Leveraging Digital Twins to Build Comprehensive Network Failure Databases for Predictive Machine Learning in Optical Transport Networks

2025 IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS) | 979-8-3315-2681-8/25/$31.00 ©2025 IEEE | DOI: 10.1109/ANTS66931.2025.11430070

Leveraging Digital Twins to Build Comprehensive Network Failure Databases for Predictive Machine Learning in Optical Transport Networks

Mashboob Cheruvakkadu Mohamed
Politecnico di Torino, Italy
mashboob.cheruvakkadu@polito.it

Renato Ambrosone
Politecnico di Torino, Italy

Rocco D'Ingillo
Politecnico di Torino, Italy

Muhammad Umar Masood
Politecnico di Torino, Italy

Gulmina Malik
Politecnico di Torino, Italy

Stefano Straullu
Links Foundation, Torino, Italy

Sai Kishore Bhyri
Nokia, India

João Pedro
Nokia, Portugal

Antonio Napoli
Nokia, Germany

Gabriele Maria Galimberti
Nokia, Italy

Walid Wakim
Nokia, USA

Vittorio Curri
Politecnico di Torino, Italy

***Abstract*—The increasing complexity and capacity demands of modern optical communication networks require advanced tools for predictive maintenance and fault prevention. This study presents a systematic methodology for generating a comprehensive network failure database using Digital Twins (DTs) based on the open-source optical network simulation framework GNPy (Gaussian Noise model in Python). GNPy enables the generation of synthetic datasets to support the development of Machine Learning (ML) models for proactive fault detection and network performance optimization. By modeling a wide range of failure scenarios such as amplifier degradation, span loss variation, and component misconfigurations, GNPy enables detailed analysis of their impact on critical transmission parameters including Generalized Signal-to-Noise Ratio (GSNR), Optical Signal-to-Noise Ratio (OSNR), and channel power. The proposed methodology allows the generation of a diverse and scalable database that reflects both normal and degraded network operating conditions. This synthetic dataset provides valuable input for training ML algorithms to detect subtle performance degradations and predict potential failures before they impact running services. By leveraging this simulation-based data generation approach, network operators can significantly enhance the reliability and resilience of their transport networks. Furthermore, the integration of GNPy with data-driven ML models lays the foundation for intelligent network management systems capable of automated fault prediction and proactive maintenance strategies in future high-capacity optical networks.**

***Index Terms*—Digital Twin, Failure Prediction, Fault Detection, Machine Learning, Network Reliability, Optical Networks, Proactive Failure Management, Soft Failures.**

I. Introduction

Optical networks form the high capacity foundation of modern communication systems, supporting ever-increasing demands for bandwidth, latency-sensitive applications, and uninterrupted service delivery. Given the critical nature of these networks, ensuring their reliability and performance is of paramount importance. Traditional reactive or time-based maintenance methods do not fully exploit the early signs that soft failures usually provide, missing a potential opportunity to mitigate service interruptions and economic losses for network operators [1]. In this context, predictive maintenance has emerged as a transformative approach, aiming to anticipate failures and take preventive actions to maintain network integrity and service continuity. Predictive maintenance can complement existing methods, contributing to improving the robustness of optical networks [2].

In recent years, Machine Learning (ML) techniques have been increasingly explored for failure management in optical networks [3]–[9]. These techniques offer the potential to learn complex patterns from historical data, enabling intelligent prediction of component degradation, link instability, or network element failures. ML enables automation of alarm analysis and performance optimization in optical networks [10]. ML-based models, such as supervised classifiers and time-series predictors, can effectively detect subtle signs of impending faults that may not be evident through traditional threshold-based monitoring. As a result, ML-driven predictive maintenance promises to reduce downtime, optimize resource utilization, and improve overall efficiency of network operations. However, the major hindrance to the practical application of ML in optical networks is the scarce/no availability of high-quality training data. Effective ML models require large volumes of labeled data capturing both normal and failure states under various network conditions. In reality, failure events in optical networks are rare and uneven and, when they do occur, the available data are incomplete, noisy, or unlabeled. Furthermore, data imbalance with far fewer failure instances compared to normal operations can hinder model generalization and lead to biased predictions. This challenge is further exacerbated by difficulties in collecting data from live

This publication has received funding from the project PNRR-NGEU (MUR–DM117/2023), the EU Horizon 2020 projects ALLEGRO (GA: 101092766) and SENSEI (GA: 101189545).

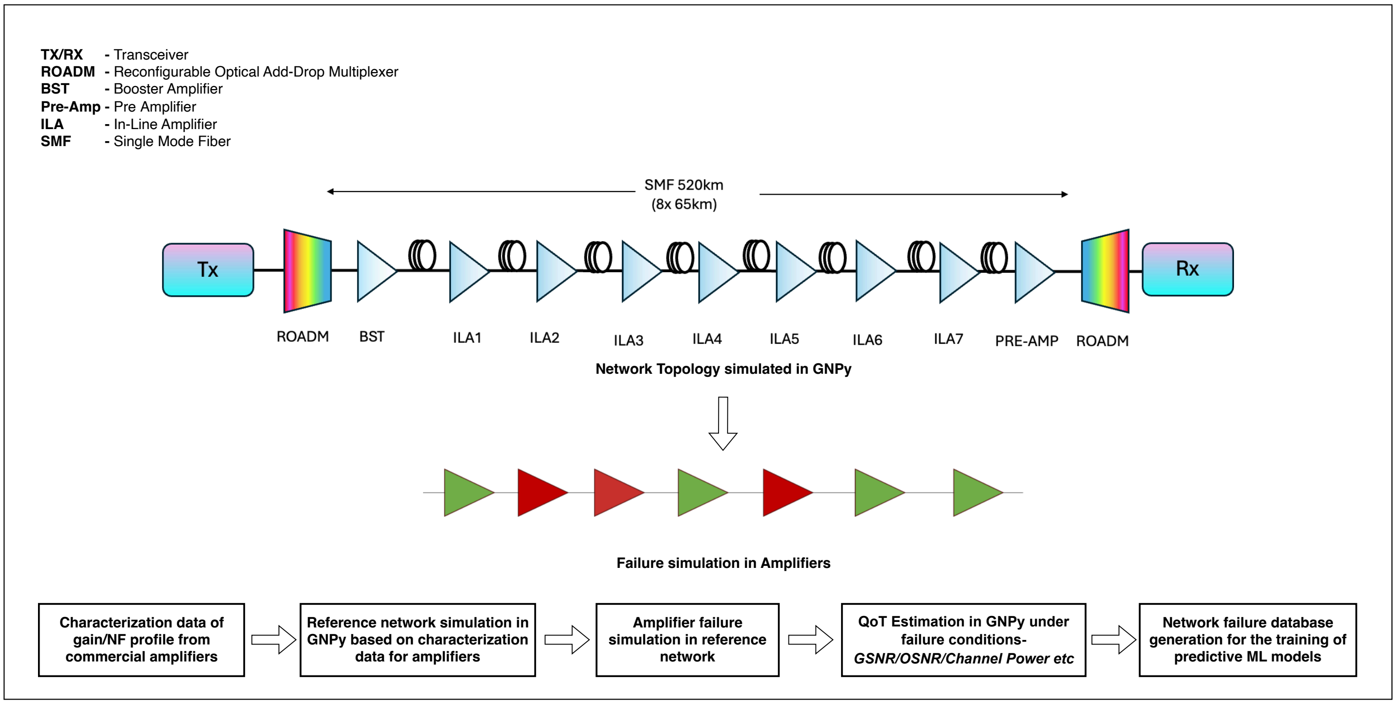

Fig. 1: Network failure database generation for the training of predictive ML models using GNPy - a schematic representation.

production networks, where operational constraints, privacy concerns, and the risk of service disruption limit the extent and granularity of data that can be captured.

To address the lack of failure data, data augmentation techniques have been explored in the existing literature [11], [12]. These techniques involve synthetically generating data samples by manipulating existing data. Although data augmentation can help improve model training in some cases, it fails to capture the full complexity and dynamics of real-world network failures. Moreover, artificially transformed data may not generalize well enough to actual network conditions, leading to unreliable predictions and overfitting. A promising alternative to overcome these challenges is to use Digital Twins (DTs), i.e. the virtual replicas of physical optical networks that simulate their behavior, topology, and operational characteristics. DTs can model the evolution of the network in various configurations, traffic patterns, and fault scenarios, enabling the generation of realistic high-fidelity datasets to train and validate ML models [13], [14]. Unlike data augmentation, DTs can incorporate physics-based modeling, domain knowledge, and historical telemetry to simulate rare or possible events with greater accuracy. ML models designed to address network failures can be built and evaluated effectively by leveraging scalable and adaptable DTs.

This article describes and demonstrates the generation of a network failure database using the open-source optical network simulation framework – GNPy (Gaussian Noise model in Python) [15]. Leveraging GNPy, a comprehensive failure dataset can be created by modeling different scenarios of failure conditions and then evaluating the impact of the failures on transmission parameters such as Generalized Signal-to-Noise Ratio (GSNR), Optical Signal-to-Noise Ratio (OSNR) and channel power. This study focuses on the prediction and analysis of optical amplifier performance in optical networks. To achieve this, the GNPy simulation tool is used to effectively model the behavior of the optical network under varied conditions, from normal operational scenarios to simulating various failure conditions of optical amplifiers. The general concept of network failure database generation for predictive ML in optical networks is illustrated in Fig. 1. By capturing both normal and faulty behaviors, the simulation provides information on network performance, thereby generating the necessary datasets representing network failures to anticipate potential issues related to amplifier degradation or malfunction. The dataset generated from the simulations serves as the foundation for the training of the ML model based on Gradient Boosted Decision Trees (GBDT). This model is trained to predict and locate faulty optical amplifiers in the network span. By analyzing patterns and variations in the transmission parameter data, the GBDT ML model learns to differentiate between normal operating conditions and those indicative of amplifier failures. This predictive capability enables proactive maintenance that contributes to improved network reliability and reduced downtime. The remainder of this paper is organized as follows. Section II provides a comprehensive overview of the GNPy simulation framework, including optical network modeling (Section II-A), characterization of optical amplifiers (Section II-B), and simulation of amplifier failure scenarios along with the generation of a corresponding failure database (Section II-C). Section III presents the ML model trained using the generated data for failure prediction and its performance evaluation metrics. Finally, Section IV concludes the paper and outlines directions for future research.

II. GNPy and Network Failure Data Generation

A. Optical Network Simulation using GNPy

The flexible and open framework of GNPy has enabled it to be a widely adopted tool for open optical networks. It provides capabilities to model and analyze transparent optical routes, check signal reachability, and evaluate the capacity and efficiency of both new and existing network setups. Its modular structure enables a detailed representation of individual disaggregated components, aligning well with the needs of modern, software-driven, and open networking environments [15], [16]. As shown in Fig. 1, DT of the reference network is constructed using the GNPy. The reference network includes transceivers, Reconfigurable Optical Add-Drop Multiplexers (ROADMs), and amplifiers on a single-mode fiber link with a total distance of 520 km, divided into eight spans, each measuring 65 km.

B. Optical Amplifier Characterization

The characterization of the Erbium Doped Fiber Amplifiers (EDFAs) was performed to evaluate both the gain profile and the Noise Figure (NF) across the relevant C-band spectrum. Each EDFA was driven under nominal operating conditions using factory-calibrated bias currents and internal temperature stabilization. Gain profile measurements were carried out by injecting a low-power continuous-wave tunable laser source into the EDFA input. The signal was swept across the C-band in steps of 0.1 nm, and the corresponding output power was measured using an Optical Spectrum Analyzer (OSA) with a resolution bandwidth of 0.05 nm. The gain $G(\lambda)$ at each wavelength was computed as:

$$G(\lambda) = P_{\text{out}}(\lambda) - P_{\text{in}}(\lambda) \tag{1}$$

where P_{in} and P_{out} are the input and output optical powers (in dBm) at wavelength λ, respectively.

For the NF profile characterization, measurements were carried out under full spectral load conditions at different target gain settings. A set of densely spaced input channels was first applied to simulate the fully loaded amplifier. Then, each channel was turned off sequentially and the corresponding out-of-band Amplified Spontaneous Emission (ASE) noise generated by the EDFA at the vacant spectral slot was measured using the OSA. This procedure allowed reconstruction of the ASE noise profile per channel $n_{\text{ASE}}(\lambda)$ under realistic operating conditions. The NF at each wavelength was then calculated using:

$$\text{NF}(\lambda) = 10 \log_{10} \left(\frac{n_{\text{ASE}}(\lambda) \cdot h\nu \cdot \Delta\nu}{G(\lambda) \cdot P_{\text{in}}(\lambda)} + 1 \right) \tag{2}$$

where h is Planck's constant, ν is the optical frequency, and $\Delta\nu$ is the resolution bandwidth of the OSA. This approach ensures an accurate estimation of the NF under realistic spectral loading and amplifier saturation conditions.

C. Optical Amplifier Failures and Database Generation in GNPy

Although optical amplifiers are integral components in long-haul and high-capacity optical communication systems, serving to mitigate signal attenuation and maintain transmission integrity over long distances, they are susceptible to a range of impairments that can adversely affect network performance and directly degrade Quality of Transmission (QoT). A prevalent issue observed in optical amplifiers is the gradual degradation of the amplifier gain. This phenomenon is typically attributed to the aging of pump lasers, reduction in pump efficiency, or gain medium saturation under high input power conditions. Over time, these factors contribute to a decrease in amplification ability, resulting in suboptimal signal levels at the output [13]. Persistent gain reduction can lead to signal degradation across multiple spans, particularly in cascaded amplifier configurations. Another significant concern is the increase in NF, primarily due to the elevated noise of ASE [17]. ASE levels may escalate as a consequence of pump laser instabilities, spectral drift, or aging-related deterioration of the pump source. The increased noise contribution from ASE adversely affects OSNR, thereby impairing system performance, especially in high-order modulation formats where OSNR margins are tight. Furthermore, optical amplifiers may exhibit output power anomalies, including unexpected fluctuations such as transient spikes or sudden drops. These irregularities are symptomatic of pump current variations, control circuitry malfunctions, or complete pump laser failure. Such deviations can lead to imbalanced channel power profiles and may trigger protection mechanisms or service interruptions in dynamic optical networks. Persistent impairments in optical amplifiers can lead to critical system-level failures, causing substantial signal degradation and network instability. As optical amplifiers are essential to maintain transmission across long-haul links in Optical Line Systems (OLS), early detection of performance degradation is necessary to prevent cascading transmission faults.

GNPy is used to simulate optical network performance by accounting for key effects of the physical layer such as ASE noise, nonlinear impairments, and dispersion accumulation. The simulation operates on a detailed network topology, as depicted in Fig. 1, encompassing fiber span properties, optical node data, and comprehensive technical specifications including amplifier characteristics and transponder parameters. Using these inputs, GNPy computes transmission quality metrics such as GSNR, OSNR, and channel power. Signal propagation is analyzed by simulating the transmission of optical signals through various network components, such as fiber links, amplification stages, and ROADMs. The framework captures detailed physical characteristics, including span attenuation, fiber lengths, wavelength-dependent gain dynamics, and nonlinear interference. By iteratively applying the Gaussian noise (GN) model to each segment of the transmission path, GNPy quantifies the cumulative degradation of signal quality resulting from these impairments.

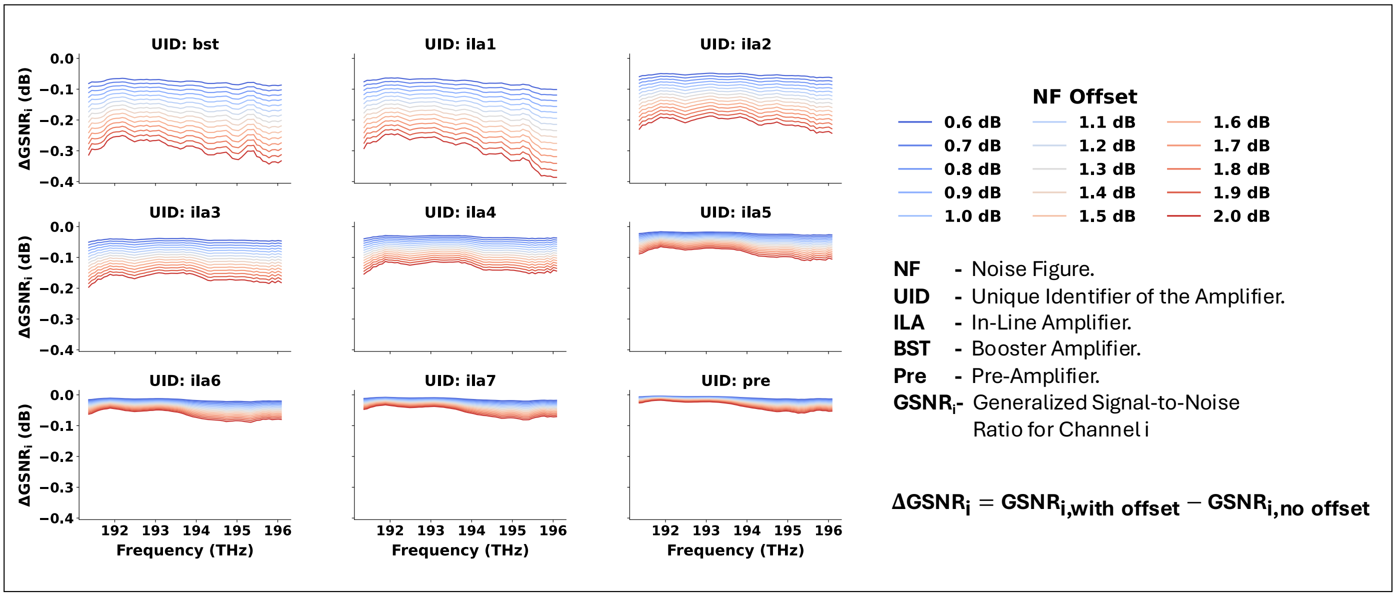

Fig. 2: Simulation results obtained using GNPy show the per-channel GSNR degradation as a function of increasing NF offsets applied to different amplifiers along the span. The plot highlights that both the position of the amplifier and the magnitude of the NF offset influence the per-channel GSNR. Lower offset values are shown in blue, transitioning to red as the offset increases.

The GNPy simulation framework was used to generate the training dataset by systematically varying critical network parameters over a large set of simulation runs. This approach enabled modeling of a broad range of network configurations and operational conditions, including infrequent fault scenarios that are often challenging to capture in real-world networks due to their rarity and the practical limitations of collecting data from live systems. By adjusting parameters such as amplifier gain settings, transponder configurations, and fault injection scenarios, GNPy enables emulation of a wide range of network behaviors and their associated impacts on signal quality metrics.

The gain and NF profiles of commercially deployed optical amplifiers, characterized as detailed in Section II.B, were integrated into the simulation framework to reflect the realistic physical layer behavior. In this work, to emulate scenarios involving elevated levels of ASE noise, a controlled NF offset was applied to the baseline NF profiles of the amplifiers deployed within each fiber span. This adjustment allowed for systematic evaluation of the resulting degradation in signal quality, quantified in terms of the per channel GSNR at the receiver end. The ASE power generated in optical amplifiers over a bandwidth Δf can be expressed as [18]:

$$P_{\text{ASE}} = 2n_{\text{sp}}h\nu(G-1)\Delta f \tag{3}$$

where n_{sp} is the spontaneous emission factor, h is Planck's constant, ν is the optical frequency, G is the amplifier gain (linear scale), Δf is the optical noise bandwidth.

The NF of the amplifier is influenced by the spontaneous emission factor n_{sp} as:

$$\text{NF} = 2n_{\text{sp}} \tag{4}$$

Assuming $G \gg 1$, increasing NF increases P_{ASE}, resulting in the degradation of GSNR, defined as:

$$\text{GSNR} = \frac{P_{\text{signal}}}{P_{\text{ASE}} + P_{\text{NLI}}} \tag{5}$$

where P_{NLI} accounts for non-linear noise contributions.

Through controlled variation of the NF of selected amplifiers in the simulated topology, GNPy enables to investigate the influence of noise behavior. Using GNPy, the generation of per-channel GSNR data under failure conditions reflects the aging of the amplifier or soft failure behavior. The simulation results, presented in Fig. 2, illustrate the variation in GSNR as a function of the NF offset across all channels. The difference in GSNR per channel received at the transceiver is obtained as follows.

$$\Delta\text{GSNR}_i = \text{GSNR}_{i,\text{no offset}} - \text{GSNR}_{i,\text{with offset}} \tag{6}$$

where:

- $\text{GSNR}_{i,\text{no offset}}$ is the GSNR for channel i without any offset applied.
- $\text{GSNR}_{i,\text{with offset}}$ is the GSNR for channel i with the applied offset.

The findings indicate a clear trend to increase the degradation of GSNR per channel with higher offset values of NF. Furthermore, the data suggest that the impact of ASE noise is not uniform across the amplifiers in the link. Instead, it is influenced by the specific position and role of each amplifier within the transmission line. This variation arises from the cascading nature of optical amplification, where noise introduced at the upstream stages accumulates and interacts with the subsequent amplification stages, leading to non-linear and span-dependent effects on overall transmission quality.

The generated database serves as a critical foundation for the development and evaluation of ML models that aim to predict optical amplifier-related faults. Leveraging the flexibility of the GNPy simulation framework, the dataset encompasses a wide range of realistic and diverse network conditions, including both nominal and degraded states. This comprehensive coverage improves the robustness and generalization capability of the models, enabling them to detect and predict faults under different operational scenarios.

III. ML Model and Performance Evaluation

To develop an effective predictive model for optical amplifier forecasting, the GBDT ML technique [19] is used. GBDT is a highly effective technique within the ensemble learning paradigm. Ensemble methodologies integrate predictions from multiple base models, enhancing generalization capabilities and reducing sensitivity to data variability. Specifically, GBDT operates by sequentially training a series of weak learners, usually shallow decision trees, each aimed at addressing the residual errors of its predecessors [20]. This iterative refinement is guided by gradient descent optimization, which systematically reduces the loss of prediction of the model. Through this process, the ensemble progressively improves its accuracy by focusing on the instances most difficult to predict in former iterations [21]. Thus, GBDT offers a reliable and scalable approach for high-precision predictive modeling in complex tasks such as predicting optical amplifier failures. For each Unique Identifier (UID) corresponding to a specific amplifier, the ML model leverages span-specific parameters and channel characteristics to perform fault prediction. The ML model outputs a binary label, where 0 indicates the absence of any fault and 1 denotes the presence of an incoming fault. Through this approach, the ML model effectively predicts and localizes faulty amplifiers in the OLS.

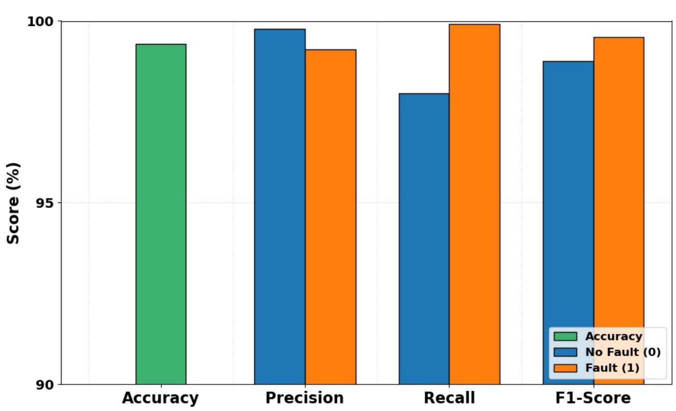

Fig. 3: Bar plot displaying Accuracy, Precision, Recall, and F1-Score, illustrating class-wise classification performance.

To evaluate the performance of the ML model, standard classification metrics are used, such as accuracy, precision, recall, and the F1 score. In binary classification, each data point

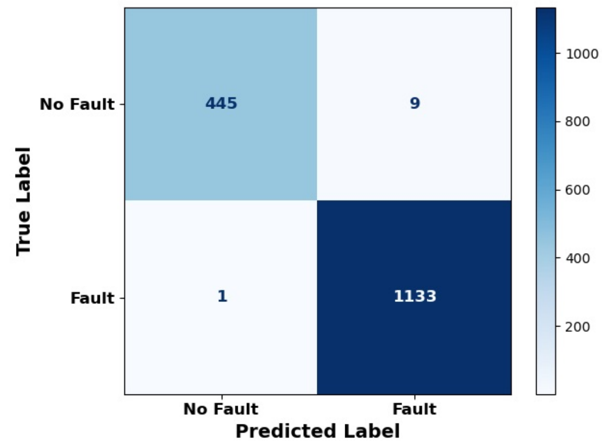

Fig. 4: Confusion Matrix showing the True vs. Predicted Class Labels, visualizing the misclassifications.

is categorized into one of two groups, labeled as positive or negative. This classification yields four possible outcomes: true positives, true negatives, false positives, and false negatives. Accuracy represents the ratio of correctly identified cases, both positive and negative, to the total number of predictions, providing a general measure of the model's correctness. Precision focuses on the quality of positive predictions, indicating the proportion of those that are actually correct and thus reflecting the model's ability to limit false alarms. Recall, also called sensitivity, captures how effectively the model detects actual positives, emphasizing its success in finding all relevant fault instances. The F1 score combines precision and recall into a single metric that balances the trade-off between identifying relevant instances (recall) and minimizing incorrect positive predictions (precision) [22], [23]. These key performance metrics of the proposed model are illustrated in Fig. 3, which highlights its robustness and reliability in predicting failures. The model achieved an impressive overall accuracy of 99%, demonstrating its high effectiveness in correctly identifying target classes. The dataset was partitioned using a 70:30 train-test split, ensuring that the model had a sufficient amount of data to learn from while maintaining a fair evaluation of unseen data. The train-test split, used in this study, follows a standard practice in ML, designed to strike a balance between allowing the model to learn effectively from the training data while also assessing its ability to generalize to new samples.

To better understand the performance of the model, the confusion matrix depicted in Fig. 4 provides further insight into the classification performance of the model at the granular level. A confusion matrix is a useful tool for evaluating classification models, as it shows the number of true positives, true negatives, false positives, and false negatives. The confusion matrix shown in Fig. 4 reveals a total of one false negative and nine false positives, indicating a strong ability to correctly classify true positive cases with minimal errors. The low number of false negatives suggests that the model rarely misses actual positive instances, which is particularly important in this use case of optical amplifier failures where such omissions could carry significant consequences. The presence of nine false positives, while minimal, suggests that the model is generally effective in distinguishing between classes, with only rarely misclassifications of negative samples as positive. These metrics show the overall performance reliability of the ML model.

IV. Conclusions and Future Work

In conclusion, this study demonstrates the feasibility and effectiveness of using GNPy open-source simulation framework to generate synthetic datasets for ML based fault prediction in optical communication networks. By simulating a wide spectrum of failure scenarios and analyzing their effects on key transmission parameters such as GSNR, OSNR, and channel power, the proposed methodology enables the creation of a scalable and diverse failure database. This dataset serves as a critical foundation for training ML models capable of detecting performance degradations and forecasting faults before they impact network service, thereby supporting the development of intelligent and proactive network management strategies. This study showcased the prediction of optical amplifier failures by training a GBDT ML model using the simulated dataset, achieving an accuracy of 99%. This result highlights the potential of simulated data-driven approaches leveraging physics-based modeling and domain knowledge to accurately identify and localize network faults.

For future work, the underlying simulation-based methodology can be extended to develop ML models for predicting failures in other critical components and subsystems of optical networks beyond the amplifier-related scenarios discussed in this paper, such as failures in ROADMs, transponders, or attenuation-related issues in different types of optical fibers. By simulating fault conditions across these components, it is possible to generate datasets that capture a wider spectrum of network degradation patterns. Such an extension would enhance the general applicability of the proposed framework and contribute to the development of fault prediction models for the development of better maintenance strategies for next-generation high-capacity and reliable optical networks.

References

[1] Sergio Cruzes. Failure management overview in optical networks. *IEEE Access*, 12:169170–169193, 2024.

[2] Mashboob Cheruvakkadu Mohamed, Muhammad Umar Masood, Imran Chowdhury Dipto, Renato Ambrosone, Gulmina Malik, Stefano Straullu, Sai Kishore Bhyri, Gabriele Maria Galimberti, João Pedro, Antonio Napoli, Walid Wakim, and Vittorio Curri. Network data based transfer learning failure prediction agent pre-trained using digital twin. In *2025 International Conference on Optical Network Design and Modeling (ONDM)*, pages 1–6, 2025.

[3] Danish Rafique and Luis Velasco. Machine learning for network automation: overview, architecture, and applications [invited tutorial]. *Journal of Optical Communications and Networking*, 10(10):D126–D143, 2018.

[4] Francesco Musumeci, Cristina Rottondi, Giorgio Corani, Shahin Shahkarami, Filippo Cugini, and Massimo Tornatore. A tutorial on machine learning for failure management in optical networks. *Journal of Lightwave Technology*, 37(16):4125–4139, 2019.

[5] Behnam Shariati, Marc Ruiz, Jaume Comellas, and Luis Velasco. Learning from the optical spectrum: Failure detection and identification. *Journal of Lightwave Technology*, 37(2):433–440, 2019.

[6] Camille Delezoide, Petros Ramantanis, and Patricia Layec. Streamlined failure localization method and application to network health monitoring. *Journal of Lightwave Technology*, 41(19):6119–6125, 2023.

[7] Jatin Babbar, Ahmed Triki, Reda Ayassi, and Maxime Laye. Machine learning models for alarm classification and failure localization in optical transport networks. *Journal of Optical Communications and Networking*, 14(8):621–628, 2022.

[8] Josh W. Nevin, Sam Nallaperuma, Nikita A. Shevchenko, Xiang Li, Md. Saifuddin Faruk, and Seb J. Savory. Machine learning for optical fiber communication systems: An introduction and overview. *APL Photonics*, 6(12):121101, 12 2021.

[9] Faisal Nadeem Khan, Qirui Fan, Chao Lu, and Alan Pak Tao Lau. An optical communication's perspective on machine learning and its applications. *Journal of Lightwave Technology*, 37(2):493–516, 2019.

[10] Danshi Wang, Yidi Wang, Xiaotian Jiang, Yao Zhang, Yue Pang, and Min Zhang. When large language models meet optical networks: Paving the way for automation. *Electronics*, 13(13), 2024.

[11] Lareb Zar Khan, Jo ao Pedro, Nelson Costa, Andrea Sgambelluri, Antonio Napoli, and Nicola Sambo. Model and data-centric machine learning algorithms to address data scarcity for failure identification. *J. Opt. Commun. Netw.*, 16(3):369–381, Mar 2024.

[12] Cheng Xing, Chunyu Zhang, Bing Ye, Danshi Wang, Yinqiu Jia, Jin Li, and Min Zhang. Failure data augmentation for optical network equipment using time-series generative adversarial networks. In *Optical Fiber Communication Conference (OFC) 2023*, page M3G.4. Optica Publishing Group, 2023.

[13] Mashboob Cheruvakkadu Mohamed, Muhammad Umar Masood, Renato Ambrosone, Gulmina Malik, Rocco D'Ingillo, Stefano Straullu, Sai Kishore Bhyri, Gabriele Maria Galimberti, Joao Pedro, Antonio Napoli, Walid Wakim, and Vittorio Curri. Machine learning agents leveraging digital twins for failure prediction in optical networks. In *2025 IEEE International Conference on Machine Learning for Communication and Networking (ICMLCN) (2025 ICMLCN)*, page 5.63, Barcelona, Spain, May 2025.

[14] Mashboob Cheruvakkadu Mohamed, Renato Ambrosone, Muhammad Umar Masood, Gulmina Malik, Stefano Straullu, Sai Kishore Bhyri, Joao Pedro, Antonio Napoli, Gabriele Maria Galimberti, Walid Wakim, and Vittorio Curri. Digital Twin-Integrated binary classifier ML model for EDFA failure prediction. In *The 33rd International Conference on Software, Telecommunications and Computer Networks (SoftCOM 2025)*, page 6, Split, Croatia, September 2025.

[15] Vittorio Curri. Gnpy model of the physical layer for open and disaggregated optical networking [invited]. *Journal of Optical Communications and Networking*, 14(6):C92–C104, 2022.

[16] Alessio Ferrari, Mark Filer, Karthikeyan Balasubramanian, Yawei Yin, Esther Le Rouzic, Jan Kundrat, Gert Grammel, Gabriele Galimberti, and Vittorio Curri. Gnpy: an open source application for physical layer aware open optical networks. *Journal of Optical Communications and Networking*, 12(6):C31–C40, 2020.

[17] Alexander William Setiawan Putra, Minoru Yamada, Hiroyuki Tsuda, and Sumiaty Ambran. Theoretical analysis of noise in erbium doped fiber amplifier. *IEEE Journal of Quantum Electronics*, 53(4):1–8, 2017.

[18] Emmanuel Desurvire. *Erbium-Doped Fiber Amplifiers: Principles and Applications*. Wiley, 1994.

[19] Jerome H. Friedman. Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5):1189 – 1232, 2001.

[20] Zhendong Zhang and Cheolkon Jung. Gbdt-mo: Gradient-boosted decision trees for multiple outputs. *IEEE Transactions on Neural Networks and Learning Systems*, 32(7):3156–3167, 2021.

[21] Ibomoiye Domor Mienye and Nobert Jere. A survey of decision trees: Concepts, algorithms, and applications. *IEEE Access*, 12:86716–86727, 2024.

[22] Lars E. Kruse, Sebastian Kühl, Annika Dochhan, and Stephan Pachnicke. Experimental validation of machine learning-based joint failure management and quality of transmission estimation. *IEEE Photonics Journal*, 15(6):1–9, 2023.

[23] Marina Sokolova and Guy Lapalme. A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, 45(4):427–437, 2009.