



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

**Comparative Analysis of Machine Learning Models for Prediction
of BTS Power System Failure**

BY:

BIRUKTAYT FENTABIL

ADVISOR:

DR. -ING. DEREJE HAILEMARIAM

A Thesis Submitted to the School of Electrical and Computer Engineering Graduate Studies
in Partial Fulfillment of the Requirements for the Degree of Master of Science in
Telecom Network Engineering

JUNE 2025

ADDIS ABABA, ETHIOPIA

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

**Comparative Analysis of Machine Learning Models for Prediction
of BTS Power System Failure**

by:

Biruktayt Fentabil

<u>Approval by:</u>	<u>Name</u>	<u>Signature</u>	<u>Date</u>
Advisor:	<u>Dr.-Ing. Dereje Hailemariam</u>	_____	_____
First Examiner:	<u>Dr. Murad Ridwan</u>	_____	_____
Second Examiner:	<u>Dr. Eng Yihenew Wondie</u>	_____	_____

Declaration

I certify that this thesis is my original work, which I completed at Addis Ababa University under the direction and supervision of Dr.-Ing. Dereje Hailemariam. To the best of my knowledge, none of the content contained here has been published anywhere else, nor does it contain any work written by others that has not been properly acknowledged with citations. No other degree, diploma, or certification from this or any other institution has received this work in whole or in part.

Signature

Biruktayt Fentabil

Date of Submission: June/2025

Abstract

Loss of power system integrity in Base Transceiver Stations (BTS) can have significant impacts on communication networks, causing service outages and loss of revenue. This study was focused on developing and testing machine learning models to predict BTS power system failures so that maintenance can be performed before system failures occur. The models tested were Hidden Markov Models (HMM), Long Short-Term Memory (LSTM), and Random Forest (RF).

The data used for this research included 46,943 records from ethiotelecom's (ET) monitoring system. ET's provided information related to variables sourced from monitoring features that impact power system failure, including environmental, load, battery, or other BTS-based metric records. All these data were pre-processed. Before feature analysis, Z-score normalization was conducted to standardize the data. After this data preparation step, principal component analysis (PCA) was undertaken to perform feature analysis. In addition, K-means clustering was also applied to categorize the hidden states ('Normal,' 'Degraded,' and 'Failure') and group the observable sequences.

The HMM was trained using the Baum-Welch algorithm, and the Viterbi technique aided state prediction. To enhance performance, a range of hyperparameter approaches were applied to the RF and LSTM models. With a 97.72% F1-score, 98.04% accuracy, 98.08% precision, and 98.04% recall, compared to the other two models, the HMM performed better when load-related parameters were observable. With an accuracy of 97.81% and an F1-score of 96.74%, the LSTM model came in second place for identifying temporal connections in the data. Despite being robust, RF's performance metrics were marginally worse, with an F1-score of 93% and an accuracy of 95%. The study's conclusions show that the best model for forecasting BTS power system breakdowns is HMM. Due to its exceptional precision and dependability, ethiotelecom can enhance network performance by facilitating proactive maintenance, reducing downtime, and increasing user satisfaction.

Keywords: BTS, power system failure, HMM, LSTM, RF, prediction

Acknowledgment

First and foremost, I thank Almighty God for his infinite mercy and kindness, which enabled and strengthened me in every challenge and success of the journey. Secondly, I also thank His Mother, Saint Mariam, whose intercession and consistent presence have been my constant comfort and support throughout my life.

I am deeply indebted to my advisor, Dr. Ing. Dereje Hailemariam, whose guidance, patience, and support have shaped this study from its inception to its completion. His thoughtful comments, words of encouragement, and confidence in this study urged me to explore new horizons and excel.

I want to thank my evaluators, Dr. Murad Ridwan and Dr. Eng Yihenew Wondie, for being helpful evaluators, offering constructive feedback, useful suggestions, and insights that greatly enhanced this research.

I truly appreciate ethiotelecom and all staff members for their generous support and the opportunity to work collaboratively and for enabling this research to have the practical and data basis needed.

I want to offer my appreciation to my amazing family. Your encouragement and compassion kept me going! My husband's steadfast understanding and support are something I wish to thank him for. He has always been dependable and stable, much like my kids, who provide me with joy and fulfillment in my life.

Lastly, I want to express my gratitude to everyone who helped with this thesis. This thesis would not have been possible without their support and encouragement.

Contents

Declaration	iii
Abstract	i
Acknowledgment	ii
Table of Contents	v
List of Table	vi
List of Figures	vi
Abbreviation	viii
1 Background and Motivation	1
1.1 Introduction	1
1.2 Statement of the Problem	3
1.3 Literature Review	4
1.4 Objectives	7
1.4.1 General Objective	7
1.4.2 Specific Objectives	7
1.5 Scopes and Limitations of the Research	8
1.6 Methodology	8
1.7 Contribution of the Research	9
2 Overview of Base Transceiver Station Power System	10
2.1 Base Transceiver Station Power System	10
2.2 Components of a BTS Power System	12
2.2.1 Main Source of Power	13
2.2.2 Backup Power Systems	14
2.2.3 Automatic Transfer Switches	15
2.2.4 Rectifier	16

2.2.5	Power Distribution Units (PDUs)	16
2.2.6	Protection Devices	17
2.3	AC vs. DC Distribution	19
2.4	Load Management	20
2.4.1	Battery Low Voltage Disconnect (BLVD)	20
2.4.2	Load Low Voltage Disconnect (LLVD)	20
2.5	BTS Power System Failure and Power Outage	21
2.6	Factors of BTS Power System Failure	23
2.7	Monitoring and Maintenance of BTS Power Systems	24
2.7.1	Alarm Systems for BTS Power Failure	25
2.7.2	Predictive maintenance using IoT and AI technologies	25
2.8	Challenges in BTS Power System	26
3	Machine Learning Algorithm	28
3.1	Hidden Markov Model	28
3.1.1	Markov Chains and the Basis of HMMs	28
3.1.2	Components of HMM	29
3.1.3	Problems in Hidden Markov Model and Their Solutions	32
3.1.4	Solutions for Decoding Problem, and Viterbi Algorithm	37
3.1.5	Solutions for Learning Problem and the Baum-Welch Algorithm	37
3.1.6	Solutions for Evaluation Problem and Back and Forward Algorithm	40
3.1.7	Significance of Hidden Markov Models in Predicting Base Transceiver Station Power System	41
3.2	Long Short-Term Memory Algorithm	41
3.3	Random Forest Model	44
3.3.1	Random Forest Classification and Regression	45
3.3.2	Random Forest Hyperparameters	46
3.4	Model Performance Evaluation Technique	47
4	Modeling BTS Power System Failure Using Machine Learning	49

4.1	System Model	49
4.2	Data Collection	49
4.3	Important Feature Selection	50
4.4	Defining the Hidden State	52
4.5	Defining the Observation Sequence	54
5	Results and Discussions of the Model	56
5.1	Model Training	56
5.1.1	The Probability of Transition Matrix	56
5.1.2	The Probability of Emission Matrix	58
5.1.3	The Probability of Initial Matrix	59
5.2	Model Evaluation and Prediction	60
5.2.1	Hidden Markov Model Evaluation and Prediction	60
5.2.2	Long Short Term Memory Model Evaluation and Prediction	65
5.2.3	Random Forest Model Evaluation and Prediction	68
5.2.4	Importance and Drawbacks of HMM, LSTM, and RF for BTS Power System Failure Prediction	69
6	Conclusion and Recommendations	72
6.1	Conclusions	72
6.2	Recommendations	73
	References	74

List of Tables

1	Transition Probability Matrix	58
2	Emission Probability Matrix	59
3	Initial Probability Matrix	60
4	Performance of the Model with Different Assumptions and Observations	63
6	LSTM Model performance	67

List of Figures

1.1	Overall Methodology of the BTS Power System Failure	9
2.2	Various Configurations of BTS Power System [28]	12
2.3	BTS power system structure [16]	13
3.4	HMM, Structure with hidden and observable states	28
3.5	HMM, Structure and component [7]	30
3.6	HMM, with three hidden and observable states	31
3.7	LSTM Architecture [29]	42
4.8	System Model	49
4.9	Feature Selection by Using on PCA	51
4.10	Importance Features	52
4.11	Hidden State Classification by using silhouette	53
4.12	Observation Sequence Using Silhouette method	55
4.13	Observation Symbols	55
5.14	Training Convergence of the Baum-Welch Algorithm	57
5.15	Confusion Matrix of Metric 1	64
5.16	Actual vs Prediction of Metric 1	65
5.17	LSTM Train Validation Loss	66
5.18	Confusion Matrix of LSTM	67
5.19	DC Output Voltage of Actual vs Prediction	68
5.20	Random Forest Model Performance	69

List of Abbreviations

2G	Second Generation
5G	Fifth Generation
AC	Alternate current
ATS	Automatic Transfer Switch
BBUV	bus bar under voltage
BLVD	Battery Low Voltage Disconnect
BSC	Base Station Controller
BTS	Base Transceiver Station
CNNs	Convolutional Neural Networks
DC	Direct Current
DG	Diesel Generator
DTs	Decision Tree
EP	Emission Probability
EPM	Emission Probability Matrix
EM	Expectation-Maximization
ET	ethiotelecom
GRU	Gated Recurrent Units
GSM	Global System for Mobile Communications
HMMs	Hidden Markov Model
HVAC	High Voltage Alternating Current
IP	Initial Probability
LLVD	Load Low Voltage Disconnect
LSTM	Long Short-Term Memory
LTE	Long Term Evolution
ML	Machine Learning
MSE	Mean Square Error
NetEco	Network and Environment Monitoring System
PDU	Power Distribution Unit
RF	Random Forest

RNN	Recurrent Neural Network
SPDs	Surge Protection Devices
SVM	Support Vector Machine
TP	Transition Probability
TPM	Transition Probability Matrix
UMTS	Universal Mobile Telecommunications System
UPS	Uninterruptible Power Supply

1 Background and Motivation

1.1 Introduction

In today's interconnected world, reliable communication networks are essential for maintaining seamless connectivity and enabling efficient data transmission. The evolution of global network infrastructures is closely linked to advancements in BTS and their power systems, which play a pivotal role in modern mobile telecommunications. As generations of mobile networks have evolved from 2G to 5G, the BTS systems have also evolved with more complex features, as mobile systems now need to provide support for the ever-increasing amount of internet traffic demands from different sectors such as banking, healthcare, and manufacturing. BTS are the main elements of any wireless telecommunication network, whereby they are responsible for transmitting and receiving signals over the air from mobile portions of the network while transmitting signals to wired portions of the network [1]. The operational efficiency of BTS depends heavily on its power supply systems, which must ensure continuous functionality to provide users with uninterrupted access to services. Such power systems are very important to prove that communication services are reliable and accessible, which in turn has a direct impact on the continuous operation of the stations.

For the BTS's uninterrupted power, the worldwide telecom industry has gone a long way. Modern BTS are composed of batteries, rectifiers, backup generators, power management systems, and many other components. This invention's infrastructure is designed to withstand even power outages or unstable electrical grids without an issue. These significant developments notwithstanding, it is a difficult task to provide a reliable power supply for these BTS [2]. Moreover, not only can hardware failures, grid outages, or severe weather rouse such situations, but the communication disruption caused by such failures may be very substantial. On top of short-term service drops, the consequences harm revenue sources, user satisfaction, and overall user happiness.

Telecom operators are placing a higher emphasis on predictive problem management and

proactive maintenance to reduce these risks. Operators of power systems can significantly reduce trouble by anticipating failures in BTS units. They might even limit the uptime losses significantly if they had preventive measures in place. Network performance is improved, and user experience is also better during service restoration, while the restoration process turnaround time is shortened with both proactive and rapid response measures. Primary and backup power systems are usually included in a BTS's power design, which applies both AC and DC components for redundancy [3]. Yet the difficulties of these systems, along with the changing conditions of the grid, make them work harder and create a greater need for distinct techniques.

Traditionally, power outage management was reactive, where operators only took action after an incident occurred. Nevertheless, the strategy is being changed by ML algorithms, which are very good at recognizing the very small trends in past data that would then be used for insights that can be unknown to human analysts. Service interruptions can be minimized by operators if they use these models to forecast service disruptions and thus perform preventive interpolation. Predictive analytics is the instrumentation that ML is using to renew industries such as healthcare, manufacturing, and energy grids [4]. This trend is also becoming more popular in telecom infrastructure. The employment of ML in failure prediction is quite advantageous. It allows operators to change the old parts before they break down, time their rest for times when the load is low, and even make sure the backup system is fully operational. HMMs have indeed been secured to be good examples in a power system that records data over time to predict failure.

This paper studies the applicability of HMMs, LSTM, and RF models for predicting power outages in the BTS network. The study compares various methods in an effort to give ET practical ways to improve power grid dependability. By doing this, it hopes to assist Ethiopia's telecom industry in satisfying the needs of a society that prioritizes digitalization, guaranteeing steady service availability, and promoting customer satisfaction in a time of speedy technical advancement.

1.2 Statement of the Problem

Base Transceiver Stations (BTS) are the foundation of contemporary cellular networks, providing extensive coverage and satisfying the ever-increasing demand for connectivity. However, their dependability depends on reliable power sources, and network instability, service outages, and financial losses might result from weakness in their power infrastructure. Operators can no longer afford to rely on reactive solutions; proactive failure prediction is increasingly essential to preserving user confidence and ensuring uninterrupted service.

Manual rules or static thresholds are frequently used in traditional failure prediction techniques. Despite being easy to put into practice, these methods are not flexible enough to handle the dynamic, linked character of contemporary power networks. Voltage fluctuations, temperature variations, and progressive equipment degradation are examples of less evident interactions between factors that they overlook, which compromise.

ML represents a significant shift in technology. Unlike traditional rule-based systems, it excels at analyzing complex and dynamic relationships in power systems. This ability allows it to generate highly accurate failure forecasts. By examining historical power system data, the model can identify and predict potential disruptions. This capability empowers operators to act before issues arise, improving system reliability. In this study, we developed a predictive model using HMM, LSTM, and RF algorithms. To evaluate their performance, we tested them on real-world power system data from ET. Notably, the HMM-based framework takes a probabilistic approach to predicting failures. It helps operators focus on critical maintenance tasks and make informed decisions about deploying backup power.

Operators would be better equipped to proactively manage risks, reduce downtime, and improve network resilience with a dependable and flexible prediction model for BTS power failure. This directly translates to enhanced consumer satisfaction, decreased operational

costs, and improved service availability. In an era when reliable communication is a must, the main aim of this study is to improve Ethiopia's telecommunications network so that there is no interruption in connectivity.

1.3 Literature Review

Many researchers have comprehensively surveyed the literature to identify the best predictors of power system failure, which is their main focus in various fields, such as electrical networks, BTSs, energy consumption, and others.

In [5], a new approach was introduced for forecasting power transformer faults based on HMM along with dissolved gas analysis (DGA). The authors utilized the Gaussian mixture model to locate stable patterns of transformer health and improve the accuracy of fault detection. It emphasizes the importance of cross-validation in confirming the model's reliability and reports high performance of the model in fault diagnosis and transformer condition assessment. Such a method illustrates the potential of combining ML with DGA data for the development of energy systems and provides a feasible instrument for the failure-free operation of the power grid.

In [6], the role of ML in the prediction of power failures in telecommunication networks is presented, and it is shown that proactive maintenance can not only reduce outages but also financial losses. Their work establishes the comparative performance of models such as convolutional neural networks (CNNs), LSTM networks, and hybrid CNNs with LSTM techniques, hence demonstrating their capacity for network reliability improvement and downtime reduction. The LSTM model was the best, of course; it reached the mean square error of 0.001 and the mean absolute percentage error of 1.194.

In [7], the authors evaluated the performance of HMMs for forecasting energy demand in residential houses and compared HMMs with the performance of artificial neural networks (ANN's), SVMs, and classification and regression trees. According to the authors, HMMs are particularly suitable for capturing the seasonal effect in energy consumption

when real data from Seoul, South Korea, is used. They further explain that the HMMs are very much in line with the model's capabilities and, if the analysis is conducted in this way, HMMs are the best fit for analyzing time series data.

In [8], a researcher utilized LSTM and Gated Recurrent Unit (GRU) models to forecast power system failures in BTS, thereby enhancing network reliability and service quality. To explain the shift in the communication industry from reactive repairs to proactive maintenance, the authors used a collection of 11 features, such as DC/AC output voltages, temperature, and bus bar voltage, from five BTS sites over 20 weeks. The study provides an example of early failure detection by comparing the performances of LSTM and GRU to demonstrate their effectiveness in reducing service interruptions, protecting revenue sources, and ensuring uninterrupted connections for users. In addition, the paper [4] discusses a common problem of power outages in BTS that not only cause disruptions in telecom services but also result in customers who are not satisfied. Their predictive maintenance model, developed by the XGBoost algorithm within the Team Data Science Process (TDSP) framework, got an accuracy of 97% in detecting power system risks for Econet Wireless in Zimbabwe. The study goes beyond technical issues and highlights the financial benefits of predictive measures, like less downtime and better service reliability; thus, the role of the advanced ML techniques in the sustainable running of the telecom industry is reaffirmed.

In [10], HMMs were applied in a pulp industry case study to forecast the condition of drying presses. The HMM is trained on sensor data sampled every minute for almost four years and finds three main operating modes: "Normal Function", "Warning Signs," and "Imminent Failure." The latter can be detected with this system quite promptly and allows the team to coordinate the repair in such a way that the downtime is minimized and the lines continue to run with no interruptions. Moreover, [9] improved on the initial studies by joining HMMs with GRUs to predict pulp press failures up to seven days ahead. The technique is an m-step sequence: first, sensor data is cleaned; then, key patterns are

identified; after that, data complexity reduction is done; subsequently, k-means clustering is applied to trend grouping; finally, HMM is trained to classify equipment health. The GRU then decides to use these facts for the extrapolation of the future. The authors of this fusing model point out that early failure detection and change of the algorithm are possible as long as they can adapt to different types of machinery and different sensors without being provided with data.

In [11], the authors discuss the issue of device failures in the industrial anode production process and suggest ML models trained with nine months of continuous sensor data as a solution. Their goal was to predict the forecast 5–10 minutes before the problem occurred and move from routine maintenance testing to an aggressive data control strategy. Decision Trees (DT) and RF rose, and risk was recognized in 60% of cases within the critical 5-minute window. Although this study highlights how even basic ML models can reduce unplanned downtime, the authors combine them with recurrent neural networks to propose sharper predictions.

In [12], the author introduces a cutting-edge approach to detecting and classifying faults in power transmission lines. Their hybrid model, combining RF, LSTM, and a tuned K-Nearest Neighbors (KNN) algorithm, achieves near-perfect accuracy—99.96% for multi-label fault classification and 99.85% for binary classification. To address data imbalances, the team uses techniques like SMOTE (Synthetic Minority Over-sampling Technique) and optimizes performance with Particle Swarm Optimization. Tested on a comprehensive Kaggle dataset, the model outperforms traditional methods, offering utilities a powerful tool to enhance grid reliability and reduce outage risks.

In [13], the paper evaluates deep learning models for multistep time series forecasting, comparing standard RNNs to advanced LSTM variants like bidirectional and encoder-decoder architectures. While basic RNNs struggle with long-term patterns, LSTMs excel at capturing complex temporal relationships, delivering superior accuracy on benchmark

datasets. However, the study notes that even advanced models lose precision when predicting further into the future.

In [14], demonstrated that LSTM neural networks excel at predicting voltage stability in power systems. In tests using the IEEE 33-bus system, LSTMs achieved near-perfect accuracy of 99.5%, outperforming alternatives like Support Vector Machines (SVMs), Naive Bayes, and Convolutional Neural Networks (CNNs) in both speed and precision. Crucially, LSTMs succeed with sequential data, such as real-time grid fluctuations, making them ideal for dynamic stability assessments. This capability allows power operators to monitor grid health instantaneously, identify risks proactively, and make faster, data-driven decisions. The findings underscore LSTMs' potential to transform voltage stability analysis, providing grid operators with precise, actionable insights to safeguard energy infrastructure.

1.4 Objectives

1.4.1 General Objective

The main aim of this research is to perform a comparative analysis of ML models for predicting power system failure in BTS, with the emphasis on the most efficient model identification.

1.4.2 Specific Objectives

To achieve the general objective, we have to formulate these specific objectives:

- Conduct an overarching study of the power sources utilized in BTS sites and the interconnection of power systems within these sites.
- Explore various power system alarms that occur in BTS sites and analyze their underlying causes.
- Find out the most important features that can help in the prediction of the BTS power system failure.

- Evaluate the potentiality of a designed model.
- Clarify the data resulting from the assessment and then provide insightful conclusions.

1.5 Scopes and Limitations of the Research

This research focuses on applying ML models to predict BTS power system failures in the context of the ET around Addis Ababa. The study uses a dataset from ET and evaluates the performance of RF, LSTM, and HMM models. It also compares their performance and identifies the best model for predicting BTS power system failure.

1.6 Methodology

The research outlined here starts with its greatest goal by reviewing various related papers: BTS power system failure prediction using historical data from the ET's network and environment monitoring system (NetEco). The dataset contains several critical features, like the stability of the mains power, performance of the DC distribution groups, environmental factors, and battery health indicators. At first, the raw data is cleaned to eliminate errors and to avoid redundancy of information, then scaled to normalize features like voltage and temperature, and finally transformed to comply with the requirements of ML feature extraction (PCA), where PCA is used to reduce the dimensionality of the data while retaining essential information. K-means clustering can divide the past data into groups that match these states, thus providing a structure to train an HMM. The HMM can recognize patterns by the Baum-Welch algorithm to find initial, transition, and emission probabilities. On the other hand, the Viterbi algorithm is like a tool that helps to draw a map of the most likely path that the hidden states will take, for instance, if the process changes from "degraded" to "failure." To better improve the performance, RF and LSTM models are also trained, making use of their advantages in dealing with nonlinear relations and changes of the past. Figure 1.1, shows the overall methodology to compare three machine learning models for predicting BTS power system failures.

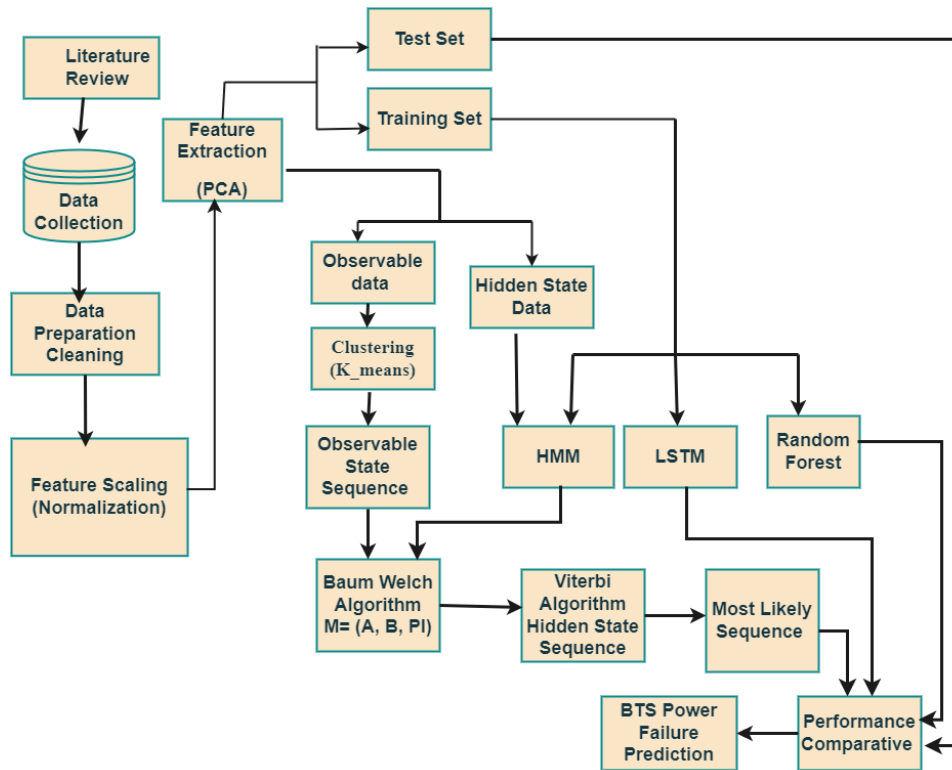


Figure 1.1: Overall Methodology of the BTS Power System Failure

1.7 Contribution of the Research

This study is designed to bring significant improvement to ET's network management and user experience. Through forecasting the power system failure in BTS before the occurrence of the issue, the suggested model will enable ET to:

- Continuity of service thus ensures that consumers can access services without any interruption because the power issue is solved proactively.
- Avoiding failures allows emergency repairs of lower cost and replacement of fewer parts.
- The reliable service with less disturbance creates trust and loyalty among users.
- Future investigation and advancement in power system administration are the research potential.

2 Overview of Base Transceiver Station Power System

2.1 Base Transceiver Station Power System

A BTS is an essential part of telecommunications systems, especially in cellular networks, where it is the main channel of communication between user equipment (UE), like mobile phones, tablets, or other wireless devices, and the telecom network infrastructure [1, 3]. It carries out the task of wireless communication by sending and receiving radio signals within the given space, here known as a cell. A BTS is most often composed of antennas, transceivers, and control units, which complement each other in carrying out the basic processes of mobile communication. The main goal of a BTS is to provide uninterrupted connectivity for voice, data, and multimedia services by regulating and controlling the transmission of signals between mobile devices and the network.

A BTS is a major contributor to the performance of several key functions. It handles sending and receiving radio signals so that users and the network can communicate efficiently. Besides, it connects to the Base Station Controller (BSC), which is responsible for directing the traffic of the core network for the necessary operations. Also, the BTS is the main party in the handover operations, thus assuring that users get uninterrupted service when they are changing cells. The BTS is also a power control unit that adjusts the strength of the signal in order to have the best performance and minimum interference. Therefore, the BTS is a fundamental element of mobile networks that support such technologies as GSM, UMTS, LTE, and 5G. The BTS units have been improved a lot to be able to support faster data transfer, lower latency, and better spectral efficiency in the course of the development of mobile networks. In the case of 4G and 5G networks, BTS units may be known as eNodeBs or gNodeBs, which represent their new features that allow them to be used in such applications as IoT devices and smart cities.

Typically, BTS sites connected to the grid experience power interruption at a high level, necessitating a reliable backup power system. Providing power to the BTS sites that

is very reliable and consistent is very important if we want to make sure that the operation of these BTS sites is not interrupted in any way. On the other hand, any power disruption can result in service outages [15]. In order to come up with solutions for this problem, we suggest a complete system framework that integrates different configurations for BTS power systems. The main goal of this framework is to make sure that power delivery to BTS sites is still uninterrupted and continuous, thus eliminating power fluctuations in the grid and maintaining reliable telecommunications services. The focus of this method is not only on redundancy, but also on the flexibility of the power solution to satisfy the different needs of the BTS operation. The different BTS power system configurations are defined in [28]:

- BTS with mains and battery storage (Figure 2.2a)
- BTS with mains, diesel generator, and battery storage (Figure 2.2b)
- BTS with mains with a Solar PV system and Battery storage (Figure 2.2c)
- BTS with mains with Solar PV system, DG, and Battery storage (Figure 2.2d)

These configurations and performance requirements for BTSs vary widely due to different factors. This includes the following considerations:

- Is the site outdoors or indoors?
- Does the site have adequate space?
- Where is the site located?

The different configurations of the BTS power systems are, as shown in Figure 2.2, the key to ensuring reliable operation under changing conditions.

The uninterrupted work of a BTS is extremely dependent on a reliable power system, since power failures might cause immediate service interruptions. Most of the time, BTS sites prefer grid electricity as their primary power source. However, the use of various backup systems such as battery banks, DG, and UPS can be of help for the continuation of operations during power outages. In addition, renewable energy sources like solar panels,

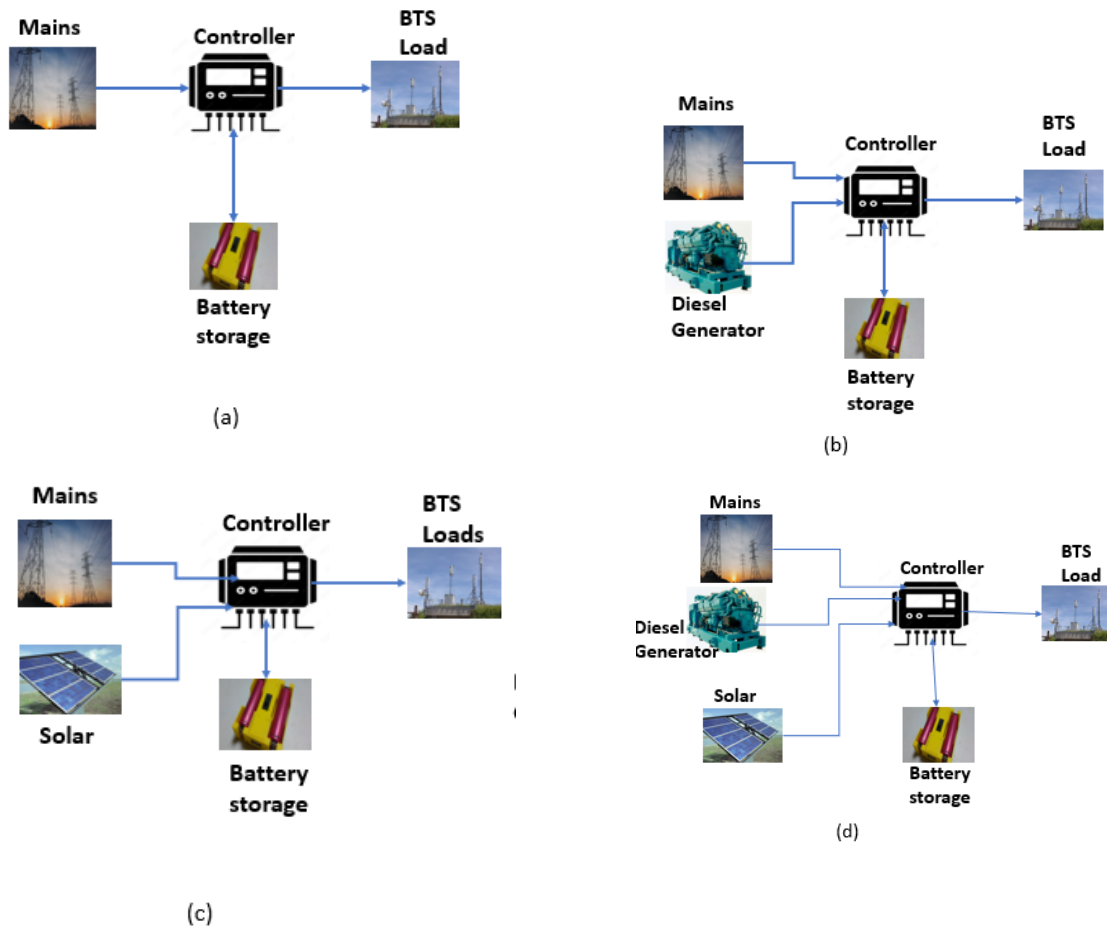


Figure 2.2: Various Configurations of BTS Power System [28]

wind turbines, and hybrid systems are the most progressive and favored techniques for supplying power in remote and off-grid areas.

Power failure at BTS sites can bring a disaster to network performance. Firstly, the most direct consequence is a complete loss of service in the affected area, followed by the discontinuity of voice, data, and multimedia communication. This risk can be critical, supposing an emergency when reliable communication is essential.

2.2 Components of a BTS Power System

The different components and their functions of the BTS power system are explained in [16].

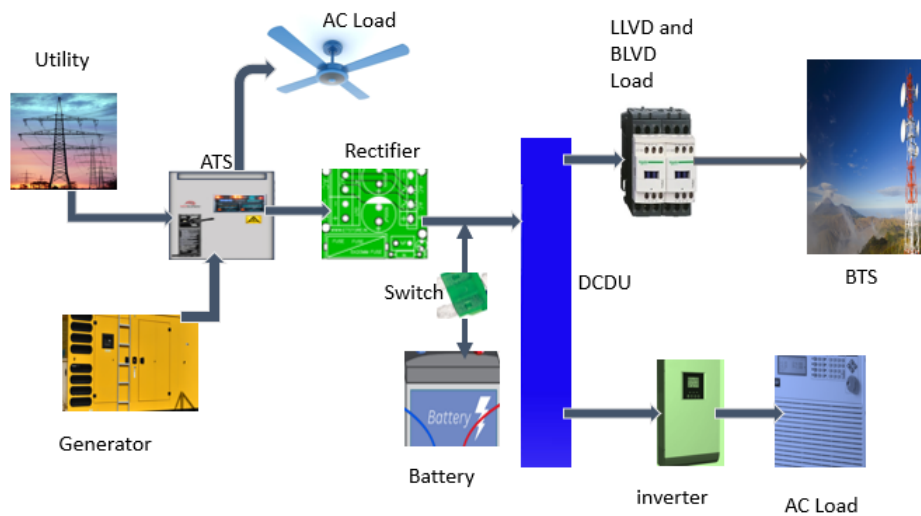


Figure 2.3: BTS power system structure [16]

2.2.1 Main Source of Power

A BTS depends mainly on grid power, which is supplied from the utility power, to ensure its operations run without interruption and maintain dependable communication services. The electrical grid delivers AC power at voltage levels of 220V or 380V, based on regional infrastructure needs and the power requirements of the BTS. Telecom industries use [16] -48VDC as a standard voltage because DC power provides stability and minimizes power fluctuations that could damage electronic devices. The stable power supply is essential for maintaining uninterrupted system performance and reducing power-related interruptions in telecommunication networks.

Power grid reliability can have multiple causes, including natural disasters like storms and earthquakes, and infrastructure breakdowns, including damaged power lines and equipment failures. Frequent power outages in certain regions can lead to significant disruptions of BTS operations, which result in network downtime and loss of connectivity for users.

2.2.2 Backup Power Systems

Telecommunication network stability and functionality remain preserved during power outages thanks to BTS's deployment of robust backup power systems [28]. BTS sites receive protection from power disruptions through backup power systems that consist of UPS units, battery banks, and diesel generators.

Diesel Generators (DG): Diesel Generators are the mainstay of the backup power infrastructure in the BTS site, making sure that the system operates perfectly during utility power failure. The power they provide can be continuous for a number of hours or even days, subject to the availability of fuel and the length of the power outage. Usually, they are set to start automatically once the power supply from the grid fails, and they continue to run until the battery has been used up or if there is a need for a longer period of power failure. They are indispensable for the continuation of BTS activities during extended outages, especially in places where the power from the grid is not reliable or is unavailable. They not only guarantee that the BTS equipment, which is used for transmission, cooling systems, and site monitoring devices, is functional but also reach the highest standards of efficiency.

Batteries: Serve as the first line of defense in a BTS during power outages. They provide instantaneous power when the grid supply fails, ensuring there is no interruption in the operation of critical telecommunication equipment such as transceivers, antennas, and signal processors. This is especially important because even brief power interruptions can disrupt network services, leading to dropped calls and data connectivity issues. Most BTS systems [16], [61] operate at a nominal -48 VDC, a telecom standard that balances efficiency and safety. To achieve this:

- **Series Connection:** For lead-acid batteries (typically 2V per cell), 24 cells are connected in series ($24 \times 2V = 48V$). For lithium-ion (e.g., 3.2V per cell for LiFePO₄), 15 cells in series yield approximately 48V ($15 \times 3.2V = 48V$).

- **Parallel Strings:** To increase capacity (ampere-hours, Ah), multiple 48V strings are connected in parallel. For example, two 100Ah strings in parallel provide a total of 200Ah at 48V.

The most commonly used batteries in BTS systems are [17]

- **Lead Acid Batteries (specifically Valve-Regulated Lead-Acid, or VRLA):** These are sealed, maintenance-free batteries that are widely used in BTS systems due to their cost-effectiveness and reliability. They are also compact, making them suitable for indoor BTS sites with space constraints.
- **Lithium-Ion Batteries:** Lithium-ion batteries are increasingly being adopted in modern BTS systems because of their higher energy density, longer life cycle, and lighter weight compared to lead-acid batteries. Additionally, they perform well under extreme temperature conditions and require less frequent maintenance.

Batteries are designed to provide short-term power during brief outages or to bridge the gap between the loss of grid power and the activation of DGs. This ensures that critical equipment remains operational during the transition period. In some cases, batteries can also serve as the sole backup power source for sites where outages are typically short and predictable.

2.2.3 Automatic Transfer Switches

Automatic Transfer Switches (ATS) play a critical role in ensuring seamless power switching for BTS, particularly in scenarios where power failure is predicted or occurs unexpectedly. An ATS, which is [16] created to sense power outages or fluctuations automatically from the main power source and efficiently change to a backup power supply like a generator or battery system without manual intervention. Network operators get a higher reliability level when they use ATS, as the system reduces the probability of a long downtime caused by a power failure. In addition, this also improves the network availability and customer satisfaction, as the users have fewer dropped calls or connectivity problems. Moreover,

ATS also diminishes the load on the maintenance teams, enabling them to concentrate on system improvements, which are proactive instead of reactive, while restoring power.

2.2.4 Rectifier

A rectifier is an essential element in AC-DC conversion since it changes AC into DC by simply allowing one direction of current to flow using diodes or similar electronic devices. In BTS power systems, rectifiers [30] carry out the conversion of the AC mains source into a DC voltage that is stable, usually -48 VDC, which is the basic requirement for the operation of both the telecommunication devices and the charging of the backup batteries. The process of changing the sinusoidal AC waveform into a pulsating DC signal through rectification is typically accompanied by the filtering step, which is intended to smooth the output and keep the voltage level constant, thus ensuring the reliable operation of BTS components.

2.2.5 Power Distribution Units (PDUs)

A Power Distribution Unit (PDU) in a telecom network is an active component that enables efficient and reliable power transfer to the Base Transceiver Station (BTS) device. The PDU [16] regulates the distribution of energy by giving each load the same amount of electrical power, thus eliminating power imbalances and overloading that can lead to the failure of the equipment.

Besides power distribution, PDUs are also very important for BTS equipment in that they provide it with protection from electrical anomalies. The surge protection feature of PDUs is the one that is targeted at the protection of the sensitive telecommunication equipment against the voltage spikes that occur due to lightning, power grid fluctuations, or switching transients. Such voltage surges may not only degrade BTS components but may also cause total outages, which will, in turn, lead to the need for costly repairs. The addition of surge protection in PDUs makes sure that the spikes will be either absorbed or redirected; hence, no harm will be done to the equipment.

Circuit breakers: Circuit breakers to safeguard the PDUs against overcurrent

situations. Short circuits, faulty devices, or temporary spikes in power are some causes of overcurrent, which, if not protected against, can lead to overheating, damage, or even fires. PDUs are equipped with circuit breakers that identify these occurrences and cut the power, preventing further damage and maintaining the safety and reliability of the BTS infrastructure.

The PDUs [16] contribute to more reliable and efficient BTS operations by providing consistent power to avoid downtime and service interruptions. They require very little maintenance, and they also protect critical equipment against surges and overcurrent situations, helping them to last longer. PDUs also lead to increased uptime and operational stability of the systems, which is critical for ensuring uninterrupted communications even in difficult situations. In general, PDUs are integral to power provisioning and therefore to the operation and dependability of telecommunication networks and wireless communication systems.

2.2.6 Protection Devices

Power system protection devices are important to maintain the reliability and efficiency of the operation of the BTS. Combined, these devices help to avoid faults, minimize damage to equipment, and keep the stations up by recognizing abnormal conditions within the system and acting upon them.

- **Surge Protection Devices (SPDs):** These are installed to mitigate transient over-voltages such as those caused by lightning strikes, switching surges, and service disturbances in the utility grid [16]. SPDs protect equipment by capturing surges and harmlessly directing them to ground or neutral, preventing equipment damage. SPDs are usually installed at these critical points, such as where utility power comes into the building, at rectifiers, and where communication lines enter. SPD's greatest benefit is instantaneous reaction time to protection against voltage spikes. They are cheap, easy to do, and most importantly, protect expensive equipment. The downside is that SPDs are consumables and degrade with each surge event. Plus, they do not provide good protection from sustained overvoltage or under-voltage

conditions and thus must be supplemented with other protection.

- **Circuit Breakers:** Circuit breakers offer overcurrent protection by interrupting the flow of electricity under overload or short-circuit conditions. These are installed in AC as well as DC distribution panels, preventing damage to the power system and associated equipment from excessive fault currents. The benefits of circuit breakers are that they are very reliable, can be reset manually or automatically after tripping, and are available in many current and voltage ratings. They are also slower than fuses in some fault conditions or need to be tested on a regular basis to make sure they operate properly.
- **Fuses:** These are another form of overcurrent protection device that is oftentimes used as a failsafe on important, high-power circuits such as those for batteries or rectifiers. It consists of a thin metal filament, which melts when current exceeds a certain level, breaking the circuit. This simple mechanism ensures rapid fault isolation, making fuses highly effective for high-speed protection. They are inexpensive, easy to install, and available in a wide range of ratings, making them suitable for diverse applications. However, they are single-use devices and must be replaced after operation.
- **Uninterruptible Power Supply (UPS):** A UPS provides instantaneous backup power to critical BTS equipment during power outages or disturbances. It bridges the gap between power loss and the activation of the backup generator. The UPS relies [30] on batteries to store energy and includes an inverter to convert the stored DC power to AC for use by the equipment. Additionally, the UPS filters voltage sags, surges, and harmonics, improving power quality. The major advantage of a UPS is its instantaneous response, which ensures the uninterrupted operation of sensitive equipment. For example, an online UPS can provide a response time of 0 milliseconds. However, the runtime of a UPS is limited by the battery capacity, making it suitable for short-term power backup only. UPS systems also require regular maintenance and battery replacement to ensure reliability.

- **Over-voltage and Under-voltage Relays:** Over-voltage and under-voltage relays protect sensitive BTS equipment from sustained voltage abnormalities. These relays continuously monitor voltage levels, and if the voltage exceeds or drops below preset thresholds, the relay disconnects the affected circuit to prevent damage. This is particularly important in ensuring the stability of equipment powered by the grid or backup systems.
- **Ground Fault Protection Devices:** Ground fault protection devices detect leakage currents caused by insulation failures or unintended contact between live conductors and the ground. These devices measure the difference between incoming and outgoing currents in a circuit. If the difference exceeds a preset threshold, indicating a ground fault, the device trips the circuit to protect personnel and equipment. This device is essential for preventing electrical fires and shock hazards. These devices respond quickly to ground faults, ensuring safety. However, they may not detect faults in low-current circuits and can cause nuisance tripping if not properly configured.

2.3 AC vs. DC Distribution

In the BTS power system, the AC and DC distributions have separate roles in carrying power to different parts of the system. AC power generally comes from the main utility grid or backup generators and is then distributed to equipment such as air conditioners, lighting, rectifiers, and other devices that consume a lot of power. DC power, however, is obtained from AC power rectification with the use of rectifiers that change AC into DC to provide low-voltage devices that are very important for BTS, such as radios, communication modules, and backup batteries. DC distribution is usually the first choice for sensitive electronics, as it has the ability to provide stable power, which consequently leads to less electromagnetic interference and hence more reliability.

2.4 Load Management

Load Control plays an important role in the optimization of power distribution in a BTS. This system helps the network in being more efficient and in maximizing the uptime, as loads are monitored and power to the critical components is what is prioritized and supplied during off times and peak load times. Battery Low Voltage Disconnect (BLVD) and Load Low Voltage Disconnect (LLVD) are protective measures in the power system of a BTS to prevent equipment damage and unproductive downtime of the BTS [8].

2.4.1 Battery Low Voltage Disconnect (BLVD)

BLVD disconnects the battery from the load when the battery's voltage level reaches a predetermined minimum point, which is another safety feature. This is a method to avoid the battery discharging below the safe operating threshold, where it starts to get damaged irreversibly, shorten the lifetime of the battery, and render it incapable of charging again. As the battery voltage decreases to an unacceptable level, the BLVD safeguards this remaining energy to be used for emergency operations or to ensure that energy isn't lost due to a total depletion of the battery, which is especially useful in scenarios like long power outage conditions.

2.4.2 Load Low Voltage Disconnect (LLVD)

The LLVD has been created to shed unessential loads when the system voltage drops below a specific value in order to protect critical operations. For example, if power is constrained in a BTS, such as in the case of grid failure, where the battery is the only power source, the LLVD will cut off non-vital apparatus like radio equipment and communication modules. Thus, core functions, e.g., transmission of signals, can continue to function even under low power availability.

A bus bar under voltage condition (BBUV) [8] is classified as when the main DC bus voltage falls below a certain threshold level, in this case 48.2 VDC. The busbar, an important component, distributes power to the various loads and equipment. This is an indication of an impending power stability issue, which can be due to a heavy load

on batteries, low charging current from rectifiers/solar systems, or battery degradation. Failure to remedy this voltage drop will lead to the swift draining of any stored energy in the battery. In this sense, BBUV is considered a precursor of critical conditions such as LLVD and BLVD, giving operators time to intervene.

According to [8], the LLVD disconnects non-essential loads at 46.2 VDC to conserve battery storage capacity for critical operations. This represents a prophylaxis that aims to prolong battery life and keep critical services alive for as long as possible. Not critical appliances, auxiliary devices, and/or HVAC systems are disconnected to minimize the total draw on the battery. This causes the voltage to decrease at a slower rate, providing operators more time to deal with it. The LLVD activates an alarm to alert the operators of the voltage drop and of the disconnection of non-critical loads. This phase keeps critical infrastructures like communications and monitoring devices in place to begin operationalization. If the voltage continues to drop and reaches a point of 45.2 VDC, the BLVD system will be turned ON, isolating all loads save for the monitoring systems. This is a protective measure to prevent the battery from deep discharge, which can cause irreparable damage. Lead-acid batteries, commonly used in telecom systems [17], have a minimum safe discharge level. Discharging below this level can lead to a chemical reaction that degrades the battery, reduces cycle life, and causes permanent capacity loss. By disconnecting loads at the BLVD threshold, the system safeguards battery health while maintaining communication between the monitoring system and the site. However, this disconnection results in the ending of most site operations, including critical services like BTS functionality.

2.5 BTS Power System Failure and Power Outage

The BTS power system failure refers to the inability of the system to supply adequate power to the BTS equipment, which could lead to degraded or interrupted cellular services. BTS power systems are designed to ensure continuous and reliable operation, but failures can occur due to various reasons. These failures [28] are generally classified into partial

power system failures and total power system failures (outages). A partial BTS power system failure occurs when some components of the power system, such as rectifiers, the battery bank, or parts of the DC distribution, fail or degrade, but the system can still function at reduced capacity. Partial failures often result in inadequate power delivery to certain BTS components, causing reduced performance, such as decreased coverage or capacity. They also act as warnings for potential larger failures. Of course, issues might be of other nature, for instance, abnormal DC output voltage drop, inconsistent DC load current, or a reduced discharge capability of the battery. By monitoring these parameters, partial failures can be detected in an early stage so that technicians can carry out preventive maintenance before they result in further issues.

Conversely, a Full shutdown of the BTS power system entails a complete shutdown of the power system, leading to a full outage on the BTS site. The result of this is generally due to significant problems, such as a completely drained battery bank, all rectifier modules being down, or the AC power being interrupted. The DC distribution system in these instances does not power the BTS equipment at all, resulting in a total outage. Partial failures, which can lead to a catastrophic failure of the power system, are usually detected by non-ignored early warnings. It is therefore crucial to detect and correct "partial failures" before they lead to total system outages.

The BTS DC distribution system is the component of the BTS power system that distributes power from the rectifiers and the battery bank to the DC loads, mainly BTS equipment like transceivers, amplifiers, and cooling systems. The main system components of a DC distribution system are the rectifier output, which converts AC power to DC and is the main source of power; the battery bank that acts as a backup power supply in case of failure of the rectifier or the AC power source; and the DC loads. These components are typically wired in parallel so that there is redundancy; the system can still operate if one of the components fails, albeit at a lower capacity. Such redundancy is important to ensure the unaffected operation in cases where the components fail.

Failure in the BTS power system can be located and recognized by analyzing certain parameters in the DC distribution system. Among these parameters are DC output voltage, DC load current, and DC load power. To illustrate, a considerable fall in the DC output voltage may lead to the rectifier going wrong, too much load, or the battery running out. Also, changes in the DC load current, like unnecessarily high or low current, may give the picture of system overloading, equipment disconnection, or short circuits. DC load power, which is the power that the BTS equipment uses additionally, can give signals about the system's condition. For instance, a figure that does not correspond to the actual consumption may indicate inefficiencies or the malfunction of some components. Hence, by being sensitive to these parameters all the time, the technician can spot the warning signs of parts or the whole failure in the system and come up with the best solutions to preserve the continuous service.

2.6 Factors of BTS Power System Failure

The BTS power system failure can affect the telecommunication network to the extent that it becomes unreliable. A typical failure is usually a mixture of electrical, environmental, operational, and maintenance factors. The BTS infrastructure [15] is a source of approximately 80% of interruptions in the cellular network caused by power outages at BTS sites. Some causes of the BTS power system failure are

- **Unstable Power Supply:** BTS power systems fail when the utility grid delivers unbalanced power, such as voltage spikes or outages. These disruptions halt BTS operations, cutting off mobile network services. Unreliable grid power is the top reason for BTS outages [8].
- **Weak Backup Power Systems:** When the grid power fails, BTS relies on backup systems like batteries or generators. If these are underpowered or poorly designed, they cannot sustain operations, leading to downtime. A study on BTS power planning compared backup configurations, finding that systems with only batteries

often failed during long outages [28].

- **Environmental Factors:** Extreme temperatures, humidity, and dust accumulation significantly reduce BTS power systems. High temperatures accelerate component aging, while humidity and dust can cause short circuits or corrosion.
- **Excessive Power Use and Poor Energy Management:** BTS systems consuming too much power, without efficient energy management, overload their power infrastructure, increasing failure risks and costs. According to a work in [31], unoptimized scheduling of power could lead to strain on systems, while optimized power management minimizes the risk of failure and cost of power.
- **Lightning and Surge Events:** BTS sites are generally in exposed locations and susceptible to lightning strikes and surges. If not properly designed or maintained, these events can impact power equipment despite having surge protection.
- **Inadequate Preventive Maintenance:** Regularly inspecting and maintaining a system would prevent issues such as corrosion, worn-out parts, and other problems that can render a system failure.
- **Human Error:** Errors in the installation, upkeep, or upgrade of the power system, like faulty wiring or configuration mistakes, can lead to a failure of the power system.
- **Equipment Aging and Wear:** The aging of components like capacitors, transformers, and relays causes them to become inefficient and unreliable. Thermal and electrical stress are aging accelerators.

2.7 Monitoring and Maintenance of BTS Power Systems

Telecommunication network reliability can only be guaranteed through the capability of BTS power systems, which need to be regularly monitored and maintained. BTS power systems that include batteries, rectifiers, and backup generators are the main components that are easily affected by environmental conditions, load changes, and the aging of parts. Such incidents can, thus, cause service interruptions.

2.7.1 Alarm Systems for BTS Power Failure

Power failure at BTS sites can lead to significant service disruptions, affecting network reliability and customer satisfaction. Alarm systems are the backbone of BTS power management by providing real-time notifications of power-related issues. They reduce downtime by alerting network operators to faults before they escalate, minimizing service interruptions. For ET [32], where power outages are a leading cause of network outages, proactive monitoring through alarms is crucial for maintaining service quality. Automated alarm systems also reduce human error and operational costs by modernized fault detection and response processes. These systems categorize alarms into critical, major, minor, and warning levels to prioritize response efforts and maintain network integrity.

- **Critical Alarms:** Indicate severe, service-affecting issues requiring immediate action. such as the BLVD alarm, battery discharge alarm, and battery fuse being broken [32].
- **Major alarms:** Serious problems that could affect service but are not as urgent as critical alarms, which frequently call for timely but planned maintenance, include air conditioner failure, rectifier failure, and DC under-voltage alarm [32].
- **Minor Alarms:** These are typically addressed during routine maintenance to prevent escalation. such as phase L1/L2/L3 failure, environmental factors, and DC overvoltage [32].
- **Warning Alarms:** Provide early notifications of potential issues. Like DG running, over-discharge [32].

2.7.2 Predictive maintenance using IoT and AI technologies

The integration of Internet of Things (IoT) and Artificial Intelligence (AI) technologies [18], in predictive maintenance for the power system of a BTS, has transformed the reliability and efficiency of telecommunications infrastructure. IoT-enabled sensors can continuously monitor critical components of the BTS power system, such as batteries, rectifiers, and

generators, collecting real-time data on temperature, voltage, current, and environmental conditions. Such data gets sent to centralized systems that employ AI algorithms to detect trends and forecast possible faults even without any signs. Furthermore, AI-based predictive models can spot irregularities and forecast the wear and tear of devices with great precision [19].

2.8 Challenges in BTS Power System

Power systems in communication have a lot of problems that arise mainly from the usage of modern technologies and the necessity for dependable, efficient, and scalable systems. These problems are made worse by the changing nature of power systems, such as the inclusion of distributed energy sources and the growing need for telecommunication services that have high bandwidth and low latency. Such solutions are very dependent on the development of technology and the deployment of more strategic plans. The succeeding sections go over the different challenges and means of power in telecommunication.

- **Grid Instability and Inadequate Power Supply:** The BTS sites often face issues with unstable power supplies, particularly in regions with unreliable grid electricity. Power instability leads to frequent shutdowns, reduced equipment lifespan, and increased maintenance costs.
- **Battery Backup System Failure:** The BTS sites often rely on backup batteries, such as lead-acid or lithium-ion, to sustain operations during power outages. However, frequent discharging and recharging cycles and high ambient temperature accelerate battery degradation. According to [33], improper battery maintenance and environmental factors are major contributors to rapid battery wear, particularly in off-grid BTS installations.
- **Diesel Generator Reliability Issue:** DGs are commonly used in off-grid or backup power scenarios for BTS sites. However, they are prone to mechanical failures, fuel supply challenges, and insufficient maintenance. [28] states that the operation and maintenance of DGs typically compensate for 35% of the total operating cost of

BTS.

- **High Energy Consumption and Cooling System Failure:** The BTS consumes significant energy, particularly for cooling systems that maintain optimal temperatures for equipment. Failures in cooling systems or inefficient energy use can lead to overheating and power system shutdown.
- **Integration Challenges with Renewable Energy Systems:** Many BTS sites still rely heavily on non-renewable energy sources, such as diesel generators, due to their simplicity and reliability. However, [34] the lack of integration with renewable energy solutions (e.g., solar or wind power) limits the sustainability and cost-effectiveness of power systems.

3 Machine Learning Algorithm

3.1 Hidden Markov Model

An HMM is a statistical model used to describe systems that transition between a set of hidden states over time, where the actual sequence of hidden states cannot be directly observed [35]. However, the actual sequence of states (hidden states) is not directly observable but can be inferred from observable data.

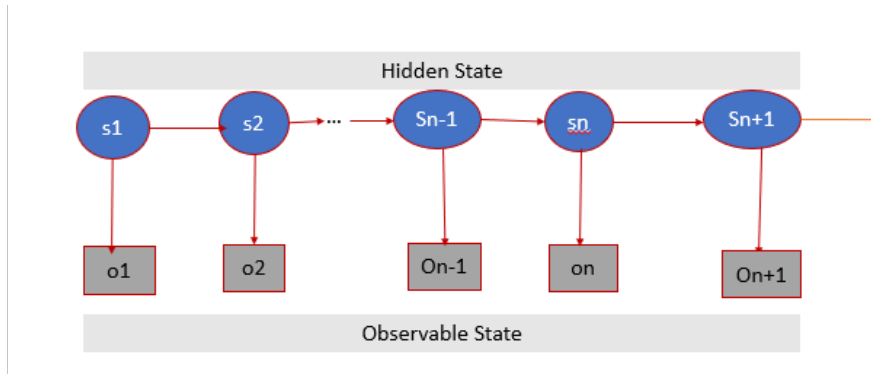


Figure 3.4: HMM, Structure with hidden and observable states

3.1.1 Markov Chains and the Basis of HMMs

A Markov chain is a stochastic model that describes a sequence of events in which the probability of transitioning from one state to another depends solely on the current state, independent of prior history [36]. Formally, a Markov chain is defined by a finite set of states $\mathcal{X} = \{x_1, x_2, \dots, x_n\}$ and a transition probability matrix $\mathbf{P} = [p_{ij}]$, where p_{ij} denotes the probability of moving from state x_i to x_j . Each p_{ij} satisfies $0 \leq p_{ij} \leq 1$, and the row sums are \mathbf{P} equal to 1, i.e.,

$$\sum_{j=1}^n p_{ij} = 1 \quad \text{for all } i, \tag{3.1}$$

A Markov chain possesses the memoryless Markov property, meaning that the subsequent state is only dependent on the present state and not on the order of states that came before it (cited by [36]). It asserts that the probability of transitioning to a future state x_{t+1} at time $t + 1$ depends exclusively on the current state x_t at time t , not on the

sequence of prior states x_0, x_1, \dots, x_{t-1} . Mathematically, this is expressed as [36]:

$$P(X_{t+1} | X_0, X_1, \dots, X_t) = P(X_{t+1} | X_t) \quad (3.2)$$

where X_t denotes the state at time t . This property simplifies analysis by reducing historical dependencies to the present state [20]. HMMs extend this concept by layering observable outputs that are probabilistically linked to the hidden states. More specifically, the mechanism of each hidden state of an HMM is that it selects an observation from the emission probability distribution, which defines, jointly with the state transition probabilities, the probabilistic structure of an HMM. The hidden states are the unobservable forces that drive the system, while the emissions are the observable parts of the system, but they are still noise, indirect representations of those states. The probabilistic setting allows HMMs to be flexible in dealing with uncertainty and variability in the data; thus, they become quite effective in handling noisy signals or ambiguous linguistic patterns. Through the fusion of the Markov property and the emission probabilities, HMMs create a sturdy framework for the sequential data analysis and the hidden structures.

3.1.2 Components of HMM

One of the defining features of a Hidden Markov Model (HMM) is that it is a statistical model, and it is complete when it is characterized by the five components that are detailed in [21].

1. **States (Hidden States):** These hidden states represent the unobservable conditions or processes underlying the system. At each time step, the system resides in one of these states, but the state itself cannot be directly observed. Instead, observations provide indirect evidence of the system's state. Denoted as $S = \{S_1, S_2, \dots, S_N\}$, where N is the number of states.
2. **Observations (Emissions):**
 - The observable outputs (or emissions) are drawn from a finite set of symbols. These outputs are generated according to a probability distribution defined for

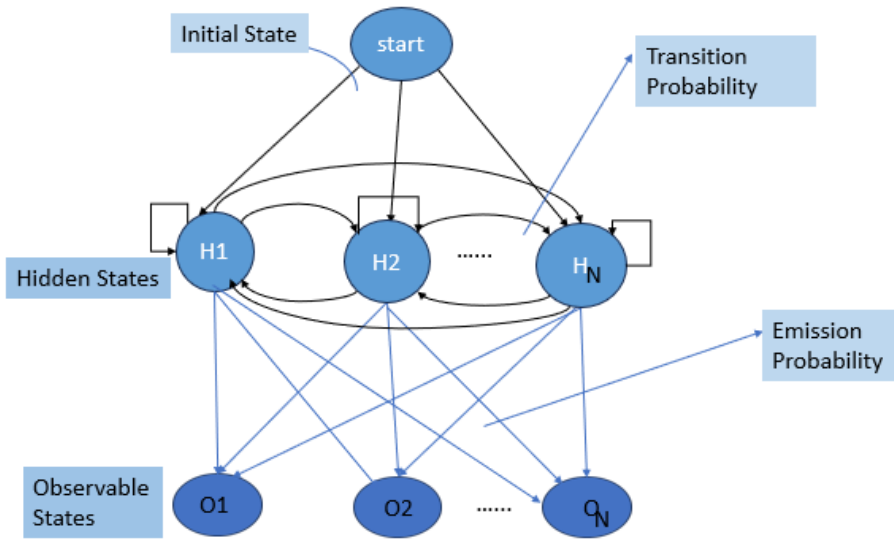


Figure 3.5: HMM, Structure and component [7]

each hidden state. Denoted as $O = \{O_1, O_2, \dots, O_T\}$, where T is the length of the observation sequence.

3. Transition Probabilities:

- A matrix A that defines the probability of transitioning from one state to another. According to [36], it is defined as

- $A = \{a_{ij}\}$, where:

$$a_{ij} = P(S_j | S_i) \tag{3.3}$$

- the rows A satisfy the condition:

$$\sum_{j=1}^N a_{ij} = 1 \quad \text{for all } i, j = 1, 2, 3, \dots, n \tag{3.4}$$

4. Emission Probabilities:

- The observation probabilities (B) specify how likely each observation symbol is to be generated from a given hidden state. In [36], ' B ' is defined as

- $B = \{b_{ij}\}$, where:

$$b_{ij} = P(O_j | S_i) \tag{3.5}$$

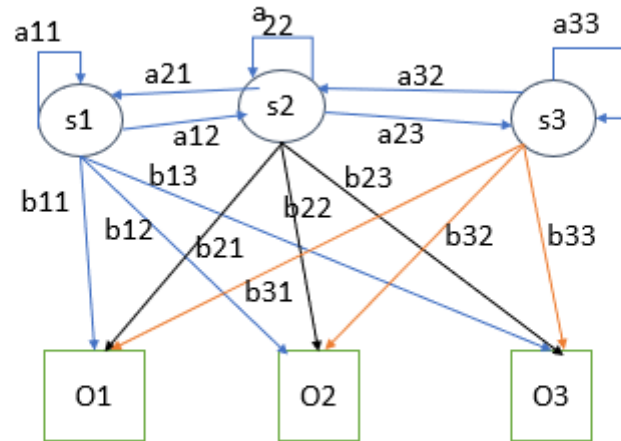


Figure 3.6: HMM, with three hidden and observable states

- The rows B satisfy the condition:

$$\sum_{j=1}^N b_{ij} = 1 \quad \text{for all } i = 1, 2, 3, \dots, S, \text{ and, } j = 1, 2, 3, \dots, n \quad (3.6)$$

where N is the number of observation symbols.

5. **Initial State Distribution:** The initial probabilities(π) define the likelihood of the system starting in each hidden state.

- A vector $\pi = \{\pi_i\}$, where:

$$\pi_i = P(S_i \text{ at } t = 1) \quad (3.7)$$

- The initial probabilities satisfy

$$\sum_{i=1}^N \pi_i = 1. \quad (3.8)$$

The structure of an HMM is described in Figure 3.5, which represented as $\lambda = (S, O, A, B, \pi)$

3.1.3 Problems in Hidden Markov Model and Their Solutions

An HMM is a statistical model that represents systems with hidden states through observable sequences. The three fundamental challenges [51], [52] associated with HMMs are:

Decoding problem: The decoding problem in HMMs [39], involves finding the most likely sequence of hidden states $Q = (q_1, q_2, \dots, q_T)$ that best explains an observation sequence $O = (O_1, O_2, \dots, O_T)$, given the HMM parameters $\lambda = (A, B, \pi)$. The goal is to maximize the posterior probability $P(Q | O, \lambda)$, which represents the probability of the hidden state sequence Q conditioned on the observations O parameters λ , mathematically, this is written as in [39]:

$$Q^* = \arg \max_Q P(Q | O, \lambda). \quad (3.9)$$

Using Bayes' theorem, the posterior probability $P(Q | O, \lambda)$ can be rewritten as

$$P(Q | O, \lambda) = \frac{P(Q, O | \lambda)}{P(O | \lambda)}. \quad (3.10)$$

Here:

- $P(Q, O | \lambda)$: The joint probability of the state sequence Q and the observation sequence O given the model parameters λ .
- $P(O | \lambda)$: The marginal probability of the observation sequence O given the model parameters λ , also known as the likelihood of the observations.

Thus, the decoding problem reduces to

$$Q^* = \arg \max_Q P(Q, O | \lambda). \quad (3.11)$$

Several factors [39] create substantial difficulties in solving the decoding problem:

1. **Exponential Growth of State Sequences:** For a sequence of length T with N states, there are N^T possible state sequences. Extensively evaluating all sequences is computationally impossible for large T or N .

2. **Conditional Dependencies:** The decoding process involves dependencies between observations and states, as well as transitions between states. Accurately modeling these dependencies is significant.
3. **Optimization Complexity:** The goal is to find the single best state sequence Q^* , which requires globally optimizing over all possible sequences.

To address these challenges, the Viterbi algorithm is used with the help of the joint probability. The joint probability $P(Q, O | \lambda)$ can be factored into two components:

- The probability of the state sequence Q , $P(Q | \lambda)$. and
- The conditional probability of the observations O given the state sequence Q , $P(O | Q, \lambda)$.

According to [39], the $P(O | Q, \lambda)$ is given by

$$P(Q, O | \lambda) = P(O | Q, \lambda)P(Q | \lambda). \quad (3.12)$$

Then, the probability of the state sequence $Q = (q_1, q_2, \dots, q_T)$ is determined by the initial state probabilities π and the state transition probabilities A . Assuming the Markov property (the probability of transitioning to a state depends only on the previous state), based on [39], we can write

$$P(Q | \lambda) = P(q_1) \prod_{t=2}^T P(q_t | q_{t-1}). \quad (3.13)$$

By substituting the model parameters,

$$P(Q | \lambda) = \pi_{q_1} \prod_{t=2}^T a_{q_{t-1}, q_t} \quad (3.14)$$

where

- π_{q_1} : The initial probability of the first state q_1 .
- a_{q_{t-1}, q_t} : The transition probability from state q_{t-1} to state q_t .

In addition, the probability of the observation sequence $O = (O_1, O_2, \dots, O_T)$ given the state sequence $Q = (q_1, q_2, \dots, q_T)$ depends on the emission probabilities B . Assuming that each observation O_t depends only on the current state q_t , according to [24], equation (3.15) is defined.

$$P(O | Q, \lambda) = \prod_{t=1}^T P(O_t | q_t, \lambda). \quad (3.15)$$

Substituting the parameters,

$$P(O | Q, \lambda) = \prod_{t=1}^T b_{q_t}(O_t), \quad (3.16)$$

where

- $b_{q_t}(O_t)$: The probability of emitting observation O_t from state q_t .

Combining the two components, the joint probability $P(Q, O | \lambda)$ can be written as

By substituting equation (3.16) and equation (3.14) in (3.12), we get

$$P(Q, O | \lambda) = \left[\prod_{t=1}^T b_{q_t}(O_t) \right] \left[\pi_{q_1} \prod_{t=2}^T a_{q_{t-1}, q_t} \right]. \quad (3.17)$$

By modifying (3.17):

$$P(Q, O | \lambda) = \pi_{q_1} b_{q_1}(O_1) \prod_{t=2}^T a_{q_{t-1}, q_t} b_{q_t}(O_t). \quad (3.18)$$

The equation (3.18) captures the joint probability of a specific state sequence Q and observation sequence O under the given HMM parameters λ .

Learning Problem: The learning problem in HMMs involves estimating the model parameters $\lambda = (A, B, \pi)$, given a set of observed sequences defined in [43]

Assume $O = \{o_1, o_2, \dots, o_T\}$ denotes the observed sequence and $S = \{s_1, s_2, \dots, s_T\}$ denote the hidden state sequence. Therefore, the learning problem involves maximizing the likelihood of the observed data

$$P(O | \lambda)$$

, which is given by

$$P(O | \lambda) = \sum_S P(O, S | \lambda), \quad (3.19)$$

where $\lambda = (A, B, \pi)$, and S is summed over all possible state sequences. Then, the joint probability $P(O, S | \lambda)$ can be expanded as

$$P(O, S | \lambda) = P(S | \lambda)P(O | S, \lambda). \quad (3.20)$$

The probability of a state sequence S is determined by the initial probabilities π and transition probabilities A :

$$P(S | \lambda) = \pi_{s_1} \prod_{t=2}^T a_{s_{t-1}, s_t}. \quad (3.21)$$

The probability of the observations O given a state sequence S is determined by the emission probabilities B :

$$P(O | S, \lambda) = \prod_{t=1}^T b_{s_t}(o_t). \quad (3.22)$$

By using (3.21) and (3.22), (3.20) becomes

$$P(O, S | \lambda) = \pi_{s_1} b_{s_1}(o_1) \prod_{t=2}^T a_{s_{t-1}, s_t} b_{s_t}(o_t). \quad (3.23)$$

Summing over all possible S , the likelihood of the observations becomes:

$$P(O | \lambda) = \sum_S \pi_{s_1} b_{s_1}(o_1) \prod_{t=2}^T a_{s_{t-1}, s_t} b_{s_t}(o_t). \quad (3.24)$$

Direct computation of, which $P(O | \lambda)$ is infeasible due to the exponential growth in the number of possible state sequences S (N^T , where N is the number of states and T is the sequence length). The forward-backward algorithm is used to handle this challenge.

Evaluation problem: The evaluation problem in HMMs involves determining the probability of an observation sequence $\mathbf{O} = (O_1, O_2, \dots, O_T)$, given the model parameters $\lambda = (\mathbf{A}, \mathbf{B}, \pi)$. This probability, denoted by $P(\mathbf{O} | \lambda)$, measures how well the model λ

explains the observations \mathbf{O} , According to [24], it is defined as:

$$P(\mathbf{O} | \lambda) = \sum_{\mathbf{Q}} P(\mathbf{O}, \mathbf{Q} | \lambda) \quad (3.25)$$

where $\mathbf{Q} = (q_1, q_2, \dots, q_T)$ represents all possible sequences of hidden states. The joint probability $P(\mathbf{O}, \mathbf{Q} | \lambda)$ can be explained as:

$$P(\mathbf{O}, \mathbf{Q} | \lambda) = P(Q_1 | \lambda) \prod_{t=2}^T P(Q_t | Q_{t-1}, \lambda) \prod_{t=1}^T P(O_t | Q_t, \lambda) \quad (3.26)$$

where:

- $P(Q_1 | \lambda) = \pi_{q_1}$: The probability of starting in the initial state q_1 , as defined by the initial state distribution π .
- $P(Q_t | Q_{t-1}, \lambda) = a_{q_{t-1}, q_t}$: The probability of transitioning from state q_{t-1} to state q_t , as defined by the state transition matrix \mathbf{A} .
- $P(O_t | Q_t, \lambda) = b_{q_t}(O_t)$: The probability of observing O_t a given state q_t , as defined by the emission probabilities \mathbf{B} .

Thus, (3.26) is defined as

$$P(\mathbf{O}, \mathbf{Q} | \lambda) = \pi_{q_1} \prod_{t=2}^T a_{q_{t-1}, q_t} \prod_{t=1}^T b_{q_t}(O_t) \quad (3.27)$$

To compute the probability of the observation sequence \mathbf{O} , (3.25) becomes:

$$P(\mathbf{O} | \lambda) = \sum_{\mathbf{Q}} P(\mathbf{O}, \mathbf{Q} | \lambda) \quad (3.28)$$

Substituting (3.27) into (2.28), (3.29) is written as

$$P(\mathbf{O} | \lambda) = \sum_{\mathbf{Q}} \pi_{q_1} \prod_{t=2}^T a_{q_{t-1}, q_t} \prod_{t=1}^T b_{q_t}(O_t) \quad (3.29)$$

3.1.4 Solutions for Decoding Problem, and Viterbi Algorithm

The Viterbi algorithm [22], named after Andrew Viterbi, who introduced it in 1967, is a dynamic programming solution efficient for the decoding problem. This algorithm [40] addresses the decoding problem by using dynamic programming to maximize the probability of the state sequence at each time step. Based on [23], it works

- **Initialization:** Defined $\delta_t(i)$ as the maximum probability of any state sequence ending in state S_i at time t , given the observations up to time t :

$$\delta_1(i) = \pi_i b_i(O_1), \quad \forall i = 1, 2, \dots, N. \quad (3.30a)$$

- **Recursion:** For $t = 2, 3, \dots, T$ and $j = 1, 2, \dots, N$:

$$\delta_t(j) = \max_{i=1, \dots, N} [\delta_{t-1}(i) a_{ij}] b_j(O_t), \quad (3.30b)$$

where a_{ij} is the transition probability from state S_i to S_j , and $b_j(O_t)$ is the emission probability of observation O_t in state S_j .

- **Termination:** At the final time step T , the probability of the most likely state sequence is:

$$P(Q^*, O | \lambda) = \max_{i=1, \dots, N} \delta_T(i). \quad (3.30c)$$

Therefore, the most likely ending state is given by

$$q_T^* = \arg \max_{i=1, \dots, N} \delta_T(i). \quad (3.31)$$

3.1.5 Solutions for Learning Problem and the Baum-Welch Algorithm

The Baum-Welch algorithm [41, 42] is a specific case of the Expectation-Maximization (EM) algorithm used to estimate the parameters of an HMM, when the underlying states are not directly observable. It is a stepwise approach that, by altering the HMM parameters,

seeks to find the maximum probability of a given sequence of emissions. The Baum-Welch algorithm's objective is to find $\lambda = (A, B, \pi)$ from only the observation sequence $O = \{O_1, O_2, \dots, O_T\}$, without having the hidden state sequence $Q = \{q_1, q_2, \dots, q_T\}$. This algorithm is based on the forward and backward variables that continuously make the likelihood of the observable given λ higher. It is stated in [43],

- **Forward variable** $\alpha_t(i)$:

$$\alpha_t(i) = P(o_1, o_2, \dots, o_t, q_t = i \mid \lambda), \quad (3.32a)$$

The likelihood of the partial observation sequence until time t and that the state at time t is i can be written as,

Recursion:

$$\alpha_t(i) = \sum_{j=1}^N \alpha_{t-1}(j) a_{ji} b_i(o_t), \quad (3.33)$$

with initialization:

$$\alpha_1(i) = \pi_i b_i(o_1).$$

- **Backward variable** $\beta_t(i)$:

$$\beta_t(i) = P(o_{t+1}, o_{t+2}, \dots, o_T \mid q_t = i, \lambda), \quad (3.32b)$$

The probability of the partial observation sequence from time $t + 1$ to T , given state i at time t is defined by,

Recursion:

$$\beta_t(i) = \sum_{j=1}^N a_{ij} b_j(o_{t+1}) \beta_{t+1}(j), \quad (3.34)$$

with initialization:

$$\beta_T(i) = 1.$$

The Baum-Welch algorithm in [56], computes:

- $\gamma_t(i)$ The probability of being in state i at time t , given the observation sequence:

$$\gamma_t(i) = P(q_t = i \mid \mathbf{O}, \lambda) = \frac{\alpha_t(i)\beta_t(i)}{\sum_{j=1}^N \alpha_t(j)\beta_t(j)}, \text{ and} \quad (3.35)$$

- $\xi_t(i, j)$ The probability of being in state i at time t and transitioning to state j at time $t + 1$:

$$\xi_t(i, j) = P(q_t = i, q_{t+1} = j \mid \mathbf{O}, \lambda) = \frac{\alpha_t(i)a_{ij}b_j(o_{t+1})\beta_{t+1}(j)}{\sum_{i=1}^N \sum_{j=1}^N \alpha_t(i)a_{ij}b_j(o_{t+1})\beta_{t+1}(j)}. \quad (3.36)$$

Generally, the Baum-Welch algorithm works as:

- Initialization: Start with initial estimates of the HMM parameters $(\mathbf{A}, \mathbf{B}, \boldsymbol{\pi})$.
- Expectation Step: Using the current estimates of the parameters, compute:
 - The forward probabilities $\alpha_t(i)$.
 - The backward probabilities $\beta_t(i)$.
 - The state occupation probabilities $\gamma_t(i)$.
 - The transition probabilities $\xi_t(i, j)$.
- Maximization Step (M-Step): Update the HMM parameters to maximize the expected log-likelihood:

- **Update initial state probabilities:**

$$\pi_i = \gamma_1(i). \quad (3.39)$$

- **Update state transition probabilities:**

$$a_{ij} = \frac{\sum_{t=1}^{T-1} \xi_t(i, j)}{\sum_{t=1}^{T-1} \gamma_t(i)}. \quad (3.40)$$

- **Update observation probabilities:**

$$b_j(o_k) = \frac{\sum_{t=1}^T \delta(o_t, o_k) \gamma_t(j)}{\sum_{t=1}^T \gamma_t(j)}, \quad (3.41)$$

where $\delta(o_t, o_k)$ is an indicator function that equals 1 if $o_t = o_k$, and 0 otherwise.

- **Iteration:** Repeat the E-step and M-step until the change in log-likelihood between iterations falls below a predefined threshold.

The Baum-Welch technique produces an optimized HMM that describes the observed data by repeating these stages until it converges to a local maximum of the likelihood function.

3.1.6 Solutions for Evaluation Problem and Back and Forward Algorithm

The forward algorithm is a recursive dynamic programming approach that efficiently computes $P(\mathbf{O} \mid \lambda)$, thereby avoiding the exponential complexity of summing over N^T state sequences. The forward variable $\alpha_t(i)$ is defined as the probability of observing the partial sequence (O_1, O_2, \dots, O_t) and being in state i at time t , given the model λ :

$$\alpha_t(i) = P(O_1, O_2, \dots, O_t, q_t = i \mid \lambda) \quad (3.42)$$

This algorithm comprises three steps [23]:

- **Initialization** (at $t = 1$):

$$\alpha_1(i) = \pi_i b_i(O_1), \quad 1 \leq i \leq N \quad (3.43)$$

Here, π_i is the initial probability of state i , and $b_i(O_1)$ is the probability of observing O_1 in state i .

- **Recursion:**

$$\alpha_t(i) = \sum_{j=1}^N \alpha_{t-1}(j) a_{j,i} b_i(O_t), \quad 1 \leq i \leq N \quad (3.44)$$

This equation updates $\alpha_t(i)$ by summing over all possible previous states j , multiplying the probability of transitioning from j to i ($a_{j,i}$), the probability of being in state j at time $t - 1$ ($\alpha_{t-1}(j)$), and the emission probability of observing O_t from state i ($b_i(O_t)$).

- **Termination:**

$$P(\mathbf{O} \mid \lambda) = \sum_{i=1}^N \alpha_T(i) \quad (3.45)$$

The total probability of the observation sequence is obtained by summing over all possible states at time T . The forward algorithm reduces the computational complexity of evaluating $P(\mathbf{O} \mid \lambda)$ from $O(N^T)$ to $O(N^2T)$:

- $O(N^2)$ operations are required at each time step t , as it involves summing over N states for each of the N current states.
- This is again done for T time steps, which results in a total complexity of $O(N^2T)$.

3.1.7 Significance of Hidden Markov Models in Predicting Base Transceiver Station Power System

In the case of BTSs, HMMs play a major role in forecasting power system performance, efficiently managing energy usage, and improving the reliability of operations. HMMs are considered to be a highly efficient means for the representation and prediction of the power system dynamics in BTSs. This results from the fact that HMMs are appropriate for random systems where the hidden processes are not directly observable but can be inferred through observable outputs.

3.2 Long Short-Term Memory Algorithm

LSTM networks are special versions of RNNs that were made to overcome the problems regular RNNs have with recognizing long-term dependencies in the sequence data, which have been caused by the vanishing gradients issue. LSTMs eat away at this problem by allowing the gradients to be transferred more optimally through the memory cell. The gating

mechanisms, as stated in [45], ensure that gradients do not get less than exponential along the time axis. In [46], an LSTM is shown to be able to gradually and selectively keep or get rid of information via the memory cell and the gating mechanism. Therefore, they are super-efficient tools for solving problems like natural language processing, time-series forecasting, and speech recognition, where the ability to capture long-term dependencies is a prerequisite to being successful. LSTMs, unlike regular RNNs, which just swap the hidden state at every time step, accomplish a different memory cell that is able to not only save but also change the information for a longer period of time through the use of special gates.

An LSTM [29], has a memory cell, an input gate, an output gate, and a forget gate. These gates work together [47] to control what information is kept, added to, or passed along. Let's explore each component with its mathematical formulation.

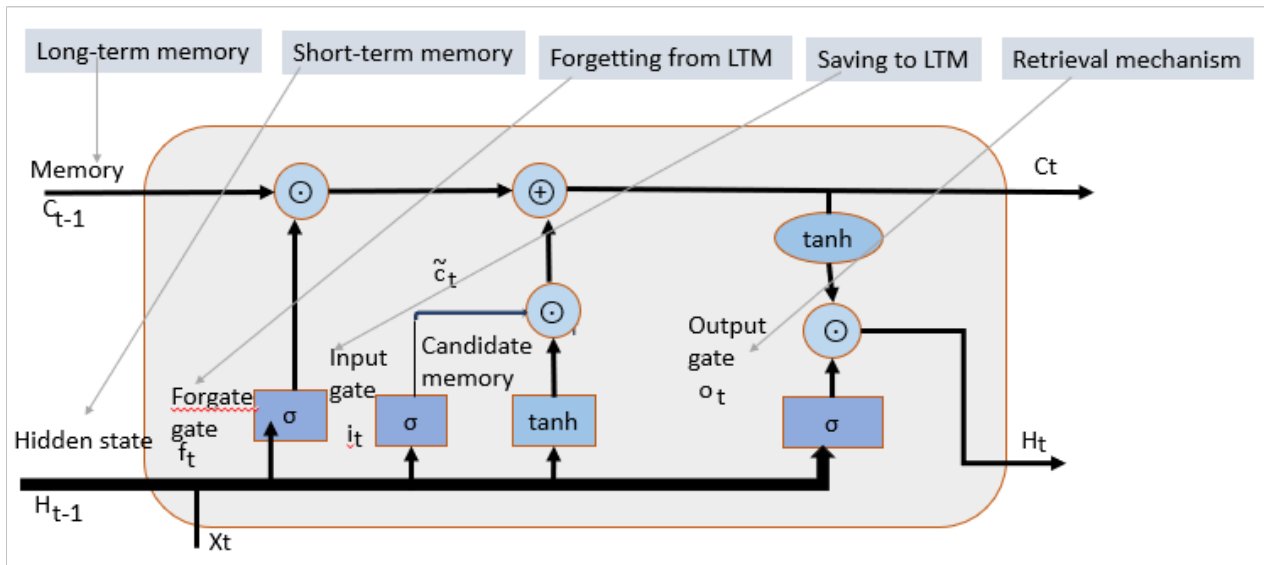


Figure 3.7: LSTM Architecture [29]

- **Cell State (C_t):** The memory cell serves as the core of the LSTM, storing information across time steps. It acts as a transporter where information flows through the network, and the gates decide what to add, update, or remove. The cell state is

updated as in [47],

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \quad (3.2a)$$

where:

- f_t is the forget gate output
- i_t is the input gate output
- \tilde{C}_t is the candidate cell state , and
- \odot denotes element-wise multiplication.

- **Forget Gate (f_t):** As the name suggests, the forget gate determines which parts of the memory cell's content should be forgotten. This gate's sigmoid output decides what fraction of the previous memory state to keep, enabling the LSTM to discard irrelevant or outdated information selectively. The forget gate decides what fraction of the previous cell state c_{t-1} should be retained in [47],

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (3.2b)$$

where:

- f_t : forget gate output, a vector of values between 0 (forget) and 1 (retain).
 - $\sigma(z) = \frac{1}{1+e^{-z}}$: Sigmoid activation function.
 - W_f : Weight matrix for the forget gate.
 - h_{t-1} : Hidden state from the previous time step.
 - x_t : Current input vector.
 - b_f : Bias vector for the forget gate.
- **Input Gate (i_t):** The input gate controls how much the incoming data (the current input and the previous hidden state) should be written to the memory cell. It features a sigmoid activation function to make a decision about the significant values, while the tanh activation assists in controlling the range of new candidate values. It

decides the extent of new data that must be incorporated in the cell state [47].

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (3.2c)$$

The candidate's memory update is

$$\tilde{c}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (3.2d)$$

- **Output Gate (o_t):** The output gate is responsible for the amount of information from the memory cell's current state that will become the LSTM's output at that particular step. The output is generated by carrying out a sigmoid activation to select the parts of the memory that are shown, and then the tanh activation is used to adjust the output. The output gate is the part of the cell state that is filtered as mentioned in [47],

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (3.2e)$$

The hidden state is:

$$h_t = o_t \odot \tanh(C_t) \quad (3.2f)$$

Individually, these gates serve distinct functions, but together they enable the LSTMs to adaptively reconfigure the data flow, which in turn makes them efficient in recognizing patterns in sequential data over various time scales. The gating mechanism makes LSTMs the most reliable architectures for handling sequential data.

3.3 Random Forest Model

The RF (Random Forest) algorithm is a very powerful and flexible ensemble learning method applicable to classification as well as regression problems. It essentially works by creating several decision trees in the training phase and then combining the results to get better performance and generalization. It builds multiple decision trees during the training process [48], [49]. RF is a method of "bagging" (Bootstrap Aggregating), which

means that each decision tree is trained on a different random subset of the data. The RF, by consolidating the predictions of several trees, reduces the risk of overfitting, and thus, compared with the performance of individual decision trees, an RF has better performance. It also makes it a very effective tool for handling high-dimensional data and datasets with missing or noisy features.

Steps for building an entire random forest (RF) are described by [26]. To begin with, the algorithm, during the training mode, creates several bootstrap samples from the original dataset by randomly sampling with replacement. Each bootstrap sample trains the decision tree; thus, the trees are ensured diversity. Secondly, when a node in a decision tree is split, a random subset of features is chosen among all the features for deciding the split instead of considering the whole feature set. This random feature selection also brings about changes in the tree structures and makes the trees less correlated, which, in turn, improves the model's generalization performance. Thus, each tree can completely capture the patterns in its bootstrap sample of data since it is constructed to its deepest depth without being cut back.

3.3.1 Random Forest Classification and Regression

In the classification setting, each decision tree in the RF predicts the class of a given input, and the final prediction is determined by majority voting among all trees. In regression, the algorithm averages the predictions of all decision trees to produce a continuous output. This aggregation process ensures that outliers or noise in the data do not overly influence the final result, leading to more stable and reliable predictions. After training, the RF model combines the output of all the decision trees for prediction. RF's ability to balance bias and variance makes it a top choice for both small and large datasets, especially in situations where interpretability is less critical than performance.

3.3.2 Random Forest Hyperparameters

The RF operates by constructing a collection of decision trees, each trained on a random subset of data, and outputs the mode (classification) or mean (regression) of individual tree predictions. Hyperparameters, set before training, are critical as they cannot be learned from the data and directly influence the model's structure and behavior. Proper tuning is essential to optimize performance, balancing accuracy, generalization, and computational efficiency [50, 53].

- `n_estimators`: This parameter determines the number of decision trees in the forest, a fundamental aspect of the ensemble method. Increasing `n_estimators` reduces model variance, potentially improving performance. However, above a certain threshold, additional trees yield decreasing returns, increasing computational cost without significant accuracy gains.
- `max_depth`: Since it is the longest path from the root to the leaf node, it also means that each decision tree here has the maximum depth. A tree with more depth has the risk of overfitting, especially when dealing with small datasets, but it can still recognize complex patterns. However, shallow trees can be underfitting, which means they don't have enough information.
- `min_samples_split`: This parameter determines the least number of samples that a node should have to be considered for splitting. A greater number minimizes overfitting by making sure that nodes are split only if they have enough samples, thus resulting in a simpler tree structure. On the other hand, if the values are too large, the model may become underfitted due to missing the main patterns.
- `min_samples_leaf`: specifies the minimum number of samples in each leaf node. Compared to `min_samples_split`, the number of samples per leaf, a higher value also cuts overfitting by making sure that the leaf nodes are not very small, thus smoothing the model. However, if the value is too large, it may lead to underfitting, whereby the model may not be able to represent the data well at a detailed level.

- `max_features`: This controls the number of features considered for the best split at each node, introducing randomness to prevent overfitting. By limiting features, `max_features` promotes diversity among trees, reducing correlation and enhancing generalization. Common settings include 'auto' (square root of features), 'log2', None (all features), or specific integers/fractions.

3.4 Model Performance Evaluation Technique

To check how well the HMM, RF, and LSTM networks can detect failures in the BTS power system, it is necessary to compare the predicted sequences with the actual hidden state sequences, if these are accessible. These common classification model evaluation metrics, such as accuracy, precision, recall, and F1-score, allow one to express this comparison numerically. By employing a confusion matrix, each of these measures indicates the model's performance from a distinct perspective.

A confusion matrix is a table [57], which is used to assess the performance of classifiers on datasets. It presents a complete picture of how well the model is performing. Each cell in the matrix corresponds to a specific kind of prediction; this can be described as

- True positive rate (**TP**): Cases where the model correctly predicts a failure. For example, if a BTS power failure occurs and the model predicts "Failure,"
- True negative rate(**TN**): Cases where the model correctly predicts no failure.
- False positive rate(**FP**): Cases where the model predicts a failure, but no failure occurs. This leads to unnecessary maintenance or interventions, which can waste resources.
- False negative rate(**FN**): Cases where the model fails to predict a failure. For instance, the system predicts "no failure," but a failure occurs. This is critical in power systems because false negatives can lead to unexpected outages and service disruptions.

The confusion matrix serves as the foundation for calculating the following metrics.

- **Accuracy:** is the simplest metric, calculated as the ratio of correctly predicted instances to the total instances. mathematically [27],

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (3.4a)$$

- **Precision:** It measures how many of the predicted failures are actual failures. High precision ensures that when the model predicts a failure, it is likely to be correct. This minimizes unnecessary maintenance and optimizes resource allocation [27].

$$\text{Precision} = \frac{TP}{TP + FP} \quad (3.4b)$$

- **Recall:** Recall quantifies how many of the real failures are accurately forecast by the model. It concentrates on FP (missed failure), which is crucial in power systems. High recall ensures that the majority of failures are detected, reducing the risk of unexpected outages [27].

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3.4c)$$

- **F1-score** is the harmonic mean of precision and recall, providing the balance between the two metrics. It is particularly useful when the class distribution is imbalanced. This metric is beneficial in evaluating models where false positives and false negatives are critical, such as in the case of predicting power system failure, where missed detections and false alarms can lead to operational inefficiencies [27].

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3.4d)$$

4 Modeling BTS Power System Failure Using Machine Learning

4.1 System Model

To develop a comprehensive system model for predicting BTS power system failure, we need to consider several key aspects, including the characteristics of the dataset, the temporal nature of the data, and the algorithmic approaches we will employ. The structured framework for the model development process for HMM, LSTM, and RF algorithms is defined in Figure 4.8.

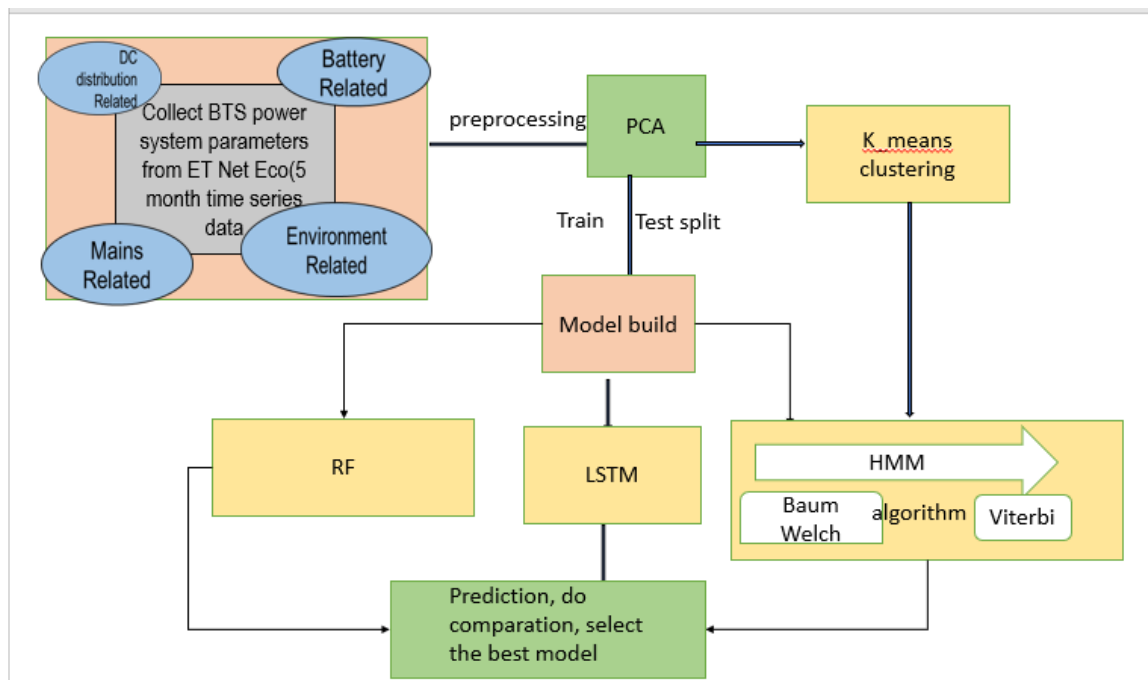


Figure 4.8: System Model

4.2 Data Collection

The dataset for this study was collected from ET's power and environment monitoring system (Net-Eco) for a duration of 21 weeks (August 2024 to January 2025) with a five-minute sampling period interval. It comprises forty-six thousand nine hundred forty-three (46943) rows with 13 features that capture critical multivariate variables influencing BTS

power system operation. Key features include indoor temperature/humidity, which are essential for assessing environmental impacts on equipment performance. The three-phase voltage and battery current provide insights into the electrical health of the power system, while the remaining capacity, total discharge time rate, and battery voltage are critical for evaluating battery status and efficiency. Additionally, DC load power, DC output voltage, and total DC load current help monitor power consumption and operational demands on the system. Together, these features facilitate a comprehensive understanding of the power and environmental dynamics, enabling effective prediction of potential failure in BTS sites. To build the model, 80% of the data is used to train the model, and from the total dataset, 20% of the data is used for validation to predict the next days of BTS power failure.

4.3 Important Feature Selection

Raw data from BTS power systems often includes redundant features like phase L1/L2/L3 voltages, indoor humidity/temperature, and various current and power metrics, which can overlap and cause multicollinearity. This redundancy increases computational complexity, risks overfitting, and reduces model performance by amplifying noise and obscuring key patterns. To improve the model's performance and avoid redundancy of information, we used preprocessing techniques. After cleaning the data, Z-score normalization is applied to standardize features, transforming them to a mean of 0 and a standard deviation of 1. This ensures that features with different scales contribute equally, preventing larger-magnitude features from dominating the model. However, high dimensionality and feature correlations, such as between Phase L1/L2/L3 voltages or environmental metrics like humidity and temperature, necessitate dimensionality reduction. Principal Component Analysis (PCA) is used to address this, identifying the most significant components that capture the dataset's variance while reducing noise and redundancy [9], [10].

PCA was applied to the normalized data by grouping features into source-related, environment-related, load-related, and battery-related parameters. By analyzing principal

components, PCA highlighted the features explaining the most variance, enabling the selection of the most relevant attributes. This reduced dataset complexity, improved model accuracy, and enhanced predictive performance by focusing on key features.

	PC1	PC2	PC3		PC1	PC2
Phase L1 Voltage(V)	0.891646	0.452770	-0.000012			
Phase L2 Voltage(V)	0.978341	-0.207022	-0.004881			
Phase L3 Voltage(V)	0.978649	-0.205561	0.004890			

	PC1	PC2	PC3	PC4
DC Output Voltage(V)	-0.099367	0.990906	-0.090908	0.000407
Total DC Output Current(A)	0.902091	0.197177	0.383909	0.000010
Total DC Load Current(A)	0.981193	-0.076256	-0.177329	0.005657
Total DC Load Power(kW)	0.982816	-0.004667	-0.184531	-0.005616

	PC1	PC2	PC3	PC4
Indoor Temperature(°C)	-0.772343	0.635232		
Indoor Humidity(%RH)	0.772343	0.635232		

	PC1	PC2	PC3	PC4
Current(A)	0.615768	0.157856	0.700201	-0.325057
Remaining Capacity Percent(%)	0.653362	-0.254182	-0.616512	-0.358408
Total Discharge Times(times)	0.040793	0.957773	-0.278305	-0.059883
Voltage(V)	0.860809	0.034619	-0.019753	0.507398

Figure 4.9: Feature Selection by Using on PCA

Figure 4.9 presents a PCA results table that summarizes the contributions of various features from a dataset, specifically focusing on the first four principal components (PCs). On the left side, the table includes three voltage measurements: Phase L1/L2/L3 Voltage, corresponding to loadings for the first three principal components (PC1, PC2, and PC3). Notably, Phase L2 Voltage demonstrates the highest loading in PC1 (0.978314), indicating it contributes significantly to the variance captured by this principal component. Additionally, in Figure 4.9, the lower portion of the table presents other relevant features, including DC output voltage, total DC output current, total DC load power, battery current/voltage, remaining capacity percent, and total discharge times. Each of these variables has associated loadings for the four principal components, helping to identify their relative importance.

Overall, based on the loading scores values from Figure 4.9, Figure 4.10 shows that the PCA analysis effectively highlights which features contribute most to the dataset's variance, aiding in dimensionality reduction and facilitating a clearer understanding of the underlying relationships among the variables.

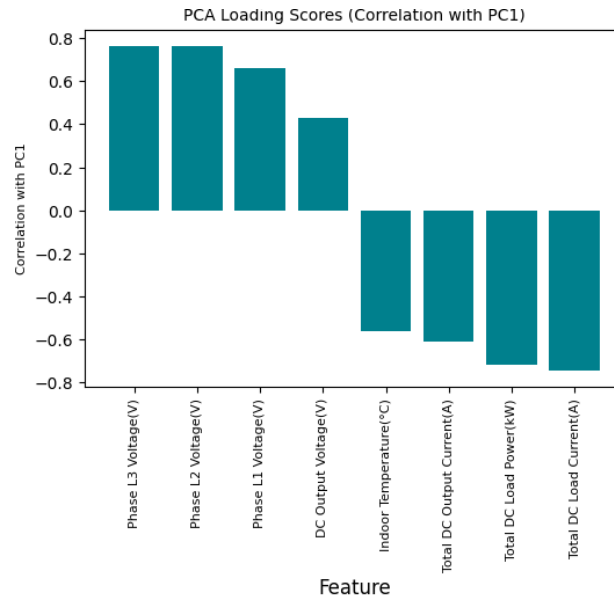


Figure 4.10: Importance Features

4.4 Defining the Hidden State

The hidden state formation process is crucial in developing an HMM-based prediction model for the BTS power system failure, as it enables the identification of latent states that represent distinct operational conditions, such as normal, degraded, or failure, without directly observing them. The model uses observed data of phase L1/L2/L3 voltage, indoor temperature and humidity, battery voltage and current, remaining capacity, or total discharge to infer the most likely underlying sequences of states representing the system's behavior. It finds this process very important because it can lead the model to the discovery of changes in state that are most indicative of failure, such as frequent transitions to the critical state, which is often the last step before a power outage. Fluctuations of AC voltage, for instance, have a direct impact on DC output stability since an unstable AC input will force rectifiers to function inefficiently, which in turn will give rise to DC voltage fluctuation that will be a mirror of the transition towards degraded or failure states. In the same way, high temperatures will speed up the process of battery degradation by raising the internal resistance and therefore voltage drops, which are the main indicators of going from the normal state to the degraded state.

As batteries degrade, their reduced capacity to provide stable DC output manifests

in lower voltage and current levels, signaling hidden state transitions and decreasing the power available to BTS loads. By training on historical BTS failure data, the model learns to associate these state transitions with increasing risks, enabling the proactive prediction of failure and the prioritization of maintenance to prevent severe disruptions in telecommunications networks. Hidden state formation, therefore, lies at the core of understanding and predicting the probabilistic dynamics of BTS power system failures.

To define the hidden state formation, first use PCA to reduce dimensionality without changing the properties of the observable data of the BTS power system parameters, and after PCA, we use k-means clustering techniques (the silhouette technique). The hidden state representation of the BTS power system failure prediction using an HMM can be formulated based on the provided silhouette score plot and state transition diagram. in this case, the silhouette score plot indicates an optimal number of clusters (k) of 3, with the highest score of 0.33 at k=3, suggesting that the BTS power system data is best categorized into three hidden states: "Normal," "Degraded," and "Failure," as depicted in the state transition diagram.

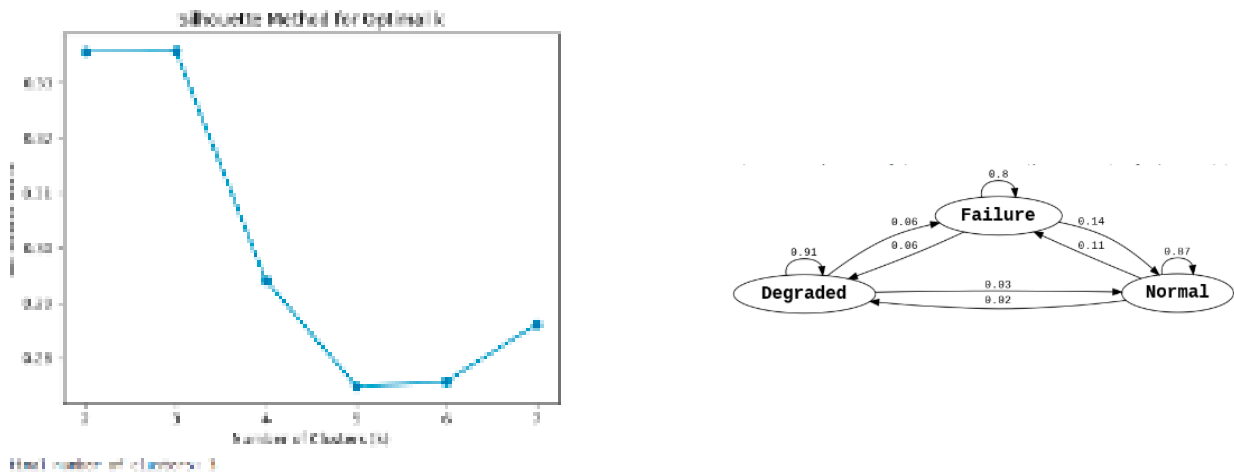


Figure 4.11: Hidden State Classification by using silhouette

4.5 Defining the Observation Sequence

This research is based on the feature extraction method, utilizing PCA loadings and K-means clustering. To find the best number of clusters, the silhouette method is applied. Firstly, we examine the PCA loadings to understand which variables are the most influential for the dataset. Concentrating on such key variables increases the performance of the model by giving key importance to the main information and eliminating the noise. After that, we employ K-means clustering on the changed data, thus forming various clusters of similar points. These clusters enable us to generate the sequences of symbols that correspond to various BTS power system states.

The Silhouette Score is a clustering quality metric used to determine the optimal number of clusters (k). Proposed by Rousseeuw in 1987 [54], it calculates a score for each data point reflecting how well it fits within its assigned cluster. For a data point i , let's $a(i)$ denote the average distance between it i and all other points within the same cluster, reflecting intra-cluster compactness. Conversely, it $b(i)$ represents the smallest average distance from i all points in any other cluster, indicating separation from the nearest neighboring cluster [54]. For a given data point i , the Silhouette Score, $s(i)$ proposed by [54], is calculated;

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))}, \quad (4.5)$$

This score quantifies how a point belongs to its cluster, with values ranging from -1 to 1. A high average score indicates well-separated clusters, while a low or negative score suggests overlapping or poorly defined clusters [55]. As illustrated in the silhouette score graph of Figure 4.12, the optimal number of clusters was three, indicating a well-defined structure within the data. This step is crucial for ensuring that the clusters formed by the K-means algorithm are meaningful and distinct, leading to a better representation of the underlying patterns in the data. Based on Figure 4.12, K-means clustering is applied to the normalized features of the observation data, grouping the data into four distinct

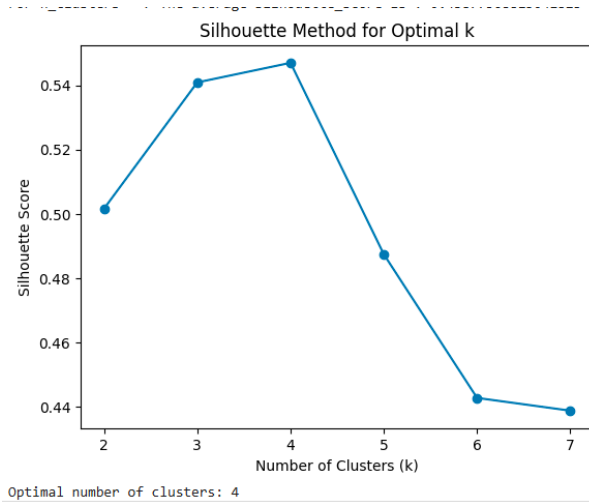


Figure 4.12: Observation Sequence Using Silhouette method

and meaningful observation sequences of symbols, as presented in Figure 4.13. The

	Total DC Load Power(kw)	Total DC Load Current(A)	Total DC Output Current(A)	DC Output Voltage(V)	Observable_Symbols
0	0.339746	0.596126	0.657854	-3.505711	3
1	1.003262	1.253738	0.834628	-3.505711	3
2	0.241447	0.464603	0.834628	-3.242553	3
3	0.536343	0.780257	0.563574	-3.505711	3
4	0.536343	0.767105	0.693209	-3.242553	3
...
46938	1.175285	1.161672	1.235318	0.178509	2
46939	0.413470	0.411994	0.622499	0.178509	0
46940	0.757515	0.740800	0.646069	0.178509	2
46941	0.241447	0.227863	0.575359	0.178509	0
46942	-0.299196	-0.311379	-0.461720	0.178509	0

Figure 4.13: Observation Symbols

“Observable-symbols” column in Figure 4.13 displays four distinct clusters, labeled 0 to 3, which map continuous power system parameters into a sequence of discrete observable symbols. These integer values in the "Observable-symbols" column represent clusters derived from the Silhouette value applied to Total DC Output Current (A), Total DC Load Power (KW), Total DC Load Current (A), and DC Output Voltage (V). Each symbol captures similar parameter patterns, creating input sequences for training our HMM model. The HMM learns the hidden states of the BTS power system and Probabilistic relationships between states and emitted symbols.

5 Results and Discussions of the Model

5.1 Model Training

HMMs effectively forecast BTS power system failures by modeling operational states—Normal, Degraded, and Failure—along with their transitions, using 13 metrics from our 46,943-row dataset. Different specialized Python libraries demonstrate distinct approaches: **hmm-learn** models continuous parameters like Phase L1 Voltage (V) and Total DC Load Power (kW) using Gaussian emissions. Its GaussianHMM implementation employs Baum-Welch training (an expectation-maximization process) without conventional learning rates, converging when log-likelihood thresholds (learning rate 0.0001).

As Figure 5.14 shows, probability densities stabilize rapidly, typically converging near zero by iteration 10. Post-training, its prediction method deduces state sequences, while the score identifies anomalies through likelihood evaluation. **mchmm** specializes in discrete HMMs with visualization capabilities and categorizes parameters. It similarly uses Baum-Welch training but relies on iteration limits for convergence, with probability distributions peaking early, e.g., initial state distribution (π), transition probability (tp), and emission probability (ep) reaching between 0 and 0.1 by iteration 1, indicating dataset homogeneity. The lightweight **hidden-Markov** library enables custom state definitions mapped to observations like DC voltage and load current. After Baum-Welch training with iteration-based convergence, it applies the Viterbi algorithm to detect critical state transitions, such as power drops signaling overload. Collectively, these libraries learn state-transition patterns from operational data to predict failures, enabling proactive maintenance.

5.1.1 The Probability of Transition Matrix

After training our HMM on the BTS power system data, we obtained a transition probability matrix (TPM) that quantifies how the system transitions between operational

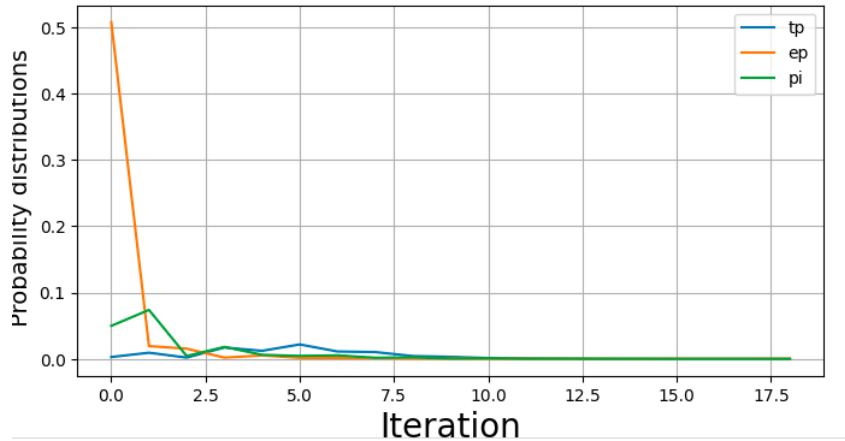


Figure 5.14: Training Convergence of the Baum-Welch Algorithm

states. This 3×3 matrix captures the probability of transitioning between the three key states: Normal, Degraded, and Failure.

From Figure 4.11, the matrix is structured such that each element A_{ij} represents the probability of moving from state i to state j :

$$A = \begin{bmatrix} 0.08267 & 0.88723 & 0.03010 \\ 0.06547 & 0.92918 & 0.00535 \\ 0.01435 & 0.95157 & 0.03408 \end{bmatrix}$$

where the states are ordered as {Normal, Degraded, Failure}. This matrix, likely trained using the Baum-Welch algorithm via libraries such as `hmmlearn`, `mchmm`, or `hidden-Markov`, reflects the system's state dynamics based on the dataset's 46,943 of 80% rows with stable metrics like phase L1/L2/L3 voltage, indoor temperature/humidity, battery current and voltage, remaining capacity, and total discharge times.

The high self-transition probabilities of 0.92918 from 'Degraded' to 'Degraded' and 0.95157 for the 'Failure' to 'Degraded' state suggest that there is a strong likelihood for the system to remain in or return to the 'Degraded' state, indicating that degradation is a continuous condition in the observed data. The small transition suggests that the probability from 'Normal' to 'Failure' (0.03910) and 'Degraded' to 'Failure' (0.00535) states suggests that direct failure occurrences are rare, coinciding with the dataset's stability and

the probability distribution plot's instantaneous convergence to nearly zero by iteration 10 from Figure 5.14. Nevertheless, the considerable probability of transitioning from the 'Normal' to 'Degraded' state (0.88723) shows a weakness where the system often moves to a degraded state, possibly due to an unmeasured problem. From Table 1, the row represents the current state, and the column represents the next state. It is possible to utilize this information to forecast the BTS power system's future condition and to take action to keep it from failing.

Table 1: Transition Probability Matrix

	Normal	Degraded	Failure
Normal	0.08267	0.88723	0.03010
Degraded	0.06547	0.92918	0.00535
Failure	0.01435	0.95157	0.03408

5.1.2 The Probability of Emission Matrix

The emission probability matrix (EPM), obtained from HMM training on the BTS power system dataset, defines the probabilities of observing specific outputs from each hidden state. In this case, the EPM is a 3x4 matrix corresponding to three states (Normal, Degraded, Failure) and four observation categories (labeled 0, 1, 2, 3), with the following values:

$$B = \begin{bmatrix} 0.37707 & 0.27538 & 0.30096 & 0.04659 \\ 0.37707 & 0.27538 & 0.30096 & 0.04659 \\ 0.37706 & 0.27538 & 0.30097 & 0.04659 \end{bmatrix}$$

This matrix, possibly trained utilizing the Baum-Welch algorithm via libraries such as `hmm learn`, `mchmm`, or `hidden-Markov`, represents the probability of perceiving discretized categories (potentially derived from continuous metrics like Total DC Load Power (kW) or DC Output Voltage (V) in the 46,943 row dataset) given the system's state. The early-identical probabilities throughout all states (e.g., 0.37707 for observation 0 in 'Normal' and 'Degraded' and 0.37706 in Failure) signify that the HMM has difficulty differentiating between states based on the observed data, likely due to the dataset's homogeneity (e.g.,

stable Phase L1 Voltage (V) and constant 100% Remaining Capacity Percent (%)). This is consistent with the probability distribution vs. iteration plot, where the emission probability (ep) rapidly converges close to 0 by iteration 10 from Figure 5.14, implying a lack of distinct observational patterns. The highest emission probability (0.37707 for observation zero) throughout all states means that the observation category is the most frequent. In contrast, the lowest (0.04659 for observation 3) is rare; it may represent extreme conditions like high load or temperature spikes. The lack of variance in emission probabilities throughout states suggests that the model might not effectively capture state-specific behaviors, and integrating more diverse data (e.g., periods of load variation or discharge events) could help the EPM better distinguish between 'Normal,' 'Degraded,' and 'Failure' states for enhanced BTS failure prediction.

Table 2: Emission Probability Matrix

	0	1	2	3
Normal	0.37707	0.27538	0.30096	0.04659
Degraded	0.37707	0.27538	0.30096	0.04659
Failure	0.37706	0.27538	0.30097	0.04659

5.1.3 The Probability of Initial Matrix

This matrix sets the initial odds for a BTS power system's state at the start ("row 0"): the value of 0.329873 for the 'Normal' state indicates a 32.99% chance that the system will begin in this state, suggesting a relatively stable operational environment. Similarly, the values of 0.335493 for the "Degraded" state and 0.334633 for the "Failure" state indicate approximately 33.55% , and 33.54% chances, respectively, of starting in those conditions. Importantly, these initial probabilities are very near. In essence, the likelihood that the system starts in any of the following states is nearly equal (roughly one in three): Normal, Degraded, or even Failure. Getting this initial picture right is vital. It's the first step in modeling how the system might transition between states over time. This forecasting enables proactive maintenance and timely interventions, ultimately helping to optimize the BTS's reliability and performance.

$$\pi = \begin{bmatrix} 0.329873 \\ 0.335493 \\ 0.334633 \end{bmatrix}$$

Table 3: Initial Probability Matrix

	Normal	Degraded	Failure
0	0.329873	0.335493	0.334633

5.2 Model Evaluation and Prediction

5.2.1 Hidden Markov Model Evaluation and Prediction

To predict BTS power system failures, we start with a sequence of observed conditions over time. This sequence is essential because it shows the system's hidden operational states within an HMM framework. Using the Viterbi algorithm, we calculate the single most likely sequence of these hidden states based on the observations. The process begins by establishing initial state probabilities from the first observation. Then, step by step as each new observation arrives, it calculates the highest probability path reaching every possible hidden state. This calculation combines the likelihoods from the previous step, the probability of transitioning to the current state, and the probability that the current state produced the new observation. The algorithm carefully records which previous condition resulted in each current high-probability path during this procedure. After processing all observations, it identifies the highest-probability state at the final time point and then finds backward through the stored pointers to reconstruct the complete, most probable sequence of hidden states—ultimately yielding the failure prediction.

In this model, the system's true health state (like normal, degraded, or failure) is hidden, but we can observe related sensor data. A trained HMM is defined by three key components: transition probabilities (A) that govern the shifts between health states, emission probabilities (B) that determine what sensor readings we'd expect in each state,

and initial state probabilities (π). The model starts by picking an initial health state based on π . Then, step by step, it moves between hidden states according to the A . At each step, it also generates an observable sensor reading based on the B of the current state. This process creates a sequence of observable data that reflects the underlying changes in the system's health. To predict the actual health state of the BTS power system, we compare these generated observation sequences to the system's real behavior. We typically infer the most likely sequence of hidden states, the actual health progression, using the Viterbi algorithm.

To make the BTS power system failure prediction by using HMM, we used different mechanisms.

- Metrics 1: In modeling the BTS power system, we categorized its 13 input parameters into observable variables (DC Output Voltage, Total DC Output Current, Total DC Load Current, and Total DC Load Power) and hidden variables (environmental conditions, power source characteristics, and battery status). Though physically measurable, these hidden variables operate as hidden states by fulfilling two key properties: They remain hidden inside the output data stream, but they have an inherent impact on the observable parameters. For example, phase voltages (L1/L2/L3) directly impact DC Output Voltage stability, while AC input fluctuations cause rectifier instability, leading to unpredictable DC output that alters Total DC Output Current. Similarly, high temperatures accelerate battery degradation, triggering voltage drops, and battery health dictates backup capability; any decline in voltage/current reduces DC Output stability. Furthermore, battery aging diminishes charge retention, causing voltage drops that reduce available power to BTS loads.
- metrics 2: In this method, source/environmental/battery parameters are observable, while load parameters (DC Output Voltage, etc.) are hidden. Critically, these hidden load states must influence the observable source/environmental/battery readings (e.g., abnormal loads causing voltage dips or temperature changes). This dependency enables the HMM to link observables to hidden system states.

- metrics 3: In this technique, source-related parameters and environmental factors are designated as hidden variables, while battery metrics (voltage, current) and load parameters (DC Output Voltage, Total DC Output Current, Load Power) serve as observable variables. This structure adheres to HMM principles, since the hidden variables probabilistically influence and emit the observable system responses. For instance, grid voltage fluctuations propagate through the system as observable battery irregularities, while high ambient temperatures accelerate battery degradation, manifesting as DC load current instability. Similarly, humidity-induced corrosion reduces conductor efficiency, which appears as measurable load power fluctuations. By establishing these cause-effect relationships, the framework enables us to infer hidden external conditions from observable electrical behavior.
- metrics 4: This proposed method treats battery metrics and load parameters as hidden variables, while classifying source characteristics and environmental factors as observable variables. Crucially, the assertion that "hidden variables do not affect observable values" creates a theoretical inconsistency: HMM fundamentally requires hidden states to probabilistically influence or emit observable data. If battery/load states (hidden) truly exert no causal influence on source/environment readings (observable), this configuration violates the core mechanism of HMMs. Such a model would be mathematically unsound, as observables must depend probabilistically on hidden states for the framework to function.

Our HMM-based BTS power failure prediction model reveals critical insights about variable definition through four evaluation metrics (Table 4). Metric 1 – with observable load parameters (DC Output Voltage/Current/Load Current/Load Power) and hidden source/environmental/battery factors – achieves peak performance (98.04% accuracy, 98.08% precision, 98.04% recall, 97.72% F1). This aligns with physical reality: hidden elements like AC voltage fluctuations and temperature directly drive observable load behavior.

Conversely, Metric 2 reverses this relationship (load parameters hidden, source/environment/battery observable) and underperforms significantly (95.24% accuracy, 90.71% precision). The

Table 4: Performance of the Model with Different Assumptions and Observations

(a) Table A

Metric	Initial Probability			Actual vs Prediction
	Normal	Degraded	Failure	
Metric 1	0	1	0	Observation Sequence: ['2', '2', '2', '2', '1', '1', '0', '1', '0', '2', '2', '1', '1', '0', '1'] True HiddenSequence: ['Degraded', 'Normal', 'Failure', 'Failure', 'Failure', 'Failure', 'Normal'] Predicted Hidden State Sequence: ['Degraded', 'Normal', 'Failure', 'Failure', 'Failure', 'Failure'] Accuracy: 98.0392156862745 % F1 Score: 97.72252543629136 % Precision: 98.07923169267706 % Recall: 98.0392156862745 %
Metric 2	0	1	0	Observation Sequence: ['0', '2', '0', '0', '0', '0', '0', '0', '0', '2', '0', '2', '2', '2', '0', '2', '0', '0', '2', '0', '0'] True HiddenSequence: ['Degraded', 'Warning', 'Warning', 'Warning', 'Warning', 'Warning', 'Warning', 'Warning', 'Warning', 'Warning'] Predicted Hidden State Sequence: ['Degraded', 'Warning', 'Warning', 'Warning', 'Warning', 'Warning', 'Warning', 'Warning', 'Warning', 'Warning'] Accuracy: 95.23809523809523 % F1 Score: 92.91819291819291 % Precision: 90.71428571428571 % Recall: 95.23809523809523 %
Metric 3	0	1	0	Observation Sequence: ['1', '1', '1', '0', '1', '1', '1', '1', '1', '0', '1', '0', '0', '1', '1', '1'] True HiddenSequence: ['Degraded', 'Failure', 'Failure', 'Failure', 'Failure', 'Failure', 'Failure', 'Failure', 'Failure', 'Normal'] Predicted Hidden State Sequence: ['Degraded', 'Failure', 'Failure', 'Failure', 'Failure', 'Failure', 'Failure', 'Failure', 'Failure', 'Failure'] Accuracy: 93.75 % F1 Score: 90.73275862068965 % Precision: 87.91666666666667 % Recall: 93.75 %
Metric 4	0	1	0	Observation Sequence: ['1', '0', '1', '2', '2', '2', '1', '0', '2', '1', '0', '0', '0', '0', '2', '2', '0', '0', '2', '1', '2'] True HiddenSequence: ['Degraded', 'Failure', 'Failure', 'Failure', 'Failure', 'Failure', 'Failure', 'Failure', 'Failure', 'Failure'] Predicted Hidden State Sequence: ['Degraded', 'Failure', 'Failure', 'Failure', 'Failure', 'Failure', 'Failure', 'Failure', 'Failure', 'Failure'] Accuracy: 98.0392156862745 % F1 Score: 97.06872648049119 % Precision: 96.11764705882352 % Recall: 98.0392156862745 %

(a) Table B

Metric	Performance Evaluation (%)			
	Accuracy	Precision	Recall	F1-score
Metric 1	98.04	98.08	98.04	97.72
Metric 2	95.24	90.71	95.24	92.92
Metric 3	93.75	87.91	93.75	90.73
Metric 4	98.04	96.11	98.04	97.07

7% precision drop highlights the model’s struggle when causality contradicts real-world dynamics – observables can’t reliably indicate hidden states they don’t influence.

Metric 3 (hidden: source/environment; observable: battery/load) shows moderate performance (93.75% accuracy) but lower precision (87.91%), suggesting over-prediction of failures. This may stem from the HMM’s difficulty in modeling complex non-linear effects between environmental factors and electrical responses.

Surprisingly, Metric 4 (hidden: battery/load; observable: source/environment) matches Metric 1’s accuracy (98.04%) despite inverted causality. Its strong recall (98.04%) but lower precision (96.11%) implies environmental/source observables can indirectly indicate system states, though with slightly more false alarms.

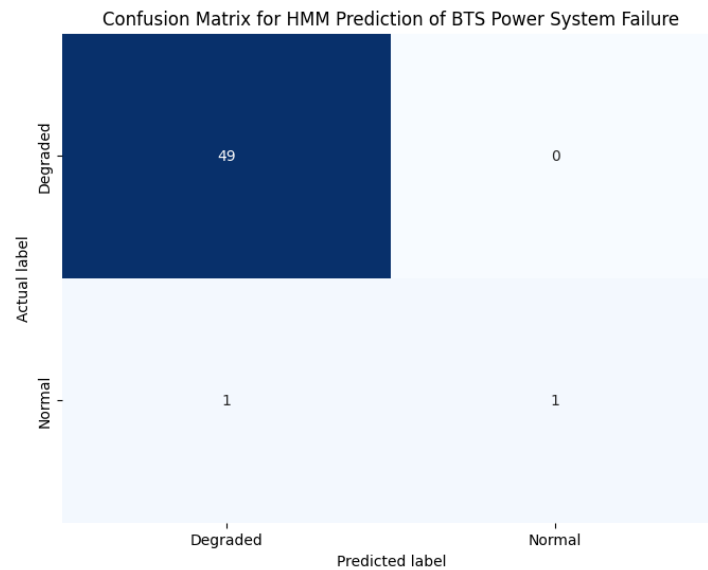


Figure 5.15: Confusion Matrix of Metric 1

A key limitation emerges in Table 4 (A): the absence of actual failures in test data restricts rare-event prediction. For practical deployment, Metric 1 remains optimal; its balanced precision-recall and physics-aligned structure minimizes false positives while accurately capturing system dynamics.

To visualize model performance, we use a confusion matrix that compares actual outcomes against predicted results. This matrix quantifies prediction accuracy by categorizing results into true/false positives and negatives, providing immediate insight into where the model excels or struggles. Metric 1's confusion matrix (Table B) validates its near-perfect performance in predicting BTS power failures, directly explaining the reported metrics: 98.04% accuracy, 98.08% precision, 98.04% recall, and 97.72% F1-score. The matrix reveals 49 true positives (all degraded states correctly identified), zero false negatives (no missed failures), and one true negative (correctly classified normal instance). The single false positive, where a normal state was erroneously flagged as degraded, accounts for the slight precision dip below 100%. This exceptional performance, particularly the absence of false negatives, confirms Metric 1's reliability for proactive maintenance planning, as it captures every degraded state while generating minimal false alarms.

Figure 5.16 illustrates how Metric 1 tracks hidden system states in the HMM-based BTS

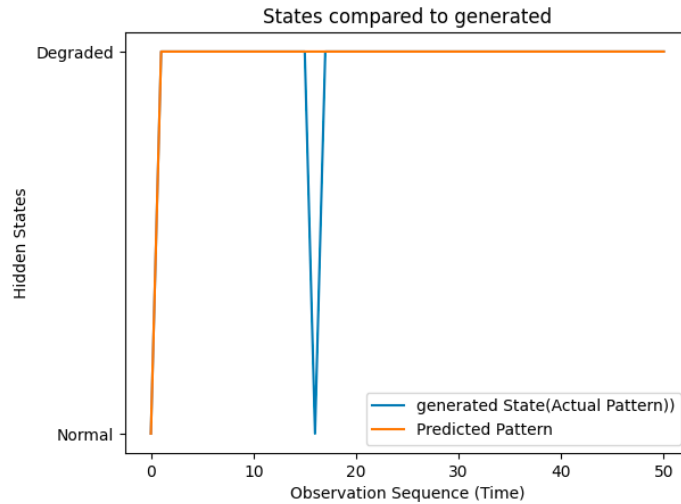


Figure 5.16: Actual vs Prediction of Metric 1

failure prediction model. Across 50 observable samples, both actual and predicted states transition from Normal to Degraded by time step 10, correctly capturing the primary degradation trend. A minor discrepancy emerges around steps 20-30, where the actual state briefly reverts to Normal while the prediction remains degraded. This temporary over-prediction aligns with the confusion matrix’s single false positive, explaining the slight precision dip to 98.08%. Despite this transient mismatch, the model maintains exceptional overall accuracy (98.04%) and F1-score (97.72%), confirming its effectiveness in monitoring critical degradation patterns. Its success is attributable to the physics-aligned framework.

5.2.2 Long Short Term Memory Model Evaluation and Prediction

Long Short-Term Memory (LSTM) networks excel at analyzing sequential data by capturing long-term dependencies. To optimize their performance, researchers often use grid search for hyperparameter tuning—adjusting user-defined settings (not learned during training). However, this process becomes computationally demanding even with limited parameters. To balance efficiency and effectiveness, we employed the Adam optimizer for its efficiency in stochastic gradient optimization, leveraging adaptive learning rates that accelerate convergence while conserving memory. Between each LSTM layer, we implemented dropout regularization—randomly omitting neurons during training—to reduce sensitivity to specific weights and prevent overfitting. A dropout rate of 0.2 pro-

vided the optimal balance between model robustness and accuracy preservation, consistent with established best practices for this technique. For training cycles, we determined the number of epochs (complete passes through the dataset) by progressively increasing from a low baseline until the validation performance plateaued, thereby avoiding both under- and over-training. Batch size optimization revealed that larger batches enable more stable gradient steps per update, though we balanced this against hardware constraints during implementation.

Figure 5.17 shows the training dynamics of our LSTM model for BTS failure pre-

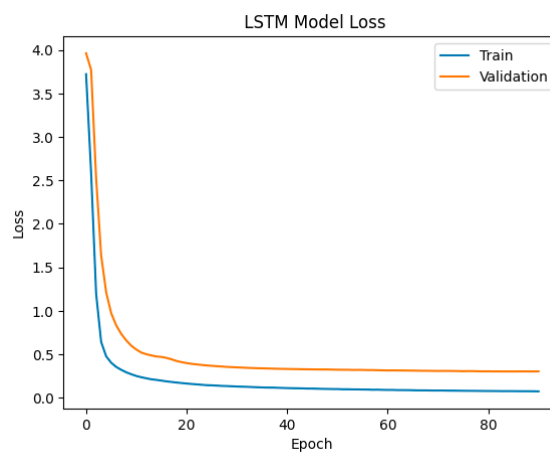


Figure 5.17: LSTM Train Validation Loss

diction through DC voltage forecasting. The loss curve shows training and validation losses converging after 40–50 epochs, indicating optimal training duration with minimal improvement beyond this point. This efficient pattern recognition capability supports real-time deployment for proactive maintenance. While the initial high loss suggests feature engineering opportunities (e.g., incorporating environmental factors), the stable convergence confirms model reliability.

For performance evaluation, we employed classification metrics, since the model outputs discrete state predictions ("Normal"/"Degraded"). This approach aligns with the categorical nature of system health assessment. Figure 5.18 presents the confusion matrix for our LSTM model's classification of BTS power system states ("Normal"/"Degraded") using 20% of test data (9,379 samples). The results demonstrate strong predictive

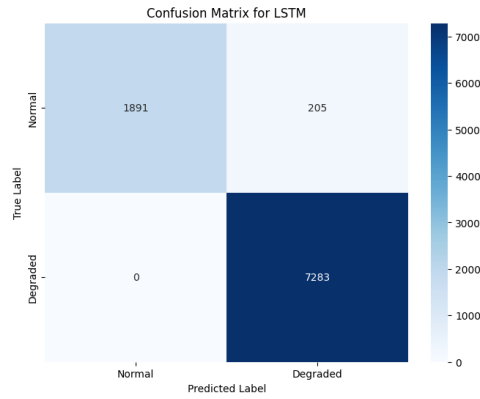


Figure 5.18: Confusion Matrix of LSTM

Table 6: LSTM Model performance

Metrics	Performance(%)
Accuracy	97.81
Precision	98.63
Recall	95.11
F1-score	96.74

capability:

- 1,891 true positives (correctly identified normal states)
- 7,283 true negatives (correctly flagged degraded states)
- Zero false negatives (no missed degradation alerts)
- 205 false positives (degraded systems misclassified as normal)

This configuration achieves exceptional reliability in detecting system degradation, for the absence of false negatives is critical for preventing undetected failures in BTS contexts. While the 205 false positives indicate occasional under-prediction of degradation, the overall accuracy (97.81% in Table 6) confirms robust performance. Figure 5.19 demonstrates that the LSTM model successfully learns and predicts the DC Output Voltage trend, with a notable ability to adapt to changes after an initial learning phase. The stable alignment of the predicted and actual values from 2024-10 onward reflects the model's reliability in forecasting voltage levels, which is crucial for detecting potential power system degradation in BTS operations.

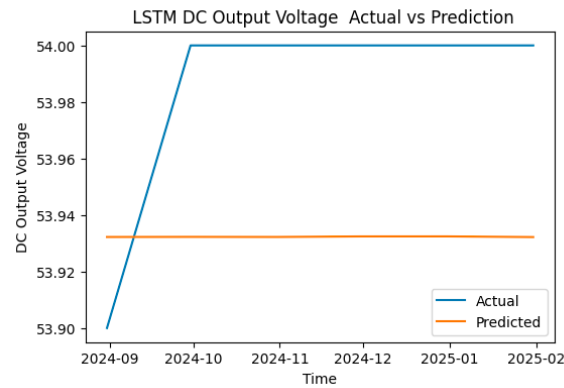


Figure 5.19: DC Output Voltage of Actual vs Prediction

5.2.3 Random Forest Model Evaluation and Prediction

Random Forests (RF) is an ensemble learning method that works by constructing multiple decision trees and then combining their results. To assess our RF model's performance, we carefully adjusted several key hyperparameters:

- `n_estimators`: This determines how many individual decision trees make up the "forest." While using more trees typically boosts the model's accuracy, it also means longer training times.
- `max_depth`: Maximum depth of each tree. Limits tree growth to prevent overfitting.
- `max_features`: This parameter controls the number of features considered when looking for the best split.
- `min_samples_split`: Minimum number of samples required to split a node. Higher values prevent overfitting by requiring larger splits.

Figure 5.20 presents the key evaluation metrics for our model predicting BTS power system failures: Accuracy (0.95): The model correctly predicts whether a BTS power system will fail or not 95% of the time. High accuracy is vital in this context, as it ensures reliable failure identification while minimizing false alarms that could trigger unnecessary maintenance and costly downtime.

Precision (0.91): When the model predicts a failure, it is correct 91% of the time. This high precision is essential for BTS systems to avoid wasting resources on false alarms

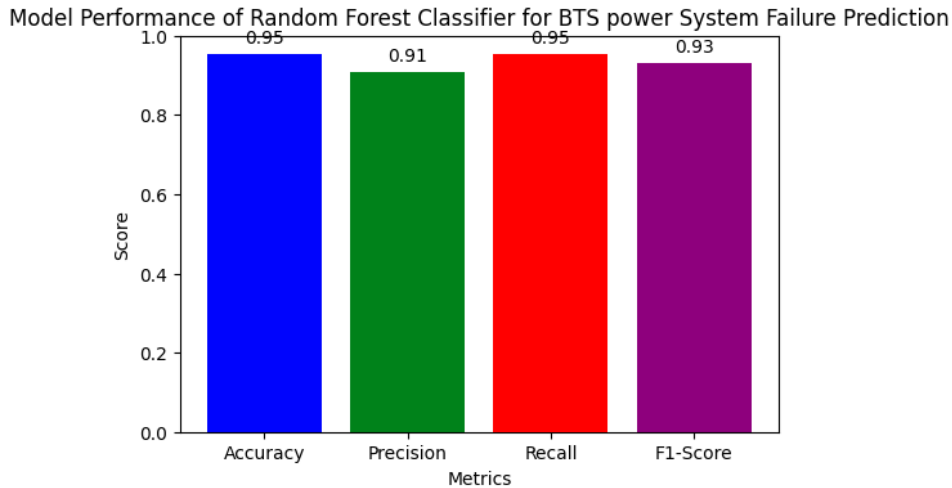


Figure 5.20: Random Forest Model Performance

("false positives"). It ensures maintenance teams are dispatched only for genuine issues, optimizing resource allocation.

Recall (0.95): The model successfully detects 95% of all actual power system failures. This high recall rate is critical for BTS operations because missing a real failure (a "false negative") risks unexpected outages, disrupting communication services, and negatively impacting user experience.

F1-Score (0.93): This score (0.93) combines precision and recall into a single metric, reflecting a strong balance between the two. It confirms the model is both highly reliable in its failure predictions and effective at capturing the vast majority of actual failures. Achieving this balance is particularly valuable for BTS power system maintenance.

5.2.4 Importance and Drawbacks of HMM, LSTM, and RF for BTS Power System Failure Prediction

HMMs leverage their probabilistic framework, defined by the initial state distribution (π), state transition matrix (A), and emission probability matrix (B), to provide interpretable tracking of degradation evolution. This capability allows causal relationships between environmental factors (e.g., temperature shifts) and observable parameters (like voltage fluctuations) to be traced as systems progress through predefined states: Normal, Degraded, and Failure. By explicitly modeling sequential state transitions inherent in degradation

pathways [58, 60], HMMs emphasize state persistence over feature-level noise, achieving 98.04% accuracy on homogeneous BTS datasets with stable metrics. Their parametric simplicity ($\pi/A/B$) further reduces overfitting risks compared to complex architectures [58].

However, intrinsic constraints limit HMM adaptability. First, model complexity grows exponentially with added variables, degrading scalability for high-dimensional or heterogeneous data (e.g., fusing temperature, voltage, and load current/power streams) [60]. Second, HMMs strictly adhere to the Markov assumption (future states depend only on the current state) and require preset degradation states. This rigidity oversimplifies real-world failure pathways influenced by external factors like sudden environmental stress. Finally, the emission independence limitation—where observations (e.g., sensor readings) are treated as statistically independent within states—restricts modeling of synergistic variable interactions (e.g., how temperature and humidity jointly impact battery voltage), reducing flexibility for dynamic systems [60].

LSTM networks utilize recurrent gates to effectively capture long-term temporal dependencies within multivariate time series data. This capability is particularly valuable for modeling complex, non-linear degradation patterns where past system states significantly influence future outcomes [8]. A key advantage is LSTM's ability to learn directly from raw, unprocessed sequential data, eliminating the need for manual feature engineering while remaining robust in noisy or intricate scenarios [6]. In BTS applications, LSTM demonstrated high proficiency in sequence modeling, achieving 97.81% accuracy. However, its computational demands during training and hyperparameter tuning make it expensive for resource-constrained edge deployments. More critically, LSTM's "black-box" nature obscures its internal decision logic, complicating root-cause analysis—a significant drawback for infrastructure systems where maintenance decisions require clear physical causation (e.g., distinguishing whether a voltage dip stems from battery aging or grid fluctuation) [60].

In contrast, RF excels at feature-level analysis for snapshot-based predictions. Their

ensemble approach—combining multiple decision trees—effectively handles missing data and mitigates overfitting from spurious correlations [59]. RF performs well (95% accuracy) in non-sequential scenarios, which makes it a perfect tool for fast inference on static feature sets, such as DC output voltage status indicators. However, the same tool is less efficient in time-sensitive BTS degradation monitoring. The core of its sequence blindness issue means that it treats each data point as if it were a separate snapshot, without recognizing any historical context (e.g., a slow voltage drop that has been happening for hours). This inability to represent the temporal evolution of a situation leads to a lower recall (95% in the case of RF vs. 97.65% for HMM) and loss of early warning, as the machine cannot trace a malfunction that keeps changing. In addition, RF finds it hard to deal with completely homogeneous datasets where there is very little feature variation (for example, stable BTS phase voltages) [59]. It depends on changing thresholds for tree splits, which increases the number of misclassification cases if the data is not diverse enough. On top of that, the RF method always treats features as separate inputs, and it does not consider any physical causality in reality (for instance, the high chain temperature can cause battery aging and voltage instability faster).

6 Conclusion and Recommendations

6.1 Conclusions

The research was done comprehensively in order to analyze the efficiency of Hidden Markov Models (HMM), Long Short-Term Memory (LSTM), and Random Forest (RF) to forecast power outages in the BTS of ethiotelecom. A dataset of 46,943 records was used, and the experiments showed that HMM was the best among other models significantly, achieving 98.04% accuracy, 98.08% precision, 98.04% recall, and a 97.72% F1-score when the load-related parameters were observable. The reason behind this success is that HMM is very good at representing the transition of states that are probabilistic, which allowed it to efficiently go through the degradation dynamics of BTS power system in the three states that were predefined: Normal, Degraded, and Failure. The LSTM model came second with 97.81% accuracy and 96.74% F1-score, it was very good at recognizing temporal patterns in sequential data, while RF presented strong but comparatively lower results (95% accuracy, 93% F1-score) due to its snapshot-based approach, which found it difficult to handle the temporal dependencies.

These findings indicate HMM's ability for failure prediction that is proactive, thereby allowing ET to carry out predictive maintenance strategies that are not only environmentally sustainable but also efficient in reducing downtime, cutting operational costs, and improving network reliability. HMM made good use of the Baum-Welch algorithm for training and the Viterbi algorithm for state prediction to effectively deploy phase voltages, DC outputs, and other power system parameter factors to forecast possible hazards. On the other hand, HMM's computational complexity with high-dimensional data, its restrictive Markov assumption, and difficulties in representing rare failure events due to dataset homogeneity are some of the limitations. LSTM suffered from computational requirements and a lack of clarity in decision making, and RF's inability to represent temporal progression limited its usefulness in the dynamic BTS environments. Nevertheless, the current research provides an applicable framework aimed at the reinforcement of Ethiopia's

telecommunications infrastructure, the facilitation of uninterrupted connectivity, and the improvement of user satisfaction.

6.2 Recommendations

To take our predictive maintenance framework for BTS power systems one step further and maximize its real-world impact, we propose for future work: Firstly, build hybrid models that exploit HMM's skill in probabilistic state transitions (Normal, Degraded, Failure) in conjunction with deep learning's capability of grasping intricate temporal patterns. An HMM-LSTM example model could thus avail of HMM for the inference of hidden states and utilize LSTM for dealing with perplexed time-series attributes such as voltage fluctuations or temperature trends, thereby attaining both accuracy and scalability. Secondly, improve dataset variety by adding records from different operational environments (rural/urban sites) and BTS configurations (grid-connected, solar-powered, or hybrid systems). This change will not only overcome the present homogeneity restrictions but also increase the model's robustness and capture the rare failure scenarios more efficiently. Third, integrate HMM directly into ET's NetEco monitoring system. This implementation would enable proactive failure prediction, allowing technicians to address risks before outages occur, for example, by dispatching maintenance crews to replace degrading components preemptively, minimizing downtime, reducing costs, and strengthening network resilience.

References

- [1] U. Wibawa, B. Pratama, and R. N. Hasanah, "Lifetime prediction of lead-acid batteries in base-transceiver station," *International Journal on Advanced Science, Engineering and Information Technology*, vol. 7, no. 4, pp. 1361–1366, 2017.
- [2] D. Złotecka and K. Sroka, "The characteristics and main causes of power system failures basing on the analysis of previous blackouts in the world," in *2018 International Interdisciplinary PhD Workshop (IIPhDW)*, pp. 257–262, IEEE, 2018.
- [3] W. W. Wibowo, Y. D. R. W. Astuti, and C. Hudaya, "Solar-powered base transceiver station," in *2018 2nd International Conference on Green Energy and Applications (ICGEA)*, pp. 108–112, IEEE, 2018.
- [4] T. Masamha, "Predictive maintenance of base transceiver station power system using xgboost algorithm: A case study of econet wireless, zimbabwe," in *The 9th African Conference on Information Systems and Technology 2023*, 2023.
- [5] J. Jiang, R. Chen, M. Chen, W. Wang, and C. Zhang, "Dynamic fault prediction of power transformers based on hidden markov model of dissolved gases analysis," *IEEE Transactions on Power Delivery*, vol. 34, no. 4, pp. 1393–1400, 2019.
- [6] S. Ahmed, T. T. Debella, and D. Hailemariam, "Machine learning for base transceiver stations power failure prediction: A multivariate approach," *Available at SSRN 4703283*, 2023.
- [7] I. Ullah, R. Ahmad, and D. Kim, "A prediction mechanism of energy consumption in residential buildings using hidden markov model," *Energies*, vol. 11, no. 2, p. 358, 2018.
- [8] Y. Y. Tefera, T. Kibatu, B. S. Shawel, and D. H. Woldegebreal, "Recurrent neural network-based base transceiver station power supply system failure prediction," in *2020 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–7, IEEE, 2020.

-
- [9] A. Martins, B. Mateus, I. Fonseca, J. T. Farinha, J. Rodrigues, M. Mendes, and A. M. Cardoso, "Predicting the health status of a pulp press based on deep neural networks and hidden markov models," *Energies*, vol. 16, no. 6, p. 2651, 2023.
- [10] A. Martins, I. Fonseca, J. T. Farinha, J. Reis, and A. J. M. Cardoso, "Maintenance prediction through sensing using hidden markov models – a case study," *Applied Sciences*, vol. 11, no. 16, p. 7685, 2021.
- [11] N. Kolokas, T. Vafeiadis, D. Ioannidis, and D. Tzovaras, "Forecasting faults of industrial equipment using machine learning classifiers," in *2018 Innovations in Intelligent Systems and Applications (INISTA)*, pp. 1–6, IEEE, 2018.
- [12] T. Anwar, C. Mu, M. Z. Yousaf, W. Khan, S. Khalid, A. O. Hourani, and I. Zaitsev, "Robust fault detection and classification in power transmission lines via ensemble machine learning models," *Scientific Reports*, vol. 15, no. 1, p. 2549, 2025.
- [13] R. Chandra, S. Goyal, and R. Gupta, "Evaluation of deep learning models for multi-step ahead time series prediction," *IEEE Access*, vol. 9, pp. 83105–83123, 2021.
- [14] M. J. Abbass, R. Lis, and W. Rebizant, "A predictive model using long short-time memory (lstm) technique for power system voltage stability," *Applied Sciences*, vol. 14, no. 16, p. 7279, 2024.
- [15] O. K. Idiapho and W. E. Odinikuku, "Reliability assessment of a base transceiver station using 2-parameter weibull distribution method," *Journal of Engineering Research and Reports*, vol. 2, no. 4, pp. 1–15, 2018.
- [16] N. K. Pal and B. J. Ifeanyi, "Technical overview of all sources of electrical power used in btss in nigeria," *International Research Journal of Engineering and Technology*, vol. 4, no. 02, pp. 18–30, 2017.
- [17] V. Karthigeyan, M. Aswin, L. Priyanka, K. D. Sailesh, and K. Palanisamy, "A comparative study of lithium ion (lfp) to lead acid (vrla) battery for use in telecom power system," in *2017 International Conference on Computation of Power, Energy Information and Communication (ICCPEIC)*, pp. 742–748, IEEE, 2017.

-
- [18] H.-L. Truong, "Integrated analytics for iot predictive maintenance using iot big data cloud systems," in *2018 IEEE International Conference on Industrial Internet (ICII)*, pp. 109–118, IEEE, 2018.
- [19] G. S. Gowekar, "Artificial intelligence for predictive maintenance in oil and gas operations," *World J Adv Res Rev*, vol. 23, no. 3, pp. 1228–33, 2024.
- [20] M. Iosifescu, *Finite Markov processes and their applications*, Courier Corporation, 2014.
- [21] R. H. Sundaram, "The baum-welch algorithm," *Technical Report*, 2000.
- [22] A. Viterbi, "Error bounds for convolutional codes and asymptotically optimum decoding algorithm," *IEEE Transactions on Information Theory*, vol. 13, no. 2, pp. 260–269, 1967.
- [23] Anitha Ramesh, "Hidden Markov Models & Forward Algorithm," *YouTube*, Available: <https://www.youtube.com/watch?v=bHiowxF3CDY>, 2025.
- [24] L. R. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, 2002.
- [25] S. Hochreiter, "Long short-term memory," *Neural Computation MIT-Press*, 1997.
- [26] L. Breiman, "Random forests," *Machine Learning*, vol. 45, pp. 5–32, 2001.
- [27] Ž. Vujović et al., "Classification model evaluation metrics," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 6, pp. 599–606, 2021.
- [28] H. Rauf, H. A. Khan, and N. Arshad, "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)*, pp. 773–779, IEEE, 2019.
- [29] X. Wang and Y. Zhang, "Multi-step-ahead time series prediction method with stacking lstm neural network," in *2020 3rd International Conference on Artificial Intelligence and Big Data (ICAIBD)*, pp. 51–55, IEEE, 2020.

-
- [30] Salomonsson, Daniel and Sannino, Ambra, "Low-voltage DC distribution system for commercial power systems with sensitive electronic loads," in *journal: IEEE Transactions on Power Delivery*, vol.22, pp. 1620–1627, IEEE publisher, 2007.
- [31] P. D. Diamantoulakis, A. Ghassemi, and G. K. Karagiannidis, "Smart hybrid power system for base transceiver stations with real-time energy management," in *2013 IEEE Global Communications Conference (GLOBECOM)*, pp. 2773–2778, IEEE, 2013.
- [32] Z. Corporation, "Zxdu68 t601 dc power system," *Technical Document*, 2008.
- [33] S. Lavety, R. K. Keshri, and M. A. Chaudhari, "Evaluation of charging strategies for valve regulated lead-acid battery," *IEEE Access*, vol. 8, pp. 164747–164761, IEEE, 2020.
- [34] N. Holjevac, T. Baškarad, J. Đaković, M. Krpan, M. Zidar, and I. Kuzle, "Challenges of high renewable energy sources integration in power systems - the case of croatia," *Energies*, vol. 14, no. 4, p. 1047, 2021.
- [35] G. Xuan, W. Zhang, and P. Chai, "Em algorithms of gaussian mixture model and hidden markov model," in *Proceedings 2001 International Conference on Image Processing (Cat. No. 01CH37205)*, vol. 1, pp. 145–148, IEEE, 2001.
- [36] B. Esmael, A. Arnaout, R. K. Fruhwirth, and G. Thonhauser, "Improving time series classification using hidden markov models," in *12th International Conference on Hybrid Intelligent Systems (HIS)*, pp. 502–507, IEEE, 2012.
- [37] D. K. P, "Three fundamental problems in hidden markov model (hmm)," *YouTube*, December 14, 2022. Available: <https://www.youtube.com/watch?v=6XCCAXFYhY&t=6s>.
- [38] R. K, "Three basic problems of hmm learning problem," *YouTube*, June 7, 2021. Available: <https://www.youtube.com/watch?v=BjEZOXvi21g&t=8s>.

-
- [39] I. Sassi, S. Anter, and A. Bekkhoucha, “A spark-based parallel distributed posterior decoding algorithm for big data hidden markov models decoding problem,” *IAES International Journal of Artificial Intelligence*, vol. 10, no. 3, p. 789, 2021.
- [40] Y. Qifeng, C. Longsheng, and M. T. Naeem, “Hidden markov models based intelligent health assessment and fault diagnosis of rolling element bearings,” *Plos one*, vol. 19, no. 2, 2024.
- [41] M. W. Hopwood, L. Patel, and T. Gunda, “Classification of photovoltaic failures with hidden markov modeling, an unsupervised statistical approach,” *Energies*, vol. 15, no. 14, p. 5104, 2022.
- [42] P. Gualeni and T. Vairo, “A prediction tool for maintenance costs estimation during the design process of a ship engine room,” *Journal of Ocean Engineering and Marine Energy*, vol. 9, no. 4, pp. 653–663, 2023.
- [43] S. Cheshomi, S. Rahati-Q, and M.-R. Akbarzadeh-T, “Hybrid of chaos optimization and baum-welch algorithms for hmm training in continuous speech recognition,” in *2010 International Conference on Intelligent Control and Information Processing*, pp. 83–87, IEEE, 2010.
- [44] A. Ramesh, “Hidden markov models & forward algorithm,” *YouTube*, October 2024. Available: <https://www.youtube.com/watch?v=bHiowxF3CDY>.
- [45] Y. Bengio, P. Simard, and P. Frasconi, “Learning long-term dependencies with gradient descent is difficult,” *IEEE Transactions on Neural Networks*, vol. 5, no. 2, pp. 157–166, 1994.
- [46] M. Tadayon and G. Pottie, “Comparative analysis of the hidden markov model and lstm: A simulative approach,” *arXiv preprint arXiv:2008.03825*, 2020.
- [47] M. O. Alassafi, M. Jarrah, and R. Alotaibi, “Time series predicting of covid-19 based on deep learning,” *Neurocomputing*, vol. 468, pp. 335–344, 2022.

-
- [48] N. Donges, M. Urwin, and S. Pierre, "Random forest: a complete guide for machine learning," *Built In*, 2023.
- [49] A. Khalyasmaa, S. A. Eroshenko, T. P. Chakravarthy, V. G. Gasi, S. K. Y. Bollu, R. Caire, S. K. R. Atluri, and S. Karrolla, "Prediction of solar power generation based on random forest regressor model," in *2019 International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON)*, pp. 0780–0785, IEEE, 2019.
- [50] T. Zhu, "Analysis on the applicability of the random forest," in *Journal of Physics: Conference Series*, vol. 1607, p. 012123, IOP Publishing, 2020.
- [51] Dr. Kuppusamy P, "Three Fundamental Problems in Hidden Markov Model (HMM)," in Available: <https://www.youtube.com/watch?v=BjEZOXvi21g&t=8s>, December 14, 2022.
- [52] Rahamathulla K, "Three Basic Problems of HMM Learning Problem Model (HMM)," in Available: <https://www.youtube.com/watch?v=6XCcAXFYyhY&t=6s>, Jun 7, 2021.
- [53] K. Naik, "Random forest hyperparameter tuning using randomisedsearchcv | machine learning tutorial," *YouTube*, October 28, 2020. Available:
- [54] Peter, J and ROUSSEEUW, Silhouettes, "A graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational and Applied Mathematics*, vol. 20, p. 53–65, 1987
- [55] SMART, "Determine the right Number of Clusters | WSS | Elbow Method | Silhouette Score | Data Science," <https://www.youtube.com/watch?v=DzrvLpxTxJw>, Accessed: March 15, 2025, 2023
- [56] Khreich, Wael and Granger, Eric and Miri, Ali and Sabourin, Robert, "On the memory complexity of the forward–backward algorithm," in *journal: Pattern Recognition Letters*, vol. 31, p. 91-99, 2010.

-
- [57] H, Muhammad and P, Muhammad Fermi and Ghani, Imran and Imran, Muhammad and Alzahrani, Mohammed Y and Budiarto, Rahmat, "Evaluating trust prediction and confusion matrix measures for web services ranking," in *journal: Ieee Access*, vol. 8, p. 90847–90861, IEEE, 2020.
- [58] Hofmann, Paul and Tashman, Zaid, "Hidden markov models and their application for predicting failure events," in *International Conference*, Amsterdam, The Netherlands, June 3–5, 2020, Proceedings, Part III 20, p. 464–477, Springer organization 2020.
- [59] Diamantoulakis, Panagiotis D and Ghassemi, Abolfazl and Karagiannidis, George K, "Lithium-ion battery state-of-health prediction for new-energy electric vehicles based on random forest improved model," in *booktitle: Applied Sciences*, p. 2773–2778, IEEE, 2013.
- [60] Delgado, Carlos J and Alfaro-Mejía, Estefanía and Manian, Vidya and O'Neill-Carrillo, Efrain and Andrade, Fabio, "Photovoltaic Power Generation Forecasting with Hidden Markov Model and Long Short-Term Memory in MISO and SISO Configurations," in *journal: Energies*, vol. 17, number.3, p. 668, 2024.
- [61] Gogoi, Swagat, "A review of the power distribution system in the telecommunications sector," *journal:Impact Fact*, vol.3, p. 143, 2018.