE IV
CNN-LSTM MODEL HYPERPARAMETERS.

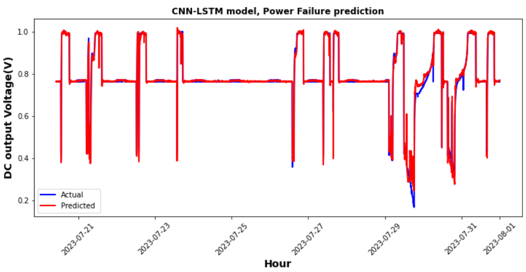

Fig. 6. Actual verses predicted Plot for CNN-LSTM Hypbrid model.

VII. CONCLUSION

This study delved into Machine Learning for Power Failure Prediction in BTS, focusing on improving communication service reliability by tackling power failures. It employed a multivariate approach to develop predictive models, including CNN, LSTM, and a hybrid CNN-LSTM. By collecting five months of multivariate data and leveraging precise feature engineering, these models demonstrated strong performance.

Among the three, CNN showed the fastest training, taking just an hour. LSTM, on the other hand, took three and a half hours but offered superior prediction accuracy. The hybrid CNN-LSTM model combined CNN's feature extraction capabilities with LSTM's time-series dependency handling, achieving a two-hour training time while maintaining LSTM-like accuracy. This synthesis provides a powerful predictive framework for anticipating power failures and enabling proactive measures.

The study highlights the potential of machine learning in telecom, enhancing service reliability, reducing downtime, and cutting maintenance costs, showcasing a promising synergy

Model Evaluation		
Model	RMSE	MAPE
CNN	0.472	2.643
LSTM	0.037	1.194
CNN-LSTM	0.038	2.528

TABLE V
MODEL PERFORMANCE EVALUATION.

between technology and power failure prediction in the BTS sector.

Future work may explore integrating real-time data streams from various BTS sources, such as network traffic and O&M records, to expand datasets for more comprehensive power failure predictions.

Furthermore, assessing the adaptability of these models across diverse BTS sites and locations could reveal insights into their transferability and robustness.

Lastly, conducting an economic analysis of the predictive system's cost-effectiveness in reducing service disruptions and downtime would offer valuable industry insights.

REFERENCES

[1] Mishra, A. R. Advanced cellular network planning and optimisation: 2G/2.5 G/3G... evolution to 4G (John Wiley & Sons, 2007)

[2] Chayanam, K. Analysis of Telecommunications Outages due to power loss PhD thesis (Ohio University, 2005)

[3] Wahid, A., Breslin, J. G. & Intizar, M. A. Prediction of Machine Failure in Industry 4.0: A Hybrid CNN-LSTM Framework. Applied Sciences 12, 4221 (2022).

[4] Fan, X., Wang, X., Zhang, X. & Yu, P. A. X. B. Machine learning based water pipe failure prediction: the effects of engineering, geology, climate and socio-economic factors. Reliability Engineering & System Safety 219, 108185 (2022).

[5] Celikmih, K., Inan, O. & Uguz, H. Failure prediction of aircraft equipment using machine learning with a hybrid data preparation method. Scientific Programming 2020 (2020).

[6] Betti, A. et al. Fault prediction and early-detection in large pv power plants based on self-organizing maps. Sensors 21, 1687 (2021).

[7] Zhang, S., Wang, Y., Liu, M. & Bao, Z. Data-based line trip fault prediction in power systems using LSTM networks and SVM. Ieee Access 6, 7675–7686 (2017).

[8] Zheng, L. et al. A fault prediction of equipment based on CNN-LSTM network in 2019 IEEE international conference on energy internet (ICEI) (2019), 537–541.

[9] Tan, Z. & Pan, P. Network fault prediction based on CNN-LSTM hybrid neural network in 2019 International Conference on Communications, Information System and Computer Engineering (CISCE) (2019), 486–490.

[10] Rao, S. S. & Amer, G. Power supply architecture for telecom application: A review in 2010 international conference on communication control and computing technologies (2010), 262–264.

[11] Wang, R., Vanga, V. L., Zaiken, Z. B. & Bennett, J. Analysis of the electric power outage data and prediction of electric power outage for major metropolitan areas in Texas using Machine Learning and Time Series Methods. SMU Data Science Review 6, 5 (2022).

[12] Kumari, P. & Toshniwal, D. Deep learning models for solar irradiance forecasting: A comprehensive review. Journal of Cleaner Production 318, 128566 (2021).

[13] Alzubaidi, L. et al. Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. Journal of big Data 8, 1–74 (2021).

[14] Sparks, R., Zucchini, W. & Coutsourides, D. On variable selection in multivariate regression. Communications in Statistics-Theory and Methods 14, 1569–1587 (1985).

[15] Bergstra, J., Yamins, D., Cox, D. D., et al. Hyperopt: A python library for optimizing the hyperparameters of machine learning algorithms in Proceedings of the 12th Python in science conference 13 (2013), 20.