POLITECNICO DI TORINO
Repository ISTITUZIONALE

Digital Twin-Integrated Binary Classifier ML Model for EDFA Failure Prediction

Original
Digital Twin-Integrated Binary Classifier ML Model for EDFA Failure Prediction / Mohamed, M.C., Ambrosone, R., Masood, M.U., Malik, G., Straullu, S., Bhyri, S.K., Pedro, J., Napoli, A., Galimberti, G.M., Wakim, W., Curri, V.. - (2025), pp. 1-6. (2025 International Conference on Software, Telecommunications and Computer Networks (SoftCOM) Split (Cro) 18-20 September 2025).

Availability:
This version is available at: 11583/3004446 since: 2025-10-24T13:52:56Z

Publisher:
IEEE

Published
DOI:

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright
IEEE postprint/Author's Accepted Manuscript

©2025 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collecting works, for resale or lists, or reuse of any copyrighted component of this work in other works.

(Article begins on next page)

27 November 2026

Digital Twin-Integrated Binary Classifier ML Model for EDFA Failure Prediction

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

Renato Ambrosone
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, USA

Walid Wakim
Nokia, USA

Vittorio Curri
Politecnico di Torino, Italy

***Abstract*—This study presents a digital twin-enabled framework integrated with a binary classification Machine Learning (ML) model for forecasting failures in Erbium-Doped Fiber Amplifiers (EDFAs). The framework utilizes GNPy, an open-source optical network planning tool, to construct a digital twin that serves as a virtual replica of the physical EDFA system. This digital twin facilitates the estimation of Quality of Transmission (QoT), along with the collection and analysis of key operational parameters. A binary classification model, based on Long Short-Term Memory (LSTM) networks, is trained on the data generated by the digital twin to predict potential EDFA failures, achieving a high prediction accuracy of 98%. This predictive capability enables early fault detection and proactive maintenance, thereby minimizing unplanned downtime and service disruptions. By incorporating real-time analytics and predictive insights, the proposed approach significantly enhances the reliability, availability, and intelligence of optical network management.**

***Index Terms*—Digital Twin, Erbium-Doped Fiber Amplifier, Fault Detection, Machine Learning, Network Reliability, Proactive Failure Management, Soft failures.**

I. Introduction

Optical amplifiers are fundamental to modern optical communications, as they allow direct amplification of optical signals over long distances, eliminating the need for complex and costly Optical-Electrical-Optical (OEO) regeneration and support high-capacity Dense Wavelength Division Multiplexing (DWDM) systems [1]. Erbium-Doped Fiber Amplifier (EDFA), operating in both the C-band (≈1530–1565 nm) and L-band (≈1565–1625 nm), is a key component in multi-band transmission architectures [2]. With the growing demand for scalable and cost-efficient solutions, today's optical network product industries are transitioning to EDFA pluggable modules for applications such as datacenter interconnects and IP over Wavelength Division Multiplexing (IPoWDM) metro networks. Given the critical role of EDFAs in both current and next-generation optical network infrastructures, their proper monitoring and failure management are vital for meeting stringent Service Level Agreements (SLAs). Early detection or prediction of EDFA failures enables network operators to maintain optimal Optical Signal-to-Noise Ratio (OSNR), which is critical for ensuring the desired Quality of Transmission (QoT). Proactive failure management not only improves network reliability but also contributes to reducing Operational Expenditures (OpEx) by minimizing service disruptions, the need for reactive maintenance, and the on-site interventions.

Machine Learning (ML) techniques enable the automation of detection and classification of soft failures, leading to more effective failure management [3]. Several studies have explored the application of ML for soft failure detection in optical networks by training models under normal operating conditions [4], [5] and considering multiple soft failures simultaneously such as an increase in EDFA Noise Figure (NF), frequency drift or drop in transmit lasers, filter misalignment or tightening [6], amplifier gain degradation, transponder power reduction, and fiber attenuation [7], etc. However, this approach presents significant challenges. Training exclusively on normal conditions can lead to inaccurate predictions if latent or undetected faults are present during data collection. Moreover, "normal" operating conditions can vary widely, making comprehensive representation in the training dataset very difficult. Furthermore, training a single ML model to detect multiple types of soft failures across various components—such as amplifiers, transponders, and optical fibers—introduces additional complexity. This increases the difficulty of optimizing model parameters (e.g., batch size, learning rate, and other hyperparameters) and often results in longer inference times. The heterogeneous nature of these failures also contributes to a higher risk of false positives and degraded detection accuracy, thereby limiting the model's practical deployment in real-time network monitoring environments. This work adopts a focused machine learning strategy dedicated solely to modeling and forecasting failures in EDFAs, thereby reducing model complexity and improving detection accuracy.

This publication has received funding and support from the project PNRR-NGEU (MUR–DM117/2023) and from the European Union's Horizon Europe research and innovation programme under grant agreement No. 101092766 (ALLEGRO Project).

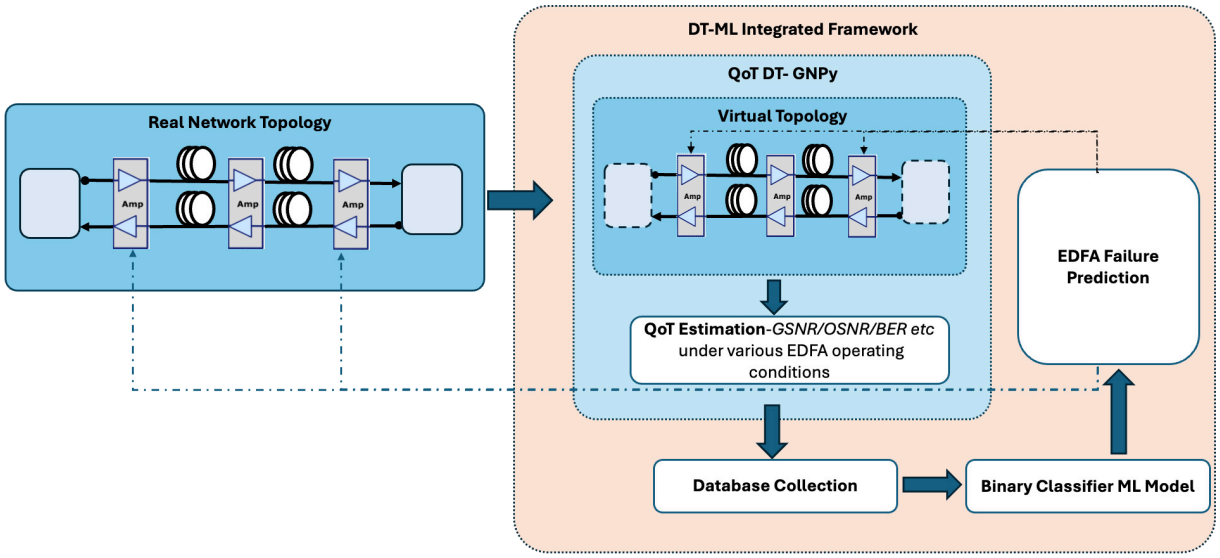

Fig. 1: DT Integrated Binary Classifier ML Model Framework for EDFA Failure Prediction - an Illustration.

This study presents an innovative methodology that leverages a virtual replica of the physical EDFA within a QoT Digital Twin (DT) environment, implemented using the open-source GNPy framework [8]. The proposed approach enables the systematic generation of a comprehensive monitoring dataset by simulating a wide range of EDFA operating conditions and parameter variations. Key transmission performance metrics—such as Generalized Signal-to-Noise Ratio (GSNR), OSNR, Bit Error Rate (BER), and channel power—are captured under both normal and degraded states to reflect real-world dynamics of EDFA behavior. This synthetic dataset forms the foundation for training a binary classification ML model aimed at the accurate prediction of EDFA failures. By integrating the digital twin-based simulation environment with the ML model, the framework facilitates proactive failure detection, reduces dependency on field data, and enhances the reliability and manageability of optical transport networks.

The framework outlining the general concept of database generation using the DT and the integration of the binary classifier ML model for the prediction of EDFA failure is illustrated in Fig. 1. In this study, we employ a Long Short-Term Memory (LSTM)-based ML model, chosen for its proven ability to learn patterns from sequential data[6], to capture the temporal evolution of EDFA failure scenarios. The model is trained on a comprehensive dataset generated through the simulation of various failure conditions using the GNPy, which emulates realistic operational and degradation behaviors of the EDFA. The following outlines the structure of the remaining sections of this paper. The implementation of DT within the GNPy framework and the process of database generation are presented in Section II. Section III details the ML model and its performance evaluation. Finally, Section IV provides concluding remarks and outlines potential directions for future work.

II. QoT Digital Twin and Database Generation

This section describes the implementation of DT of real network topology in GNPy and discusses EDFA failures and its simulation in GNPy for the generation of databases for training ML models.

A. Optical Multiplexed Section DT in GNPy

The implementation of a DT in the domain of optical networks is emerging as a transformative paradigm capable of reshaping how these complex infrastructures are designed, operated, and maintained. By offering a digital counterpart that mirrors the behavior and status of the physical optical infrastructure, DTs enable end-to-end lifecycle management. This includes functionalities such as network design and planning, real-time monitoring, performance forecasting, dynamic optimization, and even failure prevention and self-healing mechanisms. In recent years, there has been substantial progress in the development and application of DT technologies tailored specifically to optical communication systems. These efforts span from the initial stages of conceptual modeling, to rigorous simulation-based validation, and ultimately to real-world deployment and control.

A DT tailored for optical networks is typically composed of three essential and interdependent building blocks, as described in [9]:

1) **DT Model:** This represents the core simulation environment, accurately replicating the physical structure and behavior of the network. It includes all the critical components of the optical infrastructure—such as optical amplifiers, fiber spans, transponders, and Reconfigurable Optical Add-Drop Multiplexers (ROADMs)—along with their respective physical and operational parameters (e.g. gain, NF, fiber loss, dispersion, etc.). The fidelity of the digital model is key to ensuring that the DT can reliably mirror the real-world system.

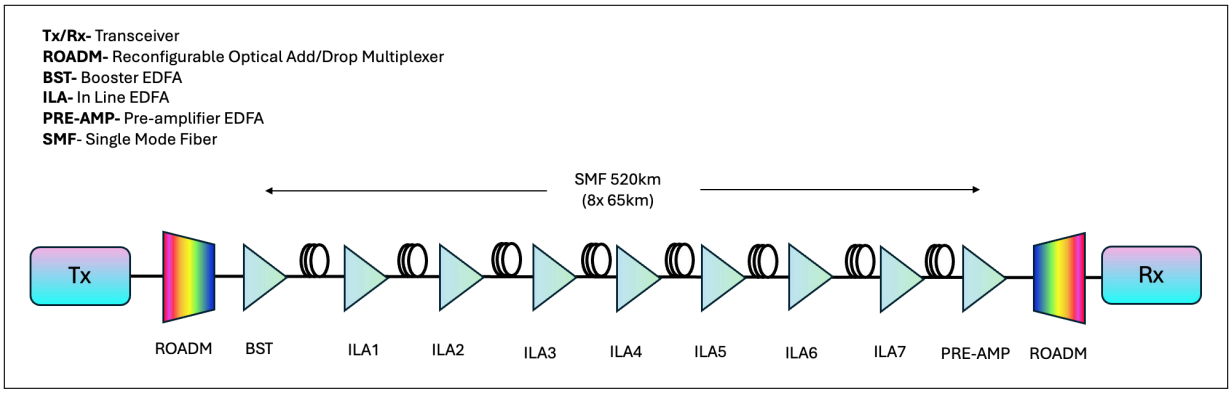

Fig. 2: Reference Network Topology Used in this Work for EDFA Failure Analysis, Comprising of 9 EDFAs (7 ILAs, 1 Booster and 1 Preamplifier).

2) **Real-Time Data Exchange:** A continuous and bidirectional flow of telemetry data is essential to synchronize the digital model with the physical network. This stream of live information, collected from optical monitors, sensors, and device controllers, allows the DT to dynamically update its internal state, accurately reflecting current network conditions. This connectivity allows the DT to evolve in real time as the network conditions change.
3) **Closed-Loop Control and Optimization:** Beyond passive monitoring, a DT can actively influence the physical network. Leveraging simulations and analytical models, the DT supports network operators in making informed decisions. It can propose or directly enact configuration changes, optimize operational parameters, detect and mitigate degradations, and ultimately enhance the QoT and overall system reliability.

Within this framework, the GNPy software library plays a fundamental role as the core modeling engine of the digital twin. GNPy is an open-source Python-based library that implements the Gaussian Noise (GN) model to evaluate the QoT of coherent optical transmission systems. It is designed to simulate optical signal propagation across complex, multi-span, and multi-vendor network topologies. By treating the transmission channel as an accumulation of Amplified Spontaneous Emission (ASE) noise and nonlinear interference modeled as Additive White Gaussian Noise (AWGN), GNPy provides an efficient and vendor-agnostic way to evaluate link and path performance in terms of Signal-to-Noise Ratio (SNR) or GSNR.

With its flexibility and openness, GNPy has become a widely adopted tool among researchers and operators in the field of open optical networking. It allows users to design and simulate transparent optical paths, validate optical reach, and estimate the maximum capacity and performance of both planned and already-deployed infrastructures [8], [10]. Its modular architecture supports the representation of disaggregated network components, making it especially suitable for modern open and software-defined environments.

In this work, we leverage GNPy to construct the digital model of an optical multiplexed section, focusing on a specific test scenario that mimics a realistic link deployment. As illustrated in Fig. 2, the network under study consists of a single-mode fiber link of 520 km total length, segmented into eight fiber spans of approximately 65 km each. The modeled digital twin includes a detailed representation of all active and passive optical elements, including seven In-Line Amplifiers (ILAs) deployed along the link, as well as booster and pre-amplifiers at both ends of the transmission direction. These amplifier models have been instantiated to mirror the behavior of commercially available devices, using accurate configuration parameters.

By embedding GNPy within the DT loop, it becomes possible to estimate GSNR values at the receiver, assess optical feasibility before service provisioning, and even simulate what-if scenarios to predict the impact of failures, degradations, or configuration changes. This framework demonstrates the potential of GNPy not only as a QoT estimation tool, but as a building block of a broader digital ecosystem for the autonomous control and optimization of optical networks.

B. Performance Degradation in EDFAs and Database Generation

EDFAs, while playing a critical role in optical communication systems by compensating for signal attenuation over long distances, are susceptible to various issues that can compromise network performance, directly impacts the QoT. One common problem in EDFA is amplifier gain reduction, which may occur over time due to aging of pump lasers, saturation of the gain media, pump efficiency degradation resulting in overall gain reduction. Another critical issue is increased noise levels, primarily from ASE [11], which can worsen due to pump laser instabilities, wavelength drift or aging of the pump laser. Additionally, anomalies in output power, such as unexpected spikes or drops, may result from variations in pump current or complete pump laser failure. These impairments not only degrade the signal quality but can also lead to significant disruptions in network operation if not detected and mitigated promptly [12].

Persistent faults in EDFAs can escalate into critical system-level failures, potentially resulting in significant service outages and degraded network performance. Given the central role that amplifiers play in Optical Line Systems (OLS)—primarily by sustaining signal strength and preserving the required performance metrics over long distances—it is imperative to monitor the performance degradation.

GNPy is employed to model the optical network and assess how physical layer effects—such as noise (ASE), non-linear signal distortion, and dispersion buildup—impact transmission quality. The simulation framework takes a detailed representation of the network as shown in Fig. 2, incorporating elements such as fiber span, node information, along with technical specifications including amplifier characteristics, noise figures, and transponder configurations. Based on these inputs, the tool computes quality indicators like GSNR, OSNR, etc. GNPy conducts optical signal propagation analysis by modeling how signals traverse and interact with various network components such as fiber spans, amplification stages, and ROADMs. It incorporates detailed physical layer attributes, including attenuation per span, length of fiber sections, gain variations across wavelengths, and nonlinear effects. Through successive application of the GN model along the transmission route, GNPy simulates the aggregate impact of these impairments on signal integrity.

The GNPy simulation framework was utilized to synthesize the training dataset by systematically varying key network parameters across numerous simulation instances. This methodology facilitated the modeling of a wide spectrum of network configurations and operational scenarios, including rare fault conditions that are typically difficult to observe in production networks due to their low occurrence rate and the inherent complexity of acquiring data from live networks.

By altering parameters such as amplifier gain, transponder configurations, and fault injection settings, GNPy facilitated the emulation of diverse network behaviors and corresponding signal quality indicators. To model gradual amplifier degradation, gain values were incrementally reduced in small steps, while the Variable Optical Attenuator (VOA) initially kept at maximum value was decremented concurrently to compensate for output power loss. This approach effectively represents silent failures—subtle degradations that cannot be detected through standard power monitoring. In practical scenarios, maintaining amplifier gain degradation within the compensatory range of the VOA is critical for avoiding service disruption, as the objective is to detect and predict soft failures. If the induced fault exceeds the VOA's corrective capability, it may escalate into a hard failure, resulting in complete loss of the optical signal.

The impact of failures on transmission parameters is evaluated by observing gradual changes. Step sizes for parameter degradation are randomly selected to ensure that the cumulative deterioration mimics realistic field-induced defects. This randomness in step size across different time intervals captures natural fluctuations observed in operational networks. The resulting time-series dataset, used for training the machine learning model, encapsulates variations in GSNR corresponding to the degradation patterns of individual amplifiers—each characterized by distinct gain and tilt profiles. This dataset reflects diverse degradation behaviors and supports the model in learning to predict performance degradation trends accurately.

The constructed dataset plays a key role in the development and evaluation of ML models for forecasting EDFA-related faults. By utilizing the adaptability of the GNPy simulation tool, a broad spectrum of practical and varied network conditions is incorporated. This diversity helps enhance the model's accuracy and ensures it performs well across multiple operating scenarios.

III. Binary Classifier ML Model and Performance Evaluation

An outline of the LSTM ML approach is provided in this section, followed by an assessment of its performance in predicting EDFA failures.

A. LSTM – The Binary Classifier

ML techniques have proven to be highly effective in handling time-series forecasting problems because of their ability to learn complex patterns from sequential data [13]. LSTM model, a specialized type of Recurrent Neural Network (RNN), which is designed to overcome the limitations of traditional RNNs by incorporating memory cells and gating mechanisms to retain relevant information over extended time intervals, is used for its capability to capture both short- and long-term temporal dependencies [6], [13]. The core architecture of LSTM includes forget, input, and output gates, which regulate the flow of information and enable the network to selectively remember or discard data in the time steps [14]. In this context, LSTM stands out for its capability to capture long-term dependencies while simultaneously accounting for short-term variations. This makes LSTM particularly effective for learning slow degradation trends in optical signal quality, where EDFA faults may develop gradually over time.

Here, the LSTM model is employed to analyze data sets generated from GNPy, focusing on detecting signal degradation due to failures in EDFAs. Specifically, the GSNR values obtained from GNPy under normal conditions, without faults in the amplifiers, serve as baseline. As slow degradation is intentionally introduced in the amplifier gain, the corresponding GSNR values reflect the impact of these changes. The model is trained using the difference between GSNR values under degradation and the baseline, alongside the duration of fault conditions. A gradual decline in GSNR values is observed as amplifier degradation increases, and these trends are utilized to enhance the model's predictive accuracy.

B. Performance Evaluation in Binary Classification

In binary classification tasks, a predictive model assigns each instance to one of two possible categories, typically referred to as "positive" (or true) and "negative" (or false). The effectiveness of such a classifier is commonly assessed through the outcomes it produces when compared to known

ground-truth labels. These outcomes can be categorized into four fundamental types[15]:

1) **True Positives (TP):** The count of instances correctly predicted as belonging to the positive class, i.e., cases where both the actual and predicted labels are positive.
2) **True Negatives (TN):** The number of instances accurately classified as negative, indicating agreement between the true label and the prediction in identifying a negative case.
3) **False Positives (FP):** The total number of instances that were incorrectly labeled as positive by the model, despite being negative in reality. This type of error is also known as a Type I error.
4) **False Negatives (FN):** Instances where the model incorrectly predicted the negative class for samples that actually belong to the positive class. This corresponds to a Type II error.

These four quantities form the basis for various evaluation metrics used to measure the classifier's performance. Accuracy represents the overall proportion of correct predictions, combining both true positives and true negatives, relative to the total number of predictions made. It serves as a basic indicator of a model's general correctness [6], [15]:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

Precision measures the reliability of positive predictions. Specifically, it is the fraction of instances predicted as positive that are actually positive. A higher precision indicates that the classifier makes fewer false positive errors:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{2}$$

Recall also known as sensitivity or the true positive rate, recall assesses the model's ability to identify all relevant instances of the positive class. It quantifies the proportion of actual positives that are correctly detected:

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3}$$

The F1-score provides a balanced measure that combines both precision and recall into a single value. It is particularly useful in scenarios with uneven class distributions or where both types of errors (false positives and false negatives) carry significant consequences. The F1-score is defined as the harmonic mean of precision and recall:

$$\text{F1-score} = 2 \times \frac{TP}{2TP + FP + FN} \tag{4}$$

This formulation ensures that the F1-score will only be high when both precision and recall are reasonably high, penalizing models that perform well on only one of the two dimensions.

The performance of the LSTM Binary classifier model is evaluated using two tools: the classification score (Accuracy, Precision, Recall and F1-score) as described above and the confusion matrix highlighting the actual and predicted EDFA failures.



Fig. 3: Heatmap Displaying Precision, Recall, and F1-Score across all Classes Predicted by the ML Model, Illustrating Class-wise Classification Performance.

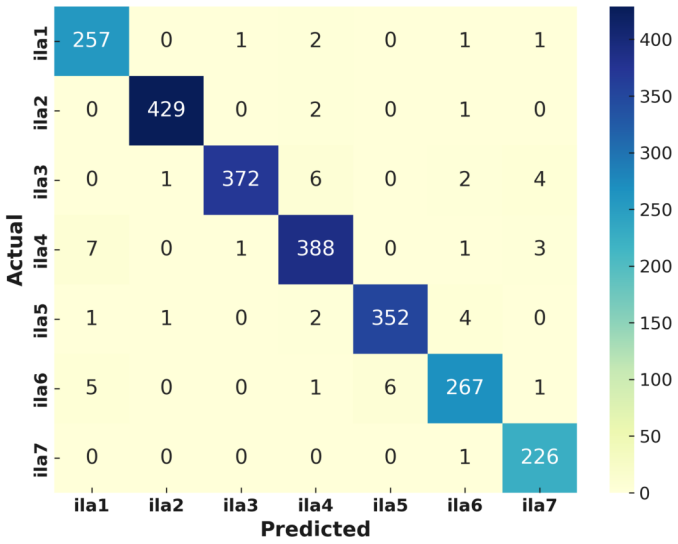

Fig. 4: Confusion Matrix Showing the Actual vs. Predicted Class Labels, Visualizing the Misclassifications.

Fig. 3 shows the three key metrics of the classification score as a heat map. Darker shades in the heat map correspond to higher values, indicating stronger performance. For instance, class ILA_2 achieves the highest Precision (1.00), Recall (0.99) and F1-score (0.99), reflecting excellent classification accuracy for this class. Class ILA_5 has all the metrics with the same score 0.98 showing consistent performance. Other classes, such as ILA_6, also show strong results, with Recall values (0.95) and Precision, F1-score of (0.96) each indicating that the model captures the majority of relevant instances and were correctly detected. Overall accuracy of the model is 98%. The heat map provides an insightful overview of the model's performance across these metrics.

Fig. 4 shows the confusion matrix which provides a detailed comparison between the actual and predicted classes, helping to identify the model's performance across these seven categories (ILA_1 to ILA_7). Each entry in the matrix shows the number of samples where the predicted class (columns) matches or differs from the actual class (rows). For example, the model correctly classified 429 samples for class ILA_2 and misclassified 2 samples as ILA_4 and 1 sample as ILA_6. This matrix gives a clear picture of the model's strengths in distinguishing between the classes, with the highest misclassifications limited to 13 samples for ILA_3 and ILA_6 while the model correctly classified 372 and 267 samples respectively for these classes.

IV. Conclusion and Future Directions

In conclusion, this paper introduced a methodology for predicting EDFA failures in optical networks through the integration of DT and ML algorithms. Leveraging the open-source GNPy framework, the proposed approach enables a vendor-neutral solution for the early detection of EDFA failures, thereby enhancing its adaptability across multi-vendor optical network environments.

The study concentrated specifically on the prediction of EDFA-related faults, aiming to optimize the practical application of ML techniques for this critical use case. By constructing DTs that capture network dynamics and training a LSTM model on the resulting time-series data, the method achieved a high prediction accuracy of 98%. This result highlights the model's effectiveness in identifying both short- and long-term behavioral trends associated with amplifier anomalies.

More importantly, the ability to forecast failures before they manifest enables a proactive network management strategy. Timely interventions based on model predictions can substantially reduce unplanned service disruptions, improving both reliability and operational efficiency within the optical layer.

Future extensions of this research could involve broadening the scope of failure scenarios modeled for EDFAs. Incorporating additional failures—such as device temperature variations, pump current anomalies, and pump laser failures—would enhance the comprehensiveness and real-world applicability of the predictive framework, leading to improved fault coverage. Furthermore, the underlying methodology can be adapted to develop specific models for failure prediction in other critical optical network elements, such as ROADMs or transceivers.

References

[1] R. G. Hunsperger, "Optical amplifiers," in *Integrated Optics: Theory and Technology*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2002, pp. 225–240, ISBN: 978-3-540-38843-2.

[2] K. Grobe, "Chapter 5 - optical wavelength-division multiplexing for data communication networks," in *Handbook of Fiber Optic Data Communication (Fourth Edition)*, C. DeCusatis, Ed., Fourth Edition, Oxford: Academic Press, 2013, pp. 85–122, ISBN: 978-0-12-401673-6.

[3] D. Wang, Y. Wang, X. Jiang, Y. Zhang, Y. Pang, and M. Zhang, "When large language models meet optical networks: Paving the way for automation," *Electronics*, vol. 13, no. 13, 2024, ISSN: 2079-9292.

[4] Silva *et al.*, "Learning long- and short-term temporal patterns for ml-driven fault management in optical communication networks," *IEEE TNSM*, vol. 19, no. 3, pp. 2195–2206, 2022.

[5] H. Lun, M. Fu, Y. Zhang, *et al.*, "A gan based soft failure detection and identification framework for long-haul coherent optical communication systems," *Journal of Lightwave Technology*, vol. 41, no. 8, pp. 2312–2322, 2023.

[6] L. E. Kruse, S. Kühl, A. Dochhan, and S. Pachnicke, "Experimental validation of machine learning-based joint failure management and quality of transmission estimation," *IEEE Photonics Journal*, vol. 15, no. 6, pp. 1–9, 2023.

[7] K. S. Mayer, R. P. Pinto, J. A. Soares, *et al.*, "Demonstration of ml-assisted soft-failure localization based on network digital twins," *J. Lightwave Technol.*, vol. 40, no. 14, pp. 4514–4520, Jul. 2022.

[8] V. Curri, "Gnpy model of the physical layer for open and disaggregated optical networking [invited]," *Journal of Optical Communications and Networking*, vol. 14, no. 6, pp. C92–C104, 2022.

[9] D. Wang, Y. Song, Y. Zhang, *et al.*, "Digital twin of optical networks: A review of recent advances and future trends," *Journal of Lightwave Technology*, vol. 42, no. 12, pp. 4233–4259, 2024.

[10] A. Ferrari, M. Filer, K. Balasubramanian, *et al.*, "Gnpy: An open source application for physical layer aware open optical networks," *Journal of Optical Communications and Networking*, vol. 12, no. 6, pp. C31–C40, 2020.

[11] A. W. Setiawan Putra, M. Yamada, H. Tsuda, and S. Ambran, "Theoretical analysis of noise in erbium doped fiber amplifier," *IEEE Journal of Quantum Electronics*, vol. 53, no. 4, pp. 1–8, 2017.

[12] S. Cruzes, "Failure management overview in optical networks," *IEEE Access*, vol. 12, pp. 169 170–169 193, 2024.

[13] Q. Wang, R.-Q. Peng, J.-Q. Wang, Z. Li, and H.-B. Qu, "Newlstm: An optimized long short-term memory language model for sequence prediction," *IEEE Access*, vol. 8, pp. 65 395–65 401, 2020.

[14] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, ISSN: 0899-7667. eprint: https://direct.mit.edu/neco/article-pdf/9/8/1735/813796/neco.1997.9.8.1735.pdf.

[15] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Information Processing & Management*, vol. 45, no. 4, pp. 427–437, 2009, ISSN: 0306-4573.