Smart Optical Fingerprint Analysis for Soft Failure Diagnosis in Transport Network

Abdennour Ben Terki*, Ahmed Triki†, Mathieu Herviou†, Maxime Laye†,
Bertrand Le Guyader†, Sébastien Buis†, Nicola Sambo*
*Scuola Superiore Sant'Anna, Piazza Martiri della Libertà, 33, 56127 Pisa, Italy
†Orange Innovation, 2 Avenue Pierre Marzin, 22300 Lannion, France
Corresponding author: abdennour.benterki@sssup.it

***Abstract*—A novel multistage ML model identifies the type, severity, and localization of failures based on their optical fingerprint. Data is sourced from a test-bed emulating two common optical network failures: connector issues and laser drift.**

***Index Terms*—soft failure, machine learning, optical network**

I. Introduction

Managing failures within the fiber optic transport network is a complex undertaking. Until now, identifying and locating faults has been a largely manual and time-consuming process. The intricate nature of optical network failures has increased the demand for a robust and reliable failure management system. This system should tackle not only hard failures (e.g., fiber cuts) but also predict soft failures, such as those caused by equipment aging, which can gradually degrade performance and result in significant consequences if not addressed promptly [1]. The use of Machine Learning (ML) techniques for soft-failure management in optical networks has been widely explored in the literature. The long-short-term memory (LSTM) neural network-based approach has shown high accuracy in anomaly detection in optical networks [2], while the authors in [3] used a hybrid model combining Convolutional Neural Networks (CNN) and LSTM to achieve higher accuracy in failure prediction compared to approaches using a single model. In [4], the authors proposed an approach based on the Deep Neural Evolution Network (DNEN) to accurately locate the link where a failure occurred. Despite these advancements, there is still a need for a comprehensive model that can not only detect the soft failures, but also evaluate their severity and accurately localize them within the optical network.

In this paper, we propose a novel multistage ML model for managing soft failures in optical networks based on XGBoost (XGB) and Random Forest (RF) algorithms. This model is capable of identifying the type, the severity and the localization of the failures solely based on the optical fingerprint associated with each failure. By *fingerprint*, we refer to the variation in Quality of Transmission (QoT) metrics resulting from these failures on the receiver side. The model was trained and tested using data generated from an industrial-grade test-bed optical network designed to emulate two of the most common types of soft failure in optical networks. In the following sections, we will describe the test-bed setup used to generate data, followed by an explanation of our ML model. Finally, we will evaluate the performance of the model.

This work has received funding from the European Union's Horizon 2020 MENTOR program under the Marie Skłodowska-Curie Grant Agreement N. 956713.

II. Test-Bed Setup

The test-bed illustrated in Fig. 1 features a transmission line designed to replicate a real operational transport network. This line includes six amplifiers with constant gain, arranged in five spans over a distance of 374 km of fiber. On the transmission side, we have two transmitters, operating at 193.1875 THz and 193.275 THz, with a bandwidth of 87.5 GHz, a symbol rate of 67 Gbaud, and using 8-QAM modulation format. The channels under study share the C-band spectrum with 33 channels, each with a bandwidth of 37.5 GHz. These channels are filled with optical noise generated by an Amplified Spontaneous Emission (ASE) comb source. All optical signals are shaped and injected into the optical line using a waveshaper, which adjusts the input power per channel based on the experimental requirements. On the reception side, a filter extracts the signals under study before they are processed by the receiver. Using this test-bed, we conduct two set of experiments involving controlled perturbations that enable the emulation and investigation of the impact of two prevalent soft failures in optical networks.

- Connector failure: is emulated by positioning a Variable Optical Attenuator (VOA) across spans 1 to 5 and varying its attenuation from 0 to 10 dB in 0.5 dB steps. Additionally, the channel power launched from the waveshaper was adjusted between -5 dBm and 5 dBm in 1 dB increments.
- Laser drift: is emulated by keeping the VOA attenuation at 0 dB while introducing a frequency shift on the transmission laser source ranges from -20 to +20 GHz in 5 GHz steps. As in the previous experiment, the channel power is adjusted between -5 dBm and 5 dBm in 1 dB increments.

In order to perform multiple efficient failure emulation experiences and an accurate data collection, an automated streaming telemetry system was implemented using gRPC/gNMI protocol. This system operates in parallel with Telegraf as the collection agent and InfluxDB as the database, capturing the receiver's QoT performance metrics (shown in Table I)

978-3-903176-67-6 © 2025 IFIP

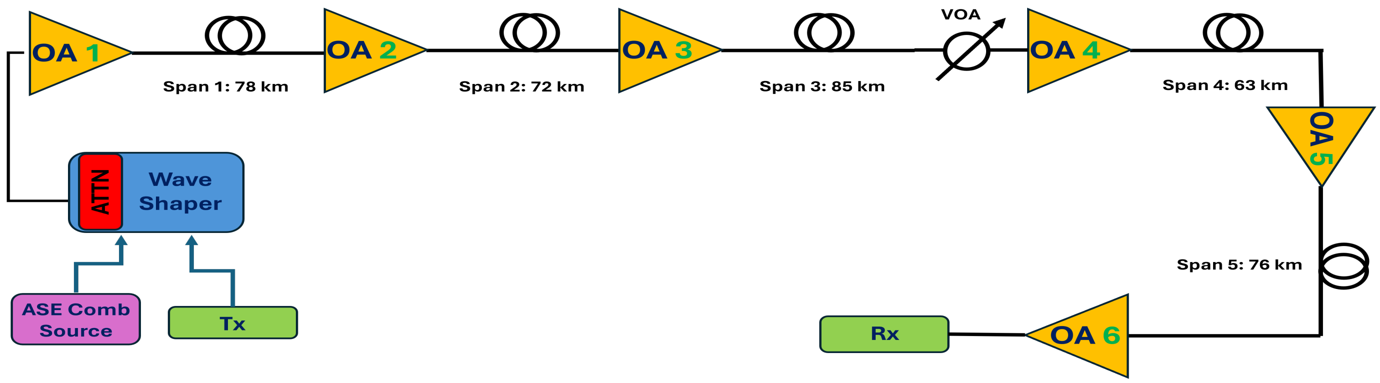

Fig. 1: Test-bed setup for emulating soft failures.

every 10 seconds during each 10-minute experiment. The parameters of the inline equipment were automatically and remotely adjusted according to the requirements of the failure use-case experiments using the SSH protocol, as noted in Table I The proposed multistage ML approach was trained exclusively using QoT metrics collected from the optical terminal transponders without relying on the inline equipment parameters. The dataset also includes supplementary information such as timestamps, experiment descriptions, and failure characteristics (VOA attenuation, VOA localization, and frequency shift). The collected QoT metrics demonstrated that each failure scenario exhibits a distinctive optical performance fingerprint, characterized by unique patterns in the monitored performances, thereby associating specific fingerprints for each type of failure.

TABLE I: Dataset parameters.

Transponders parameters and QoT metrics
P_{in}, P_{out}, CD Pre-FEC BER Q-Margin, ESNR
Experiment parameters
VOA attenuation, Frequency shift Label, Timestamp, Failure location

III. Multistage Failure Management ML Approach

The solution proposed in this paper is designed for diagnosing soft failures (Fig. 2), specifically targeting the two previously mentioned cases that are the focus of our test-bed experiments. Our multistage ML model is described in Table II and aims to meet the following objectives:

1) Soft failure identification: is based on a multi-class Random Forest (RF) classifier enable to determine whether a connector failure, laser drift, or no failure is present.
2) Severity estimation: is an XGB classifier for multi-class classification to estimate the severity of the failure (i.e., low, moderate or severe) according to the failure type identified earlier in the previous stage. For our two use-cases under study - connector failure and laser drift - the severity is determined by the extent of power attenuation and the frequency shift, respectively.
3) Failure localization: is based on a multi-class RF classifier. This stage pinpoints the location of the failure.

The usage of multistages approach combining RF and XGB models enables specialized performance in each stage, which potentially leads to improving the accuracy of the failure management approach and taking advantage of the strength of each algorithm. RF showed better performance in handling high-dimensional data one-to-one classification justifying its selection for identification and localization stages. Meanwhile, XGB's gradient boosting approach is more capable of capturing nuanced patterns in the severity estimation stage [5]. The model continuously receives streaming telemetry data. In each iteration, as described in Table II, the first stage analyzes the optical fingerprint to determine if any previously learned failures are present. If a failure is detected, the model then uses the second and third stages to assess its severity and localization. It is important to note that the latter two stages are trained on data specific to the identified failure, while the first stage is trained using a combination of data from all failures.

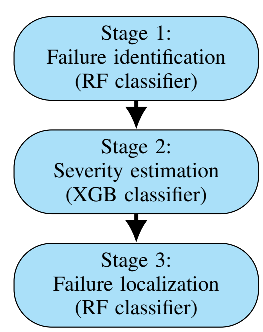

Fig. 2: The proposed ML approach flowchart.

IV. Results Analysis

The models were trained and tested on dataset collected from the test-bed, considering two failure types: connector failure and laser drift under different scenarios and with variety of optical inline parameters. Fig. 3 shows the proposed ML-approach performances in the form of confusion matrices. The Random Forest classifier used for failure identification

TABLE II: The proposed ML stages description.

Stage	UC 1: Connector Failure	UC 2: Laser Drift
Failure Identification	Determine if there is a failure and its type or not	
Severity Estimation	The severity depends on the attenuation value in dB	The severity depends on the frequency shift in GHz
	Low: [1-3] Moderate: [4-6] High: [7-10]	Low: [-5, 5] Moderate: [-10, 10] High: [-20, 20]
Failure Localization (Specific to UC 1 only)	The span where the failure occurred	Not applicable

achieved an overall accuracy of 95.12%. It demonstrated high precision in distinguishing connector failures (97%) and laser drift (98%) from no-failure scenarios. However, there was some misclassification between no-failure and connector failure cases. For severity estimation, the XGBoost classifier achieved an accuracy of 94.30%. It showed strong performance across all severity levels, with the highest accuracy for high-severity cases at 97.6%. There was minor confusion between adjacent severity levels, particularly between low and moderate severities. The model demonstrated minimal misclassification between non-adjacent severity levels, indicating its ability to distinguish between significantly different severity degrees. The Random Forest classifier for failure localization in connector failure scenarios achieved an impressive accuracy of 99.13%. It showed exceptional performance across all spans, with correct classifications ranging from 1675 to 1756 instances per span. There were minimal misclassifications between adjacent spans, and near-perfect distinction between non-adjacent spans.

V. Conclusions

In this paper, we introduce a novel multistage machine learning approach for diagnosing soft failures in optical networks. Our solution integrates Random Forest and XGBoost algorithms to achieve three key objectives: failure identification, severity estimation, and localization. By utilizing streaming telemetry data from limited set of QoT metrics, the model can detect failure fingerprints (i.e., variations in QoT caused by failures) thereby minimizing the need to collect data from multiple line equipment ports and focusing telemetry on the transponder's reception side. Using data from an industrial-grade and fully automated test-bed that emulates connector failures and laser drift scenarios, our model demonstrated impressive performance, achieving accuracy rates of 95.12% for failure identification, 94.30% for severity estimation, and 99.13% for failure localization. We are currently working to replicate additional failure scenarios with the same test-bed to expand our model's capability to address a wider range of soft failures in optical networks.

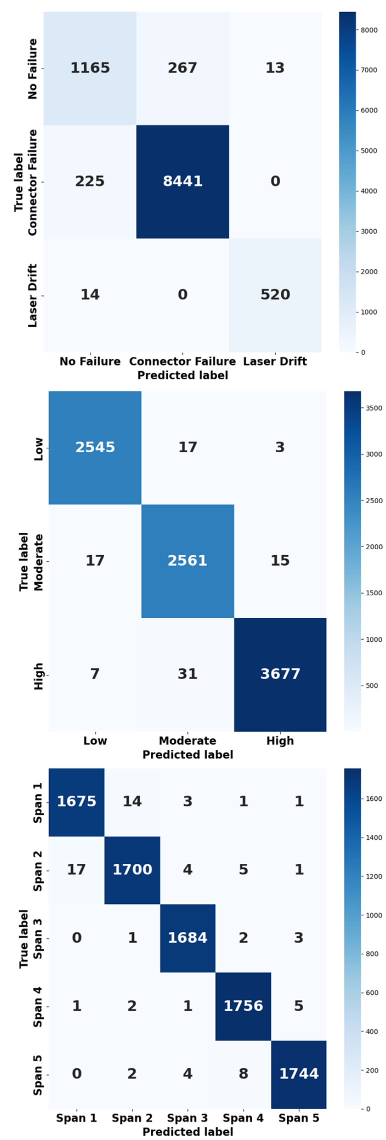

Fig. 3: Confusion matrices for all stages

References

[1] D. Wang et al. "a review of machine learning-based failure management in optical networks". *SCIENCE CHINA Information Sciences*, 65(11):211302, 2022.

[2] S. Behera et al. "machine learning for real-time anomaly detection in optical networks". In *2023 23rd International Conference on Transparent Optical Networks (ICTON)*, pages 1–4, 2023.

[3] Z. Tan et al. "network fault prediction based on cnn-lstm hybrid neural network". In *2019 International Conference on Communications, Information System and Computer Engineering (CISCE)*, pages 486–490, 2019.

[4] X. Zhao et al. "accurate fault location based on deep neural evolution network in optical networks for 5g and beyond". In *2019 Optical Fiber Communications Conference and Exhibition (OFC)*, pages 1–3, 2019.

[5] B. Wohlwend. "decision tree, random forest, and xgboost: An exploration into the heart of machine learning". https://medium.com, 2023. Accessed: 2024-10-15.