Addis Ababa Institute of Technology

School of Electrical and Computer Engineering

Telecommunication Engineering Graduate Program

Machine Learning Based Soft Failure Detection by Exploiting Optical Channel Forward Error Correction Data

By: Dawit Tadesse

Adviser: Dr. Yalemzewd Negash

A Thesis Submitted to the School of Graduate Studies of Addis Ababa University in Partial Fulfilment of the Requirements for the Degree of Master Science in Telecommunication Network Engineering.

September 2023

Addis Ababa, Ethiopia

Addis Ababa Institute of Technology

School of Electrical and Computer Engineering

Telecommunication Engineering Graduate Program

Declaration

I, the undersigned, declare that the thesis covers my work in compliance with internationally accepted practices; I have fully acknowledged and referred all materials used in this thesis work.

Name: Dawit Tadesse

Signature__________

As a university advisor, I approved the submission of this thesis for examination.

Name: Dr. Yalemzewd Negash

Signature__________

Addis Ababa Institute of Technology

School of Electrical and Computer Engineering

Telecommunication Engineering Graduate Program

Machine Learning Based Soft Failure Detection by Exploiting Optical Channel Forward Error Correction Data

By

Dawit Tadesse

Signed by the Examining Committee:

Yalemzewd Negash (PhD)	_____________	__ _________
Advisor	Signature	Date
Tsegamiak Terefe (PhD)	_____________	__ _________
Examiner 1	Signature	Date
Dereje Hailemariam (PhD)	_____________	__ _________
Examiner 2	Signature	Date

Abstract

The performance of optical channel degrades because of soft failures, such as filter failures, laser drift, and system aging. If such degradation is handled correctly and promptly, soft failures will not affect services. A crucial element in the protection against failures in optical channels is soft failure detection. Traditional approaches, however, find it difficult to complete this task because of their limitations in adapting the dynamic behavior of soft failure, requirements for professional manual intervention, and vastly increased optical performance data.

The aim of this thesis is to detect soft failures based on machine learning by exploiting optical channel forward error correction data. Soft failures detection is explored using three ML algorithms: support vector machines (SVM), artificial neural networks (multilayer perception), and random forests (RF).The input of ML algorithms is Pre forward error correction bit error rate (Pre-FE-BER) that captured from forward error correction data on real optical channels. We implemented feature labelling and extraction based on the behavior of a time series window. We use stratified shuffle split method in cross-validation approaches to optimize and validate algorithm performance in terms of confession matrix, accuracy, and building time. As a result, RF with significant features, which has a validation accuracy of 99.2% and a standard deviation of 0.49%, is the best method. Beside, a lower computational complexity of 12 features and a building time of 17.5ms were determined.

Keywords: Quality of transmission, soft failure detection, support vector machine (SVM), artificial neural network (multilayer perception), and random forest (RF),

Contents

List of Figures

List of Tables

List of Acronyms

ANN	Artificial neuron network
ASE	Amplifier Spontaneous Emission
BER	Bit error rate
CNN	Convolutional neural network
CD	Chromatic dispersion
DWDM	Dense wavelength division multiplex
DSP	Digital signal procreating
EDFA	Erbium-doped fiber amplifier
EPDM-BPSK	Polarization-multiplexed binary phase shift keying
FEC	Forward Error Correction
FEC-CE	Forward Error Correction - Corrected Errors ()
FEC-UB	Forward Error Correction - Uncorrected Blocks ()
FOADM	Fixed optical add/drop multiplexer
GMI	Generalized mutual information
FS	Filter Shift
FT	Filter Tightening
IQR	Interquartile range
MI	Mutual information
INI	Inter-channel Interference
ML	Machine Learning
NLI	Non-linear Impairments
NN	Neural network
OCh	Optical channel,

ODU3	Optical data for 40 GB
OP	Optical Received Power
OSNR	Optical Signal to Noise Ratio
OTN	Optical Transport Network
PMD	Polarization mode dispersion
Pre-FEC BER	Pre-Forward Error Correction Bit Error Rate
PSD	Power spectrum density
Q-factor	Quality factor
QoT	Quality of Transmission
RA	Raman amplifier
ReLu	Rectified linear units.
RF	Random forest
ROADM	Reconfigurable optical add and drop multiplexer
SNR	Signal-to-Noise Ratio
SVM	Support Vector Machine
TT	Trouble ticket

Acknowledgment

First, I would like to express deepest appreciation to saint virgin marry and almighty God for granting me the capacity to complete my thesis. Apart from my efforts, I take this opportunity to thank my class mate, teachers and everyone who contributed to the complication of this thesis.

I would like to express my genuine gratitude to Yalemzewd Negash (PhD) **as** the thesis advisor. I cannot say thank you enough for his grateful, continuous help and support. Without his help and guidance, this thesis would not have been realized. I like to take this opportunity to thank my defence examiners Tsegamiak Terefe (PhD) and Dereje Hailemariam (PhD) for their generous support, constructive criticism, and ideas during my progressive defence.

I am grateful to acknowledge Ethio telecom staff members, particularly those from transmission maintenance, humane resource and management staff, who have shown heart-warming support in my study

Finally, and y, I incredibly thankful to my beloved family my mom, Berhane Badada, my wife Yewbdar Shimels, and my four beautiful children Natan, Eldana, Eyob, and Nafkaot Dawit for their truthful love and inspiration thought throughout academic year and the completion of this thesis\

Chapter 1

1. Introduction

Backbone optical transport network failures divide into soft and hard failures. Hard failures are unpredictable because they are triggered by sudden disruptive events (such as fiber cuts, equipment failures, and power failures). Soft failures degrade the quality of transmission**.** Such degradation occurs from millisecond gradually over a large time scale (hours or days), sometimes leading to service interruptions (hard failure). The most common soft failures are filter shift (FS), filter tightening (FT), amplifier spontaneous emission (ASE), inter-channel interference, power attenuation, and non-linear impairments (NLIs) [1]. Soft failure is potentially predictable and anticipated by adopting proactive recovery mechanisms, such as network protection and reconfiguration [2]. In order to perform recovery mechanisms, the first soft failure has to be detected by the optical channel to the failure. Soft-failure detection is the most important issue the survivability strategy.

Soft-failure detection is a process that notices abnormal behaviors without pinpointing the faulty element. It is typically implemented by continuously monitoring the quality of transmission parameters. Such as pre-forward error correction bit error rate (Pre-FEC BER), optical received power (ORP), optical signal- to-noise ratio (OSNR), and quality Factor (Q-factor) [3], the dedicated monitors installed at optical transmitters and receivers, regenerators, or intermediate nodes monitor and collect quality of transmission performance parameters. Those data used to build data-driven classification models with different machine learning algorithms.

In recent years, machine learning (ML) has become one of the most famous and powerful algorithms studied for soft failure detection. Machine learning is a subfield of artificial intelligence that delivers the ability to learn from data, patterns recognition, improve from experience without sticking to explicit instruction, and settle an issue with minimal human involvement. The large amount of data collected and elaborated on in each node and centralized location can be an input source to automate soft-failure detection based on ML, which is used to handle more complex and large-scale failure problems with intelligence, efficiency, and powerful solutions [4].

1.1 Statement of the Problem

The quick development in consumer and service demand generated a demand for high capacity and effective optical networks. However, soft failures in the optical transport network are causing to have lower quality of transmission. Those failures can occur because of system aging, fiber loss, filter failures (filter shrinking or misalignment), and equipment malfunction. If soft failure detection is not handled properly and timely, soft failures will not only cause direct revenue loss, but they also increase maintenance costs and degrade service quality.

The fourth generation optical network present fault management system (FMS) is based on conventional pre-defined thresholds made by laborious threshold-setting operations. The monitoring team will manage the alarm based on the alarm type or the network equipment by creating a trouble ticket (TT) and assigning it to the maintenance staff who have direct responsibility to handle it if a performance event exceeds the pre-defined thresholds. The required background knowledge, technical aptitude, familiarity with the network, and equipment knowledge for maintenance personal.

There are restrictions on how well this manual soft failure detection can adapt to a complicated and dynamic circumstance where expert manual intervention is needed. High labor expenses and unavoidable human error are the results. It also depends on the previous data, reactive failure management system, and knowledge of the monitoring crew. Because of being primarily designed to detect hard failures, this results in longer troubleshooting durations and a lack of ability to detect soft problems. In contrast to soft failures, hard failures are simple to detect and have a set meantime to repair and network downtime. Therefore, it is preferable to incorporate an intelligent machine learning-based soft-failure detection modal to replace these manual and ineffective failure management approaches.

A machine-learning algorithm and several large data analysis methods are employed for a soft failure detection system. These methods make it possible to track, gather, and evaluate data about optical network soft failures in real time. Machine learning provides strong capabilities for soft failure detection. The algorithm generates intelligent, precise, and affordable solutions, since it has the ability to learn from intricate mapping between samples or attributes retrieved from data. This thesis proposes machine learning-based backbone optical network soft failure detection.

1.2 Objective

1.2.1 General Objective

The general objective of this thesis is to detect soft failures based on machine learning by exploiting optical channel forward error correction data to improve the soft failure management system

1.2.2 Specific Objectives

The specific objectives of the thesis are:

- To exploit and determine soft failure, quality of transmission parameters, and optical network element behaver.
- To monitor, collect and analysis quality of transmission parameters from real optical channels scenario.
- To select relevant usage fields of quality of transmission parameters
- To build a data set using suitable data preparation techniques.
- To evaluate and compare an effective machine-learning algorithm based on their performance.
- Finally, to recommend the best performed soft failure detection modal.

1.3 Scope and Limitation

This research work focuses on soft failure and its detection mechanism. Thus, this analysis is limited to three months’ quality of transmission parameters on a real Ethio telecom backbone optical network.

1.4 Contributions of the Thesis

The result of this thesis is deliberating scientific contribution on soft failure detection. Besides, the following list of contributions is considered.

- It helps to telecom operators in preventing service outage by degradation of the quality of transmission.
- The feature label approaches contribute as an input for other related studies.
- It provides benefit to researchers to use as a citation and further research area on soft failure detection.

1.5 Methodology

In order to accomplish the objectives of this research, we followed the following methodologies as duplicated on Figure 1.

- Literature review: review of literature and related studies.
- Data collection: collect legitimate data from the backbone optical network management system.
- Data preparation: transform data in to correct machine-learning format.
- Modal evaluation: perform optimization, simulation setup, and evaluate model's performance to select better performance modal.
- Detection

1.6 Related Works

Soft failure detection introduced based on machine learning methods in optical networks recently.

The authors in [5] focused their work on three different methods of machine learning to detect and identify soft failures in the 380-km optical channel on the optical network. The authors experimentally tested their models on synthetic BER degradation caused by filter misalignment and an undesired amplifier-gain reduction. The features considered by the authors are statistical parameters (maximum, minimum, mean, and standard deviation) of the BER. The labeled failures were defined by applying BER changes to a threshold detected during observation periods. The results show that, for failure detection cases, higher performance is achieved by low sampling time or longer sampling time with longer windows, and the RF algorithm outperformed all other algorithms in terms of prediction accuracy. In the failure identification case, we need much smaller sampling period is needed compared to the failure detection case.

While in [6], the authors proposed identification of soft failure identification scheme using a convolutional neural network (CNN) running in the transmitters and receiver DSP. The input of the CNN is the power spectrum density (PSD) extracted from a coherent receiver. The output contains the identified causes of soft failures along with their probabilities. The author studied four failures affecting the QOT of an optical connection. Those factors are filter shift (FS), filter tightening (FT), increased amplified spontaneous emission (ASE), and the nonlinear impairment (NLI).

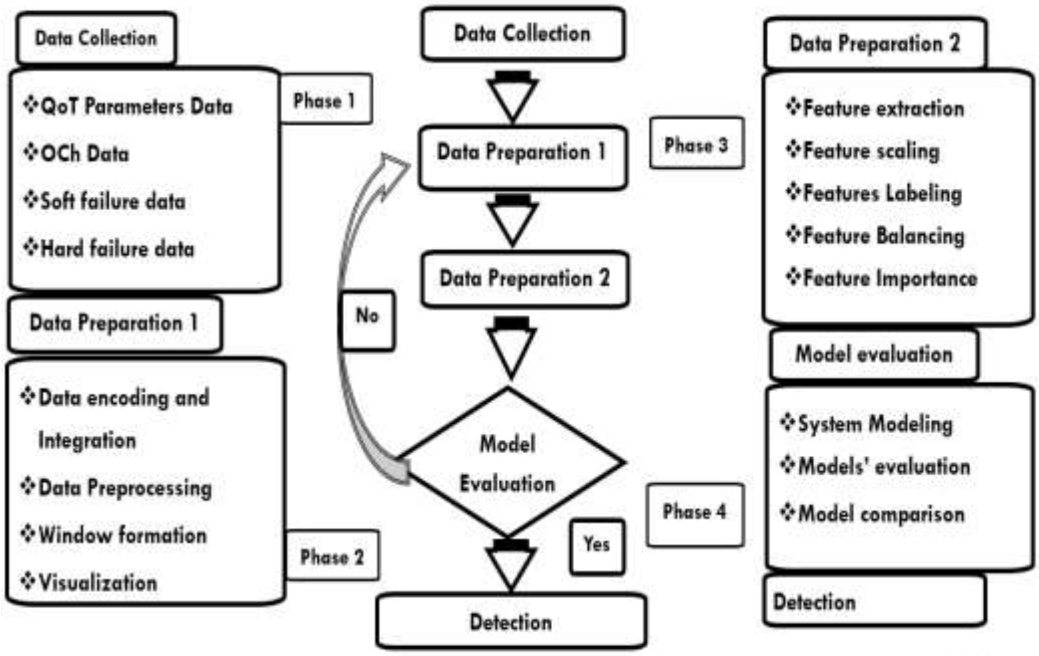

Figure 1: Soft failure detection methodology.

In the result, the authors presented that, when only one soft failure type exists, they achieved excellent accuracy. Whereas, when more than one soft failure types exist, they achieved less accuracy compared to when one soft failure type exists. The authors also considered one additional soft failure, which is fiber NLI.

Receiver DSP can compensate for chromatic dispersion (CD) and polarization mode dispersion (PMD).

To confirm the transponder's proper functionality, the authors of [7] looked at soft failures on the optical channel. By identifying the BER brought on by FS, FT, and laser drift. The authors of the study suggested two different algorithms. The BANDO (BER anomaly detection) method, which is based on BER shift detection in optical channel, is described as am FSM (finite state machine) to track the metered BER and issue notifications in case of rapid BER changes. The BER anomalies that are established using a set threshold are referred to as the labels. It was derived using the observed BER data's standard deviation and mean for monitored windows ranging from two minutes to ten days.

The LUCIDA (Failure Identification approach), the second approach, is based on identifying a probabilistic algorithm that analyzes time series from monitoring and notification data and delivers the all like hood failure class together with its probability will produce in all like hood failure sequence. When BANDO functions inside network nodes and delivers a message to the LUCIDA algorithm, which can run on a centralized controller, the authors also suggested quick degradation detection. They conducted experimental tests on two different configurations of optical network to determine the BER and power parameters needed for later simulations to generate fictitious findings. However, how well the parameters are set up determines how accurately the algorithm will find the ideal setup parameters are difficult, since optical networks are heterogeneous.

The author [8] focused on soft failure localization by proposing two techniques for active monitoring during commissioning testing and passive monitoring of operations. The first technique, Testing Optical Switching at the Setup Time (TISSUE), is based on specific low-cost optical test channel (OTC) modules deployed with collect BER measurements from each traversed node and then compare them to theoretical BER values. In cases of significant discrepancies, they display failure alarm at the considered span. The second technique, failure source localization, is suggested for optical networking (FEELING), which uses a machine-learning-based (decision tree) identification and localization platform that operates at optical channel by taking advantage of continuous monitoring of the optical spectrum using cost-effective OSAs installed in the optical nodes. FEELING predicts if component failed or not, and with failure, estimates the magnitude of the failure. The results showed that feeling identifies or localizes laser drift with 100% accuracy. With filter-related failures, FEELING can identify or localize a failure with an accuracy above 90%.

The authors also presented that the optical spectrum becomes asymmetrical with FS, and its edges are noticeably rounded in the case of FT. These irregularities allow for distinguishing optical spectra suffering from such failures from the properly configured ones. One of the challenging tasks is to differentiate between FT failure and degradation due to filter cascading effects. Both phenomena show a similar property in the shape of the optical spectrum. They identified and prevented this misclassification of the failed signal as the properly configured signal.

In this paper [9] proposes an SVM-based model to detect anomalies in the signal-to-noise ratio (SNR) of an optical channel. In their study, they defined the anomalies using the interquartile range (IQR) by considering the unique patterns of the anomalies. Rather than relying on the fixed threshold that is provided by the operators. The authors tested their models on six optical channels carried on a 100- to 1400-kilometer real network.

1.7 Thesis organization

The organization of the thesis is:

- Chapter 1: address the thesis's background, the problems it attempts to address, the goals and scopes, related works, and methods of study.
- Chapter 2: This chapter summarize network elements, quality of transmission, and soft failure types.
- Chapter 3: Deal with the machine learning classifiers and ML performance metrics.

- Chapter 4: Data preparation: describe the simulation setup and the data preparation process, including data preparation, feature labeling, and feature scaling, and training and testing data split.
- Chapter5: Present optimization, simulation setup, and discuss performance of the proposed ML algorithm using a performance measure matrix.
- Chapter 6: Conclusion Section: summarizes the objectives achieved from the assessment of the result, and we discussed possible future works.

Chapter 2

2. Basic of Soft Failure

Soft failures gradually affect the quality of transmission over a large time scale of hours or days. This section provides an overview of the network elements from which performance data is generated. It also explains what the observed parameters track and describes the soft failures generated for evaluation.

2.1 Optical Network Elements

Optical network elements include all the physical and optical section components involved in the soft failure detection of optical networks. This section summarizes provides an overview of the network elements from which performance data is generated.

2.1.1 Optical channel:

An optical channel is a portion of an optical network through which light passes unaltered. It is made up of wavelength that connects two network nodes and is sent via one or more intermediary nodes. The devices in optical channel are optical line amplifiers, fixed optical add and drop multiplexers, and reconfigurable optical add and drop multiplexers. A system-level management strategy for the key performance indicators of the system is soft failure detection on optical channels [10].

2.1.2 Optical amplifier

Optical amplifier: amplifies the power of the multiplexed optical signals to extend the transmission distance. Even though there are different amplification technologies, the erbium-doped fiber amplifier (EDFA) is the most widely applied commercial amplifier used in the current optical networks. It provides excellent energy conversion, stable high gain, little crosstalk, and low-cost power amplification system [11]. However, EDFA presented optical surge problem (noise figure), gain unflatted's, and it operates on fixed gain range.

In the long-haul dense wavelength division multiplex (DWDM) system, the front most end of the reception station contains the Raman amplifier (RA), a second type of amplifier. RA amplifies the optical signals while they are being sent by using distributed amplification of the backward and forward pumps. It benefits from tunable gain, little noise, and fewer nonlinear effects.

The combination of RA and EDFA offers wideband flat gain, a non-linear effect, less noise, and lengthy transmission distance. Gain decrease, pump deterioration, amplifier spontaneous emission (ASE), power anomaly, and power excursion are the soft failure types that are present in optical amplifiers.

2.1.3 Optical filter

A tool with wavelength-dependent transmission or wavelength-reflectance is an optical filter. Filters come in a variety of forms, including FOADM, ROADM, and optical multiplexer and demultiplexer.

- Multiplex and demultiplexer optical signals over various wavelengths using an optical multiplexer and demultiplexer
- ROADM: used to add, path through, drop any single or multi wavelength signals from a multiplexed signal, and route the signals to any port in any order, achieving flexible wavelength grooming in multiple directions. The key element of a ROADM is the wavelength selective switch (WSS), which performs dynamically route, block, and attenuates all the wavelengths in a fiber.
- FOADM: used to drop individual optical signals from a multiplexed signal and send these optical signals to the customer side. Besides, FOADM adds individual optical signals to multiplex optical signals and send to optical power amplifiers.

Filter failures occur frequently and such as filter shifting, filter tightening, and insertion loss [12]

2.1.4 Optical fiber and connector

A physical route called an optical fiber uses total internal reflection to transmit optical signals along its axis. Fiber is categorized into multi-mode and single-mode, which can send many lights and one light, respectively, based on the number of modes of light ray propagation. The most frequent and significant soft failure source is fiber failure namely, fiber bending, fiber aging, fiber breakage, and fiber nonlinearity [13].

Depending on the connectors and optical terminal type, optical connectors are passive parts used in fiber connectors. Reflected fault, extra loss, and angular fault are examples of connector soft faults can occur.

2.2 Quality of Transmission parameters

For optical networks to operate reliably and to provide required quality of service, effective soft failure detection is essential. It is set into practice by using continuously monitoring quality of transmission metrics. These parameters are deferential group delay (DGD), optical received power (ORP), pre forward error correction bit error rate (Pre-FEC BER), optical signal-to- noise ratio (OSNR), and Q-factor [3].

2.2.1 Deferential Group Delay (DGD)

The difference in propagation times between a signal's two Eigen modes, X and Y polarizations, is known as Deferential Group Delay (DGD). Because different modes' group velocities are frequently different, it happens. When using multimode fibers, it is typically a limiting issue for the transmission bandwidth that can be obtained and, for the transferrable data rate.

2.2.2 Pre-Forward Error Correction Bit Error Rate (Pre-FEC BER)

The total number of optical channel errors up to the hard symbol decision divided by the total number of optical channel bits is Pre-FEC BER [14]. Before the maximum BER that the equipment can correct is achieved, Pre-FEC BER is used to monitoring performance optical channel [7]. Generalized mutual information

(GMI) provide bit rate at receiver side by considering Gaussian channel modal assumption. GMI gives the maximum achievable rate for a modulation format and a bit to symbol mapping. To improve the quality of transmission in optical networks, forward error correction (FEC) is used to repair bit errors in the data that has been received.

Following a hard symbol decision, Pre-FEC BER introduced from the following FEC data.

- Forward Error Correction: Corrected Errors (FEC-CE) is the correction of byte errors using Forward Error Correction.
- The number of uncorrected blocks within a specific amount of time is Forward Error Correction: Uncorrected Blocks (FEC-UB).
- The bit rate for the optical data unit (ODU3) service type is 40319218 Kbits per second and 329492 blocks per second.

2.2.3 Optical Received Power (ORP)

The input power of the receiver at the receiving end of the optical channel is known as optical received power (ORP).

2.2.4 Optical Signal-to-Noise Ratio (OSNR)

The net signal power to the net noise ratio, given in dB, is defined as the optical signal-to-noise ratio (OSNR). Both passive (such as connectors and fiber) and active (such as lasers and amplifiers) devices can introduce noise. This nose is the main contributor to OSNR degradation. The relationship between OSNR and the signal-to-noise ratio (SNR) given by [15]:

$$\mathrm{OSNR[dB]} = \mathrm{SNR[dB]} + 10\log_{10}\left(\frac{\rho B_e}{2B_o}\right) \quad \text{Equation 1}$$

Where p is the number of polarizations, be is the bandwidth of electrical signal and B_o= the reference bandwidth.

$$\mathrm{SNR[dB]} = \frac{P_s}{P_n} \quad \text{Equation 2}$$

Where P_s is the signal power and P_s the noise power

2.2.5 Q-factor

The Q-factor is the signal-to-noise ratio at the decision circuit in voltage or current units. The Q value directly shows the network performance. The higher the Q-factor and OSNR, the lower the BER. Hence, it gives the absolute quality of an optical signal, and the parameter is dimensionless [16]. Mathematical relation between Q-factor and BER is: [17]

$$\mathbf{BER} = \frac{1}{2}\mathbf{erfc}\left(\frac{Q}{\sqrt{2}}\right) \qquad \text{Equation 3}$$

Whereas

$$\mathbf{erfc}\left(\frac{Q}{\sqrt{2}}\right) = \frac{2}{\sqrt{\pi}}\int_0^{\frac{Q}{\sqrt{2}}} e^{-t^2}\,dt \qquad \text{Equation 4}$$

2.3 Soft failure type

Soft failures degrade the quality of transmission, and sometimes these soft failures lead to hard failures. In a real optical network, there are many soft failures, which affect the quality of transmission of the signal. Most frequently occurs soft failures presented here.

2.3.1 Filter failure

If an optical signal is accurately set, it should have a central frequency that is close to the center of the designated spectrum slot to prevent filter failure and be symmetrical regarding that frequency in the optical spectrum of an optical channel. In Figure 2, when a filter fails, the spectrum is warped, and there are two types of filter failures, which can cause different type of distortion [18] and [19].

- Filter shift (FS) involves changing one edge of the signal while leaving the other unchanged, which destroys the symmetry of the signal and reduces both the spectral symmetry and the affected channel energy.
- Filter tightening (FT): Although the symmetry of the signal is preserved, the energy of the spectrum will still drop accordingly of tighter edge filtering. Because the bandwidth of the filter is smaller than the slot width designated for

the signal, the optical spectrum seems overly rounded in comparison with expected by the number of filters.

These changes aid in separating properly equipped optical spectra from those with filter failures. Therefore, a robust soft failure detection method is necessary for fast method of service protection.

2.3.2 Amplifier Spontaneous Emission (ASE)

ASE noise is the primary cause of linear degradation in the optical channel. In erbium-doped material, it results from the spontaneous decay of electrons from higher energy levels to lower energy levels. Because of the power attenuation that occurs during optical signal travel, optical amplifiers increase the power of optical signals. As the optical signal passes through additional optical amplifiers, ASE noise, which is measured by the noise figure (NF) of the optical amplifiers, increases. The form of the spectrum's symmetry is unaltered. Figure 2 demonstrates that ASE the noise has increased with noise floor also increased [20].

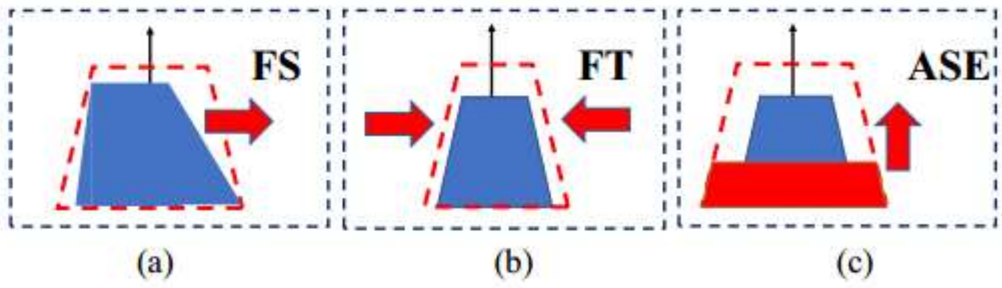

Figure 2 Diagram of :(a) FS, (b) FT and (c) increased ASE noise [6].

2.3.3 Inter-channel Interference(ICI)

When two channels present allocation of too much or too little spectrum, inter-channel interference results. As a result, for the other channel, the spectrum symmetry will diminish and the spectrum energy will increase. This most frequently occurs consequently of the core frequency shift of the laser aging [21].

2.3.4 Power attenuation

When an optical power signal travels through an optical fiber, it experiences power attenuation or loss. The main cause of power attenuation are Rayleigh scattering, Rayleigh reflection, refraction, bending losses, and absorption [22]. Usually, optical amplifiers are used to make up for power attenuation. Following passage through an optical cable at L distance, the output power (P_{ouf}) becomes:

$$\mathbf{P_{out} = P_{in} e^{\alpha L}} \quad \text{Equation 5}$$

Whereas P_{in} is the input power, and α is the fiber attenuation coefficient. The attenuation constant depends on the wavelength of the signal. The current low-loss optical fibers have a minimum loss of about 0.2 dB/km near 1550 nm.

Gradual power degradation can appear due to filter failure, power attention, or gradual equipment failure because of aging light waves, or ASE [6].

2.3.5 Non-linear Impairments (NLIs)

The non-linear effects are caused by the optical fibers' material characteristics (loss, refractive index, etc.), as well as by the intensity optical power. Cross-phase modulation (XPM), four-wave mixing (FWM), phase modulation (SPM), and phase modulation (SPM) are three different impairments brought on by the Kerr effect, which results from the relationship between the optical power and refractive index. The total optical channel power, type of fiber, and channel spacing are some elements that contribute to the rise in non-linear impairments. When the optical power level is high, the interaction of the fiber materials with the optical channels results in a scattering effect and it severity increase beyond certain optical power threshold value,

Chapter 3

3 Basic of Soft Failure Detection

3.1 Introduction of Machine Learning

Machine learning, which refers to computational representation, is said to learn and improve from experience with the execution of a task at a given ascertain performance [23]. Machine learning can also be defined as the process of solving a practical problem by.

- Gathering a dataset
- Algorithmically building a statistical model based on that dataset. That statistical model used somehow to solve the practical problem.

3.2 Types of Machine Learning

ML approaches categorize based on the data type and objectives of the learning task, or the type of tasks that they are intended to solve. These objectives are target pattern identification for classification and prediction, learning for action, or inductive learning methods. ML algorithms classified at a high level in the following categories [24].

Supervised ML algorithms: An algorithm builds an ML model iteratively by changing the weights of its parameters depending on the mapping of a set of input values (features) to their corresponding outputs, namely the labels. Depending on whether discrete, categorical, or numerical output values are used, supervised ML algorithms may be further divided into classification and regression tasks. Numerous supervised ML methods exist, including Naive Bayes, support vector machines (SVMs), random forests, decision trees, neural networks, and logistic regression.

Unsupervised ML algorithms: non-labelled data are assigned to the model building process of ML algorithms, which tries to create a representation of a given data set. To do this, correlations between features can be found through association rule discovery or grouping data into similar groups. K-means, hierarchical clustering, and other unsupervised machine learning techniques are examples.

Reinforcement ML algorithms: These algorithms train the computer to make particular decisions. Here, the algorithm continuously improves itself with feedback and trial-and-error techniques. This machine attempts to gather the finest knowledge it can in order to make informed decisions by learning from its prior experiences. An illustration of reinforcement learning is the Markov decision process. Here, learning data provides feedback so that the system can adapt to changing circumstances in order to fulfill a specific goal. Based on the feedback responses, the system assesses its performance.

Semi-Supervised ML algorithms: The given data are a mixture of mostly unclassified data and a small number of classified data points. A viable model for data categorization is developed using this combination of labeled and unlabeled data. The aim of semi-supervised classification is to build a model more accurately than a model created from labelled data.

3.3 Types of classification

Machine learning algorithms that learn how to assign a class label to instances originating from the problem regions, such as soft failure detection, churn prediction, and so on, can be used to classify instances. The four main categories of classification problems are binary classification, multi-class classification, multi-label classification, and unbalanced classification [25].

Binary classification: is implemented to solve categorization issues. It has two labels, one demonstrating the normal state and the other the abnormal one. Some Well-liked binary classification methods include artificial neural networks random forest support vector machine logistic regression decision tree and regression.

Multi class classification: are under the category of two or multiple class difficulties. The methods used are decision trees, random forest, k-nearest neighbors, and gradient boosting.

Multi Label Classification: when class of one or more labels is predicted for each case, a classification task more than one label is multi label classifiers. For instance, we may group movies into genres like action, romance, adventurer, and so forth. In this instance, a single film might be categorized as both a romance and an adventurer. Multi label classification of random forests, decision trees, and gradient boosting introduced rather than multi class classification or binary classification.

Imbalanced Classification: **Imbalance** classification tasks, the distribution of examples within each class are not balanced. Binary classification tasks that have an uneven distribution of instances between the normal and abnormal classes in the training dataset are known as unbalanced classification problems. The majority class may need to be under sampled or the minority class may need to be over sampled in order to modify the sample composition in the training dataset, despite the fact that these tasks are categorized as binary classification problems.

3.4 Types of Classification algorithms

For optical network soft failure detection, algorithms for classification are techniques based on supervised learning that divide the data into two classes. SVM, RF, and ANN are the classifiers used for this thesis. They are compared, and the classifier with the best classification performance is recommended. Following are further details discussed.

3.4.1 Support vector machine (SVM)

SVM algorithm's major application is binary classification issues, and its goal is to use a hyperplane to separate the database's classes. This hyperplane serves as a division line between two groups of related features. According to Figure 3, there are classes zero (normal) and one (soft failure). The boundary of the hyperplane may be linear or nonlinear. The design aims to maximize the separation space between the two classes. The margin is the separation space between the classes. Support vectors are related to hyperplane with maximum-margin. Such a maximum margin is found using SVMs for the decision boundary.

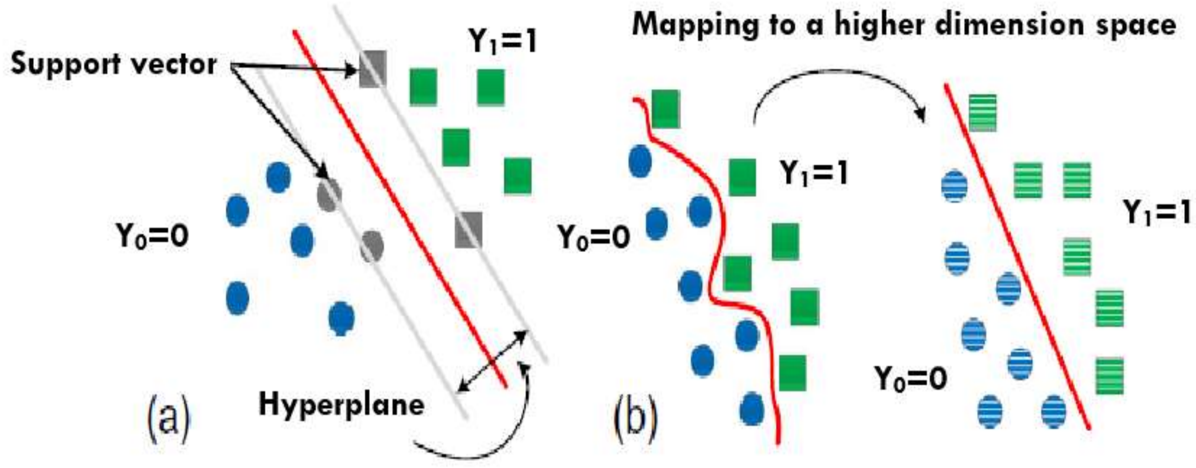

Figure 3 Support vectors for a binary SVM classifier a) linear b) kernel.

Hyper-parameters are settings within an algorithm that can be changed to enhance the classification method's effectiveness. The best way to find the ideal settings is to test numerous combinations and assess how well each model performs. The three important hyper-parameters for SVM are Regulation (C) parameter, gamma (G) parameter, and kernel.

Regulation (C) parameter: used to prevent or minimize misclassification; higher C values result in smaller margins and stricter classification. The margin expands with decreasing C, leading to a broader classification or misclassification. In the calculation of the hyperplane, the gamma (G) parameter stands in for the points defining the support vectors. Support vectors are located distant from the hyperplane because of the low gamma parameter. High gamma indicates support vectors that are near the hyperplane. Gamma functions as a regularization hyper-parameter: you should increase it if your model is under-fitting and decrease it if it is over-fitting.

Kernel: In the original low-dimensional space of the attributes, the input data might not be separable; however, they frequently do so in an induced high-dimensional space. The kernel function in Figure 3 is the function that converts an observation of the original attributes into a higher-dimensional space [26]. With SVMs, several kernel functions, such as polynomials and radial basis functions can be used. One common feature of machine learning is that the randomly sampled. High dimensional vectors are generally very small in number, increasing the risk of over-fitting and making it very difficult to identify patterns in the data without sufficient training data.

There are two methods SVM classifiers can be implemented. The optimal algorithm for linear SVMs is implemented by the lib linear package, on which the linear SVC class is based. Although it does not support the kernel trick, the SVC class is built on the “libsvm” library, which does. Despite being slower than other algorithms, SVMs frequently result in accurate classifiers. SVMs are mostly used for failure prediction, failure detection and root cause analysis.

3.4.2 Artificial neural network (ANN)

A computational model called an artificial neural network ANN is based on the composition and operation of biological neural networks. It is regarded as a nonlinear statistical data modeling tool as well. Figure 4 presented the linked components neurons that are arranged in layers configuration an ANN. Input, which corresponds to the columns of a data frame and weights, hidden which contains the summation and activation function, and output layers [27].

Input layer's sum of products of inputs and its corresponding weights are passed on to the neurons. In the following layer, ANN output is determined by the activation function. In terms of activation functions, sigmoid, hyperbolic tangent, and rectified linear units (ReLu) are the three most used forms. When dealing with binary classification issues, the output and hidden layers use the Sigmoid and “ReLu” functions.

The back-propagation algorithm is used to estimate the weights between the layers of a multilayer perception ANN. We may calculate the gradient descent of the cost function using the back-propagation equations. It determines the gradient of the error function relative to the weights of the neural network. The output error is back-propagated through the network starting with the final layer, and weights are changed to reduce the error rate.

The number of hidden layers, the number of neurons in each hidden layer, the activation function utilized in each hidden layer and the output layer, batch size, and epochs make up the list of hyper-parameters. Batch size determines the number of back propagate the error values so that individual node weights can be adjusted, and epochs, the number of times we want to run the entire test data over again to tune the weights. If the MLP over-exhausts the training data, this can be handled by reducing the number of hidden layers and neurons per hidden layer.

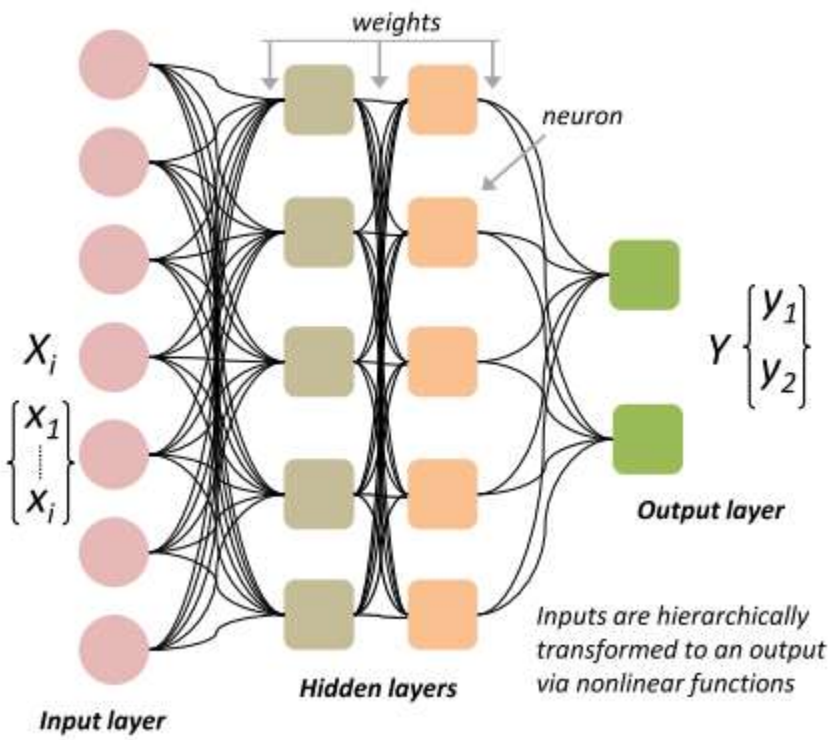

Figure 4: ANN (multilayer perception) architecture [24].

3.4.3 Random Forest (RF)

Random Forest is a branch of supervised ML algorithm. It commonly used for solving a problem of classification and regression. RF builds decision trees, starting with bootstrap aggregation. Bootstrap aggregation, commonly referred to as bagging, selects a random sample from the dataset. As a result, each model is created using row sampling, which uses the bootstrap samples provided by the original data with replacement. Bootstrapping is the term for this step of row sampling with replacement. Currently, each model is trained separately, producing results. After merging the outputs of all the models, the final decision is made based on a majority vote and average. Aggregation is the process of aggregating all the results and producing a result based on a majority vote for classification and average in the case of regression.

The two most important hyper-parameters for random forest classification are number of trees and tree depth [28].

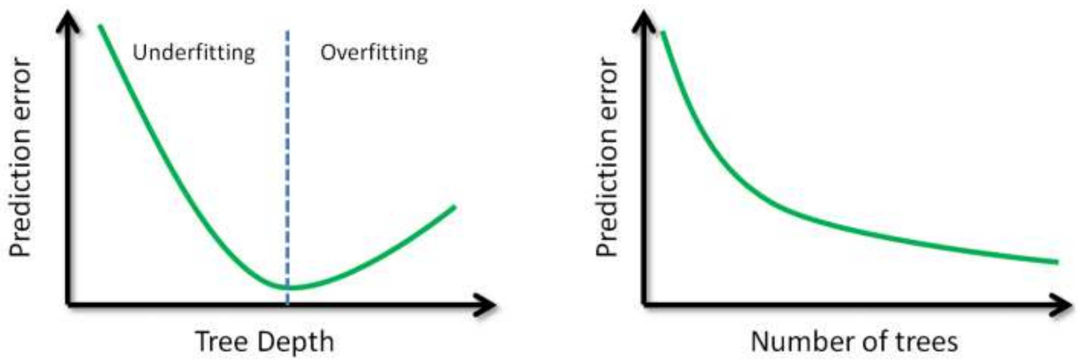

Figure 5: Hyperparameters of random forest properties [29]

In Figure 5, increasing the number of trees corresponds to a monotonic decrease in error. While increasing tree depth, the error curves decreases until it reaches a minimum and then increases again. The best performance of modeling is provided at the minimum point, and this point corresponds to the optimal tree depth value [29].

3.5 Classification Metrics [24]

In this section, we describe the metrics used to evaluate classification performance. Classification metrics are important to measure how well our classification models predict the desired outcome when building and optimizing ML modals. There are different types of metrics available to evaluate model performance. However, the most important classification metrics are identified and described here.

Confusion matrix

A confusion matrix is a tabular representation of a 2x2 matrix used as an evaluation approach for binary classification. The individual tabular data is used to calculate various other classification metrics. The confusion matrix working principle is clearly given in the confusion matrix **Table 1**. Let us consider the binary classification of soft failure detection, which has two classes. The classes represent a set of classes zero (normal or negative) and one (soft failure or positive).

		Predicted value	
		Class 0 (Normal/Negative)	Class 1 (Soft failure /Positive)
Actual Value	Class 0 (Normal/Negative)	True Negative (TN)	False Positive (FP)
	Class 1 (Soft failure/ Positive)	False Negative (FN)	True Positive (TP)

Table 1: Confusion matrix of binary classification

Where,

A true positive (TP) correctly predicts the class soft failure instance.

A true negative (TN) correctly predicts the normal instance.

A false negative (FN) incorrectly predicts the normal instance.

A false positive (FP) incorrectly predicts a soft failure instance.

Accuracy is one of the metrics used to measure the performance of a classification model. It is the ratio of the number of correct predictions to the total number of input samples. A formula to find accuracy:

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+FP+TN+FN}} = \frac{\text{Number of correct predictions}}{\text{Total number of input samples}} \quad \textbf{Equation 6}$$

Sensitivity, or true positive rate (TPR), measures the ability of a classifier to predict all positive instances or gives the probability that a true outcome is actually true. Its values range from zero to one. The formula for sensitivity is derived from the confusion matrix.

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP+FN}} \quad \text{Equation 7}$$

Specificity, or true negative rate (TNR), measures the ability of a classifier to predict all negative instances calculated as:

$$\text{Specificity} = \frac{\text{TN}}{\text{TN+FP}} \quad \text{Equation 8}$$

Precision, or positive predictive value (PPV), determines the fraction of actual positives out of the total predicted positives. It is also referring to the proportion of true outcomes given a set of outcomes classified as true and given by:

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \quad \text{Equation 9}$$

False positive rate (FPR): measures the performance of the model for incorrect prediction of soft failure (positive) instances per class.

$$\text{FPR} = \frac{\text{FP}}{\text{FP+TN}} = 1 - \text{Specificity} \quad \text{Equation 10}$$

False negative rate (FNR) measures the performance of the model for incorrect prediction of normal (negative) instances per class and is calculated as:

$$\text{FNR} = \frac{\text{FN}}{\text{FN+TP}} \quad \text{Equation 11}$$

The receiver operating characteristic (ROC) curve is obtained by plotting the true positive rate (TPR) with the false positive rate (FPR) and varying the class decision thresholds. A perfect classifier provides complete separation of the two classes and curve of ROC passes through the top left.

The area under the curve (AUC) represents area below the ROC curve, it measure the exact binary classifier quality, If AUC equal to one, then the classifier perfectly predicts all normal and soft failure classes. If AUC is equal to zero, then the classifier predicts all normal as soft failure.

Chapter 4

4 Data perpetration

4.1 Data Collection

The whole analysis is performed using Pre-FEC BER traces of the twenty-one optical channels carried in the real Ethiopian telecom backbone optical network. The optical channel is deployed on a coherent long-haul transmission system with similar route numbers, single-mode fiber types, and a number of intermediate nodes (amplifier nodes) ranging from 6 to 12 and a total distance of 450 km up to 932km. Filter sites range from 4 to 7. All the optical channels are 40 GB/s polarization-multiplexed binary phase shift keying (EPDM-BPSK) channels with the same coding type.

4.2 Data encoding and Integration

The process of encoding and integrating data from multiple stored Excel file formats with various field types into a single, unified Excel file known as data encoding and integration. Based on the time identifier index, optical channel identifier index, and quality of transmission identifier index, each piece of data is encoded and aggregated.

4.3 Data Preprocessing

Data preprocessing is resolving mismatched data formats and removing unnecessary data fields so that it may be fed into a machine learning model. It performed on three stages, which are duplicate, missing value, outlier and hard failure handling, window formation, and visualization.

Handling duplicate time series data

In a data set, duplicate rows are ones where all of the attribute values with another one or more rows are the same. Also time series analysis, duplicate rows with equal time stamps cause the data becomes unreasonably large as a results, it must be deleted

Missing Values and Outlier Detection

The timestamp index must be evenly spaced across the time series in order for the data to be in chronological order for time series analysis. We use the timestamp index to sort the data frame and compare consecutive timestamps to ensure that they are in chronological order. We will have missing values if the distances between the data are more than one.

Outliers: Values that deviate greatly from other observations represented by outliers in time series data. Their existence will significantly affect statistical analyses like mean and standard deviation as well as the performance of the ML model. They may be brought on by problems like when there values are grater then threshold values, hard failure, and human mistake in optical networks. An extremely steep decline in optical receiver power can have an immediate effect on the Pre-FEC BER value and be brought on by a hard failure (a fiber cut, equipment malfunction, or card defect). As a result, outliers were found far from the center of the data.

We must decide what to do with the outliers after locating them. Unfortunately, because it relies on the significance of the outliers and the objectives of the thesis, there is no clear-cut "best" technique for dealing with outliers. Form many methods for dealing with outliers; linear interpolation can be used to substitute outliers as though they were missing values with one condition, when more than one adjacent

Handling Missing Values

Imputation is the process of replacing missing data with statistical methods. It is advantageous to maintain circumstances by replacing missing data with an approximated value based on other available data. When more than one adjacent data value is absent, such performance data sequences are unavailable and should be eliminated because too much information is lost.

Finally, less than 5% (3.56%) of the 339 values are missing and more than thousands are deleted. As a result, we calculated the missing values using the linear interpolation method, which is frequently applied to time series data [30] and [31]. So that the modal building phase's time series data would become reliable.

4.4 Windows Formation

A time series is a set of data points that are captured over an equally spaced period [32]. On the other hand, machine learning algorithms work in feature space, which describes the various features of each Optical channel data point. Therefore, we need to map this data sequence in a way suitable for analysis with machine learning algorithms. One way of doing this mapping is by using a moving window of defined window size (number of data points) and sampling period (time distance between the two data points) [33]. Figure 6 presents an example of creating a window "a" with an optical channel that contains samples from t_0 to t_{29} of data points on time; a consecutive window “b” contains samples from t_1 to t_{30} and soon.

During this transformation, our data is split into a set of windows, and those highly overlapped windows would be correlated, which would only add more noise to the ML model rather than add additional information on the labels. It can be done independently at the train and validation phases, respectively [10]. Thus, a correlation analysis can be performed for different training and test sets using the Pearson test [5].

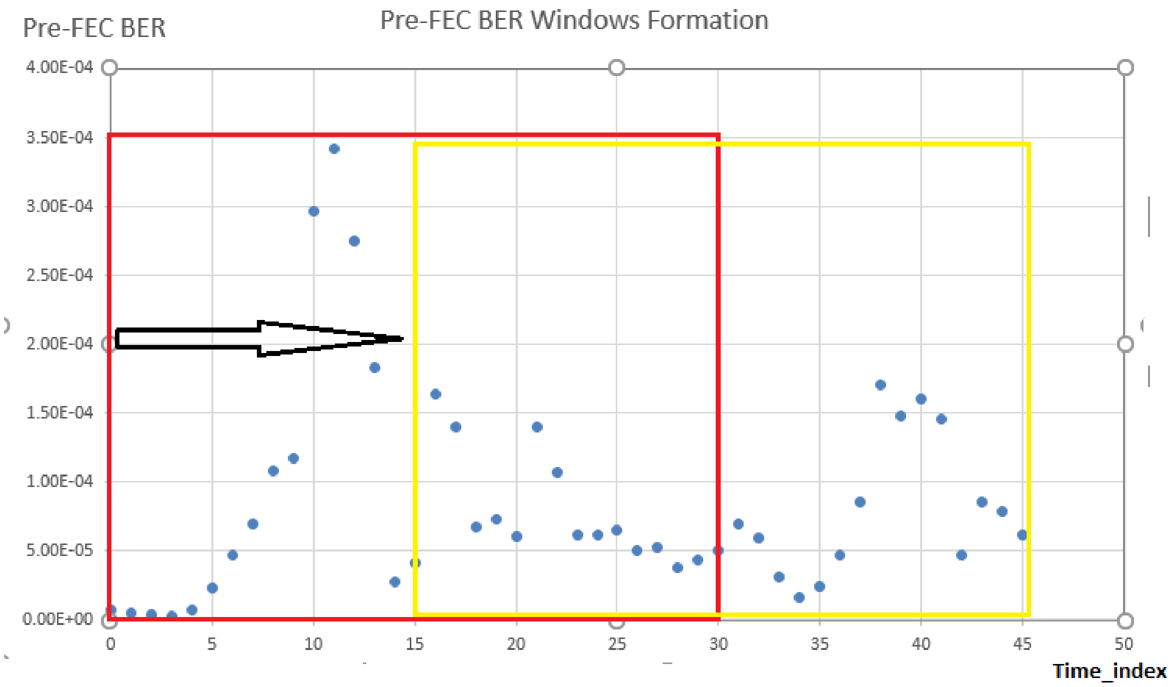

Figure 6: Mechanism of window formation.

4.5 Features extraction

In ML and pattern recognition, an individual measurable property is a feature, characteristic of a phenomenon being studied. Choosing the right and important features is a key step in achieving successful ML algorithms [5] and [33]. To train the various ML algorithms, several Pre-FEC BER windows are used, each characterized by a feature vector X. For each Pre-FEC BER window, we consider the following 17 features (i.e.$X = \{X_1, X_2, - - -, X_{17}\}$):

- X_1 = the minimum value of Pre-FEC BER;
- X_2 = the maximum value of Pre-FEC BER;
- X_3 = mean of value Pre-FEC BER;
- X_4 = the standard deviation value of Pre-FEC BER,
- X_5 = X2-X1 =Peak-to-Peak of Pre-FEC BER,
- X_6 = the root mean square value of Pre-FEC BER;

- X_7 = Variance of Pre-FEC BER ,
- X8-X17: the ten strongest spectral components in the window, extracted by applying the Fast Fourier Transform (FFT) on the Pre-FEC BER

The Fourier transform is a great tool for extracting the different seasonality patterns from a time series variable. It also allows us to describe a time series as a frequency rather than as a function of time [32]

4.6 Feature scaling

The majority of machine learning algorithms behave much better if features are on the same scale, so feature scaling is an important step in the feature engineering process. There are two common methods for bringing different features onto the same scale: normalization and standardization. Normalization is the process of scaling individual samples to have unit variance, while standardization refers to the process of bringing different features onto the same variance. For this research, we performed feature normalization to obtain features ranging from zero and one [10].

$$X_{\text{normalization}} = \frac{X - X_{Min}}{X_{Max} - X_{Min}} \quad \textbf{Equation 12}$$

Whereas X_{max} and X_{min} is maximum and minimum value

4.7 Data Labeling

Labeling is done based on the quality of transmission parameters that degrade and normal behavior, namely, normal, concept shift, small variation, trend, and other behavior. [7] [9] [34].

Algorithm I

Y ← X<time series>

For all windows

Error-Free Zone and **Pre-FEC BER** Normal threshold=$1*10^{-10}$

If the 95th percentile (Y) ≤ the normal threshold (N), then return false.

#Concept shift zone and **Pre-FEC BER s**oft threshold = 1*10-6

If 95th percentile (Y) ≤ soft threshold (S), then return true.

Else if max(Y) - min(Y) ≤ 1*10-9-1*10-10

Normal stationary behavior

If the* 95th percentile (Y) is ≤1*10-9, *then return false.

Soft stationary behavior

Else,* return true*.

Trend behavior

Decrement Trend behavior (performance improvement)

If the Pearson coefficient (Y) is ≤ -0.5 and the Spearman rank (Y) is ≤ -0.75, then return false.

Incremental trend behavior (performance degradation)

If the Pearson coefficient (Y) is ≤ 0.5 and the Spearman rank (Y) is ≤ 0.75, then return. True

Next behavior (periodicity*, cyclic*, random peaks*, and* undefined behavior)

Otherwise, return true.

Algorithm-1 was implemented to find feature labeling based on the behavior of time series data, and the procedure is described below.

- The algorithm receives an object X (a time series window of **Pre-FEC BER** data for each window).
- Normal behavior: This step is used to find error-free normal (class 0) data.

If 95% of the window sample value is less than or equal to the normal threshold value, there is no soft failure, and then it returns false.

- Concept shift behavior: the normal state changed over time, and it is not possible to find degradation behavior, but the individual window sample value is less than the maximum **Pre-FEC BER** that equipment can correct.

If 95% of the window sample value is greater than or equal to the soft threshold value, then return true.

- Small variation behavior is represented by normal and soft failure instances. The difference between maximum and minimum in Y is compared with the threshold. If it is lower than the threshold, go to small variation behavior to find normal and soft stationary behavior. Otherwise, the algorithm carries out an analysis to characterize the next behavior.

- Trend behavior: A trend is an increase or decrease that is observed for some significant period, starting at any point in the time series. This behavior is also represented by normal and soft failure instances if Y presents a decrement or incremental monotonic evolution in terms of the Pearson correlation coefficient and Spearman rank.
- Otherwise, cyclic, periodicity, random peaks, and undefined behavior

The Pearson Correlation Coefficient (PCC)

Pearson correlation coefficient (PCC) uses to determine the degree of a linear relationship between two random variables. The coefficient's value lies between negative one and positive one. There is no link between the two variables, as indicated by a value of zero Positive associations have values greater than zero, meaning that if one variable's value rises, so does the value of the other. A result that is less than zero denotes a negative connection, meaning that when one variable's value rises, the value of the other variable falls. PCC is the normalizing of covariance by each random variable's standard deviation.

$$PCC(X,Y) = \frac{COV(X,Y)}{SD_X * SD_Y} \quad \text{Equation 13}$$

Whereas X, Y: Two random variables, COV (): covariance, SD: standard deviation X_{mean} and Y_{mean}: mean of random variable X and Y respectively and n: length of random variable

$$\boldsymbol{COV(X,Y) = \frac{1}{n}\sum_{i=1}^{n}\left((X_i - X_{mean}) * (Y_i - Y_{mean})\right)}$$

$$\boldsymbol{SD_X = \sqrt{\frac{\sum_{i=1}^{n}\sqrt{(X_i - X_{mean})^2}}{n}} * SD_Y = \sqrt{\frac{\sum_{i=1}^{n}\sqrt{(Y_i - Y_{mean})^2}}{n}}}$$

Spearman rank

A measure of correlation known as Spearman rank reveals the statistical relationship between the ranks of two variables. It establishes how well a monotonic function may capture the relationship between the two variables. It measures monotonic, linear, and non-linear relationships between two variables using the Pearson correlation of the rank.

$$R_S(X,Y) = \frac{COV(R_X,R_Y)}{SD_{R_X} * SD_{R_Y}} \qquad \textbf{Equation 14}$$

Whereas X, Y: two random variables; COV () stands for covariance; SD for standard deviation R_X and R_Y: rank of X and Y variables, respectively

It can take a real value in the range $-1 \leq R_S(X, Y) \leq 1$. Its greatest value, $R_S(X, Y)$ = one, represents the scenario in which the relationship between x and y has a monotonically increasing. In other words, larger x values correspond to larger y values, and vice versa. Its minimum value, $R_S(X, Y)$ = negative one, represents monotonically decreasing function. To put it another way, higher x values are matched by lower y values and vice versa. RS (X, Y) is less than zero value.

4.8 Feature Balancing

Because soft failure positive cases occur less frequently than negative examples of typical behavior, there is an imbalance in failure data between the classes. The majority prejudices the minority class if there is an imbalance in the data when creating a classification model, which leads to anomalous model training and inaccurate prediction performance outcomes on modal validation [35].

index	Feature	Importance	index	Feature	Importance	index	Feature	Importance
0	X_8	0.1962	13	X_{13}	0.0051	26	X_{30}	0.0028
1	X_3	0.1842	14	X_{32}	0.0047	27	X_{21}	0.0028
2	X_1	0.1690	15	X_{25}	0.0047	28	X_{26}	0.0027
3	X_2	0.1675	16	X_{24}	0.0042	29	X_{17}	0.0027
4	X_6	0.1523	17	X_{27}	0.0040	30	X_{16}	0.0026
5	X_9	0.0096	18	X_{22}	0.0040	31	X_{26}	0.0024
6	X_4	0.0092	19	X_{28}	0.0040	32	X_{33}	0.0022
7	X_{10}	0.0085	20	X_{29}	0.0038	33	X_7	0.0019
8	X_{11}	0.0073	21	X_{19}	0.0036	34	X_{35}	0.0012
9	X_{12}	0.0063	22	X_{20}	0.0033	35	X_{34}	0.0011
10	X_{31}	0.0056	23	X_{18}	0.0032	36	X_{36}	0.0007
11	X_5	0.0055	24	X_{15}	0.0031			
12	X_{23}	0.0052	25	X_{14}	0.0029			

Table 2 Modal important feature

4.9 Feature Importance

The feature importance describes which features are more relevant to the solution model. It can lead to model improvements by employing feature selection. **Table 2** represents the importance of random forest features Modal output.

Chapter 5

5 Result and Discussion

This section presents the details of the experiments carried out for model building as well as a comparison analysis among three different classifiers.

5.1 Modal building

Modal construction, which entails data separation, techniques of validation, and hyper-parameters optimization, is a crucial step in soft failure detection. This thesis used the stratified shuffle split approach to create a 70/30 split between the training and validation data sets. Samples fairly and equally represent the full dataset

The models are trained using the training datasets, which are also utilized to optimize the hyper-parameters that will be employed in the models. Cross-validation (CV), a statistical technique, is used to compare and assess the performance of ML models throughout the optimization process. Grid search CVs and random search CVs are the two types. A search space is defined by a bounded region in the random search CV, and points within that domain are sampled at random. Grid search CV, on the other hand, is a search technique that examines every combination in the search space rather than selecting samples at random from a distribution. Following is a discussion of each of the three algorithms' performance outcomes following a thorough search for the ideal values for the hyper-parameters.

SVM

The optimization values set to be varied

- The C (Regulation) is equal (0.1, 1, 10, 100, 500, and 1000).
- Gamma parameter is equal (1, 0, 1, 0.01, 0.001, and 0.0001).
- Kernel (linear, polynomial, and RBF)

The optimal hyper-parameters found with the grid search CV method and radial basis function (RBF) score for the best-performing kernel. Because the relationship between the soft failure features is nonlinear, Whereas C equally to 1000 and G is 0.1.

ANN

In order to prevent over-fitting, hyper-parameters like the number of hidden layers and the number of neurons in those layers are calculated. We assess the precision of various arrangements of those hyper-parameters. One hidden layer with nine hidden neurons (half of the input feature) makes up the best-performing ANNs. For output layer and hidden layers, rectified linear units and sigmoid activation functions were applicable. Batch size equal to 25 for ideal hyper-parameters and epoch is 150.

RF

The two most important hyper-parameters for random forest classification are the number of trees and tree depth. Based on Section 3.4.3, the RF model was evaluated using the optimum and effective search method of grid search on a smaller search space, based on the best space provided by a random search that allowed us to narrow down the search pace and time. The optimal hyper-parameters of the number of trees equal to nigne9 and tree depth is 32.

5.2 Modal Comparison and discussion

The models were evaluated using a relatively balanced class of normal and soft failure instances. The comparison of the proposed three algorithms is discussed based on confusion matrices, accuracy, and building time, which have been measured when building and testing modes.

5.2.1 Confusion matrix

Form confusion matrix, actual soft failure samples correspond to FN+TP, while actual normal samples correspond to TN+FP. However, predicted soft failure samples correspond to TP +FP, while predicted normal samples correspond to TN + FN.

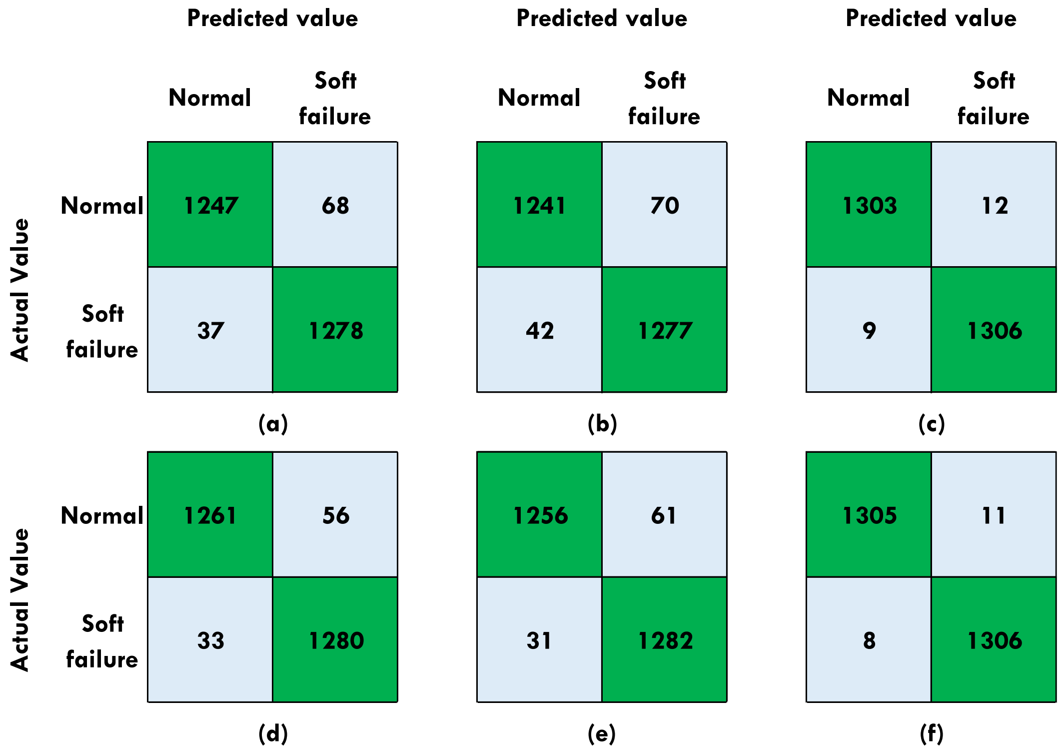

Figure 7: Confusion matrices for static and spectrum features (a) ANN model, (b) SVM model, (c) RF model. Confusion matrices for importance features d) ANN model, e) SVM model, (f) RF model.

Figure 7, the confusion matrices derived for each model provide illuminates the better performance of the models based on greater true positive and true negative values. The RF with important features model is the best model for recognizing instances of soft failure, in contrast to the lower false negative and false positive values for detecting soft failure and the normal class, with the lowest false negative (6) and the highest true positive (1306) values. It is still the strongest model for identifying typical occurrences, even though it has the lowest true positive rate (11) and the greatest true negative rate (1305). Therefore, the main goal of this thesis is to emphasize occurrences with lower FN and greater TP. For the true positive (1277) and false negative (42)

5.2.2 Accuracy

We evaluate the effectiveness of the proposed approaches based on their accuracy, which is calculated as the proportion of correctly identified failures over all failures. In Figure 8, when classifiers are based on RF with significant features, training and validation accuracy are 100% and 99.2%, respectively. When the classifier is built using SVMs with static and spectral information,

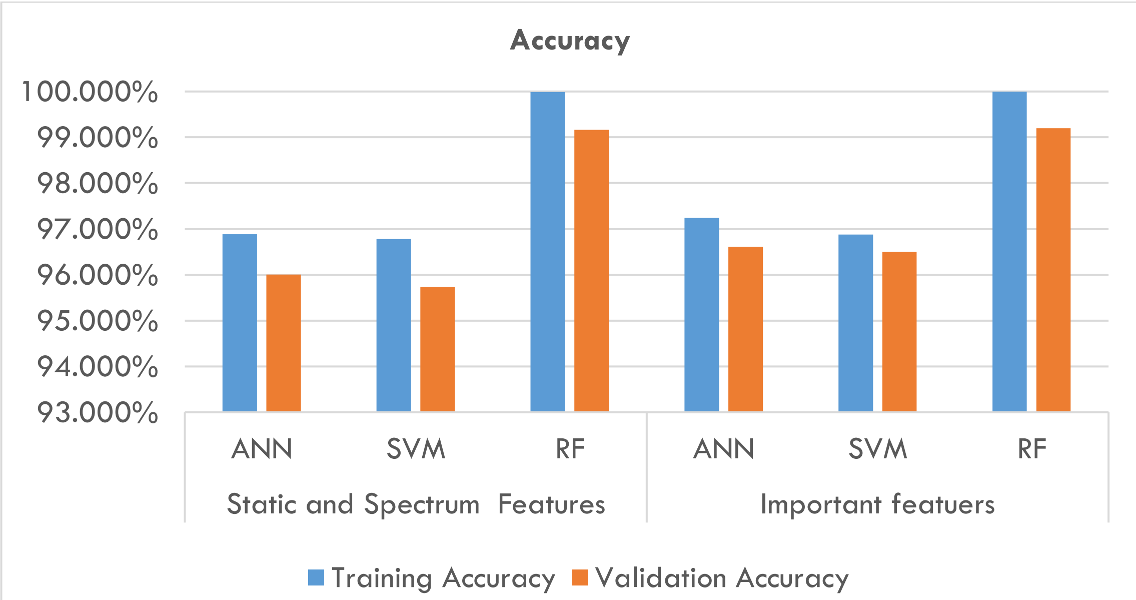

Figure 8 Comparison of three classifications model-based accuracy.

We also contrast how well the suggested methods perform in terms of accuracy standard deviation, as shown in Table 3. It gauges the degree of variation among each fold's accuracy values. In all models, no notable variation has been found. When classifiers are built on SVMs with static and spectral features, the maximum validation accuracy standard deviation recorded is 1.95%. The best-performing classifier, based on two performance matrices, is RF with lower features, even if the minimum for two classifiers based on RF is 0.42% and 0.49%., based on two performance matrices, the best-performing classifier is RF with lower features.

Metrics	Static and Spectrum Features			Important featuers		
	ANN	SVM	RF	ANN	SVM	RF
Std of Training Accuracy	0.39%	0.18%	0.05%	0.36%	0.17%	0.06%
Std of Validation Accuracy	1.45%	1.95%	0.42%	1.11%	0.86%	0.49%

Table 3 Comparison of three classifications models with accuracy standard deviation

5.2.3 Modal build Time

In **Table 4**, the lowest modal build time is obtained with the RF static and spectrum feature approaches, but it also provides the highest number of features (15). All three modals decrease misclassifications of soft failure and are normal when the number of features decreases. On the other hand, only RF modals increase modal building time when the number of features decreases.

When the number of characteristics reduces, all three modals become normal and help to reduce soft failure misclassifications. On the other hand, when the quantity of features reduces, only RF modals lengthen modal building time. The RF with important features, which offers the lowest misclassification of soft failure and normal (0.61%, 0.84%) with lower computational complexity (12 features and 17.5ms of building time), represents the ideal compromise between misclassification of soft failure and normal with complexity. Table 4 contrasts three categories with modal builds. Time comparing three classes using various classification matrices is shown in Table 5.

Metrics	Static and Spectrum Features			Important features		
	ANN	SVM	RF	ANN	SVM	RF
FPR	5.17%	5.34%	0.91%	4.25%	4.63%	0.84%
FNR	2.81%	3.18%	0.68%	2.51%	2.36%	0.61%
Modal build time(ms}	178.00	22.70	7.90	171.50	15.00	17.50
Number of features	17	17	17	12	12	12

Table 4 Comparison of three classification with modal build Time

Metrics	Static and Spectrum Features			Important featuers		
	ANN	SVM	RF	ANN	SVM	RF
Precision	96.03%	95.76%	99.20%	96.63%	96.52%	99.28%
Recall	96.01%	95.74%	99.20%	96.62%	96.50%	99.28%
F1-score	96.02%	95.75%	99.20%	96.62%	96.51%	99.28%
Sensitivity	97.19%	96.82%	99.32%	97.49%	97.64%	99.39%
Specificity	94.83%	94.66%	99.09%	95.75%	95.37%	99.16%
ROC	96.01%	95.74%	99.20%	96.62%	96.50%	99.28%
kappa_statistic	0.92	0.91	98.40%	93.23%	93.00%	98.56%
log_loss	1.44	1.53	0.28	1.22	1.26	0.25
R_squared	1.44	0.83	0.97	0.86	0.86	0.97
MSE	0.84	0.04	0.01	0.03	0.03	0.01
RMSE	0.04	0.20	0.09	0.18	0.19	0.07
MAE	0.20	0.04	0.01	0.03	0.03	0.01

Table 5 Comparison of three classification with different classification matrix

5.2.4 Modal Interpretation

- Random forest is the most accurate algorithm. The main drawback of the RF algorithm is taking a significant time to build its model compared to SVM, but it is faster relative to ANN'.
- Number of features affects ANN and SVM but not significant on RF case.
- All three modal decrease misclassifications of soft failure and normal when number of feature decrease (RF case not significant).
- **Pre-FEC BER** is vendor dependent and need sample cross check of BER test using loop test.
- Very dormant feature is FFT (fundament) with 19.6% and second mean with 18.42% value.

Chapter 6

6 Conclusion and Feature Work

6.1 Conclusion

This thesis focused on detecting soft failure on the optical channel of the optical layer by continuous monitoring and exploiting of the FEC data. We explored the performance data from which network elements are generated, soft failure from which performance data is detected, and soft failure detection from which classification ML is built. Through data preparation and modal building, which starts with a map time series data sequence in a window, then algorithm-based feature labelling was implemented based on the behaviour of a window. From the window, feature extraction is performed using statistical and Fourier transform in a way suitable for ML (Data Set 1). Reduced the list of features to the most important by random forest feature important modal to remove redundant features. (Data set 2).

Cross-validation techniques with Stratified Shuffle Split method are used to validate the performance of algorithms. Modal performance is measured in terms of confession matrix, accuracy, and complexity of modal. As a result, RF with Important features is best performed algorithm by 99.2% validation accuracy with standard devotion 0.49% and lower complexity

To conclude, we demonstrate anticipation in soft failure detection with high accuracy and good performance of the proposed scheme. In addition, we demonstrate feature labelling based on the behaviour of a time series window.

6.2 Future work

Optical channel of 40 GB/s polarization-multiplexed binary phase shift keying (EPDM-BPSK), advanced FEC code, and Pre-FEC BER quality of transmission parameters based on classical offline supervised learning approach applied in this thesis. Further work will focus on

- Optimization of ML models based on multi-class scenario, different input data, different type of quality of transmission parameters, feature labelling method and complex patterns of quality of transmission parameters.
- Unsupervised or semi-supervised ML by taking into account reduce size of data sets, unbalanced feature labelling, and lack of feature labelling data..
- Changing network conditions, traffic type, modulation type, and data rate.
- Soft failure detection in practical systems like fiber, amplifier and optical fitter with different and complex soft failure type.
- Discovering relevant patterns and useful monitoring information in case of large-scale data sets.

Appendix-I

CV result

Fold	Accuracy Training	Accuracy Validation	Precision Normal	Precision Soft	Precision weighted avg	Recall Normal	Recall Soft	Recall	F1 score Normal	F1 score soft	F1 score weighted avg	ROC
0	96.91%	95.06%	96.09%	94.07%	95.08%	93.89%	96.21%	95.06%	94.98%	95.13%	95.06%	95.05%
1	97.00%	95.06%	96.12%	94.03%	95.08%	93.94%	96.18%	95.06%	95.02%	95.09%	95.06%	95.06%
2	97.46%	93.92%	96.75%	91.43%	94.08%	90.84%	96.97%	93.92%	93.70%	94.12%	93.91%	93.90%
3	96.49%	96.58%	96.21%	96.95%	96.58%	96.95%	96.21%	96.58%	96.58%	96.58%	96.58%	96.58%
4	96.83%	94.68%	96.83%	92.70%	94.77%	92.42%	96.95%	94.68%	94.57%	94.78%	94.67%	94.69%
5	96.23%	98.86%	100.00%	97.76%	98.88%	97.73%	100.00%	98.86%	98.85%	98.87%	98.86%	98.86%
6	96.74%	96.96%	96.97%	96.95%	96.96%	96.97%	96.95%	96.96%	96.97%	96.95%	96.96%	96.96%
7	96.78%	95.44%	96.85%	94.12%	95.48%	93.89%	96.97%	95.44%	95.35%	95.52%	95.44%	95.43%
8	97.50%	96.58%	99.19%	94.24%	96.71%	93.89%	99.24%	96.58%	96.47%	96.68%	96.58%	96.57%
9	96.91%	96.96%	96.27%	97.67%	96.97%	97.73%	96.18%	96.96%	96.99%	96.92%	96.96%	96.96%
	96.89%	96.01%	97.13%	94.99%	96.06%	94.83%	97.19%	96.01%	95.95%	96.06%	96.01%	96.01%

Table 6 ANN with data set 1 and different classification matrices 1

Fold	Modal build time	kappa statistic	log loss	R_squared	Mean squared error	Root mean squared error	Mean absolute error
0	0.17	0.90	1.78	0.80	0.05	0.22	0.05
1	0.17	0.90	1.78	0.80	0.05	0.22	0.05
2	0.17	0.88	2.19	0.76	0.06	0.25	0.06
3	0.19	0.93	1.23	0.86	0.03	0.18	0.03
4	0.18	0.89	1.92	0.79	0.05	0.23	0.05
5	0.18	0.98	0.41	0.95	0.01	0.11	0.01
6	0.18	0.94	1.10	0.88	0.03	0.17	0.03
7	0.19	0.91	1.64	0.82	0.05	0.21	0.05
8	0.19	0.93	1.23	0.86	0.03	0.18	0.03
9	0.18	0.94	1.10	0.88	0.03	0.17	0.03
	0.18	0.92	1.44	0.84	0.04	0.20	0.04

Table 7 ANN with dataset 1 and different classification matrices 2

Fold	Accuracy Training	Accuracy Validation	Precision Normal	Precision Soft	Precision weighted avg	Recall Normal	Recall Soft	Recall	F1 score Normal	F1 score soft	F1 score weighted avg	ROC
0	96.95%	95.44%	96.85%	94.12%	95.48%	93.89%	96.97%	95.44%	95.35%	95.52%	95.44%	95.43%
1	96.45%	98.48%	98.47%	98.48%	98.48%	98.47%	98.48%	98.48%	98.47%	98.48%	98.48%	98.48%
2	96.70%	91.63%	92.25%	91.04%	91.64%	90.84%	92.42%	91.63%	91.54%	91.73%	91.63%	91.63%
3	97.00%	96.20%	99.19%	93.57%	96.37%	93.13%	99.24%	96.20%	96.06%	96.32%	96.19%	96.19%
4	96.78%	96.58%	99.20%	94.20%	96.71%	93.94%	99.24%	96.58%	96.50%	96.65%	96.58%	96.59%
5	96.91%	95.06%	96.09%	94.07%	95.08%	93.89%	96.21%	95.06%	94.98%	95.13%	95.06%	95.05%
6	96.91%	94.68%	93.98%	95.38%	94.69%	95.42%	93.94%	94.68%	94.70%	94.66%	94.68%	94.68%
7	96.66%	98.10%	100.00%	96.35%	98.17%	96.18%	100.00%	98.10%	98.05%	98.14%	98.10%	98.09%
8	96.57%	94.68%	94.66%	94.70%	94.68%	94.66%	94.70%	94.68%	94.66%	94.70%	94.68%	94.68%
9	96.91%	96.58%	96.92%	96.24%	96.58%	96.18%	96.97%	96.58%	96.55%	96.60%	96.58%	96.58%
	96.78%	95.74%	96.76%	94.82%	95.79%	94.66%	96.82%	95.74%	95.69%	95.79%	95.74%	95.74%

Table 8 SVM with dataset 1 and different classification matrices 1

Fold	Modal build time	kappa statistic	log loss	R_squared	Mean squared error	Root mean squared error	Mean absolute error
0	0.02	0.91	1.64	0.82	0.05	0.21	0.05
1	0.02	0.97	0.55	0.94	0.02	0.12	0.02
2	0.08	0.83	3.02	0.67	0.08	0.29	0.08
3	0.02	0.92	1.37	0.85	0.04	0.19	0.04
4	0.02	0.93	1.23	0.86	0.03	0.18	0.03
5	0.02	0.90	1.78	0.80	0.05	0.22	0.05
6	0.02	0.89	1.92	0.79	0.05	0.23	0.05
7	0.02	0.96	0.69	0.92	0.02	0.14	0.02
8	0.02	0.89	1.92	0.79	0.05	0.23	0.05
9	0.01	0.93	1.23	0.86	0.03	0.18	0.03
	0.02	0.91	1.53	0.83	0.04	0.20	0.04

Table 9 SVM with dataset 1 and different classification matrices 2

Fold	Accuracy Training	Accuracy Validation	Precision Normal	Precision Soft	Precision weighted avg	Recall Normal	Recall Soft	Recall	F1 score Normal	F1 score soft	F1 score weighted avg	ROC
0	100.00%	99.62%	100.00%	99.25%	99.62%	99.24%	100.00%	99.62%	99.62%	99.62%	99.62%	99.62%
1	99.92%	98.48%	97.74%	99.23%	98.49%	99.24%	97.73%	98.48%	98.48%	98.47%	98.48%	98.48%
2	99.92%	99.24%	100.00%	98.51%	99.25%	98.47%	100.00%	99.24%	99.23%	99.25%	99.24%	99.24%
3	99.92%	99.62%	100.00%	99.24%	99.62%	99.24%	100.00%	99.62%	99.62%	99.62%	99.62%	99.62%
4	100.00%	99.62%	100.00%	99.25%	99.62%	99.24%	100.00%	99.62%	99.62%	99.62%	99.62%	99.62%
5	99.87%	98.86%	98.50%	99.23%	98.86%	99.24%	98.47%	98.86%	98.87%	98.85%	98.86%	98.86%
6	99.92%	98.86%	99.23%	98.50%	98.86%	98.47%	99.24%	98.86%	98.85%	98.87%	98.86%	98.86%
7	99.87%	98.86%	99.24%	98.48%	98.86%	98.48%	99.24%	98.86%	98.86%	98.86%	98.86%	98.86%
8	99.83%	99.24%	98.51%	100.00%	99.25%	100.00%	98.47%	99.24%	99.25%	99.23%	99.24%	99.24%
9	99.87%	99.62%	100.00%	99.24%	99.62%	99.24%	100.00%	99.62%	99.62%	99.62%	99.62%	99.62%
	99.91%	99.20%	99.32%	99.09%	99.21%	99.09%	99.32%	99.20%	99.20%	99.20%	99.20%	99.20%

Table 10 RF with dataset 1 and different classification matrices 1**.**

Fold	Modal build time	kappa statistic	log loss	R_squared	Mean squared error	Root mean squared error	Mean absolute error
0	0.01	0.99	0.13	0.98	0.00	0.06	0.00
1	0.01	0.97	0.53	0.94	0.02	0.12	0.02
2	0.01	0.98	0.26	0.97	0.01	0.09	0.01
3	0.01	0.99	0.13	0.98	0.00	0.06	0.00
4	0.01	0.99	0.13	0.98	0.00	0.06	0.00
5	0.01	0.98	0.39	0.95	0.01	0.11	0.01
6	0.01	0.98	0.39	0.95	0.01	0.11	0.01
7	0.01	0.98	0.39	0.95	0.01	0.11	0.01
8	0.00	0.98	0.26	0.97	0.01	0.09	0.01
9	0.01	0.99	0.13	0.98	0.00	0.06	0.00
	0.01	0.98	0.28	0.97	0.01	0.09	0.01

Table 11 RF with dataset 1 and different classification matrices 2

Fold	Accuracy Training	Accuracy Validation	Precision Normal	Precision Soft	Precision weighted avg	Recall Normal	Recall Soft	Recall	F1 score Normal	F1 score soft	F1 score weighted avg	ROC
0	96.91%	97.34%	99.21%	95.59%	97.41%	95.45%	99.24%	97.34%	97.30%	97.38%	97.34%	97.35%
1	96.53%	96.58%	96.92%	96.24%	96.58%	96.18%	96.97%	96.58%	96.55%	96.60%	96.58%	96.58%
2	96.91%	94.30%	96.80%	92.03%	94.42%	91.67%	96.95%	94.30%	94.16%	94.42%	94.29%	94.31%
3	95.94%	97.72%	96.99%	98.46%	97.73%	98.47%	96.97%	97.72%	97.73%	97.71%	97.72%	97.72%
4	96.95%	95.82%	97.62%	94.16%	95.88%	93.89%	97.73%	95.82%	95.72%	95.91%	95.82%	95.81%
5	96.78%	96.20%	96.92%	95.49%	96.21%	95.45%	96.95%	96.20%	96.18%	96.21%	96.20%	96.20%
6	96.66%	96.96%	96.27%	97.67%	96.97%	97.73%	96.18%	96.96%	96.99%	96.92%	96.96%	96.96%
7	97.12%	95.82%	95.49%	96.15%	95.82%	96.21%	95.42%	95.82%	95.85%	95.79%	95.82%	95.82%
8	96.49%	97.72%	98.46%	96.99%	97.73%	96.97%	98.47%	97.72%	97.71%	97.73%	97.72%	97.72%
9	96.95%	97.72%	100.00%	95.62%	97.82%	95.45%	100.00%	97.72%	97.67%	97.76%	97.72%	97.73%
	97.25%	#REF!	97.47%	95.84%	96.66%	95.75%	97.49%	96.62%	96.59%	96.64%	96.62%	96.62%

Table 12 ANN with dataset 2 and different classification matrices 1

Fold	Modal build time	kappa statistic	log loss	R_squared	Mean squared error	Root mean squared error	Mean absolute error
0	0.19	0.95	0.96	0.89	0.03	0.16	0.03
1	0.15	0.93	1.23	0.86	0.03	0.18	0.03
2	0.15	0.89	2.06	0.77	0.06	0.24	0.06
3	0.15	0.95	0.82	0.91	0.02	0.15	0.02
4	0.15	0.92	1.51	0.83	0.04	0.20	0.04
5	0.21	0.92	1.37	0.85	0.04	0.19	0.04
6	0.19	0.94	1.10	0.88	0.03	0.17	0.03
7	0.16	0.92	1.51	0.83	0.04	0.20	0.04
8	0.22	0.95	0.82	0.91	0.02	0.15	0.02
9	0.17	0.95	0.82	0.91	0.02	0.15	0.02
	0.17	0.93	1.22	0.86	0.03	0.18	0.03

Table 13 ANN with dataset 2 and different classification matrices 2

Fold	Accuracy Training	Accuracy Validation	Precision Normal	Precision Soft	Precision weighted avg	Recall Normal	Recall Soft	Recall	F1 score Normal	F1 score soft	F1 score weighted avg	ROC
0	97.00%	95.44%	96.85%	94.12%	95.48%	93.89%	96.97%	95.44%	95.35%	95.52%	95.44%	95.43%
1	96.83%	96.20%	96.21%	96.18%	96.20%	96.21%	96.18%	96.20%	96.21%	96.18%	96.20%	96.20%
2	96.95%	97.34%	98.45%	96.27%	97.36%	96.21%	98.47%	97.34%	97.32%	97.36%	97.34%	97.34%
3	96.83%	96.20%	98.40%	94.20%	96.29%	93.89%	98.48%	96.20%	96.09%	96.30%	96.20%	96.19%
4	97.21%	96.20%	96.21%	96.18%	96.20%	96.21%	96.18%	96.20%	96.21%	96.18%	96.20%	96.20%
5	96.61%	97.72%	99.21%	96.32%	97.76%	96.18%	99.24%	97.72%	97.67%	97.76%	97.72%	97.71%
6	96.95%	96.58%	96.95%	96.21%	96.58%	96.21%	96.95%	96.58%	96.58%	96.58%	96.58%	96.58%
7	96.95%	95.06%	97.60%	92.75%	95.19%	92.42%	97.71%	95.06%	94.94%	95.17%	95.05%	95.07%
8	96.78%	97.34%	97.71%	96.97%	97.34%	96.97%	97.71%	97.34%	97.34%	97.34%	97.34%	97.34%
9	96.70%	96.96%	98.44%	95.56%	97.00%	95.45%	98.47%	96.96%	96.92%	96.99%	96.96%	96.96%
	96.88%	#REF!	97.60%	95.48%	96.54%	95.37%	97.64%	96.50%	96.46%	96.54%	96.50%	96.50%

Table 14 SVM with dataset 2 and different classification matrices 1

Fold	Modal build time	kappa statistic	log loss	R_squared	Mean squared error	Root mean squared error	Mean absolute error
0	0.01	0.98	0.39	0.95	0.01	0.11	0.01
1	0.06	0.99	0.13	0.98	0.00	0.06	0.00
2	0.01	0.98	0.39	0.95	0.01	0.11	0.01
3	0.01	0.98	0.39	0.95	0.01	0.11	0.01
4	0.01	0.99	0.13	0.98	0.00	0.06	0.00
5	0.01	0.98	0.39	0.95	0.01	0.11	0.01
6	0.04	1.00	0.00	1.00	0.00	0.00	0.00
7	0.01	0.98	0.39	0.95	0.01	0.11	0.01
8	0.01	0.98	0.26	0.97	0.01	0.09	0.01
9	0.01	1.00	0.00	1.00	0.00	0.00	0.00

Table 15 SVM with dataset 2 and different classification matrices 2

Fold	Accuracy Training	Accuracy Validation	Precision Normal	Precision Soft	Precision weighte d avg	Recall Normal	Recall Soft	Recall	F1 score Normal	F1 score soft	F1 score weighte d avg	ROC
0	99.92%	98.86%	98.50%	99.23%	98.86%	99.24%	98.47%	98.86%	98.87%	98.85%	98.86%	98.86%
1	99.87%	99.62%	99.24%	100.00%	99.62%	100.00%	99.24%	99.62%	99.62%	99.62%	99.62%	99.62%
2	100.00%	98.86%	99.24%	98.48%	98.86%	98.48%	99.24%	98.86%	98.86%	98.86%	98.86%	98.86%
3	99.96%	98.86%	100.00%	97.76%	98.88%	97.73%	100.00%	98.86%	98.85%	98.87%	98.86%	98.86%
4	99.96%	99.62%	99.25%	100.00%	99.62%	100.00%	99.24%	99.62%	99.62%	99.62%	99.62%	99.62%
5	99.96%	98.86%	99.24%	98.48%	98.86%	98.48%	99.24%	98.86%	98.86%	98.86%	98.86%	98.86%
6	99.83%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
7	99.96%	98.86%	98.50%	99.23%	98.86%	99.24%	98.47%	98.86%	98.87%	98.85%	98.86%	98.86%
8	100.00%	99.24%	100.00%	98.51%	99.25%	98.47%	100.00%	99.24%	99.23%	99.25%	99.24%	99.24%
9	99.87%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
	99.93%	#REF!	99.40%	99.17%	99.28%	99.17%	99.39%	99.28%	99.28%	99.28%	99.28%	99.28%

Table 16 RF with dataset 2 and different classification matrices 1

Fold	Modal build time	kappa statistic	log loss	R_squared	Mean squared error	Root mean squared error	Mean absolute error
0	0.01	0.98	0.39	0.95	0.01	0.11	0.01
1	0.06	0.99	0.13	0.98	0.00	0.06	0.00
2	0.01	0.98	0.39	0.95	0.01	0.11	0.01
3	0.01	0.98	0.39	0.95	0.01	0.11	0.01
4	0.01	0.99	0.13	0.98	0.00	0.06	0.00
5	0.01	0.98	0.39	0.95	0.01	0.11	0.01
6	0.04	1.00	0.00	1.00	0.00	0.00	0.00
7	0.01	0.98	0.39	0.95	0.01	0.11	0.01
8	0.01	0.98	0.26	0.97	0.01	0.09	0.01
9	0.01	1.00	0.00	1.00	0.00	0.00	0.00
	0.02	0.99	0.25	0.97	0.01	0.07	0.01

Table 17 RF with dataset 2 and different classification matrices 2

References

[1] D. Wang, C. Zhang, W. Chen, H. Yang and M. Zhang, "A Review of Machine Learning-based Failure Management in Optical Networks," vol. 1, pp. 1-4, 23 Aug 2023.

[2] W. Kebede, "Constraint-Based Hybrid Resiliency Mechanisms for Better Resource Utilization and Service Performance Quality in ASON SLA," Addis Ababa University, Addis Ababa , November 2018.

[3] I. F. Dick, "Machine Learning based Fault Detection Algorithms for Long Haul Elastic Optical Networks," Germany, 2021.

[4] G.Ruoxuan, L.Lei., L.Xiaomin, L.Huazhi, Y.Lilin., H.Weisheng and Z.Qunb, "An overview of ML-based applications for next generation optical networks," *Science China Information Sciences,* vol. 63, 2020.

[5] S. Shahkarami, F. Musumeci, F. Cugini and M. Tornatore, "Machine-Learning-Based Soft-Failure Detection and Identification in Optical Networks," *Optical Fiber Communication Conference {OFC),* 2018.

[6] H. Lun, M. Fu, X. Liu, Y. Wu, L. Yi, W. Hu and Q. Zhuge, "Soft Failure Identification for Long-haul Optical Communication Systems Based on One-dimensional Convolutional Neural Network," *Journal of Lightwave Technology,* vol. 38, p. 2992–2999, 2020.

[7] A.P.Vela., M.Ruiz., F.Fresi, N.Sambo., F.Cugini, G.Cugini., L.Potì., L.Velasco. and P.Castoldi, "BER Degradation Detection and Failure Identification in Elastic Optical Networks," *Journal of Lightwave Technology,* vol. 5, pp. 4595-4604, 2017.

[8] A. P. Vela, B. Shariati, M. Ruiz, F. Cugini, A. Castro, H. Lu, R. Proietti, J. Comellas, P. Castoldi, S. J. B. Yoo and L. Velasco, "Soft Failure Localization during Commissioning Testing and Lightpath Operation," *IEEE/OSA J. of Optical Communications,* vol. 10, p. A27–A36, 2018.

[9] S. Allogba, B. Laure, M. Yaméogo and C. Tremblay, "Extraction and Early Detection of Anomalies in Lightpath SNR Using Machine Learning Models," *Journal of Light wave Technology,* vol. 40, pp. 1864-1872, 2022.

[10] F.Musumec and G.V.Velasco, "Domain Adaptation and Transfer Learning for Failure Detection and Failure-Cause Identification in Optical Networks Across Different Lightpaths," *Optical Communications and Networkin,* vol. 14, pp. A91-A100, 2022.

[11] L.Kruse and S.Pachnicke, "Soft-Failure Detection and Lifetime Prediction based on Spectral Data using 1-D Convolutional," *Proceedings of 22th ITG Symposium,* pp. 1-6, 2021.

[12] B.Shariati, M.Ruiz, J.Comellas and L.Velasco, "Learning from the Optical Spectrum: :Failure Detection and Identification," *Journal of Lightwave Technology,* vol. 37, pp. 433-440, 2019.

[13] K.Abdelli, H.Griee., P.Ehrle. and e. al, "Reflective fiber fault detection and characterization using long short-term memory," *Journal of Optical Communications and Networking,* vol. 13, pp. E32-E41, 2021.

[14] R. T. Jones, "Machine Learning Methods in Coherent Optical Communication Systems," Technical University of Denmark, Kongens Lyngby, 2019.

[15] R. Essiambrt, G. Kramer, P. J. Winzer, G. J. Foschin and B. , "Capacity limits of optical fiber networks," *Journal of Lightwave Technology,* vol. 28, p. 662–701, 2010.

[16] M.Bouda, S.Oda, O.Vassilieva and e. al., "Accurate prediction of quality of transmission based on a dynamically configurable optical impairment model," *Journal of Optical Communications and Networking,* vol. 10, p. A102, 2018.

[17] B. Nebendahl, R. Schmogrow, T. Dennis, A. Josten, D. Hillerkuss, S. Koenig, J. Meyer, M. Dreschmann, M. Winter, M. Huebner, W. Freude and J. C. KoosLeuthold, "Quality Metrics in optical modulation analysis: EVM and its relation to Q-factor, OSNR, and BER,," *Communications and Photonics Conference,* 2012.

[18] B. Shariati, A. P. Vela, M. Ruiz and L. Velasco, "Monitoring and Data Analytics: Analyzing the Optical Spectrum for Soft-Failure Detection and Identification," 2018.

[19] L. Shu, Z. Yu, Z. Wan, J. Zhang, S. Hu and K. Xu, "Low Complexity Dual Stage Soft Failure Detection by Exploiting Digital Spectrum information," *European conference Optical Communication (ECOC),* 2019.

[20] H. L, Z.Qunbi, F.Mengfan, W.Yiwen, L.Qiaoye, C.Meng, Z.Xiaobo and H.Weisheng, "Soft failure identification in optical networks based on convolutional neural network," *European Conference on Optical Communication,* pp. 1-3, 2019.

[21] L. Shu, Z. Yu, J. Z. Z. Wan, S. Hu and K Xu, "Dual-Stage Soft Failure Detection and Identification for Low-Margin Elastic Optical network by Exploiting Digital Spectrum Information," *IEEE/OSA Journal of Lightwave Technology,* 2019.

[22] C. V. Saradhi and S. Subramaniam, "Physical layer impairment aware routing (PLIAR) in WDM optical networks: issues and challenges," *IEEE Communications Surveys Tutorials,* vol. 11, p. 109–130, 2009.

[23] P. Dönmez, Introduction to Machine Learning, Cambridge: MA The MIT Press2010., 2013.

[24] D.Rafique and L.Velasco, "Machine Learning for Optical Network Automation: Overview, Architecture and Applications," *EEE/OSA Journal of Optical Communications and Networking (JOCN),* vol. 10, 2018.

[25] K.Tewabe, "A Termination Prediction for Postpaid Mobile Service using Machine learning: the case of ethio telecom," Addis Ababa University, Addis Ababa, 2022.

[26] H. T, R.Tibshirani and J.Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, New York: Springer series in statistics, 2008.

[27] W. Mo, C. L. Gutterman, Y. Li, S. Z. G. Zussman and D. C.Kilper, "Deep Neural Network Based Wavelength Selection and Switching in ROADM Systems," *Journal of Optical Communications and Networking,* 2018.

[28] A.Criminisi, J.Shotton and E.Konukoglu, "Decision forests for classification, regression, density estimation, manifold learning and semi-supervised learning," Microsoft Research Cambridg, UK, 2011.

[29] P.Olivier, "Random Forests for Medical Applications," Technical University Munich , Munich , 2012.

[30] Neil.Y.Yen, Jia-Wei.Chang, J.-Y. Liao. and ou-Ming.Yong, "“Analysis of interpolation algorithms for the missing values in IoT time series: a case of air quality in Taiwan.," *Journal of Supercomputing,* vol. 6475–6500, 2020.

[31] N. Y. Yenl, J.-W. Chang, Y. L. Liao and M. Y. Yong, "Analysis of interpolation algorithms for the missing values in IoT time series: a case of air quality in Taiwan.” In: The Journal of Supercomputing," *Journal of Supercomputing,* vol. 76., p. pp. 6475–6500., 2020.

[32] J.D.Hamilton, Time Series Analysis. Princeton, NJ, USA: Princeton Univ. Press,, 2020.

[33] F. Musumeci, C. Rottondi, G. Corani, S. Shahkarami, F. Cugini and M. Tornatore, "A Tutorial on Machine Learning for Failure Management in Optical Networks," IEEE/OSA Journal of Lightwave Technology, vol. 37, p. 4125–4139 , 2019.

[34] S. Barzegar, M. Ruiz, A. Sgambelluri, F. Cugini, A. Napoli and L. Velasco, "Soft-Failure Detection, Localization, Identification, and Severity Prediction by Estimating QoT Model Input Parameters," IEEE Transactions on Network and Service Managemen, vol. 18, pp. 2627-2640, 2021.

[35] M. E.Prabhakar, "A hybrid combined under-over sampling method for class imbalanced datasets," International Journal of Research and Advanced Development (IJRAD) , p. 27, 2018.