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

Root Cause Analysis of Optical Transport Network Channel Failure

By:

Aziza Ahmed

Advisor:
Dr. Yalemzewd Negash

A Thesis Submitted to the School of Gradiauate Studies of Addis Ababa University in Partial Fulfillment of the Requirement for the Degree of Master's Science in Communication Engineering

May, 2024

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

Root Cause Analysis of Optical Transport Network Channel Failure

by:

Aziza Ahmed

Approval by Board of Examiners	Signature	Date
Chairman, School Graduate Comittee:	____________	____________
Dr. Yalemzewd Negash:	____________	____________
Internal Examiner's Name:	____________	____________
External Examiner's Name:	____________	____________

Declaration

I declare that this thesis is entirely my own work and does not include any content from other educational institutions without proper acknowledgement. To the best of my knowledge, it does not contain previously published material by another person without recognition.

Aziza Ahmed	Signature	Date
______________	____________	____________

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

Dr. Yalemzewd Negash	Signature	Date
______________	____________	____________

Addia Ababa, Ethiopia

ABSTRACT

As demands for high bandwidth surge due to 4K/8K streaming, 5G networks, and cloud-based applications, robust optical transport networks (OTNs) are critical. OTN failures can significantly impact service quality, network availability, and service level agreements (SLAs). While research has addressed fault prediction and localization, a gap exists in root cause analysis (RCA) for OTN channel failures. This study proposes a novel approach utilizing machine learning (ML) to pinpoint the root cause of these failures efficiently.

This study compared four ML classifier models to analyze real data from ethio telecom's network. The data included eight key features that influence OTN channel performance. Extreme Gradient Boosting (XGBoost) emerged as the superior performer, achieving an impressive 99.91% accuracy and a high F1-score of 97.5%. Furthermore, it excelled in efficiency, with training times of just 5.42 seconds and testing times of 0.2 seconds. Interestingly, the model identified minimum input optical power (Min IOP) as the most critical factor, suggesting that extrinsic loss within the fiber optic cables is a major cause of OTN channel failures.

This study explores a novel ML system for OTN RCA, enabling faster and more precise root cause identification. This empowers network operators to proactively address issues and ensure optimal performance, significantly boosting network reliability and efficiency in the high-bandwidth age.

Key Words: OTN, RCA, Channel Failure, Feature Importance, XGBoost, Min IOP, SM Errored Second, Extrinsic loss.

Acknowledgments

I would like to begin by expressing my deepest gratitude to my Creator ALLAH (S.W), for guiding me through this journey and granting me the strength and wisdom to complete this thesis. Without His blessings and grace, none of this would have been possible. I am also thankful to all those who have contributed to the completion of this thesis.

First and foremost, I am deeply thankful to my adviser, Dr. Yalemzewud Negash, for his invaluable guidance, support, and encouragement throughout this research journey. His expertise, patience, and constructive feedback have been instrumental in shaping this work.

I am also deeply grateful to my examiners for their insightful comments and suggestions during the thesis review process. Their feedback was invaluable in helping me improve my work.

Special thanks to my family for their unwavering love, understanding, and encouragement. Their support has been my source of strength and motivation.

I would like to extend my appreciation to my friends and colleagues for their encouragement, helpful discussions, and moral support.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AI	Artificial Intelligence
APS	Automatic Protection Switching
ANN	Artificial Neural Network
ATM	Asynchronous Transfer Mode
APS	Automatic Protection Switching
AUC	Area Under the Curve
BER	Bit Error Rate
BET	Board Environment Temperature
BW	Band Width
CF	Channel Failure
CM	Confusion Matrix
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
DCM	Dispersion Compensation Mode
DT	Decision Trees
DWDM	Dense Wavelength-Division Multiplexing
DW	Digital Wrapper
EMI	Electromagnetic Interference
EMS	Element Management System
FAS	Frame Alignment Signal
FEC	Forward Error Correction
FMEA	Failure Mode and Effect Analysis
FPR	False Positive Rate
FTA	Fault Tree Analysis
GB	Gradient Boosting
HF	Hard Failure
IAE	Incoming Alignment Error
IOP	Input Optical Power
IoT	Internet of Things
IP	Internet Protocol

ITU	International Telecommunications Union
KPI	Key Performance Indication
LASSO	Least Absolute Shrinkage and Selection Operator
LBC	Laser Bias Current
LR	Linear regression
LOS	Loss of Signal
MDI	Mean Decrease in Impurity
ML	Machine Learning
ML	Multi-Protocol Label Switching
MSE	Mean Squared Error
MTTR	Mean Time to Repair
mW	milliwatts
NA	Network Availability
NCE	Network Cloud Engine
NE	Network Element
NMS	Network Management System
NN	Neural Network
OCH	Optical Channel layer
ODU	Optical Data Unit
OH	Over Head
O&M	Operation and Maintenance
OMS	Optical Multiplex Section
OPU	Optical Payload Unit
OTN	Optical Transport Network
OTS	Optical Transmission Section
OTU	Optical Transport Unit
OOP	Output Optical Power
PM	Path Monitoring
PMD	Polarization mode dispersion
Q-Factor	Quality Factor
QoS	Quality of Service
QoT	Quality of Transmission
RCA	Root Cause Analysis
RF	Random Forest
RFE	Recursive Feature Elimination

ReLU	Rectified Linear Unit
ROC	Receiver Operating Characteristic
SDH	Synchronous Digital Hierarchy
SF	Soft Failure
SLA	Service Layer Agreement
SM	Section Monitoring
SMOTE	Synthetic Minority Over-Sampling Technique
SONET	Synchronous Optical Network
STM	Synchronous Transport Module
SVM	Support Vector Machines
TCM	Tandem Connection Monitoring
TPR	True Positive Rate
UEC	Uncorrected Error Count
WDM	Wavelength Division Multiplexing
XGBoost	Extreme Gradient Boosting

Chapter 1: **Introduction**

1.1 Background

The ever-increasing demand for high-bandwidth applications like 4K streaming, 5G, and cloud services is driving the need for innovative solutions. OTNs are at the forefront, providing the robust infrastructure needed to unlock the full potential of these exciting advancements. The OTN is commonly referred to as a 'digital wrapper' because it encapsulates each client/service in a transparent container for seamless transport across optical networks, maintaining the native structure, timing information, and management data of the client [1]. The enhanced multiplexing capabilities of OTN enable various types of traffic, such as Internet Protocol (IP), Ethernet, storage, digital video, and Synchronous Optical Network(SONET)/Synchronous Digital Hierarchy (SDH), to be transmitted over an OTN framing structure, which is a primary driver for its adoption [1].

Since its inception in 2001, OTN has evolved beyond a simple SONET/SDH wrapper. Today, OTN has been optimized to support Ethernet, which is one of the most prevalent client services ranging from 1G to 400G. OTN-enabled technology often forms the foundation of next-generation optical networks due to its capability to accommodate flexible packet technologies, including new Ethernet interfaces, Multi-Protocol Label Switching (MPLS), Segment Routing, and Time-Sensitive Networking (TSN), among others. Currently, OTN technology has seen widespread implementation in networks worldwide, expanding its reach across a diverse range of applications. Hundreds of thousands of OTN ports have been deployed, effectively transporting essential data from network edges to metropolitan and core areas, and even in submarine communication applications [1].

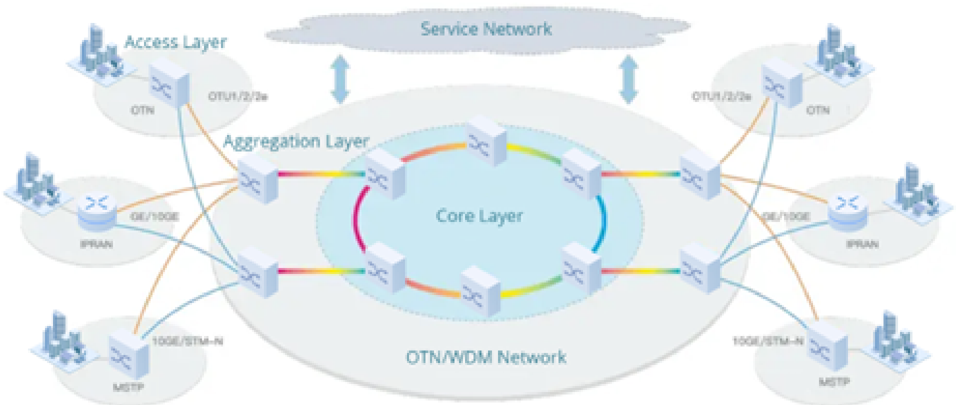


Figure 1.1: Optical Network Hierarchy Diagram [1]

OTN failures can be the silent saboteurs of your network. These disruptions can come in two forms: sudden outages (hard failures) or gradual performance degradation (soft failures). Both can be caused by issues within the optical lines or network equipment itself. These problems can range from minor annoyances that reduce service quality to complete outages with data loss. Unfortunately, pinpointing the root cause of these failures can be a time-consuming process, further impacting service availability and potentially jeopardizing agreements with your customers [2].

RCA is a detective's workhorse, used to uncover the true culprit behind issues in complex systems like OTN networks. Traditionally, this involves human experts meticulously combing through data to find the root cause, which can be a time-consuming and subjective process [3]. However, the growing complexity of networks demands faster, more precise solutions. Here's where Artificial Intelligence steps in, particularly ML. By analyzing network alarm logs and key performance indicators (KPIs), ML can assist in identifying the root cause of failures more efficiently, leading to quicker network restoration and improved overall service quality [3].

1.2 Statement of the problem

The ever growing complexity of fiber optic networks, particularly OTNs, poses a significant challenge to traditional methods of RCA for channel failures. The sheer volume of variables and intricate interactions within these networks create hurdles in data handling, pattern recognition, and extracting meaningful insights. This translates to a time-consuming and potentially inaccurate process of identifying the root cause of failures, leading to:

- **Delayed Service Restoration:** Traditional RCA methods can significantly delay pinpointing the root cause of a failure. This delay disrupts communication, impacting businesses, government operations, and individual users by causing service outages and communication delays.

- **Negative Impact on Network Availability and Service Quality:** The time consuming nature of traditional RCA methods can negatively impact network availability and service quality. Delayed identification of root causes extends downtime, leading to lost revenue for businesses and service providers. Additionally, it can potentially violate SLAs with customers.

Existing RCA methods for OTN channel failures fall short due to their limitations. Manual analysis, while thorough, is subjective, time-consuming, and prone to human error, especially for complex situations. While faster, rule-based systems lack flexibility and miss root causes outside their pre-defined rules. This research addresses this gap by proposing ML algorithms specifically designed for OTN networks. By leveraging ML's capabilities, we aim to significantly improve the speed and accuracy of RCA, ultimately leading to faster service restoration, enhanced network availability, and improved service quality for OTN users.

1.3 Objective

1.3.1 General Objective

Develop a more efficient and accurate approach for identifying the RCA of channel failures in OTNs.

1.3.2 Specific Objectives

- Improve RCA speed and accuracy in complex OTN networks.
- Leverage ML algorithms to identify root causes of OTN channel failures.
- Investigate ML algorithms suitable for OTN RCA.
- Implement and evaluate an ML-based RCA system for OTNs.
- Reduce impact of channel failures on service quality, network availability, and SLAs.

1.4 Scope and Limitations

1.4.1 Scope

This thesis delves into the RCA of OTN channel failures within the ethio telecom backbone transmission network, focusing specifically on failures experienced by ZTE network (excluding huawei). The analysis will be conducted on 12 network elements out of 118, encompassing 49 individual channels and eight KPIs will be utilized. To identify the root causes, four supervised ML classifier models will be evaluated using comprehensive performance metrics such as the confusion matrix, accuracy, precision, recall, F1-score, and execution time.

1.4.2 Limitations

This thesis faced two main limitations. Firstly, exporting the necessary input data for the ML models was time-consuming due to the limited historical data available, spanning only a few days. Ideally, a longer timeframe would have provided a richer dataset for analysis. Secondly, a significant gap was identified in existing research on RCA specifically focused on OTN channel failures. This lack of prior work highlights the potential novelty and contribution of this study.

1.5 Methodology

To attain the objectives of the research, the methods to be employed are:

- **Define Problem and Objectives:** Our research starts with a clear understanding of the problem we're tackling. The problem statement concisely explains the issue, its importance, and why it deserves investigation. This lays the foundation for our objectives, which are specific, measurable goals that break down the problem into actionable steps.

- **Literature Review:** To gain a comprehensive understanding of existing approaches to RCA for OTN failures, we conducted an extensive literature review. This review revealed a significant gap on the limited use of ML algorithms in this area. Recognizing this critical need, our thesis delves into the application of ML models for RCA in OTN. By carefully evaluating various ML models proposed in the literature for similar tasks, we aim to bridge this gap and offer a novel approach to RCA in OTN.

- **Data Collection:** To build the foundation for analysis, We first identified KPIs that are most relevant to OTN channels within Ethio-Telecom's backbone transmission network. Following this, We proceeded to extract the data directly from their Network Management System (NMS), ensuring its authenticity and accuracy for the subsequent analysis.

- **Data Preprocessing:** During this phase, essential methodologies are employed to preprocess the data, ensuring its compatibility with the applied models to enhance their performance. This includes verifying and addressing missing values through selected mechanisms for quantitative data. Additionally, steps such as identifying and removing duplicate values, as well as eliminating irrelevant features, are undertaken. Furthermore, the categorization of channel failure and non-failure is conducted numerically as part of this process.

- **Model Building:** We utilized Python to simulate the RCA of OTN channel failures, employing a variety of ML models. Through this process, We observed the fluctuation in the performance of the ML models. Subsequently, We conducted an in-depth analysis of the results, leveraging techniques such as feature importance and correlation analysis. This analysis enabled me to pinpoint the optimal KPIs for addressing OTN channel issues effectively.

- **Model Evaluation and Selection:** We conducted a comparison of ML model performance for the RCA of OTN channel failures by employing various classifier performance evaluation metrics. Subsequently, we identified and selected the model that outperformed the others based on the evaluation results.

- **Result and Discussion:** We delved into the implications of our findings concerning the RCA of OTN channel failures. Subsequently, We proposed preventive solutions aimed at addressing the identified root causes.

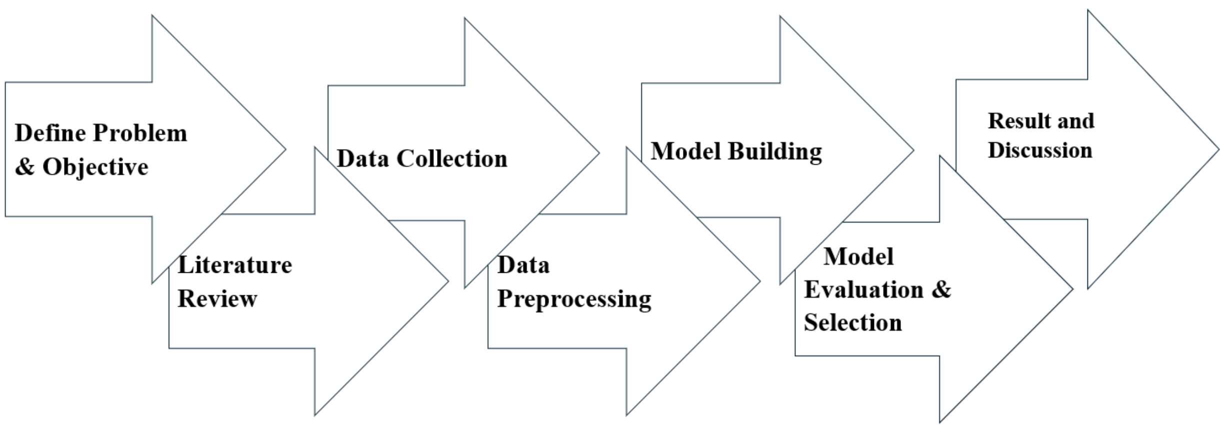


Figure 1.2: Methodology

1.6 Literature Review

While significant progress has been made in predicting and locating faults within OTNs, a key gap remains on analyzing root causes behind these failures. A literature survey of some selected papers are presented here.

A study conducted in [6], proposes a cause-aware failure detection scheme for OTN boards using an interpretable ML algorithm called XGBoost. They compare XGBoost with Support Vector Machine (SVM) and Neural Networks (NN) based on metrics like accuracy, F1-score, false positives, and false negatives. Notably, the study employs SHAP (SHapley Additive exPlanations) to understand feature attribution and identify the key factors contributing to different failure types. Their findings suggest XGBoost achieves high detection accuracy and F1-score (over 98%) and reduces both false negatives and false positives compared to other models. By analyzing feature importance through SHAP values, they link high temperatures (average and maximum) to equipment failure. While industry experience suggests fan malfunctions as a potential cause, this study identifies a gap in exploring XGBoost's performance in terms of execution time.

A study conducted in [7], explores the root causes of persistent fiber cuts in Ghana's telecom industry. Focusing on MTN Ghana's Western and Central regions, they aim to identify causes and propose solutions that can improve service quality for various network operators. Their research, targeting reduced fiber cuts, lower operational costs, and ultimately, increased customer satisfaction, employed a two-phase approach. Phase one utilized the Pareto and Ishikawa analytical tools on primary data from MTN Ghana to understand the root causes. Phase two involved a targeted survey to gather additional information. By analyzing both sets of data, the authors categorized all fiber cut incidents into six groups based on the underlying causes. This categorization approach allowed them to identify and scrutinize the key factors responsible for the recurring issue of fiber cuts within the network.

1.7 Contribution

This paper significantly advances the scientific understanding of OTN channel failure RCA by introducing a novel ML-based system. The research demonstrates that this approach is both faster and more accurate than traditional methods, while also identifying XGBoost as a particularly effective ML model for this task. These findings contribute to the development of more efficient and accurate RCA techniques in the field of OTN communications.

Consequently, the proposed ML-based system offers a valuable tool for network operators. By enabling faster and more accurate root cause identification, the system can significantly improve network reliability and efficiency, particularly in today's high-bandwidth environments. Additionally, accurate identification of root causes like fiber attenuation allows for targeted corrective actions, ensuring optimal network performance and reducing operational costs. This research offers a practical solution with real-world benefits for the telecommunications industry.

1.8 Thesis Organization

The remaining sections of the thesis are structured as follows: Chapter 1 encompasses the research introduction, problem statement, thesis objectives, research methodology, and review of related literature. Chapter 2 delves into the fundamental concepts of OTN. Chapter 3 explores RCA, RCA methods, ML for RCA, data preprocessing techniques such as feature selection and correlation, and introduces ML algorithms and evaluation methods. Chapter 4 details the data preprocessing steps and model building process. Chapter 5 summarizes the thesis results and their implications. Finally, Chapter 6 presents conclusions drawn from the results and offers recommendations for future research endeavors.

Chapter 2: **Basics of OTN**

2.1 Introduction

This chapter delves into the fundamentals of OTN technology. We will embark on a journey to understand its layered structure, the KPIs that measure an optical channel's health, and the potential causes of channel failures. We'll also explore the mechanisms employed to monitor OTN channels and ensure their smooth operation.

2.1.1 OTN concept

OTN is outlined in the International Telecommunications Union (ITU-T) G.709 Network Node Interface for the OTN [9]. The primary objective of the OTN is to merge the advantages of SONET/SDH technology with the scalability of Dense Wavelength Division Multiplexing (DWDM). In essence, OTNs leverage the operations, administration, maintenance, and provisioning (OAM&P) functionalities of SONET/SDH within DWDM optical networks. This recommendation, also known as digital wrapper (DW), extends the capabilities of single-wavelength SONET/SDH technology, enabling the establishment of transparent, wavelength-manageable multi-wavelength networks. Additionally, the integration of Forward Error Correction (FEC) into OTN offers network operators the opportunity to reduce the number of regenerators employed, thereby lowering network costs [9].

2.1.2 Properties of OTN

The aim of the OTN is to enable the multiservice transport of packet-based data and legacy traffic, while DW technology accommodates non-intrusive management and monitoring of each optical channel assigned to a particular wavelength [9]. The "wrapped" overhead (OH) facilitates the management and control of client signal information. Figure 2.1 illustrates how OTN management capabilities are achieved with the addition of OH at several positions during the transport of the client signal [9]. OTNs present several advantages to network operators including:

- **Protocol transparency:** OTNs seamlessly carry various data formats, like Ethernet or SONET/SDH, without any alterations.
- **Backward compatibility:** Existing network protocols can be smoothly integrated into the OTN infrastructure.
- **FEC coding advantage:** FEC strengthens the signal, allowing for longer transmission distances and fewer regeneration points.
- **Reduced regeneration:** Flexible OTN network designs can minimize the need for regeneration equipment, leading to cost savings.

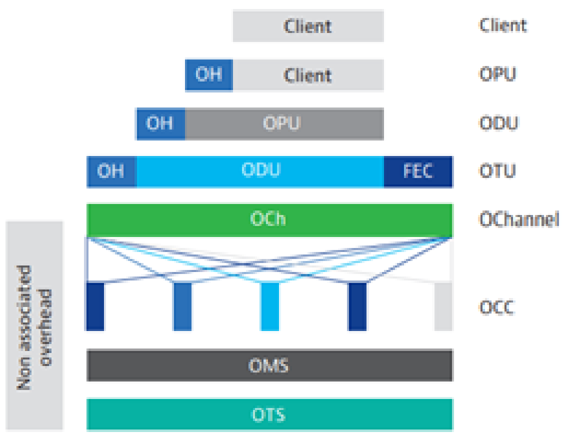


Figure 2.1: Basic OTN transport structure [9]

The client signal undergoes the addition of multiple OH sections. These, along with the FEC, collectively constitute the Optical Transport Unit (OTU). Subsequently, this composite unit is transmitted via a single wavelength, forming what is known as an Optical Channel (OCh). As multiple wavelengths are transported over the OTN, an overhead must be added to each to enable the management functionality of the OTN [9]. The optical multiplexing sections and optical transmission sections are constructed using the additional OH together with the OCH [9].

2.2 OTN Layer Structure

OTN is transmission protocol for high speed and high capacity. Can transmit up to 400 gbps over single channel. It utilize a layered structure defined by the Telecommunication Standardization Sector ITU-T. This structure facilitates modularity, clear delineation of functionalities, and interoperability between different vendors' equipment. Here's a breakdown of the key layers: The basic OTN layers are visible in the OTN transport structure and consist of the OCH, optical multiplex section (OMS), and optical transmission section (OTS).

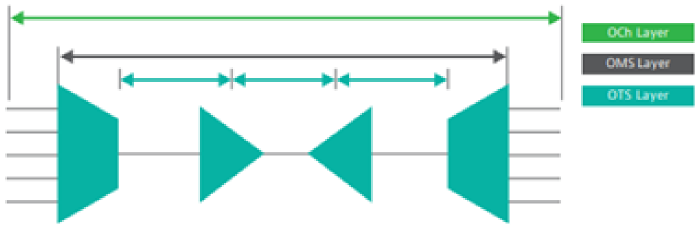


Figure 2.2: OTN layer structure [9]

2.2.1 OCH Layer

OCh is an information structure consisting of the information payload (OCh_PLD) with a certain bandwidth and non-associated OH (OCh_OH) for management of the OCH. Non-associated OH means the overhead information is transported on a physically separate channel. The OCh-layer network enables end-to-end transmission of OCHs, allowing for the transparent conveyance of client information in various formats (e.g., SDH STM-N, gigabit Ethernet, cell-based ATM), including OCH connection rearrangement for flexible network routing [9].

Transport of a client signal in the OTN follows the procedure outlined below and in Figure 2.3:

- OH is added to the client signal to form the OCH payload unit (OPU).
- OH is then added to the OPU thus forming the OCH data unit (ODU).
- Additional OH plus FEC are added to form the OTU.
- Adding further OH creates an OCh, which is carried by one color.

- Additional OH may be added to the OCh to enable the management of multiple colors in the OTN. The OMS and the OTS are assembled. The result is an OCh comprising an OH section, a client signal, and a FEC segment.

OCH Overheads

OCh OH, facilitating OTN management functionality, comprises four substructures: Optical Payload Unit Overhead (OPU), Optical Data Unit(ODU), Optical Transmission Unit (OTU), and Frame Alignment Signal (FAS).

OPU OH: is appended to the OPU payload and serves to accommodate diverse client signals. It governs the mapping of multiple client signals and furnishes details regarding the type of signal being transported.

ODU OH: allows the user to support tandem connection monitoring (TCM), path monitoring (PM) and automatic protection switching (APS). End-to-end path supervision and client adaptation are facilitated through the OPU, as previously described. The ODU OH provides two important OHs: the PM OH and the TCM [9].

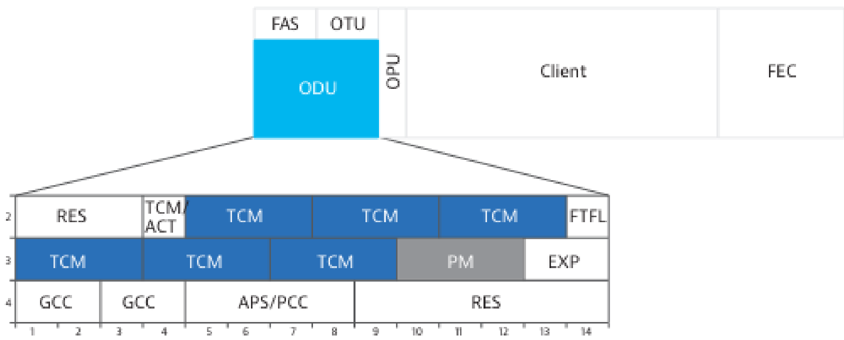


Figure 2.3: Overhead structure of ODU [9]

The Optical Data Unit Path Monitoring Overhead (ODU PM OH) allows for the monitoring of specific sections within the network and fault localization via the OH bytes detailed in the PM OH [9]. The PM OH is configured in row 3, columns 10 to 12 to facilitate PM. The PM field structure comprises subfields, as illustrated in Figure 2.8.

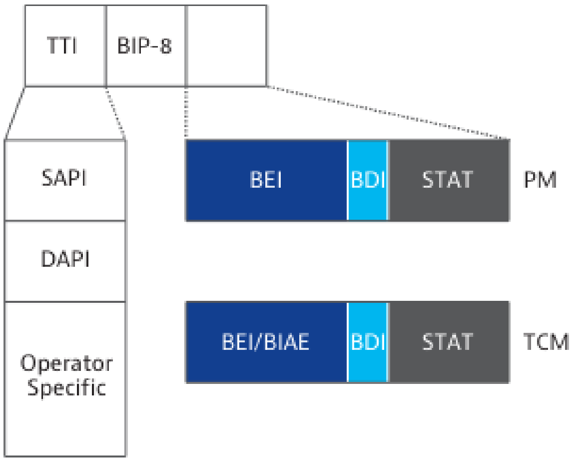


Figure 2.4: TCM and PM overhead structures [9]

The OTU within the OTN facilitates transport across one or more OCH connections and specifies frame alignment as well as FEC [9].

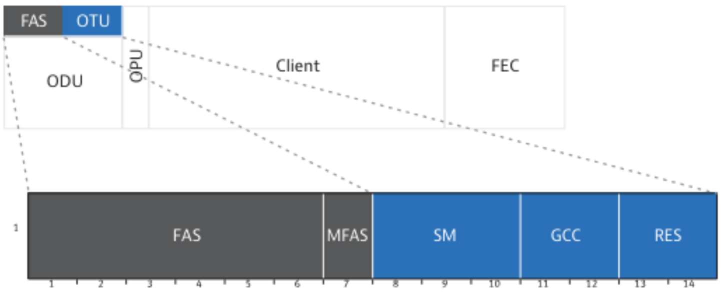


Figure 2.5: Frame alignment and OTU OH structure [9]

The Section Monitoring (SM) OH includes the same subfields as described for the path monitoring OH, with the exception of the Incoming Alignment Error (IAE) bit. This bit enables the ingress point to notify the egress point of a detected alignment error in the incoming signal. IAE is set to "1" when an error occurs, otherwise it is set to "0" [9].

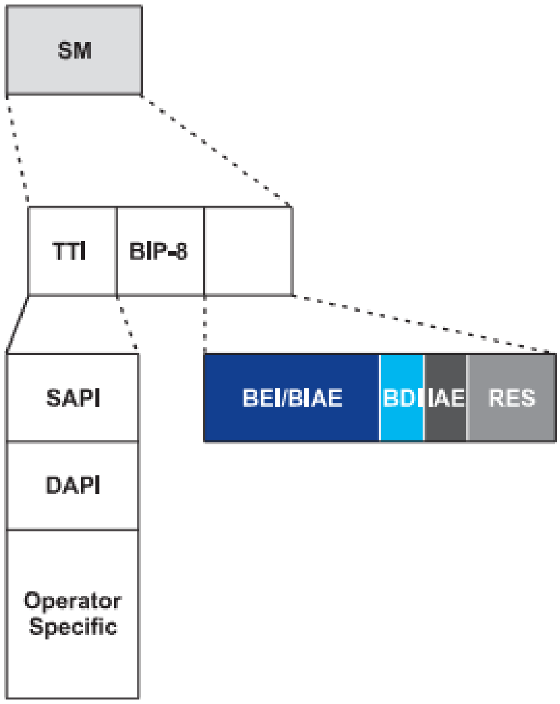


Figure 2.6: Section monitoring OH [9]

The client signal or actual payload to be transported could be of any existing protocol, that is: SONET/SDH, GFP, IP, GbE, as shown in Figure 2.5.

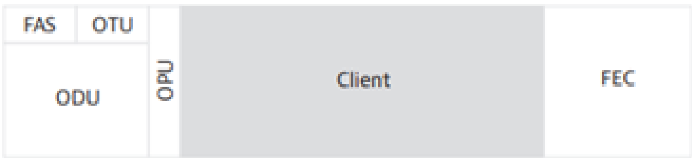


Figure 2.7: Client in an optical channel [9]

FEC: In tandem with the OCh OH of the DW envelope, extra bandwidth is introduced, in this instance, for FEC. The employed FEC algorithm facilitates the detection and rectification of errors within an optical link, as depicted in Figure 2.8.

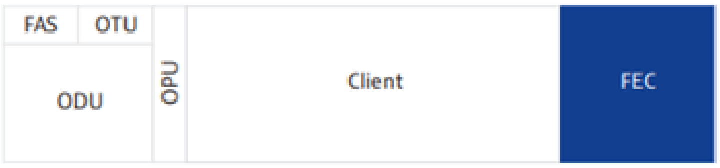


Figure 2.8: FEC structures [9]

FEC enables the identification and rectification of bit errors stemming from physical impairments in the transmission medium. These impairments encompass linear effects such as attenuation, noise, and dispersion, as well as nonlinear effects like four-wave mixing, self-phase modulation, and cross-phase modulation. By employing FEC in a network link, the network operator can tolerate a lower quality signal within the link, thereby mitigating potential errors [9].

2.2.2 OMS Layer

The OMS plays a crucial role in optimizing bandwidth utilization within OTNs. It essentially deals with the multiplexing and demultiplexing of various OCHs onto a single optical fiber using a technique called Wavelength Division Multiplexing (WDM).

Multiplexing: Combines multiple individual OCHs, each carrying separate data streams, onto a single fiber using different wavelengths of light. This allows for efficient utilization of the fiber's immense bandwidth capacity.

Demultiplexing: At the receiving end, the OMS separates the multiplexed signals back into their individual channels, enabling them to be routed to their respective destinations.

Benefits of OMS Layer

- Increased Network Capacity: By enabling multiple channels on a single fiber, the OMS significantly boosts the overall capacity of an OTN compared to traditional single-channel transmission.
- Scalability: The OMS allows for easy network expansion by adding more channels without the need for additional fiber infrastructure.
- Cost-Effectiveness: Utilizing existing fiber infrastructure with WDM through the OMS proves to be more cost-efficient compared to deploying new fibers for each additional channel

2.2.3 OTS Layer

OTS serves as the workhorse layer in OTNs, ensuring the reliable transmission of light signals over long distances. It focuses on the physical aspects of light propagation within the optical fiber to maintain signal integrity and minimize signal degradation.

Key Functionality of The OTS Layer

- **Amplification:** To combat the natural weakening (attenuation and dispersion) of light in optical fibers, OTNs strategically deploy optical amplifiers like EDFAs and Raman amplifiers. These amplifiers act like signal boosters, periodically regenerating and strengthening the light pulses for extended, reliable travel across long distances.
- **Regeneration:** In addition to amplification, the OTS layer might also perform signal

regeneration. This involves converting the received optical signal back to its electrical form, cleaning it up (removing noise and distortion), and then converting it back to a clean optical signal for further transmission. This ensures the signal maintains its original quality over long distances.

- **Dispersion Compensation:** Optical fibers can introduce different forms of dispersion, which cause the various components of a light pulse to travel at slightly different speeds, leading to signal distortion. The OTS layer might utilize techniques like dispersion compensating fibers (DCFs) or chirped gratings to mitigate these effects and maintain signal integrity.

- **Monitoring:** The OTS layer plays a crucial role in monitoring various transmission parameters like signal power, optical noise levels, and dispersion characteristics. This information is vital for proactive maintenance and troubleshooting potential issues within the network.

2.3 OTN Channel Parameters

Fiber-optic technology serves as the cornerstone of contemporary communication systems, furnishing swift, secure, and dependable data transmission across extensive distances. However, the signal quality of an optical link is subject to various impairments, such as attenuation, dispersion, and noise [14]. To evaluate the signal quality numbers of parameters are used.

2.3.1 Input Optical Power

In OTNs, the strength of the light signal entering a device, measured in milliwatts (mW) or dBm, is crucial. This "input optical power" needs to fall within a specific range for each receiver to function properly and avoid errors. The acceptable range depends on factors like the OTN technology, distance traveled, and the receiver itself.

2.3.2 Output Optical Power

Within OTNs, "output optical power" refers to the strength of the light signal launched into the fiber. Measured in mW or dBm, it significantly impacts network performance. OTN equipment often provides controls to adjust this power, allowing for optimization based on the specific fiber cable and network needs.

2.3.3 Board Environment Temperature

One often overlooked factor in OTNs is board environment temperature. Just like with any electronic device, heat can significantly impact performance and reliability. Luckily, OTN equipment manufacturers specify a safe operating temperature range to ensure optimal functionality and lifespan. Most OTNs even have built-in cooling systems like fans and heat sinks to keep things from getting too hot, with the specific design depending on the equipment's power needs and surrounding environment [35].

2.3.4 Laser Bias Current

In OTNs, a laser's bias current acts like a faucet for light. This electric current controls the intensity and other features of the light signal emitted by the laser diode. A stable bias current is crucial, as fluctuations can cause variations in signal strength and introduce noise, ultimately affecting network performance.

2.3.5 Section Monitoring Errored Second (SM Errored Second)

It is a one-second interval during which one or more errors are detected in the section monitoring layer and is a crucial performance metric in optical networks that helps in monitoring and maintaining the quality and reliability of the physical transmission layer. By tracking SM Errored Second, network operators can detect, diagnose, and address physical layer issues to ensure optimal network performance.

2.3.6 Path Monitoring Errored Second (PM Errored Second)

A one-second interval during which one or more errors are detected in the path monitoring layer and is a critical metric for ensuring the performance and reliability of optical networks. By continuously monitoring and managing PMES, network operators can maintain high-quality data transmission, quickly identify and address issues, and ensure compliance with performance standards and SLAs. This proactive approach helps in maintaining the overall health and efficiency of the optical network, ensuring uninterrupted and reliable service to end-users.

2.3.7 Quality Factor (Q-Factor)

Q-factor is a measure of the quality of a digital signal in an optical communication system. The Q-factor is determined by the Bit Error Rate (BER), signal power, and noise power. A higher Q-factor value indicates superior signal quality (Kane, year). Calculated as the ratio of the distance between the average signal levels of two adjacent symbols to the standard deviation of the noise. The formula for calculating Q-factor is as follows:

$$Q_factor = SignalLevel1 - SignalLevel2/NoiseRMS \tag{2.1}$$

2.3.8 Optical Signal-to-Noise Ratio (OSNR)

OSNR quantifies the signal quality of an optical link by delineating the extent to which the signal power surpasses the noise power. A higher OSNR value signifies superior signal quality within the optical link The OSNR is calculated as the ratio of the optical signal power to the noise power within a specific bandwidth [14]. The formula for calculating OSNR is as follows:

$$\text{OSNR(dB)} = 10\log_{10}\left(\frac{\text{Signal Power}}{\text{Noise Power}}\right) \tag{2.2}$$

2.3.9 Bit Error Rate After FEC

In OTNs, BER after FEC represents the number of errors remaining in the data stream after the FEC decoder has attempted to fix them. It's a crucial metric for evaluating the effectiveness of FEC and the overall quality of the transmitted signal.

Ideally, FEC significantly reduces the BER after correction compared to the pre-FEC BER (BER before correction). The effectiveness of FEC depends on the chosen code and its error correction capability. Stronger FEC codes can handle higher pre-FEC BERs and achieve lower post-FEC BERs.

2.3.10 FEC Uncorrected Error Count

In OTNs, the FEC Uncorrected Error Count is a key metric for network health. It tracks errors that snuck past the FEC and couldn't be corrected, potentially indicating signal quality issues. A rising count warrants investigation to prevent data transmission problems.

2.4 Causes of Channel Failure in OTN

Channel failure in an OTN disrupts data transmission on a specific optical channel, potentially impacting services carried on that channel. It's crucial for network operators to understand the causes of channel failures, identify them promptly, and implement effective recovery mechanisms.

2.4.1 Physical Layer Issues

- **Optical Fiber Cable Cut**

  An optical fiber cable cut is a significant disruption in an OTN. Since these cables transmit data using light pulses, a physical severing of the fiber completely halts signal transmission on the affected channels. This can lead to service outages for any communication services carried on those channels.

Fiber optic cable cuts can stem from various threats: accidental damage during construction or by animals, environmental factors like storms or landslides, and even equipment malfunctions or poor installation practices.

- **Optical Fiber Loss or Attenuation**

  Optical fiber loss, also known as attenuation, refers to the weakening of light signal strength as it travels through an optical fiber cable. This weakening occurs due to various physical phenomena within the fiber itself, and it's a crucial factor to consider in designing and operating OTNs. These losses can be categorized into intrinsic and extrinsic types based on whether they arise from inherent fiber properties or external operating conditions [28].

  (a) Intrinsic fiber loss, or cable attenuation arises from the inherent properties of the fiber itself, independent of external factors. It is measured in decibels (dB) per kilometer with the relation:

  $$A(\text{dB}) = 10 \log \left(\frac{P_{\text{in}}}{P_{\text{out}}} \right) \tag{2.3}$$

  Where P_{in} and P_{out} refer to the optical power going into and coming out of the fiber. Intrinsic losses in the fiber itself are relatively small. Modern optical fibers are designed to have low intrinsic losses to ensure efficient signal transmission over long distances [28].

Light traveling through optical fibers faces a three-pronged attack: material absorption by the core itself, scattering from microscopic imperfections that weaken the signal, and various forms of dispersion that cause light pulses to spread and distort. Understanding these enemies is crucial for maintaining strong data transmission.

(b) Extrinsic losses: Extrinsic losses pertain to all components besides the fiber optic cable itself. These losses arise at the interfaces between fiber components and can notably impact the overall system performance. Depending on the loss mechanism, these extrinsic losses are described as either insertion loss or return loss [28].

Insertion loss: also referred to as connector losses, refers to the loss of optical power that occurs when light is transmitted through a component, such as a connector, splice, coupler, or any other device that introduces an additional optical path [28]. It is given by:

$$Insertionloss = -10 \log \left(\frac{P_{\text{out}}}{P_{\text{in}}} \right) \tag{2.4}$$

Where insertion loss is typically in dB, P_{out} is the output optical power, P_{in} is the input optical power.

Return loss: denotes the optical power loss when light reflects back towards the source at an interface of an optical component. It is usually quantified by the magnitude of the reflected signal. Return loss can apply to all fiber optic components including couplers, splitters, splicers, connectors, and attenuators

[28].

$$Returnloss = -10\log\left(\frac{P_0}{P_1}\right) \tag{2.5}$$

Where return loss is typically in dB, P_0 represents the reflected optical power, and P_1 represents the input optical power.

This insertion and return losses are primarily caused by misalignment, reflection, scattering, positioning errors, or absorption at the interface between the fiber and the component [28].

2.4.2 Environmental Factors

- Extreme Temperatures: Excessive heat or cold can impact the performance of optical components, potentially leading to channel failures.
- Electromagnetic Interference (EMI): Strong electromagnetic fields can introduce noise into the signal, causing errors and potentially leading to channel failures.

2.4.3 Software or Configuration Issues

- Software Bugs: Software bugs within OTN equipment can lead to unexpected behavior and potentially disrupt data transmission on a specific channel.
- Misconfiguration: Incorrect configuration of OTN equipment can cause various issues, including channel failures.

Faulty or bad cable, short haul cable is used for long haul or long-haul cable is used for short haul, momentary sync loss, loose cable connection at one or both ends, FEC type mismatch and improper GBIC or SFP connection at one or both ends are also the causes for OTN channels failure.

2.5 OTN Channel Failure Monitoring Mechanism

OTN equipment typically incorporates alarm systems that trigger alerts in case of a channel failure. These alarms can indicate loss of signal (LOS), high BER, or other anomalies. Additionally, network monitoring tools constantly track signal strength, error rates, and how well error correction is functioning. Any major changes in these readings can serve as early warnings of potential channel problems.

Chapter 3: **RCA Methods**

3.1 Introduction

This chapter dives deep into RCA, exploring both traditional models and the exciting potential of ML. We'll compare and contrast their strengths and weaknesses, then delve into various ML models and the metrics used to evaluate their effectiveness. Finally, we'll shed light on the overall system model, its operational principles, and the fascinating world of feature selection, correlation, and importance mechanisms within the ML framework.

3.2 What is RCA?

RCA is the process of uncovering the fundamental reasons behind problems to pinpoint suitable solutions. It operates on the premise that addressing underlying issues systematically is far more effective than merely treating immediate symptoms or reacting to crises. RCA encompasses a set of principles, techniques, and methodologies, all of which can be employed to unearth the root causes of an event or trend. The primary objective of RCA is to uncover the fundamental cause of a problem or event. Subsequently, the secondary aim is to gain a comprehensive understanding of how to rectify, mitigate, or extract valuable lessons from the underlying issues identified within the root cause. Lastly, the tertiary goal involves utilizing the insights gleaned from the analysis to implement systematic measures aimed at preventing future occurrences of similar issues or replicating successful outcomes. It is imperative to recognize that the effectiveness of the analysis hinges on the actions taken based on its findings, making the third goal of RCA particularly significant. We can also use RCA to modify core process and system issues in a way that prevents future problems [15].

3.3 Importance of RCA

- RCA can be used in a variety of setting, including information technology, manufacturing, healthcare, and government.
- The goal of RCA is to identify the root cause of a problem, not just the symptom.
- Once the root cause has been identified, corrective action can be taken to prevent the problem from recurring.
- RCA can be a time-consuming process, but it is worth the investment in order to improve quality and reliability.

3.4 RCA in OTN Channel Failure

OTNs are crucial for high-speed data transmission. However, channel failures can disrupt communication and lead to significant downtime. Identifying the root cause of these failures is essential for efficient network management and troubleshooting.

The Impact of OTN Channel Failures:

- Disrupted Communication: When an OTN channel fails, data transmission gets disrupted, leading to service outages and communication delays. This can impact businesses, government operations, and individual users, causing frustration and lost productivity.

- Financial Losses: Network downtime translates to lost revenue for businesses and service providers. Additionally, the troubleshooting and repair process itself incurs costs.

- Data Loss: In some cases, OTN channel failures can lead to data loss, particularly if the failure occurs during data transmission. This can be especially critical for sensitive information or real-time applications.

3.5 RCA Methods

There are two main approaches to RCA modeling: traditional methods and those powered by ML. Let's explore both options.

3.5.1 Traditional RCA Method

There are several well-established traditional RCA models;

1. Pareto Charts

   Pareto charts are the first RCA tool that are bar graphs in most cases, and they show the ordered frequency of counts of data. This means that the charts can be used to show which areas need your attention first for improvement purposes[38]. The length of the bars stands for the cost or frequency (money or time). The longest bars usually are arranged at the left and go down as you move forward across the graph. So, by looking at this chart, you can immediately see what needs your attention first and balance where you spend your time and money [38].

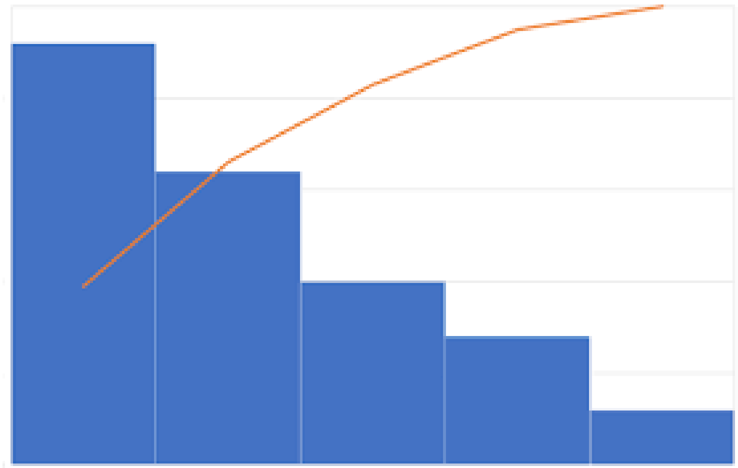


Figure 3.1: Pareto chart[38]

2. Failure Mode and Effect Analysis (FMEA)

   The FMEA, focuses on failures that happen within a particular system. The two parts include Failure Mode as well as Effect Analysis. With Failure Mode, what you are looking at is identifying the ways in which something can fail. With the Effect Analysis, you are looking at the effects of the failure modes.

3. 5 Whys

   The 5 Whys is an investigative method. This method helps you get down to the bottom of a problem by fully investigating it. You just have to keep asking, “why?”.

This tool is great for exploring and investigating problems that do not require quantitative analysis. You can also combine this tool with others, like the Pareto chart, to really dig into an area that needs more attention. At the fifth "Why", we transition to the lowest element of root cause, the latent cause. This method is not suitable for complex problems because it is limited to a single causal chain[38].

4. Ishikawa Fishbone Diagram

   This diagram looks at a big problem and figures out all the possible causes. Then, it breaks them down into subcategories that link back to the main issue being investigated[38]. Each part of the fishbone diagram would break down the potential causes into categories like: material, method, machine, measurement/medium and man/mind power. Then, potential causes would be investigated within those categories until a resolution could be found. A fishbone diagram is used when there is no known root cause and major brainstorming has to take place[38].

5. Fault Tree Analysis (FTA)

   FTA is another kind of graph you can build to investigate how a top fault (also known as an abnormal condition or failure) happened. It is used to promote reliability, maintainability, and safety analyses [38].

   FTA is broken down into parts including four major steps:

   - Scoping to define the event and determine the scope.
   - Developing the tree with relevant causes and items.
   - Validating the tree with qualitative or quantitative information, supply chain

information or other data.

- Verifying the information with qualitative or quantitative data.

3.5.2 ML-based RCA Method

Modern applications are evolving faster, growing more complex and becoming increasingly distributed. But they are also failing in new ways, and while most organizations are able to tell when an application breaks, finding the root cause is another matter. Traditional troubleshooting approaches rely on a combination of human instinct and slow, manual searches in conjunction with observability tools. But this is reactive and time intensive, hurting productivity and Mean-Time-To-Resolution (MTTR).

ML for RCA is the state-of-the-art application of algorithms and statistical models to identify the underlying reasons for issues within a system or process. Rather than relying solely on human intervention or time-consuming manual investigations, ML automates and enhances the process of identifying the root cause [39].

When to Use ML for RCA?

- Complex System Failures: For systems where issues can arise from a multitude of factors and applications.
- Recurring Issues: When traditional methods have failed to identify repetitive problems.
- Predictive Maintenance: To forecast potential system failures before they occur.

- Large Scale Operations: Where manual analysis would be too time-consuming or impractical.

3.6 Types of ML Models

There are three major types of ML models. While all ML modeling techniques work on a common purpose, their way of approaching a data problem differs.

3.6.1 Supervised Learning

Supervised learning trains ML models using labeled data. This data includes both the input and the desired output, essentially providing the model with the answer key. The model learns to map features from the input data to the corresponding outputs. During training, the model is presented with new data points and tries to predict the correct output based on what it has learned [40].

Data scientists refine the model by iteratively adjusting its internal parameters until it achieves a high degree of accuracy. Common supervised learning tasks include tasks like optical character recognition, spam filtering, and voice recognition.

This type of learning can be further broken down into different techniques for analyzing data. Two prominent ones are classification and regression. Classification involves sorting data points into predefined categories, while regression focuses on predicting a continuous output value.

Supervised Classifier Models

There are two main types of supervised classifiers:

Binary classification: These models deal with problems where there are only two possible outcomes, like classifying emails as spam or not spam.

Multi-class classification: These models handle problems with more than two categories, like classifying images as containing cats, dogs, or neither.

When it comes to supervised classifier models, there are several contenders, each with its own strengths:

(a) Random Forest (RF)

A RF classifier is a powerful ensemble learning method that combines multiple decision trees to improve overall accuracy and robustness. Each tree is built on a random subset of features and random data points, preventing overfitting. During prediction, unseen data is passed through all the trees, and the most frequent class prediction is considered the final output. This approach effectively reduces variance and handles large datasets well.

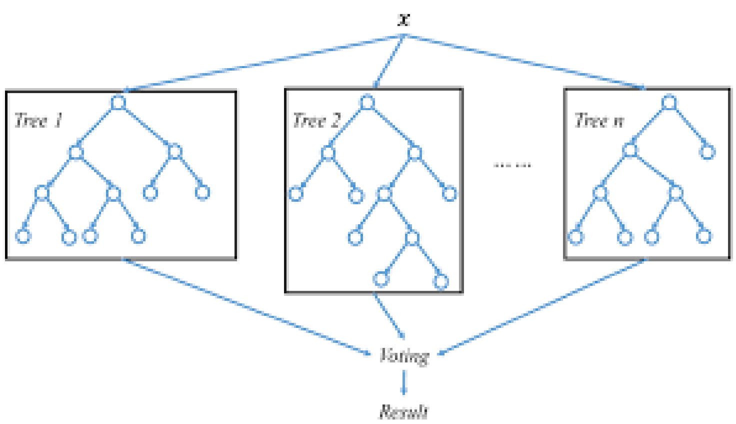


Figure 3.2: Model architecture for RF algorithm [29]

(b) Extreme Gradient Boosting (XGBoost)

XGBoost is a powerful classifier known for its speed and accuracy. It works by sequentially building decision trees, where each tree learns to correct the errors of the previous one. This ensemble approach leads to high performance. Additionally, XGBoost utilizes regularization techniques to prevent overfitting and handles missing data efficiently, making it a robust choice for various classification tasks.

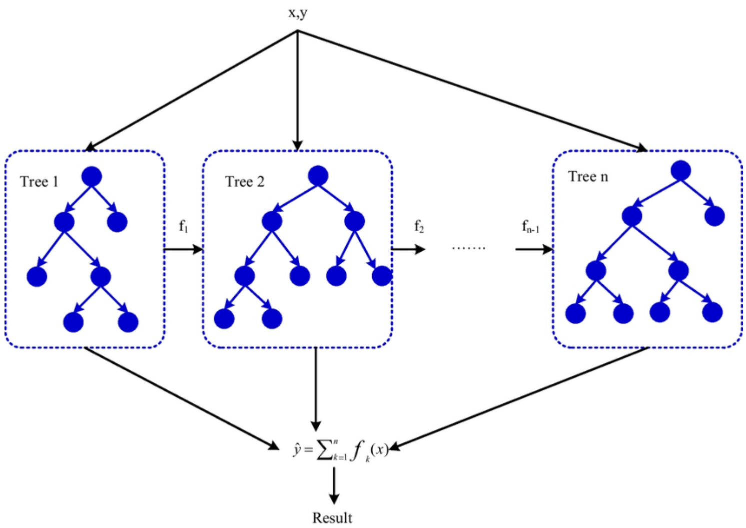


Figure 3.3: Model architecture for XGBoost algorithm [41]

(c) Support Vector Machine (SVM)

SVMs are a powerful classification algorithm that excels at finding the optimal hyperplane (decision boundary) to separate data points belonging to different classes. In simpler terms, SVMs identify the widest margin between classes, focusing on data points closest to the boundary (support vectors) to define the best separation. This approach leads to good generalization and works well with high-dimensional data.

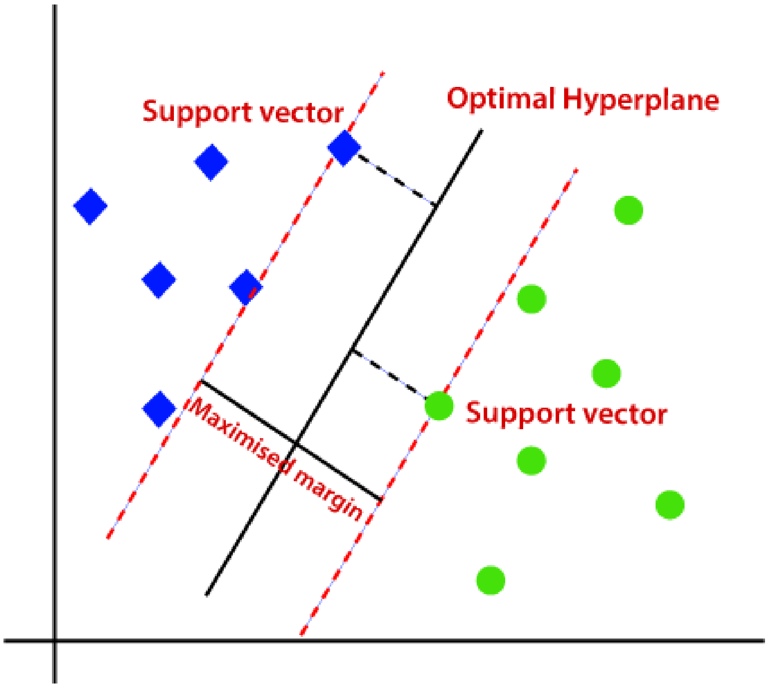


Figure 3.4: Model architecture for SVM algorithm

(d) Artificial Neural Network (ANN)

ANNs are powerful classifiers inspired by the structure of the human brain. They consist of interconnected layers of nodes (artificial neurons) that process information. Each layer transforms the input data, allowing ANNs to learn complex, non-linear patterns. This makes them highly versatile for tackling intricate classification problems like image recognition and spam filtering. Despite their complexity, ANNs can achieve high accuracy when trained with sufficient data.

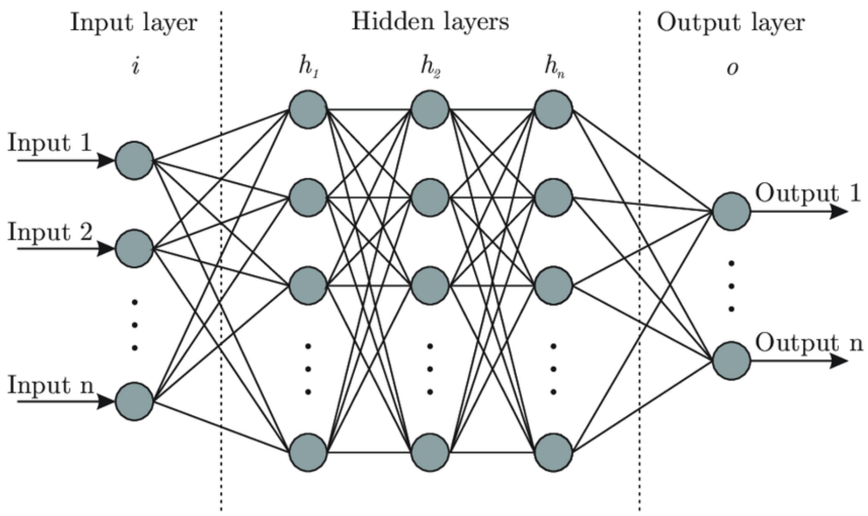


Figure 3.5: Model architecture for ANN algorithm [37]

3.6.2 Unsupervised Learning

Unlike supervised learning, unsupervised ML works with unlabeled data. This means the data doesn't have predefined categories or classifications. The model's job is to uncover hidden patterns and relationships within the data itself. By analyzing and comparing data points, unsupervised models can group similar data together or identify anomalies. The more data the model is exposed to, the more refined its understanding of the underlying structure becomes. This allows unsupervised models to be self-sufficient, learning and improving without the need for constant human intervention.

One common unsupervised technique is clustering. This process involves grouping data points into distinct categories based on their similarities. Data scientists can then use these clusters for further analysis or classification tasks. There are various clustering algorithms available, each with its own strengths. Choosing the right method depends on the specific data and the desired outcome.

3.6.3 Reinforcement Learning

Reinforcement learning takes a different approach compared to supervised and unsupervised learning. Instead of relying on labeled data or simply identifying patterns, reinforcement learning involves an intelligent agent interacting with an environment. The agent learns through trial and error, receiving rewards for desired actions and penalties for mistakes. This feedback loop allows the agent to gradually improve its decision-making and achieve a specific goal [40].

Q-learning, SARSA, and Deep Q-networks are some popular reinforcement learning algorithms. These algorithms enable agents to learn optimal behavior in complex environments, making them well-suited for tasks like game playing, robot control, and resource management.

3.7 Traditional Vs. ML-based RCA Methods

Both methods have their strengths and weaknesses, making the choice depend on the specific situation.

Table 3.1: Traditional Vs. ML RCA Methods

Aspect	**Traditional Method**	**ML Method**
Methodology	Manual	Automated and data driven
Efficiency	Time consuming	Fast and automated
Scalability	Limited for complex system	Highly scalable
Accuracy	Prone to human error	Objective and improved over time
Data Requirement	Often qualitative	Requires large amount of data
Interpret-ability	Easy to understand and explain	Challenging to interpret

3.8 ML Models Evaluation Metrics

When evaluating ML models, we use various metrics to assess their effectiveness. These metrics differ depending on the task at hand.

Regression: For tasks like predicting continuous values, metrics like Mean Squared Error (MSE) compare the model's predictions on training and test data to the actual values. Lower MSE indicates a better fit.

Classification: For tasks involving classifying data points into categories (e.g., positive or negative), the confusion matrix plays a key role. It summarizes the model's performance by showing how many data points were correctly or incorrectly classified.

(a) Precision: corresponding to the ratio of correctly predicted positive values to the total number of predicted positive values.

(b) Recall, also called true positive rate (TPR) corresponding to the ratio of correctly predicted positive values to the total number of positive values in the dataset.

(c) False Positive Rate (FPR) corresponds to the proportion of negative values predicted incorrectly.

(d) Accuracy, corresponding to the number of correctly predicted values divided by the total number of predicted values.

(e) Receiver Operating Characteristic (ROC): is a powerful tool for evaluating binary classification models. It provides a visual representation of the model's ability to distinguish between positive and negative classes, independent of the chosen classification threshold. This makes it ideal for situations where the optimal threshold might vary depending on the application.

(f) Confusion Matrix: is a valuable tool for evaluating how well an algorithm predicts labels in binary classification tasks. It compares the model's predictions (0 for negative, 1 for positive) against the actual labels, revealing four categories: true positives, true negatives, false positives, and false negatives. By analyzing these categories, we can calculate various metrics like precision and recall to understand the model's effectiveness.

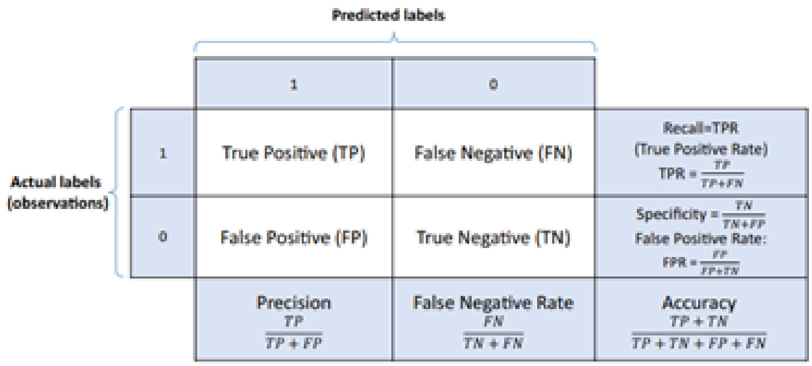


Figure 3.6: Confusion matrix for two-class problems

By analyzing these metrics, we can gain valuable insights into the strengths and weaknesses of our machine learning models, allowing us to compare different models or fine-tune them for optimal performance.

3.9 Feature Selection

Regarding feature selection techniques, it's important to note that these methods are commonly employed to reduce the dimensionality of data before model training, retaining only features pertinent to the task at hand. These techniques encompass wrapper methods [18], which evaluate various subsets of features and gauge their relevance based on model performance metrics. Additionally, filter methods are utilized to eliminate irrelevant features as a preprocessing step prior to model training. Generally, wrapper methods yield the optimal feature subset for the specific model and task, albeit at the expense of increased computational complexity.

3.10 Feature Importance

Feature importance refers to techniques that compute a score for all input features in a given model. These scores indicate the "importance" of each feature. A higher score means that the specific feature will have a larger effect on the model that is being used to predict a certain variable [23]. There are several methods to determine feature importance in a machine learning model. Here are some common methods:

I. Coefficient Magnitude: For linear models like Linear Regression or Logistic Regression, the magnitude of the coefficients can indicate the importance of the corresponding feature. Larger magnitudes typically signify more important features.

II. Decision Tree-based Methods:

- Gini Importance: For decision tree-based models like Random Forest or Gradient Boosting Machines, the Gini importance or mean decrease in impurity (MDI) can be used to measure the importance of each feature.

- Permutation Importance: This method measures the increase in the model's prediction error after permuting the feature's values, thus providing a measure of the feature's importance.

III. LASSO Regression: LASSO (Least Absolute Shrinkage and Selection Operator) is a regression analysis method that performs both variable selection and regularization. The non-zero coefficients in the LASSO model indicate important features.

IV. SHAP Values: SHAP (SHapley Additive exPlanations) is a unified measure of feature importance that is based on cooperative game theory. It provides a unified measure of feature importance that considers the interaction between features. SHAP values explain how each feature contributes to a specific prediction. They offer a more nuanced understanding of feature importance compared to the built-in methods, considering feature interactions and providing local explanations for individual predictions.

V. Recursive Feature Elimination (RFE): This method works by recursively removing the least important features and building a model until the specified number of features is reached. The ranking of feature elimination can indicate feature importance.

3.11 Feature Correlation

Correlation is a fundamental statistical concept that researchers utilize to examine relationships within their data. It helps us to Understand the Relationship Between Variables. It is important for machine learning engineers to understand the correlation between variables in their models [24]. Understanding feature correlations can help in tasks such as:

- Feature selection: Identifying highly correlated features can help in selecting the most relevant features for a model.
- Multicollinearity detection: Identifying highly correlated features in linear regression can help in detecting multicollinearity, which can affect the model's performance.
- Data exploration: Visualizing feature correlations can provide insights into the relationships between different variables in the dataset.

Measures of Correlation

(a) Pearson's correlation coefficient measures the linear relationship between two variables and ranges from -1 to 1, Where:

  - 1 indicates a perfect positive linear relationship.
  - 0 indicates no linear relationship.
  - -1 indicates a perfect negative linear relationship.

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} \tag{3.1}$$

Where, r=Pearson Correlation Coefficient x_i=x variable samples x^-=mean of values in x variable y_i= y variable sample y^-=mean of values in y variable.

(b) Spearman's correlation coefficient assesses how well an arbitrary monotonic function describes the relationship between two variables, rather than specifically testing for a linear association [24]. A monotonic relationship is one whereas one variable increase or decreases, the other variable consistently increases or decreases. This allows

Spearman's correlation to identify nonlinear relationships that may not be evident when using Pearson's r.

$$r_s = 1 - \frac{6 \sum D^2}{n(n^2 - 1)} \tag{3.2}$$

3.12 Environment and Tools Used

For running the ML algorithms, we employed the Python programming language, specifically version 3.11. Python serves as the primary language for developing artificial intelligence applications due to its ease of installation, interpreted nature, speed, and lightweight footprint. The Python distribution utilized is Anaconda, which encompasses a comprehensive set of tools and libraries essential for machine learning tasks, including Numpy, Pandas, Matplotlib, Scikit-learn, Jupiter, Spider, and others.

Chapter 4: **RCA Model Building**

4.1 Introduction

This chapter explores the application of ML classifier models for RCA in OTN channel failures. It delves into the data collection process, employed pre-processing techniques, and the architectural design of various classifier models. After evaluating and selecting the optimal model, the analysis focuses on the most critical feature identified. Leveraging this key feature, the chapter concludes by exploring the root causes underlying these OTN channel failures.

4.2 Proposed RCA System Model

This diagram illustrates the proposed system model for RCA of OTN channel failures. It provides a high-level overview of the system's components and their interactions. The specific details and inner workings of each block will be explained in more depth throughout this chapter.

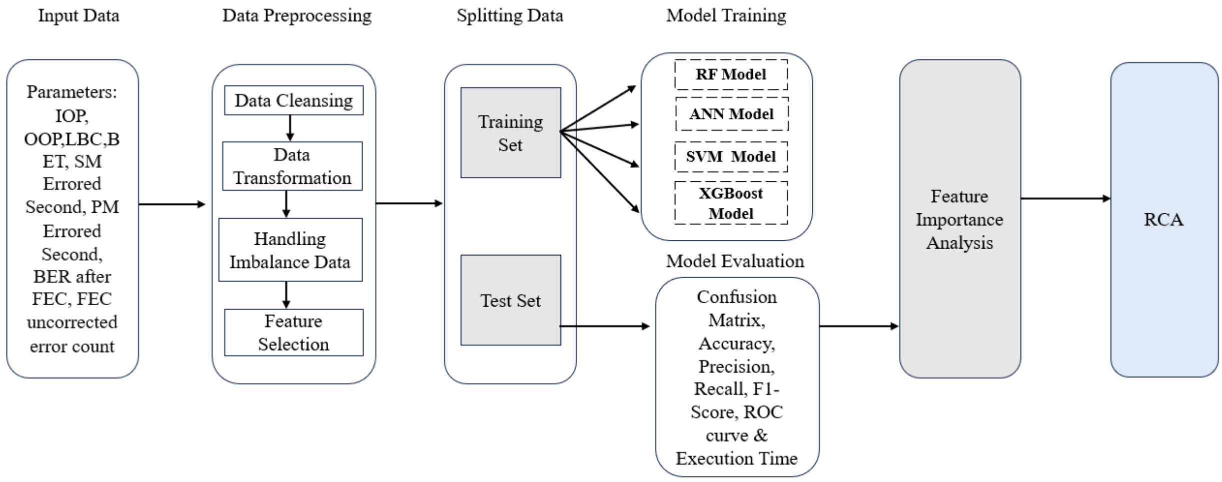


Figure 4.1: Proposed RCA System Model

4.3 Data Collection

Performance data from the history of 12 nodes and 49 channels within ethio telecom's backbone transmission network OTN channels has been extracted from the Element Management System (EMS) database. This extraction was conducted at 15-minute intervals over a span of 29 days, from March 4th, 2024, to April 1st, 2024. Below is a sample of raw historical data for the selected features of the OTN channel, collected to train the model, as depicted in Figure 4.2.

Index	Start Time	End Time	Query Gr	NE Loc	MO Locatior	NE Loca	MO Loca	Max Va	Min Val	Averag	Max Va	Min Val	Averag	Max Bo	Min Bo	Averag	Max Va	Min Val	Averag	BER Befor	SM Errc	BER Aft	PM Err
1	4/14/2024 7:30	###########	15Minute	Microw	LS4[0-6-52]-	BN:ME{	BN:ME{151948734},EQ={/r=0/sh=6/sl=52},AID={/aid=							45	44	44.61							
2	4/14/2024 7:30	###########	15Minute	Microw	LS4[0-6-52]-	BN:ME{	BN:ME{1	15.04	15.03	15.03	-4.98	-5.39	-5.13				145.2	145.2	145.2				
3	4/14/2024 7:30	###########	15Minute	Microw	LS4[0-6-52]-	BN:ME{	BN:ME{1	15.04	15.03	15.03							145.2	145.2	145.2				
4	4/14/2024 7:30	###########	15Minute	Microw	LS4[0-6-52]-	BN:ME{	BN:ME{1	-1.99	-1.99	-1.99													
5	4/14/2024 7:30	###########	15Minute	Microw	LS4[0-6-52]-	BN:ME{	BN:ME{151948734},EQ={/r=0/s				-4.98	-5.39	-5.13							6.80E-04			
6	4/14/2024 7:30	###########	15Minute	Microw	LS4[0-6-54]-	BN:ME{	BN:ME{151948734},EQ={/r=0/sh=6/sl=54},AID={/aid=							50	50	50							
7	4/14/2024 7:30	###########	15Minute	Microw	LS4[0-6-54]-	BN:ME{	BN:ME{1	15.03	15.02	15.02	-2.35	-2.49	-2.43				123.9	123.9	123.9				
8	4/14/2024 7:30	###########	15Minute	Microw	LS4[0-6-54]-	BN:ME{	BN:ME{1	15.03	15.02	15.02							123.9	123.9	123.9				
9	4/14/2024 7:30	###########	15Minute	Microw	LS4[0-6-54]-	BN:ME{	BN:ME{1	-1.99	-2.01	-1.99													
10	4/14/2024 7:30	###########	15Minute	Microw	LS4[0-6-54]-	BN:ME{	BN:ME{151948734},EQ={/r=0/s				-2.36	-2.48	-2.43							1.30E-05			
11	4/14/2024 7:30	###########	15Minute	Dilla&C	L4Kx4-S[0-6-	BN:ME{	BN:ME{151948803},EQ={/r=0/sh=6/sl=20},AID={/aid=							51	50	50.38							
12	4/14/2024 7:30	###########	15Minute	Dilla&C	L4Kx4-S[0-6-	BN:ME{	BN:ME{151948803},EQ={/r=0/s				--	--	--										
13	4/14/2024 7:30	###########	15Minute	Dilla&C	L4Kx4-S[0-6-	BN:ME{	BN:ME{151948803},EQ={/r=0/s				--	--	--										
14	4/14/2024 7:30	###########	15Minute	Dilla&C	L4Kx4-S[0-6-	BN:ME{	BN:ME{151948803},EQ={/r=0/s				--	--	--										

Figure 4.2: Raw history performance data of OTN channels

Performance events are divided into counting and analog performance events according to the properties of physical parameters based on the performance events. The counting performance events mainly includes various error parameters and analog performance events metrics which have clear physical units. Generally, the count performance event is the direct reflection of transmission quality of OTN system, while the analog performance event is the indirect manifestation [6]. Below in the table, the performance events used to measure the performance of optical transport channel with their description.

Table 4.1: OTN channel KPIs description

OTN channel KPI	**Description**
BET(Celsius)	Board Environment Temperature
IOP(dBm)	Input Optical Power
LBC(mA)	Laser Bias Current
OOP(dBm)	Output Optical Power
BER After FEC	Bit error rate After Forward error correction
PM Errored second (s)	Path Monitoring Errored second
SM Errored second (s)	Section Monitoring Errored second
FEC UEC	Forward Error Correction uncorrected error count

4.4 Data Preprocessing

Most ML algorithms are designed to handle high-dimensional datasets. Hence, derived features from the existing data are often included, such as log-transformed data, products, and ratios of features, or more advanced combinations. Such data transformation is an important preprocessing step that can have a profound effect on the model performance.

In the data preprocessing stage of RCA employing ML classifier models, several crucial tasks need to be executed to ensure that the data is appropriate for training and evaluation. Here are the primary steps we employed in data preprocessing for RCA.

4.4.1 Data Cleaning

Data cleansing, also known as data cleaning, is an essential step in the data preprocessing pipeline. It involves identifying and correcting errors, inconsistencies, and inaccuracies in the dataset to ensure that the data is accurate, complete, and suitable for analysis.

- Handle Missing Values: Identify and handle missing data by either imputing missing values or removing rows/columns with missing data, depending on the impact of missing values on the analysis. We identified entries with missing information and chose to remove entire rows containing these missing values to ensure the accuracy of our analysis.
- Handling Duplicate Data: Duplicate rows in the dataset are identified and removed.
- Handling Inconsistent Data: Inconsistencies in the data are identified and resolved, such as different representations of the same category ("Too Low Value" and -30dBm in optical input value).

4.4.2 Data Transformation

- Scale Features: To ensure all features contribute equally to the model's training, we applied a technique called standardization to our numerical data. Standardization rescales the features to have a common mean of 0 and a standard deviation of 1. This prevents features with larger values from overshadowing those with smaller ones, leading to a more balanced training process.

- Labeling: We've labeled our data based on a specific variable called "SM Errored second." is a one-second interval during which one or more errors are detected in the section monitoring layer. Here's how we've categorized it:

  SM Error second = 0: This indicates everything is functioning normally, with no channel failure detected. (We label this as 0).

  SM Error second > 0: If this value is greater than zero, it signifies a channel error has been ongoing for some time. The duration of the error is reflected in the number of seconds. (We label this as 1).

  By labeling the data in this way, we can train a model to identify patterns that are more likely to be associated with channel failures.

4.4.3 Handling Imbalanced Data

Building accurate ML models often requires balanced data, where all classes are represented fairly. In our case, the initial data for this work had an uneven distribution of classes in the target variable. This imbalance can lead to models that perform poorly on the minority class, which is often the one we care about most.

To address this challenge, we employed a technique called Synthetic Minority Oversampling Technique (SMOTE). SMOTE helps create a more balanced dataset by artificially generating new data points for the underrepresented class. This results in a more even class distribution, allowing the model to learn effectively from all classes.

- Class distribution before oversampling: Counter(0: 52323, 1: 230) and
- Class distribution after oversampling: Counter(0: 52323, 1: 52323).

The application of SMOTE has effectively addressed the class imbalance in our dataset. Originally, the data classes were not represented equally. SMOTE successfully balanced the class distribution, resulting in a significant increase in the number of instances for the minority class. Now, both classes (0 and 1) have an equal number of data points, at 52,323 instances each. This balanced dataset of 104,646 total instances will provide a stronger foundation for our ML model, allowing it to learn more effectively from all classes.

4.4.4 Feature Selection

To build the most accurate model possible, we carefully selected the features that would have the greatest impact on its ability to identify root causes. This process is called feature selection, and it's a crucial step in the machine learning pipeline.

We leveraged two key approaches in our feature selection:

- Domain Expertise: We leveraged industry insights from domain experts. Their deep understanding of the problem domain helped us identify features that are most likely to contribute to the model's ability to pinpoint root causes. This focus on industry-specific knowledge ensures the model considers the most impactful factors, leading to more effective results.

- Correlation-Based Feature Selection: To address multicollinearity, which can occur when features are highly correlated with each other and can confuse the model, we employed a correlation-based feature selection method. This method analyzes the correlation between features using techniques like Pearson's correlation coefficient. Features with exceptionally high correlations are then carefully reviewed, and in some cases, one might be removed to avoid redundancy in the model's training data.

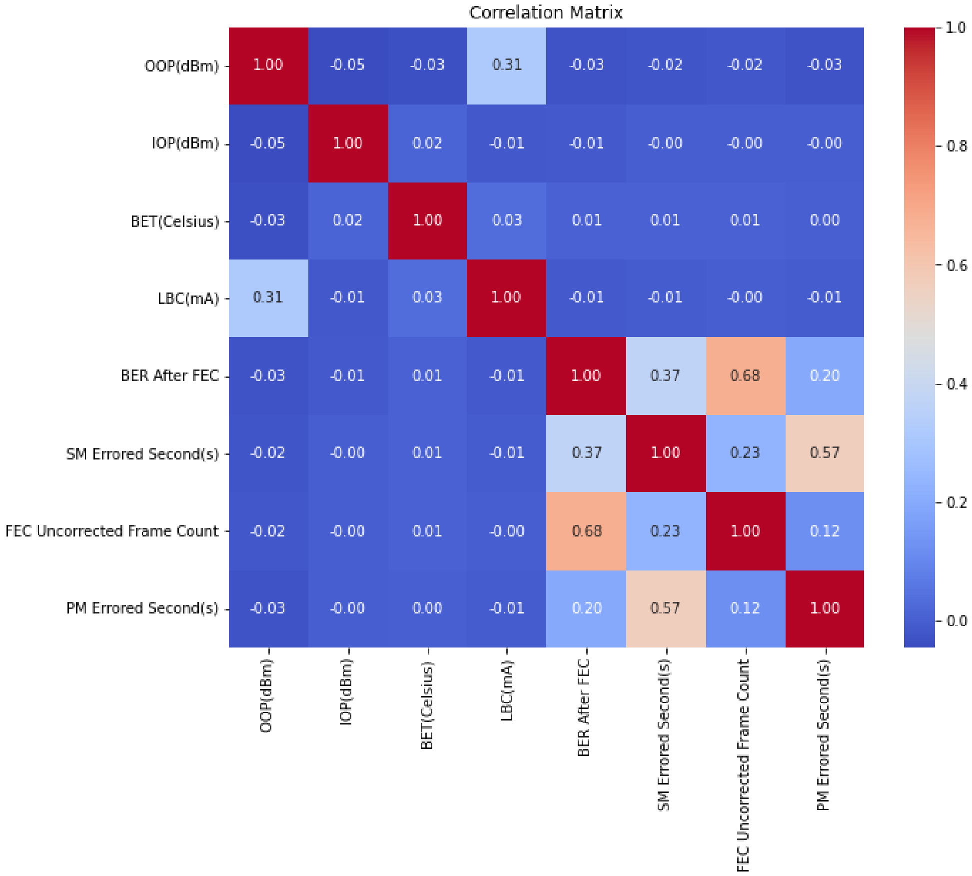


Figure 4.3: Correlation between the dataset features

Our analysis of feature correlations using techniques Pearson's coefficient revealed that all analog parameters ((IOP, OOP, BET and LBC) exhibit minimal correlation (coefficients close to zero). Consequently, we can confidently include all these analog parameters in our model as features with low correlation provide the model with independent information. However, for count events, we identified a higher degree of correlation. To avoid redundancy and potential issues like multicollinearity, we carefully selected 'SM Errored Second' as the representative feature for this category. This choice ensures our model leverages the most informative count data while maintaining a set of features with minimal overall correlation. By combining domain expertise with a data-driven approach, we ensured that our model considered the most relevant features, ultimately leading to more accurate root cause identification.

4.5 Splitting Data

In RCA, splitting data is particularly important. It helps ensure your model isn't simply memorizing the training data and can effectively pinpoint root causes in unseen scenarios. By evaluating the model on the testing data, you gain confidence in its ability to generalize its learnings and identify root causes in real-world situations.

- Training Data (80%): This larger portion acts as the detective's training ground. The model analyzes these data points (evidence) to learn the patterns and relationships between features and the target variable.

- Testing Data (20%): This unseen data represents new cases the detective hasn't encountered before. By evaluating the model's performance on this data, we can assess its ability to generalize its learnings (identify root causes) to new situations.

There's no universally perfect split ratio, but 80/20 is a well-established convention for reasons of balance between training and evaluation and computational efficiency

4.6 Model Development

4.6.1 Deriving Statistical Features for ML-based RCA

Following data preprocessing, we extract valuable insights from physical events, input/output optical power, temperature readings, and laser bias current using statistical measures like minimum, maximum, and average values. These statistics play a crucial role in our ML approach to RCA, providing us with a comprehensive understanding of system behavior.

Minimums highlight potential anomalies and dips, while maximums expose spikes and unusual peaks. The average value serves as a baseline, and deviations from it can indicate performance changes or root causes of problems. By leveraging these statistics in ML models, we gain a deeper understanding of the data and enhance our models' ability to pinpoint root causes effectively.

Table 4.2: Input features and labels of output

	KPI	**Unit**
Features	SM Error second	s
	Board Environment Temperature	Celsius
	Input Optical Power	dBm
	Output Optical Power	dBm
	Laser Bias Current	mA
Target	SM Error second (s)> 0	CF state ("1")
	SM Error second (s)$=0$	Normal state ("0")

4.6.2 Hyperparameter Tuning for Better Classification

Hyperparameters are essentially the settings that control how these models learn from the data. By carefully tuning these parameters through techniques like grid search or randomized search, we aimed to identify the configuration that best suits our specific data set and classification task. This process allowed us to build robust models capable of accurately identifying root causes.

To achieve optimal performance from our ML models, we explored a range of hyperparameters for various classifier algorithms including RF, SVM, ANN, and XGBoost.

(a) Model-1: RF classifier model

In our RF model, we experimented with different combinations of hyperparameters to find the one that performs the best at classifying root causes.

Our RF architecture leverages an ensemble approach with 100 decision trees (n_estimators=100) working together to improve classification accuracy. For reproducibility, we set a fixed random state (random_state=42), ensuring consistent results when retraining the model.

To capture complex relationships within the data, each tree is allowed to grow to its full potential by setting the maximum depth (max_depth=None). Finally, we address potential class imbalances by assigning balanced class weights (class_weight='balanced'), giving more weight to under-represented classes during training. This combination of parameters helps create a robust and informative Random Forest for tackling your classification problem.

Table 4.3: RF Model Hyperparameters

Parameter Name	**Value**
Number of Tree	100
Reproducibility	42
Tree Depth	None
Class_Weight	balanced

(b) Model-2: ANN Classifier Model

Our second model was an ANN, which works like a complex web of interconnected processing units called neurons. Our ANN architecture is designed for effective classification. It utilizes a two-layered structure, with each hidden layer containing 64 neurons. This layered approach allows the network to learn complex patterns within the data. However, it's important to note that the optimal number of layers and neurons can be further optimized through a process called hyperparameter tuning.

To enhance learning and introduce non-linearity, we leverage ReLU (Rectified Linear Unit) functions in the hidden layers. For the final layer, dealing with binary classification (normal state vs. channel failure), we use a sigmoid function. This function outputs values between 0 and 1, making it ideal for predicting class probabilities. Additionally, the Adam optimizer is employed to efficiently navigate the training process, while the binary cross-entropy function assesses the model's performance by measuring the discrepancy between predictions and actual data. Finally, the network undergoes 100 epochs of training, with each epoch representing one complete pass through the entire training dataset. This iterative process allows the network to continuously refine its decision boundaries and improve classification accuracy.

Table 4.4: ANN Model Hyperparameters

Parameter Name	**Value**
Number of Hidden Layer	2
Number of Neurons	64
Activation Function	ReLU
Optimizer	Adam
Loss Function	Binary cross entropy
Training cycle (Epochs)	100

(c) Model-3: XGBoost Classifier Model

Our third model used a powerful technique called XGBoost. Imagine it as a team of decision trees working together, but in a more sophisticated way. Our XGBoost architecture is designed for robust classification using gradient boosting. We leverage an ensemble of 100 decision trees (n_estimators=100) to make predictions, reducing the risk of overfitting. These trees learn sequentially (learning_rate=0.1), with each tree focusing on improving upon the errors of the previous one. To prevent over-complexity, we limit the maximum depth of each tree to 3 (max_depth=3), ensuring the model captures the most important patterns without becoming overly specific to the training data. Consistency is maintained through a fixed random state (random_state=42), guaranteeing repeatable results. Finally, we configure XGBoost for a binary classification task (normal vs. failure) by setting the objective function to "binary:logistic". This guides the model towards optimizing its ability to classify data points into the correct class. This architecture provides a strong foundation for XGBoost, but remember that hyperparameter tuning can significantly improve performance.

Table 4.5: XGBoost Model Hyperparameters

Parameter Name	**Value**
Number of Tree	100
Learning Rate	0.1
Tree depth	3
Reproducibility	42
Task Type	Binary: logistic

(d) Model-4: SVM Classifier Model

Our last model used a powerful technique called SVM. Imagine SVM as a tool that draws a clear line (or "decision boundary") between different categories in our data (normal state vs. channel failure). Our SVM architecture is designed for effective classification, particularly when dealing with potentially complex data. We achieve this through a combination of carefully chosen hyperparameters. The penalty parameter (C=100) balances the model's ability to fit the training data while avoiding overfitting. Gamma (gamma=0.1) influences the smoothness of the decision boundary, focusing on capturing broader patterns. To handle non-linear relationships within the data, we leverage the radial basis function (rbf) kernel (kernel='rbf'). This allows the model to effectively classify data points that wouldn't be easily separable in their original form. Finally, to address potential class imbalances in the data (normal vs. failure), we employ class weights ('balanced'). This ensures the model pays more attention to the under-represented class during training, preventing bias towards the majority class. This architectural design provides a solid foundation for SVM classification, but remember that hyperparameter tuning, especially for C and gamma, can significantly improve the model's performance.

Table 4.6: SVM Model Hyperparameters

Parameter Name	**Value**
Penalty for Mistake	100
Influence Range	0.1
Data Representation	3
Reproducibility	rbf
Balancing the Data	Balanced

4.7 Model Performance Evaluation

After determining the optimal parameters of the proposed algorithm, the XGBoost, SVM, RF and ANN algorithms will be retrained on the training set, and evaluate their performance with a fixed test set.

4.7.1 Confusion Matrix

In ML classification, where models predict which category a new data point belongs to (normal vs. channel failure), the confusion matrix acts like a scorecard for how well our model performs as a detective.

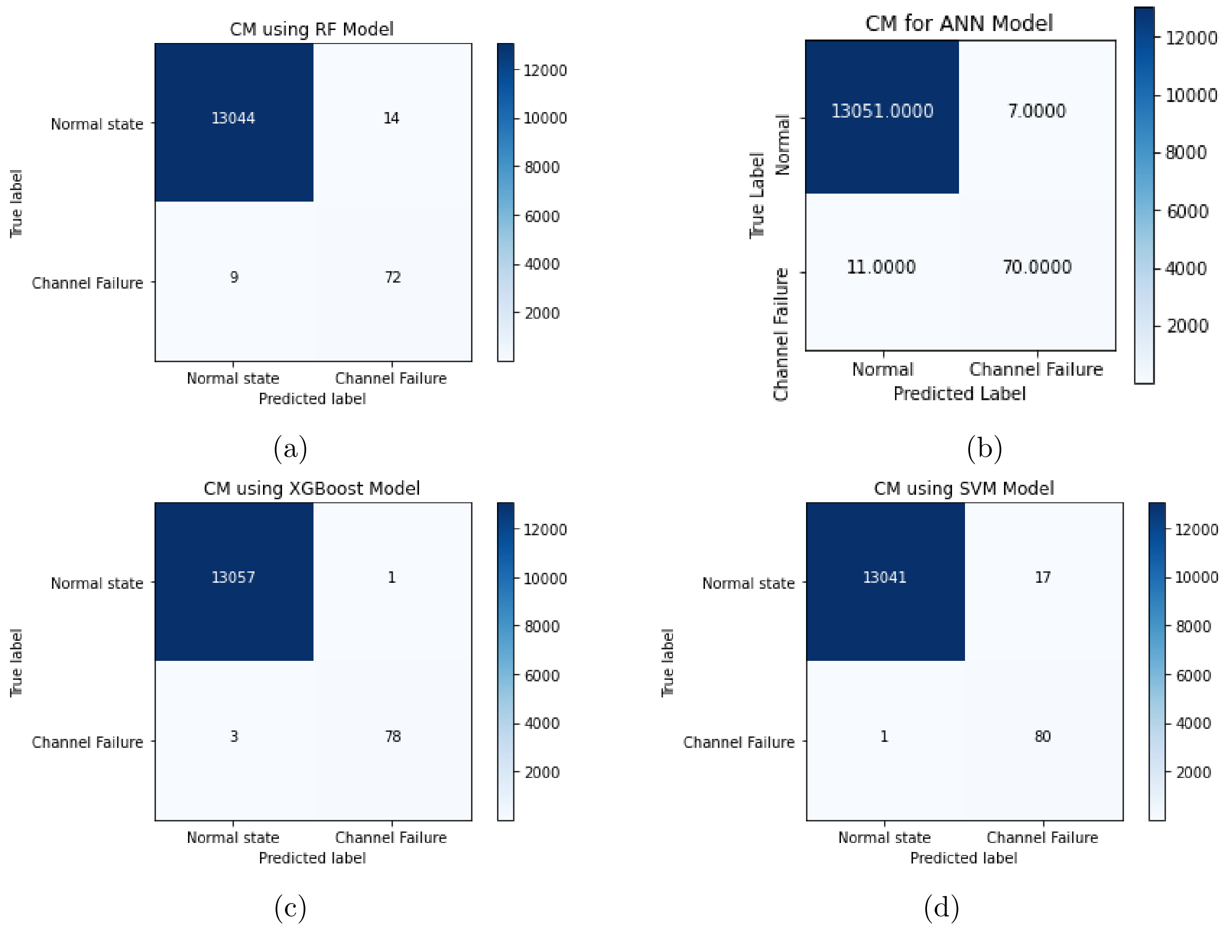


Figure 4.4: Confusion Matrix in (a)RF (b)ANN (c)XGBoost (d)SVM

In figure 4.4 the results demonstrate the model's strength in differentiating normal channel states from channel failures. It achieves a low false positive rate (identifying very few normal states as failures) and a high true positive rate (correctly identifying most failures). We are noting that the true positive value is significantly less than the true negative value in the confusion matrix for the balanced testing data. This imbalance in the true positive and true negative values suggests that the model's performance may be skewed towards the negative class.

In a balanced dataset, we would ideally want the model to have a similar level of performance for both classes. However, in this case, the imbalance in the true positive and true negative values indicates that the model may be more accurate at predicting the negative class (class 0) than the positive class (class 1).

To further evaluate the model's performance, we may want to consider additional metrics such as precision, recall, F1 score, and area under the ROC curve (AUC-ROC) to gain a more comprehensive understanding of how well the model is performing for both classes.

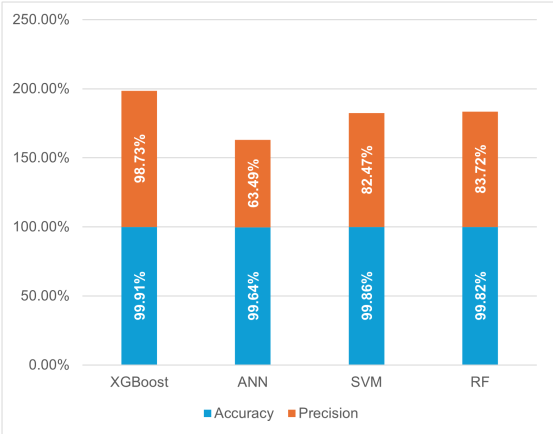
4.7.2 Accuracy,Precision, Recall and F1 Score Results

The performance of considered ML models are evaluated using classifier model performance evaluation metrics:

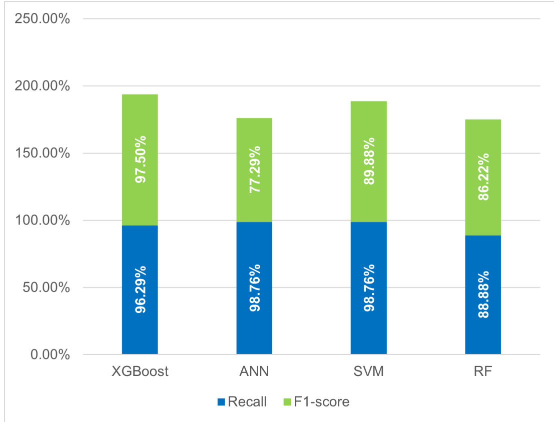
- Accuracy: $(TP + TN)/(TP + TN + FP + FN)$
- Precision: $TP/(TP + FP)$
- Recall (Sensitivity): $TP/(TP + FN)$
- F1 Score: $2 * (Precision * Recall)/(Precision + Recall)$

Table 4.7: Models performance

Model	**Accuracy**	**Precision**	**Recall**	**F1-score**
RF	99.82%	83.72%	88.88%	86.22%
SVM	99.86%	82.47%	98.76%	89.88%
XGBooost	99.91%	98.73%	96.29%	97.5%
ANN	99.64%	63.49%	98.76%	77.29%



(a)



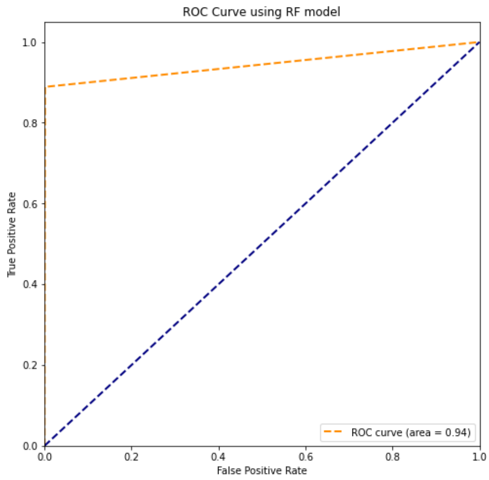
(b)

Figure 4.5: Four classifier models (a) Accuracy and Precision, (b) Recall and F1 score.

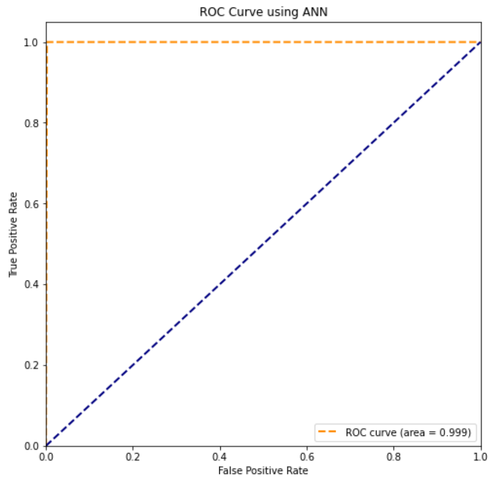
Our analysis revealed that the XGBoost model emerged as the strongest performer across all evaluation metrics. With an impressive accuracy of 99.91%, it correctly classified a near-perfect portion of the data. XGBoost also excelled in precision (98.73%), indicating a high reliability in its positive predictions (identifying true issues). While other models achieved comparable recall (ability to find actual positive cases) and XGBoost's balanced F1-score (97.50%) solidify its position as the most effective model for this task.

4.7.3 ROC Curve

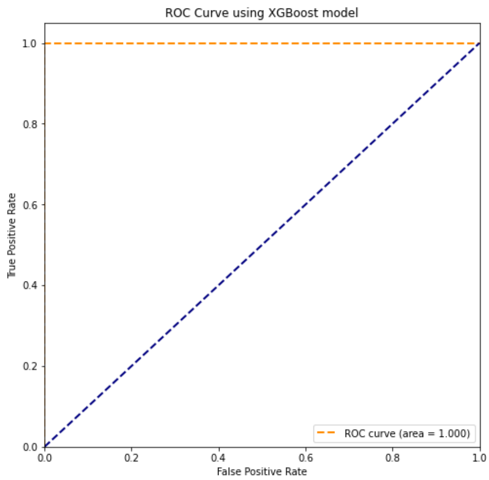
The ROC curve demonstrates the models' ability to discriminate between classes.



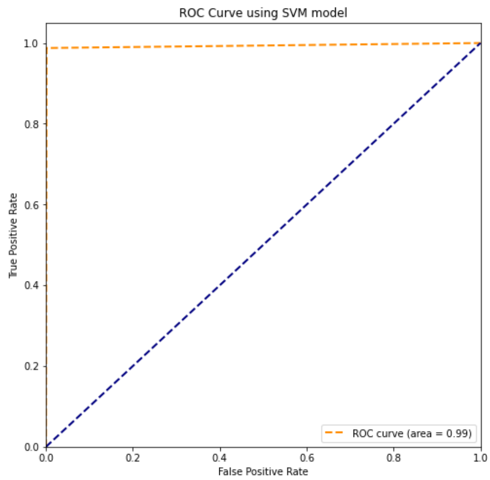
(a)



(b)



(c)



(d)

Figure 4.6: ROC curve in (a)RF (b)ANN (c)XGBoost (d)SVM

Figure 4.6 shows a helpful ROC curve that summarizes the performance of various models in classifying between normal and failure states. This curve allows us to compare how well each model distinguishes between these two classes. Among the models tested, XGBoost and SVM achieved the best performance, demonstrating their exceptional ability to differentiate between normal and failure data.

4.7.4 Model Execution Time

We also assessed the models' efficiency by calculating the average time it takes to train and test data instances. These results are presented in Table 4.8.

Table 4.8: Average times taken for the models in the training and test data instances

	RF	**SVM**	**XGBoost**	**ANN**
Av.Training Time (Sec.)	14.91	10.19	5.42	23.04
Av.Test Time (Sec.)	0.60	0.34	0.20	0.76

Table 4.8 clearly shows that the XGBoost model trains and tests significantly faster than other models. This efficiency, combined with its high accuracy and superior performance across all metrics, makes XGBoost the optimal choice for this dataset.

4.8 Model Selection

A comprehensive evaluation using various metrics revealed XGBoost as the standout performer among the evaluated classifier models. This model excelled in pinpointing the root cause of OTN channel failures, achieving exceptional accuracy and a robust F1-score. XGBoost's strength wasn't limited to just these key metrics. Its performance remained consistently balanced across a diverse range of evaluation criteria. This well-rounded performance across the board solidified its overall effectiveness for this specific task. To further solidify its position as the optimal choice, XGBoost boasted remarkably rapid training and testing speeds, making it an ideal solution for real-world applications.

XGBoost at the forefront, we can now leverage its capabilities for the crucial task of selecting the most influential features for RCA. This paves the way for a more targeted and efficient approach to identifying the root causes of channel failures, ultimately leading to improved network performance and reliability.

4.9 Feature Importance

For a comprehensive analysis of feature importance, this work employs two distinct methods in XGBoost.

4.9.1 Feature Importance Using Built-in Method

XGBoost offers a toolbox for understanding feature importance. This work focuses on the concept of 'cover,' which reveals how frequently a feature is used to make splitting decisions during the construction of decision trees. In simpler terms, cover tells us which features have the most influence in guiding the model's decision-making process.
XGBoost considers each feature one by one and evaluates how well it separates data points into groups with distinct failure probabilities. Here's the breakdown of cover:

- Splitting on a Feature: XGBoost analyzes each feature and identifies a value (threshold) that best separates the data.

- Impact on Data Points: XGBoost calculates how many channels are affected by this split based on the chosen feature and its threshold. In simpler terms, it counts how many channels fall on either side of the split line created by that feature.

- Cover Score: Higher cover indicates a feature is used to split data points more frequently across the entire tree. This suggests the feature has a broader impact on the model's decision-making process.

4.9.2 Feature Attribution Based on SHAP Value

SHAP values explain how each feature contributes to a specific prediction. They offer a more nuanced understanding of feature importance compared to the built-in methods, considering feature interactions and providing local explanations for individual predictions. Here's a breakdown of how SHAP values work to explain feature importance in XGBoost, going beyond the built-in methods:

- Individual Predictions: Instead of looking at the entire model, SHAP focuses on explaining a single prediction at a time. It analyzes how each feature in that specific instance influences on failure or non-failure for of OTN channel.

- Feature Interactions: Unlike built-in methods that treat features independently, SHAP considers how features interact with each other.

4.10 Model Performance Under Unbalanced Data

Our analysis of channel failures in OTN relied on imbalanced data. To address this, we used a technique called SMOTE to create a more balanced dataset. We then compared the performance of the XGBoost algorithm in classifying channel failures under both balanced and imbalanced conditions. The features used for the classification model remained the same. To assess the impact of data imbalance, we focused on the F1 score metric.

While the XGBoost algorithm achieved a high accuracy of 96.79% on the imbalanced data, its F1 score was lower at 95.3%. This indicates that the classification performance suffered when using the imbalanced data, compared to the results presented in section 4.7.2 where the data was more balanced.

Chapter 5: **Result and Discussion**

5.1 Result

The preceding chapters laid the groundwork for a robust approach to identifying the root cause of failures in OTN channels. We meticulously explored the potential of ML in this domain.

A core aspect of our investigation involved a comprehensive evaluation of four distinct ML classifier models. We fine-tuned their hyperparameters, to identify the model that excelled in accurately pinpointing the root cause of OTN channel failures. Our analysis revealed that XGBoost was the best performer.

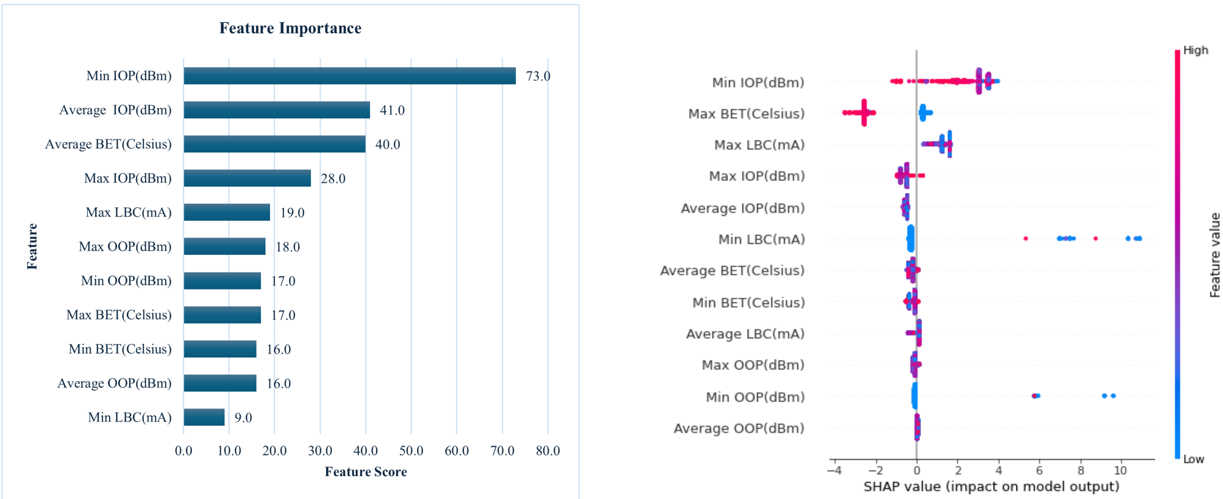
Table 5.1: Summary of XGBoost Model performance

Metrics	**XGBoost**
Accuracy	0.9991
Precision	0.9873
Recall	0.9629
F1-score	0.9750
AUC	1
Av.Training Time (Sec.)	5.42
Av.Test Time (Sec)	0.20

XGBoost model achieved an impressive accuracy of nearly 99.9% and precision of almost 99% when classifying channel states. This means it could very accurately distinguish between normal operations and failures.

5.1.1 Important Feature Analysis

We then delved deeper to understand what factors contributed most to these failures. Two methods, built-in analysis within the XGBoost model and SHAP values, both identified Min IOP as the most critical factor. This suggests that significant drops in the light received by the channel are a strong indicator of potential problems.



(a) (b)

Figure 5.1: Feature Importance Based on (a) Cover value (b) SHAP value

Figure 5.1(a) showcases the initial analysis using Cover. Here, "Min IOP," representing the minimum received light level, emerges as the standout feature. This suggests that when the model first investigates a potential channel failure, it prioritizes how significantly the light level has plummeted. Consequently, based on Cover analysis, Min IOP appears as the most critical factor, followed by average IOP(dBm), for identifying potential failures within the network.

Delving deeper with SHAP analysis in Figure 5.1(b), we observe that "Min IOP" remains the most influential feature. Removing it would have the most significant impact on the model's predictions. This reinforces the importance of checking input optical power as a crucial first step when diagnosing failures. Overall, the SHAP analysis provides valuable insights into our XGBoost model's decision-making process for channel failure detection. It highlights the critical role of minimum input optical power as a strong indicator of potential failures based on significant drops in received light. Additionally, the model prioritizes extremely high board environment temperatures (Max BET(Celsius)) as another key factor. While other features like average IOP(dBm) and LBC(mA) show some influence, they appear less important. Notably, features related to OOP(dBm) have minimal influence based on their SHAP values. This analysis is instrumental in prioritizing data collection efforts and model development by focusing on the most crucial features for accurate failure detection.

The following sections delve deeper into the intricate relationship between channel failures and the two most critical features identified by both analysis methods. This comprehensive analysis aims to shed light on how these features interact and contribute to network issues. By dissecting these interactions, we can gain valuable insights into the underlying mechanisms that trigger failures, ultimately paving the way for more effective preventative measures.

(a) Interaction of OTN Channel Failure with Min IOP

To visualize the relationship between Min IOP and channel failures, we employed a scatter plot. This graphical representation revealed a clear and compelling connection: whenever there was a significant dip in the Min IOP readings, the counts of "SM Errored Seconds" spiked dramatically. This pattern strongly suggests that a decrease in the minimum received light power directly correlates with an increase in channel errors, potentially leading to failures.

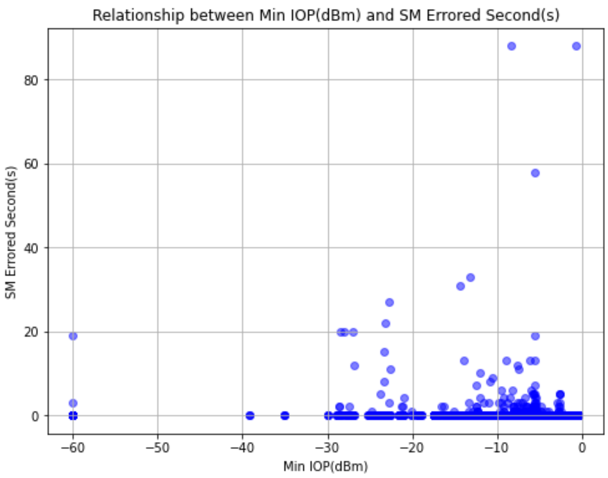


Figure 5.2: Interaction of SM Error second with Min IOP

(b) Interaction of OTN Channel Failure with Average IOP

Our built-in cover score method also highlighted 'Average IOP' as the second most influential feature. This metric captures the average level of light received by the channel over a specific period. The following figure takes a deeper dive into the potential interaction between Average IOP and channel failures. We'll explore how variations in this average light level might influence the "SM Errored Seconds" parameter, the indicator we use to identify channel issues. Higher counts of "SM Errored Seconds" directly correlate to a higher occurrence of channel problems.

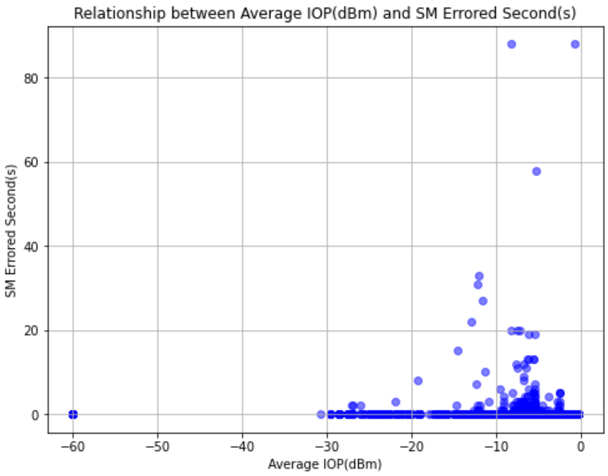


Figure 5.3: Interaction of SM Error second with Average IOP

(c) Interaction of OTN Channel Failure with Max BET

While multiple factors can contribute to OTN channel failures, our SHAP analysis identified Max BET as a critical second feature. To understand its influence, we specifically plotted the interaction between this temperature and "SM Errored second," a key indicator of channel issues. This visualization helps us identify potential patterns. Analyzing such interactions between environmental factors and failure indicators is crucial for pinpointing root causes and ultimately developing preventive measures for OTN channel failures.

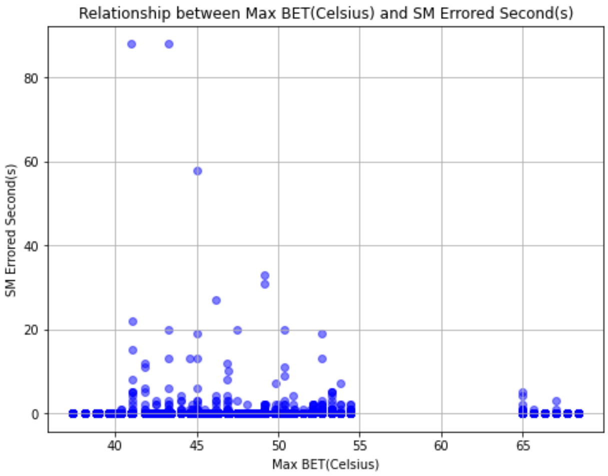


Figure 5.4: Interaction of SM Error second with Max BET

5.1.2 RCA Based on the Important Feature

We employ two approaches to analyze the root cause of OTN channel failures based on the important features.

(a) Potential Causes for Low Min IOP

Identify potential causes that might contribute to low Min IOP. Here are some possibilities:

- Fiber Degradation: Over time, fiber optic cables can degrade due to micro-bends, dust, or moisture ingress, leading to power loss.
- Connector Issues: Dirty, damaged, or misaligned connectors can cause significant power loss at connection points.
- Source Power Fluctuations: Instability in the power source feeding the transmitter can lead to variations in Min IOP.

- Component Malfunction: Failure of optical components like transmitters or amplifiers can lead to reduced power output.

(b) Domain Knowledge and Verification for Cause of Low Min IOP

Maintaining optimal performance in OTN channels relies heavily on a metric called Min IOP. Industry experts have identified external factors like splicing losses due to rough handling, connector wear, and tight fiber bends that can all contribute to low Min IOP. This aligns with the findings from XGBoost, a powerful ML tool, which highlights Min IOP as the most critical factor for predicting OTN channel failures. The link between low power levels and network issues is clear. Furthermore, by combining XGBoost's data analysis with real-world knowledge, we can pinpoint extrinsic fiber losses as a significant culprit behind low Min IOP. This deeper understanding empowers us to effectively address the root causes of OTN channel problems.

5.2 Discussion

Understanding the most important features that influence failures is crucial for improving network performance. XGBoost helped us identify Min IOP as the key factor. Industry knowledge aligns with this finding, highlighting external factors like splicing loss and connector issues that can contribute to low Min IOP.

By focusing on maintaining optimal Min IOP levels and understanding the reasons behind low Min IOP, we can take proactive measures to prevent failures and improve the overall performance of our OTN channels.

These extrinsic losses contribute to the overall signal attenuation, potentially causing the received power to fall below the Min IOP threshold. When this happens, signal quality degrades, leading to errors, increased noise, and ultimately, communication failures.

Practical Implications for Mitigating Extrinsic Fiber Loss

The findings of this study highlight the importance of proper fiber optic cable handling and maintenance practices to minimize extrinsic fiber loss and ensure sufficient Min IOP. Here are some key takeaways for achieving this:

- Following Industry Best Practices: Implementing established industry guidelines during fiber optic cable installation and maintenance is crucial.
- Proper Techniques: Utilizing proper techniques for splicing, connector cleaning, and overall cable handling helps minimize potential damage and misalignment.
- High-Quality Equipment: Utilizing low-loss connectors and adhering to recommended bend radius specifications further reduces signal attenuation.

By implementing these mitigation measures, network operators can ensure reliable data transmission and minimize the risk of communication failures caused by insufficient Min IOP.

Chapter 6: **Conclusions and Future Work**

6.1 Conclusions

Our analysis identified XGBoost as the optimal model for classifying OTN channel failures. While the initial model struggled with imbalanced data, using SMOTE to create a balanced dataset yielded excellent performance metrics, including high accuracy and a balanced F1 score.

XGBoost's feature importance analysis revealed Min IOP as the most critical factor influencing failure prediction. This highlights the importance of monitoring light levels within the OTN channel for early detection of potential issues. Further insights were gained through SHAP analysis, which confirmed Min IOP's significance and identified Max BET as another key factor.

By prioritizing these critical features, we can optimize data collection and model development for even more effective failure detection. Additionally, the identified potential causes for low Min IOP, such as fiber Splicing loss, fiber cable bending, and connector issues, provide valuable guidance for implementing preventive maintenance strategies and ensuring optimal network performance.

6.2 Future Work

This study focuses on RCA of OTN channel failures in general. However, it is important to conduct separate analyses for soft and hard failures. Additionally, while this study utilizes performance data, future research could explore RCA of OTN channel failures using alarm correlation methods.

Bibliography

[1] M.Dorland, *Why you should Understand the Optical Transport Network Evolution and its Implications on your Business Now* (Noord Holland, Netherlands, BS, 2022).

[2] C.Mas and P.Thiran, *An Efficient Algorithm for Locating Soft and Hard Failures in WDM Networks* (IEEE J. on Selected Areas in Commun. Malaysia, 2000).

[3] M.Lucke, M.Chioua, C.Grimholt, M.Hollender, and N.Thornhill, *Advances in alarm data analysis with a practical application to online alarm flood classification* (Journal of Process Control, Ladenburg, Germany, 2019).

[4] D.Michel, T.Clovis, T.Christian, A.Mamadou, M.Sone, *Machine Learning-Based Alarms Classification and Correlation in an SDH/WDM Optical Network to Improve Network Maintenance* (University of Yaoundé I, J. of Computer and Commun., Cameroon, 2023.

[5] L.Baptista, *A Comprehensive Analysis of Alarm Root Causes in an Optical Fiber Network* (FACULTY OF ENG. OF THE UNIV. OF PORTO, Portugal,2023).

[6] C.ZHANG, D.WANG, L.WANG, L.GUAN, H.YANG, Z.ZHANG, X.CHEN, AND M.ZHANG, *Cause-aware failure detection using an interpretable XGBoost for optical networks* (Lab. of Info. Photonics and Optical Commun., Beijing , China, Sep 2021).

[7] T.Hayford-Acquah, B.Asante,*Causes of Fiber Cut and the Recommendation to Solve the Problem* (Jornal of Elect. and Commun. Eng., Ghana, 2017),PP 46-64.

[8] J.Brownlee *"Feature Importance and Feature Selection With XGBoost in Python"*, August 27, 2020.

[9] A.Schubert*G.709 – The Optical Transport Network (OTN)* ,VIAVI Solutions Inc, 2021.

[10] X.Jiang, *"Optical performance monitoring in optical transport network"*, Advanced Techniques for Next-Generation Photonic Networks, Germany, 2010.

[11] Larry, Wilmington, *"The Layers of Optical Transport Network: Core, Aggregation, and Access Layer"*, 380 Centerpoint Blvd, United States, Jan 27, 2024.

[12] Trebino, "Ethernet Interfaces User Guide for Routing Devices", Sunnyvale, California , United States, 12-Feb-24.

[13] M.TECH, Quick note on Pre-FEC , Post-FEC ,BER and Q relation, mapyourtech, Sydney, Australia September 2, 2020.

[14] M.TECH, OSNR & Q-factor Dependency: Understanding The Impact On Signal Quality And Network Performance, mapyourtech, APRIL 16, 2023.

[15] M.Nelson, *root-cause-analysis*(Article of tableau.com, California, United States, Aug 15, 2021).

[16] S.Badillo1, B.Banfai1, F.Birzele, I.I.Davydov, L.Hutchinson, T.Kam-Thong, J.Siebourg-Polster1, B.Steiert and J.D.Zhang, "An Introduction to Machine Learning",Innovation Center Basel, Basel, Switzerland, January 15, 2020.

[17] Dorogush, A.V., Ershov, V. & Gulin, A., "CatBoost: gradient boosting with categorical features support", 2018).

[18] I. Guyon and A. Elisseeff, *An introduction to variable and feature selections* (Journal of Machine Learning Research, 2003).

[19] Sarker.H, *Ai-driven cyber security: an overview, security intelligence modeling and research directions*,SN Comput Sci. 2021.

[20] Han J, Pei J, Kamber M, *Data mining: concepts and techniques*,Amsterdam: Elsevier, 2011.

[21] Mohammed.M, Khan.M, Bashier.M.B *Machine learning: algorithms and applications*,(CRC Press, 2016).

[22] B.Mahesh, *Machine Learning Algorithms - A Review*, (International Journal of Science and Research (IJSR), 2018).

[23] T.Shin,Understanding Feature Importance in Machine Learning, Nov.07, 2023.

[24] A.Ashraf, Correlation in machine learning, Data Analytics, Data Science Tech., Medium, Sep 22,2023.

[25] C.Lee, Feature Importance Measures for Tree Models — Part I,Published in Veritable; Medium, Oct 28, 2017.

[26] D.Tech, *Connectrix - MDS Series Cisco: Unable to bring port online due to exceeded Bit-Error threshold*, (Article Number: 000043834, DELL Technology, 13 Jun 2021)

[27] Mellanox, Bit Error Rate (BER), InfiniBand Fabric Diagnostic Tool User Manual v2.8.0, Mellanox Technologies Ltd. in the U.S Last updated on May 23, 2023.

[28] M.optics, *Optical Fiber Loss and Attenuation*, (Acadamic press, meetoptics, optics and photonics researchers, Barcelona).

[29] G.geeks, Random Forest Classifier using Scikit-learn, geeksforgeeks.org,31 Jan, 2024

[30] Z. Wang, M. Zhang, D. Wang, C. Song, M. Liu, J. Li, L. Lou, and Z. Liu, "Failure prediction using machine learning and time series in optical network",Opt. Express 25(16), 18553–18565 (2017).

[31] J. Borland, "Analyzing the internet collapse",MIT Technol. Rev., 2008.

[32] J.R. Thompson and Rajarshi Roy, Phys. Rev. A **43**, 4987 (1991).

[33] G.Singh,*Introduction to Artificial Neural Networks*,(Automated Data Analysis in Astronomy, New Delhi, India, March 2001).

[34] G.blog, "Introduction to XGBoost Algorithm in Machine Learning", analyticsvidhya, India, Apr, 2024.

[35] S.R.Vadalia, S.S.Chavan, K.R.Bhere, F.Nabi and A.Z.Pate, "Waste Detection and Water Quality Assessment inAquatic Environments", (A Comprehensive StudyUsing YoloV8 and XGBoost), Amity University Maharashtra, Mumbai, India, January 4th, 2024).

[36] M.Shinkfield, "What is the Impact of Temperature on Network Equipment?", Melbourne, Jul 14, 2023.

[37] F.Bre and J.M.Gimenez *Prediction of wind pressure coefficients on building surfaces using Artificial Neural Networks*,(Article Energy and Buildings 158, Argentina, November 2017).

[38] rockcontent *The Top 8 Root Cause Analysis Tools for Troubleshooting*,(Rockcontent, Mar 6, 2022).

[39] R.S.Mulpuri*Machine Learning for Fast and Accurate Root Cause Analysis*,(5201 Great America Parkway, Suite 320, Santa Clara, CA 95054, Sep 21, 2023).

[40] S.Mattoo *What Are Machine Learning Models? How to Train Them*,February 9, 2023.

[41] Y.Wang *A hybrid ensemble method for pulsar candidate classification*, Chinese Academy of Sciences , August 2019.

.