

# Failure Management in Optical Networks with ML: a Tutorial on Applications, Challenges and Pitfalls [Invited]

FRANCESCO MUSUMECI\*, MASSIMO TORNATORE

*Politecnico di Milano, Milano, Italy - \*corresponding author: francesco.musumeci@polimi.it*

*Compiled March 9, 2026*

---

This tutorial identifies and discusses the main design choices and challenges arising in the application of Machine Learning (ML) to optical network failure management (ONFM), including Quality of Transmission estimation, failure detection, prediction, root-cause identification, localization and magnitude estimation. We focus on input data preparation and on interpreting and validating model outputs, tackling data scarcity, data confidentiality, model explainability, uncertainty quantification and other critical factors, in order to highlight the potential risks for practitioners when adopting ML-based solutions for ONFM. An overview of publicly available datasets is also provided.

<http://dx.doi.org/10.1364/ao.XX.XXXXXX>

---

## 1. INTRODUCTION

Addressing failure management in optical networks is of utmost importance, as optical networks represent the backbone of modern high-speed communications and optical technologies are extremely pervasive in all segments of communication networks, from access/metro and up to core and data center network segments. Therefore, the ability to prevent or quickly react to failures is crucial to provide high network availability.

Traditional techniques for optical networks failure management (ONFM) have relied heavily on manual intervention and rule-based algorithms, often resulting in slow response times and suboptimal recovery actions in the face of network failures. To address these shortcomings, Machine Learning (ML) has emerged as a promising solution for automating and optimizing failure diagnosis and consequently improve management and recovery processes in optical networks. More generally, the application of ML to various optical networking use cases has been extensively studied in the recent literature, also investigating the role played by advanced AI tools, such as Large Language Models (LLM), generative AI, meta-learning, etc., in fully automating optical networks.

For the specific case of ONFM, different surveys and tutorials have already appeared in literature [1–6], demonstrating the successful application of ML to address various failure management use cases. These existing surveys have highlighted the power of data-driven algorithms to predict and analyze failures, adapt to dynamic network conditions, and offer more efficient and proactive strategies to maintain network performance. However, as with any emerging technology, the integration of ML into

optical network management comes with its own set of challenges, mainly including issues related to data quality, model interpretability and explainability, and scalability.

This paper reports and extends the tutorial in [7] and aims to provide a comprehensive overview of these challenges, focusing on the application of advanced ML frameworks to ONFM, namely, active, continual, federated and transfer learning, generative-AI-based data augmentation, explainable AI, uncertainty quantification.

We systematically discuss how input data should be selected, collected and shared, and how ML model outputs should be interpreted and validated. To this aim, after an overview of ML-driven failure management applications, we identify key areas where recent work has brought substantial improvements in 1) input data refinement and 2) model output interpretation, also emphasizing potential gaps in existing research.

The remainder of this paper is organized as follows. In Sec. 2 we briefly discuss the main applications of ML-based ONFM, reporting the most relevant literature for various cases. In Sec. 3 we identify the most relevant problem design choices that need to be considered when dealing with ML-based ONFM. Sec. 4 overviews the currently available datasets used for ONFM. The most relevant challenges, related to improving the quality of input data and properly interpreting and validating ML model outputs, are discussed in Sec. 5. We conclude the paper in Sec. 6, where we also identify other open research directions.

## 2. ONFM APPLICATIONS OVERVIEW

ONFM involves a wide set of use cases, targeting different objectives along the lifetime of an optical network, a lightpath or a set of optical devices. Table 1 reports a summary of the different ONFM use cases together with references to a selection of the most recent papers covering them. The different ONFM applications are also overviewed in the following subsections.

**Table 1.** Different ONFM use cases.

Use case	Addressed question	Refs.
QoT estimation	Can we prevent a failure?	[8–14]
Detection	Is there a failure now?	[14?–40]
Prediction	Will there be a failure?	[11, 41–44]
Identification	What is the failure root-cause?	[14, 17, 23, 26, 28, 29, 31, 34, 45–53]
Localization	Where is the failure?	[9, 14, 15, 19, 26, 30, 32, 37, 38, 54–68]
Magnitude estimation	What is the failure severity?	[1, 23, 55]

### A. QoT estimation

When accommodating a new lightpath request to transport a given amount of traffic between two points of the network, different choices should be made concerning the used wavelength and bandwidth, route, adopted modulation scheme, coding rate, etc. For each candidate option, assessing lightpath feasibility, i.e., performing **QoT estimation**, even before it is established is crucial to avoid service disruption and also use spectrum resources effectively.

Today's most common approach to estimate lightpath QoT, e.g., in terms of received Optical Signal-to-Noise Ratio (OSNR), Q-factor, SNR or Bit Error Rate (BER), is by leveraging analytical models, that take into account main physical-layer impairments, such as nonlinear interference, amplified spontaneous emission (ASE) noise, and optical filtering [69, 70]. However, real optical networks often encompass different sources of uncertainty, e.g., related to unknown connector losses, EDFA gain ripple, fiber type and precise fiber link length. Therefore, analytical models account for these uncertainties by introducing fixed QoT margins, which might induce waste of spectrum and transponders resources.

An alternative approach is to perform ML-based QoT estimation by leveraging historical data. As illustrated in Fig. 1, historical data that maps the characteristics (i.e., the input features, such as route, modulation, wavelength etc.) of previously-established lightpaths with the corresponding *measured* QoT (e.g., the SNR) can be leveraged by ML models to learn the effects of physical layer impairments, *including uncertainties*, on lightpaths QoT with no need for accurately estimating the individual sources of uncertainty. Then, trained ML models can provide a more accurate QoT estimation for future candidate lightpaths given the knowledge of their characteristics.

An intermediate approach, that is gaining momentum in these last years, is what we refer to as *input refinement*. In this

### Phase 1: Model development

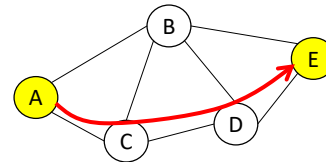
Historical data from previous/existing lightpaths

Route	Wavelength	Modulation	SNR
A-C-B	1550 nm	BPSK	18dB
C-B-E	1553 nm	8-QAM	23dB
A-C-D-E	1556 nm	QPSK	16B

Lightpath features

QoT metric

### Phase 2: Model deployment



New lightpath between nodes A and E  
Candidate solution: 1553 nm, A-C-D-E, QPSK  
What is the *estimated* SNR?

**Fig. 1.** ML-based QoT estimation.

case, analytical models are preserved, but estimation of the uncertain model parameters is refined using ML [71].

Independently of the chosen approach, QoT estimation allows operators to make more informed (hence, less conservative) decisions on the QoT margins to be used to protect the system against outage due to QoT fluctuations, resulting in decreased resource waste.

### B. Failure detection and prediction

During lightpath lifetime, even though the QoT estimated before its deployment is adequate to correctly receive the transported information, several factors can contribute to deteriorate the original quality, such as the presence of newly-deployed lightpaths, equipment aging, deterioration or malfunctioning, fiber bending, etc., possibly causing service disruption. In this context, ML can be used to extract useful information from monitored parameters, as shown in the example of Fig. 2. In the example, we assume that BER (as an example of a QoT metric) is continuously monitored at the lightpath receiver and that the lightpath is experiencing a deterioration due to which, once BER increases above a certain intolerable threshold, a failure occurs. Exploiting ML, the idea is to identify a *signature* in the BER *behavior* that goes beyond the simple comparison between raw BER values and the specific intolerable threshold, e.g., by analyzing how BER statistics (e.g., mean, standard deviation, peak, etc.) vary over time. Doing so we can predict the failure occurrence in advance (**failure prediction**) or in its immediate proximity (**failure detection** or **early-detection**), and allow enough time to prevent or limit service disruption, e.g., via lightpath reconfiguration, traffic rerouting, modulation format downgrade etc.

### C. Failure identification, localization and magnitude estimation

Once a failure has been detected over a lightpath, further information might be useful to limit service disruption and reduce the cost of reparation, especially concerning the failure root-cause, its location and its severity. Addressing these issues is commonly referred to as failure identification, failure localization and failure magnitude estimation, respectively.

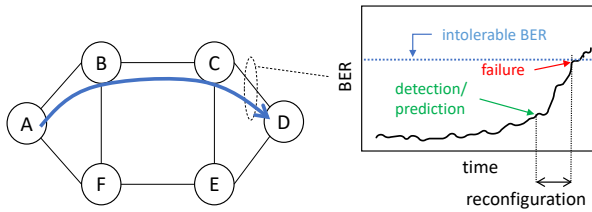


Fig. 2. ML-based failure detection and prediction.

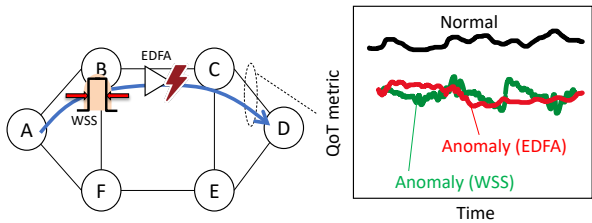


Fig. 3. ML-based failure identification.

Concerning **failure identification**, note that different root-causes can be source of the failure, e.g., a filter shift or tightening in a Wavelength Selective Switch (WSS) within a ROADM, a malfunctioning EDFA that provides an unexpectedly low gain, a laser drift at the transmitter, a bended fiber, a deteriorated connector, etc. Discriminating between the different root-causes allows to focus the intervention on the subset of specific devices that can be the source of the failure, thus reducing the time required for troubleshooting. Each of the aforementioned root-causes typically has a different impact on the affected lightpath, which can be discriminated by, e.g., observing some QoT metric at the receiver, as illustrated in the example of Fig. 3 for the cases of anomalous QoT caused by a malfunctioning WSS or EDFA, respectively. It is important to note that, in general, any anomaly may provide significantly different QoT compared to a normal situation where a lightpath is not affected by a failure. However, simply looking at the raw QoT values may not be enough to discriminate between the failure causes, as shown in the figure, where the QoT curves related to the two different sources of failures overlap. With ML, instead, it is possible to extract useful information by observing the signature of the failure cause on the monitored QoT metric.

Similar considerations can be drawn for the cases of **failure localization** and **magnitude estimation**, with the idea that a failure originated by a certain root-cause has different effects on the monitored QoT at the receiver (or even on other metrics collected through the entire lightpath), according to the specific location (e.g., whether hitting a WSS at one ROADM or another) and/or different magnitude levels (e.g., a WSS filter tightening of 10 GHz or 20 GHz).

Therefore, knowing the exact location of the failure allows to directly operate on the affected network component and, consequently, reduces the time required for failure recovery. Additionally, estimating the failure magnitude is important in case different recovery actions are possible. For example, a weaker failure can be solved by simply restarting/reconfiguring a device, while a more severe failure may necessitate in-field intervention, e.g., to physically fix or even replace a malfunctioning device. Evidently, the ability to discriminate between such different situations allows to limit unnecessary operational costs and reduce (or even avoid) service disruption.

It is important to note that most of the existing studies focus on one of the aforementioned objectives at a time. However, the rigid distinction between different failure scenarios, as considered during ML model development, is not always an adequate representation of all the possible conditions in the real network. As an example, when developing a ML model for failure localization, available data typically includes samples of a same failure type "X" and, in some cases, also with a same magnitude, occurring on different locations (e.g., a 3-dB gain reduction in EDFAs located in different links of the network). Therefore, rigorously, the developed ML model has been trained to distinguish between the occurrence of *that specific failure "X"* (with that specific magnitude) on the different links, and there is often no guarantee that the same model will provide similar localization performance in case of different failure types and/or different failure magnitudes.

Therefore, considering the combined effect of different failure causes, different failure locations and different failure magnitudes is not trivial and, to the best of our present knowledge, it is largely unaddressed in current literature.

### 3. DESIGN CHOICES

To address ONFM, especially when considering ML-assisted approaches where network data is essential, several design choices and intertwined decisions must be taken into account already before data collection starts, and of course also during ML model development, as detailed below.

#### What type of data should we collect?

Modern networks support extensive data collection and telemetry targeting different objectives beyond the application to failure management. Extremely diverse types of information are extracted from the network at physical, network and even application levels, including transmission quality parameters, bit/packet losses, data rate, service latency etc. Considering this plethora of information, it is not trivial to determine *a priori* which data will reveal more useful for ONFM and for the particular objective at hand. This decision has several implications on 1) the way data shall be transferred from the different monitors towards a location where data is processed, 2) the necessary memory and computing resources to store and process collected data, 3) the ML model selection, i.e., whether to analyze time-series data or tabular data has an impact on the type of algorithm to use, 4) model complexity and training duration.

#### Where to set monitor location?

This aspect is evidently intertwined with the above choice of data collection. After selecting the type(s) of data to use, the amount of monitors and their locations impact the trade off arising between data transfer/storage requirements and the performance of the ML models as, on one hand, collecting and processing more data implies higher communication and computing costs, while, on the other hand, larger datasets generally favor ML models performance.

#### Which network domains and technologies are of interest?

The penetration of photonics in modern Internet is undoubted and covers all networks segments, from core to metro and access segments and up to intra-datacenter networks, and as well it involves multiple technologies and operational standards, such as Wavelength Division Multiplexing (WDM), Flexi-grid and Elastic Optical Networks (EONs), Optical Transport Network (OTN), Passive Optical Networks (PONs), Fiber-to-the-Home and similar technologies (FTTx), Free Space Optics (FSO) etc. Each of these domains and technologies is subject to failures and,

of course, operators have the interest in failure management across all of them. However, in case a choice is necessary, an operator may privilege domains/technologies where a highest probability of failure is expected, or where the highest amount of traffic would be affected by a failure, or also where the cost of ineffective failure recovery is higher (i.e., time-consuming or causing unnecessary resource waste). Consequently, focusing on some domains/technologies rather than others has an impact on data collection, data storage and processing, and how to leverage ML model outputs to repair the failure.

#### What types of failures do we target?

Since multiple types of failures can affect a network, one may want to concentrate on specific ones, e.g., the most frequent, the hardest to handle, the ones potentially affecting the highest amount of traffic, etc. Another major distinction between failures concerns the difference between hard failures, consisting of sudden and typically unpredictable events (e.g., fiber cuts), and soft failures, involving gradual degradation of signal quality. Moreover, the possibility of having concurrent failures should be taken into account, which may lead to different conclusions compared to the cases where failures are handled one at a time (e.g., different impacts on the monitored parameters, different classes of failures to be distinguished with the ML model, etc.). Selecting which types of failures is of interest, among all the types above, mainly impacts data collection, especially considering the inherent failure data scarcity from real operational networks, and ML model training, as a model optimized using some type of failures data for training can provide poor performance if, at the deployment phase, the type of failures to handle is different. In this case, one can aim at training a generalized ML model including multiple failure types, at the cost of possibly lowering the global model performance and with the intrinsic limitation of collecting more diverse data before training.

#### Who is in charge of ONFM decision-making?

Regardless the physical locations where data is collected and how ML models combine the different sources of information, a main issue arises when deciding who is in charge of failure monitoring and handling.

For example, an operator may have the visibility of data collected in the whole network, and therefore develop ML models fed with all the data generated in its network. Conversely, it is also possible that the operator may require specialized assistance from the vendors to interpret the meaning and behavior of data retrieved by some specific devices. On the other hand, vendors may leverage in-house domain knowledge on their equipment, but often cannot freely access data of all other equipment in the same network, especially in disaggregated network scenarios where equipment of multiple vendors co-exist.

Therefore, a collaborative environment, where operators and vendors cooperate to monitor the network and address failures is often preferable to centralized scenarios where a single player takes decisions in an isolated manner. For the same reasons, fully automating failure management is not always possible, but in some cases a hybrid human-machine interaction might be necessary, also considering the possible uncertain outputs retrieved by a "black box" ML model.

All these aspect impact in particular the way network data is collected and shared among the different players.

#### Is the data used for ONFM confidential?

In most cases, transfer of sensitive ONFM data is necessary to develop ML models, due to the fact that ML models are typically trained in centralized servers that are not co-located with the monitoring locations of necessary data and, more important, the

data collector and data user are not always the same entity. This aspect has an impact on the type of data that can be shared with third-parties, possibly via a public network, and hence, on the information that can be leveraged at training phase. Data encryption can be leveraged to exploit valuable information from network data while still preserving its privacy, which implies additional storage and computing requirements that need to be taken into account when developing the ML models.

#### What is the proper time horizon for failure management?

The time dimension plays a fundamental role in all phases of ONFM, from data collection to model training during model development phase, and up until inference frequency and data/model updates at deployment phase.

More specifically, how often should data be collected is not trivial, especially considering that each type of data can provide valuable information at different time scales. For example, routing information (and similarly, all other lightpath "design" characteristics, such as modulation, central wavelength, etc.) can be collected less frequently as this information is typically stable over time, whereas signal quality indicators (e.g., OSNR, BER, input/output power in EDFAs), instead, vary much more frequently and so provide useful information if continuously collected. Moreover, it is not trivial to decide when performing ML model training, e.g., when a sufficient amount of data has been collected. Similarly, in case updating the ML model is necessary, e.g., due to a change in input data distributions and/or in the inputs/outputs relationship over time (known as *data drift*), deciding the moment when performing such update impacts the trade-off between model performance and storage/computing requirements. Finally, the appropriate frequency to perform inference is also crucial, especially in ONFM where detecting failure presence and determining failure root-cause, location, etc. is fundamental to reduce or avoid service downtime. For higher frequency, more aggressive ONFM can be pursued with the objective of reducing service downtime as much as possible; however, at the same time, a higher computational effort as well as more extensive data collection is necessary to perform inference.

The key questions discussed in this Section are summarized in Tab. 2 together with some possible options for each of them.

## 4. AVAILABLE DATASETS

Although a very wide set of papers have addressed ONFM in these years, to the best of our knowledge, only few datasets are publicly available at the time of writing. Indeed, most of the datasets are obtained either from private testbeds or from operational networks and are not published for confidentiality reasons. Other data sources are developed by generating synthetic datasets with commercial simulators or analytical models. For example, MATLAB code for datasets generation suitable for QoT estimation is available in [72], which allows to monitor various optical parameters including the lasers launch power, total input power of each fiber, attenuation values at Variable Optical Attenuators (VOAs), EDFA gains, etc.

Among the publicly available datasets, the "Optical Failure Dataset" in [73] includes data collected from a laboratory testbed in Scuola Superiore Sant'Anna (Pisa, Italy), where different types of failures are emulated over a lightpath by altering WSS characteristics. Moreover, among the collected metrics for the various failure scenarios, there are OSNR and BER traces at lightpath receiver, as well as optical input/output power at intermediate optical amplifiers.

**Table 2.** Different design choices in ONFM.

Key parameters	Options
Collected data	Power, BER, OSNR, spectrum, alarms, OTDR, latency, physical vs. network/app-level info
Monitor location	Receivers/transmitters, ROADMs, EDFAs, links, spans, all devices
Network domain	Single lightpath vs. network, single- vs. multi-domain, core/metro/access/datacenter
Network technology	WDM, OTN, EON, PON, FTTx, FSO
Failure type	Hard vs soft failures, single vs. multiple failures
Decision maker	Operator, vendor(s), centralized vs. distributed, collaborative, fully-automated vs. human assisted
Data ownership	Shared vs. private data
Time horizon	Continuous vs. periodic data collection and inference

Moreover, the dataset described in [74] includes a variety of transmission scenarios in a WDM system developed in a laboratory at Fraunhofer Institute for Telecommunications (Berlin, Germany), and suitable for various use case, including QoT estimation. In this dataset, OSNR, BER, SNR, EVM, Q-factor are collected as QoT metrics.

Furthermore, the dataset in [75] consists of optical time domain reflectometry (OTDR) traces and includes various fiber failures, i.e., fiber cut, eavesdropping, dirty connector and bad splice, making the dataset suitable for failure detection, identification, localization and other ONFM applications. Similar OTDR datasets are also available in [76], which includes failures due to reflective events, and [77], where reflective, and non-reflective events, as well as a mixture of them (mainly, fiber cuts and fiber bends) are included. Finally, the dataset in [78] includes reflective failures with different magnitudes.

## 5. MAIN CHALLENGES IN INPUT DATA AND OUTPUT DECISIONS

Even though nowadays it is possible to develop extremely efficient ML models, as demonstrated by the wide research in this area, and considering the large amounts of tools, libraries, and documentation available for fast ML software development, still, the most critical challenges, and possible sources of pitfalls in ML-assisted ONFM, reside in how to obtain *valuable input data* and how the *outputs predictions* of ML models are interpreted, as graphically shown in Fig. 4. For all these challenges, there exist specific ML frameworks that can be adopted and that, in some cases, have been already considered in recent literature, as discussed below.

### A. How to collect good training data?

As discussed in Sec. 3, different aspects need to be considered already during data collection, including deciding the type of collected data, data sampling frequency, etc. Additionally, even

after addressing these issues, still there is no *a priori* guarantee that the ML model developed using such data will be the best possible model to deploy in the field, regardless any possible optimization that can be done on the ML algorithms and on the selection of their hyperparameters. This is mainly due to the quality of input data, which plays a crucial role in the ML model performance. In particular, even assuming that the collected data and considered features are the most informative out of all the possible combinations, input data should still incorporate sufficient diversity to be fully representative of the real-life data distributions that the ML model will encounter at inference phase. Consider the example of Fig. 5, where we show data points expressed through a generic 2-dimensional features space distinguishing classes A and B (red circles and green crosses, respectively). Assume that, in an initial phase, no data has been collected in the gray-shaded area of the features space. In such condition, a ML model, even properly trained, may lead to a wrong decision boundary (see dashed line in Fig. 5), just because the initial training set is missing important data for training.

By means of *active learning*, useful data (e.g., light-red circles and light-green crosses in Fig. 5) can be collected, labeled and added to the training dataset so that the model can be retrained in order to improve the decision boundary of the ML classifier.

Sometimes, even visually inspecting how the available data is distributed over the selected feature space might be sufficient to identify regions where collecting new data might provide additional information. Note that the reasons for missing labeled data can be different, e.g., it might occur that few data points are available for certain features space areas or even that such data points have not been labeled. In the first case, using *probes* (i.e., forcing the system to generate specific data in a specific area) is necessary. In any case, additional effort might be required to annotate the newly generated data with a proper label, which is typically costly and time-consuming as it is performed by human experts. Considering this, and assuming that a (wide) set of unlabeled data is available, still, which data should be annotated first must be decided. A structured and widely adopted approach to do so consists of training a preliminary model with the available labeled data and then perform inference on the entire set of unlabeled data, assigning to each prediction also a *score* (e.g., based on entropy, uncertainty or other metrics) to measure how much the model is *confident* on this prediction. An example of the application of this approach for failure management, although in the context of microwave networks, is available in [79], but no similar approaches exist in optical networks, to the best of our knowledge.

An early work addressing QoT estimation [13] has leveraged active learning to optimize and decrease the amount of probe lightpaths needed to build a training dataset capable of representing the variety of lightpaths configurations in an optical networks. Since then, most of the literature has relied on other strategies to increase and obtain good quality data, such as analytical models and developing network Digital Twins (see Sec. 5.B for more details).

### B. How to deal with data scarcity?

The availability of training data is a key factor in developing ML models for ONFM, as it is difficult to collect failure data from real optical networks/systems due to the fact that failures rarely occur, compared to normal operating conditions. To address this issue, which is tightly intertwined with the availability of *good* training data as discussed in Sec. 5.A), three main approaches have been already adopted in literature, i.e., 1) perform *data*

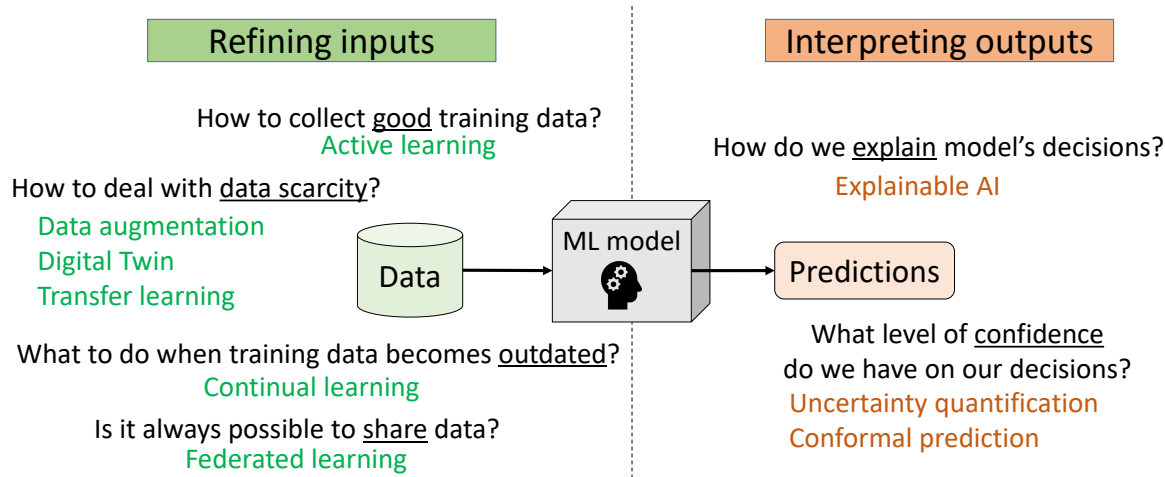


Fig. 4. Main challenges and pitfalls in ML-assisted ONFM.

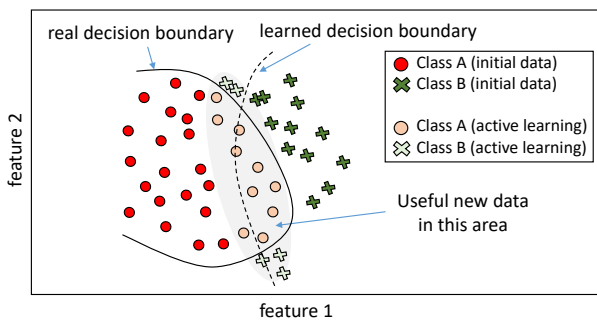


Fig. 5. Input refinement through active learning.

augmentation via synthetic data generation, 2) develop *Digital Twins (DTs)* of a real network, consisting of a digital simulated counterpart of a real network or even a physical replica (e.g., in a laboratory environment) of a part of the network, where failures are simulated/enforced with the objective of generating data [80, 81], and 3) using *Transfer Learning (TL)* to leverage the knowledge on a source domain (e.g., a different network), where larger amount of data is available, and transfer it to the domain (e.g., the network) of interest, known as target domain.

Synthetic data generation has been used in [29, 46, 47], where the authors leverage Variational Autoencoder (VAE) [82] and generative adversarial network (GAN) [83], [84] to address data scarcity for failure detection and identification and showing that a very limited amount of real data is needed to reach significant classification performance when using such data augmentation approaches, and that even unknown failure causes can be identified. A variant of GAN, i.e., Conditional-Tabular GAN (CT-GAN), has been used in [48] together with another simpler approach, known as Synthetic minority oversampling technique (SMOTE), to improve the classification performance of failure identification under extreme data imbalance, which is a very critical issue in real networks as some failure causes may occur rarely.

The concept of DT has been used in [23] to generate data in a simulated environment, by modeling the propagation of optical signals from the transmitter to the receiver in the optical time domain, and address failure detection, identification and magnitude estimation. Moreover, authors in [66] use an exper-

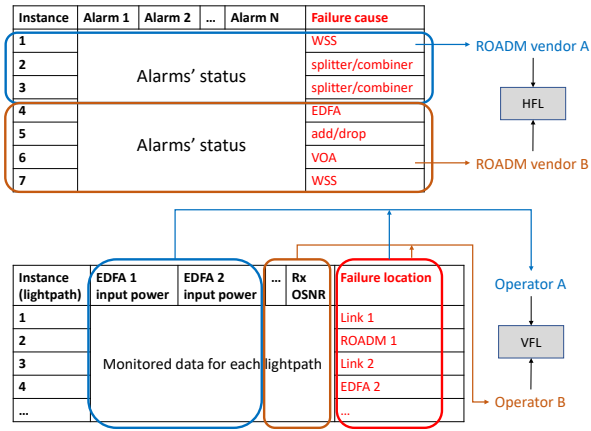
imental testbed as a DT to augment a base of field data with additional single-failure data and focusing on failure localization. Furthermore, in [56, 59] a DT is used to generate multiple mirror models of a real network where different types of failures are simulated to enlarge the knowledge base for future ML developments targeting failure localization.

Both synthetic data generation and use of DTs aim at addressing data scarcity with additional data. Still in this area, semi-supervised learning can be adopted to increase dataset size by training a ML model in a supervised manner in order to automatically annotate unlabeled data. This approach has been used in [18], where the authors perform failure detection in OTN equipment.

TL, instead, has been used in [24], to address data scarcity when developing a system capable of identifying vibrations in optical networks, which are the most relevant precursor of fiber cuts. Moreover, in [30, 31] TL has been leveraged to address data scarcity in the context of failure detection, identification and localization, focusing on transferring the knowledge acquired on one lightpath onto another lightpath [31] and also across different ONFM applications [30]. A particular flavor of TL, known as domain adaptation, has been used in [12] and compared with active learning to address data scarcity for QoT estimation.

### C. What to do when training data becomes outdated?

During a network lifetime, the distribution of input data and the relationship between input features and their corresponding outputs (i.e., the labels) may change over time, e.g., due to equipment aging or changes in network status. This implies that new data should be collected to have an up-to-date representation of network conditions and, consequently, update ML models considering new data distribution. This approach, known as *continual learning*, has been used in some recent papers addressing QoT estimation, such as [85], where the authors leverage continuous model updates in response to shifts of data distribution over time. Nonetheless, we believe that further research is needed in this context, not only to cover ONFM applications beyond QoT estimation, but especially to address key aspects such as defining the appropriate criteria for model retraining (i.e., how frequently), the amount of new data needed to perform model update, how to avoid catastrophic forgetting of older data distributions upon model retraining with new data.



**Fig. 6.** Collaboration between vendors or operators in ONFM with HFL and VFL.

#### D. Is it always possible to share data?

Failure data is typically sensitive information that operators and vendors are not always willing to share through a public network. Therefore, instead of transferring data to centralized locations where ML models are trained, adopting *Federated Learning (FL)*, possibly combined with encryption strategies such as homomorphic encryption [86], enables privacy preservation of failure data to reach ONFM objectives. This paradigm also allows multiple parties, e.g., different vendors, operators, providers, etc., to cooperate and build ML models that leverage multiple types of data sources, possibly collected by different equipment types and/or different failures scenarios, thus representing complementary pieces of information.

Two different flavors of FL are possible, namely, Horizontal and Vertical FL (HFL and VFL, respectively), illustrated in Fig. 6 through an example. In the HFL case, the different parties own information on different data instances through the same set of features. The example of the application of HFL in ONFM (see the upper part of Fig. 6) is the case of different vendors of a same type of equipment (e.g., a ROADM) that collect failure data, leveraging the same (or a similar) set of alarms status, to perform failure identification. In this case, assuming that data for different failure causes are present in local datasets at each vendor, each vendor has the interest in cooperating to develop a model capable of discriminating between multiple types of failure causes leveraging the knowledge of the other vendor. Conversely, in the VFL case, the different parties own only a subset of features pertaining to the same data instances. As an example (see the lower part of Fig. 6), information pertaining a same lightpath that traverses different monitor locations (e.g., OSNR at the receiver, input/output power of EDFAs, etc.) in a multi-operator scenario, can be collected by the different operators, each controlling a subset of the monitors. In this case, the different operators can cooperate to share information and perform failure management over the same lightpath, e.g., to determine failure location.

Note that, in both HFL and VFL cases, the different parties are assumed to own not only a subset of data instances (HFL) or a subset of the features (VFL), but also the outputs (i.e., the labels) corresponding to the various data instances.

Cooperative learning based on HFL has been used in [35] by collecting failure data from multiple network domains with the aim of improving the generalization capability of failure

detection while preserving data confidentiality of the different domains. Moreover, authors of [54] leverage VFL with homomorphic encryption in order to preserve data privacy of different operators managing different equipments traversed while performing failure localization.

It is worth noting that FL is not the only privacy-preserving method to address failure management. For example, the authors in [20, 27, 33] perform failure detection by adopting data scrambling, also combined, in some selected scenarios, with homomorphic encryption and distributed application of Principal Component Analysis (PCA), that allow training on subsets of scrambled data without losing the original information contained in the ordered (i.e., unscrambled) data.

#### E. What level of confidence do we have on our decisions?

Accurate outputs provided by a ML model can be extremely critical for subsequent decisions in any ONFM application. As an example, assume that a ML model identifies a specific failure cause as the most likely to have happened after a failure occurs on a lightpath, and that failure severity has been determined (possibly, by a different ML model) to be such that an in-field intervention is needed to substitute a ROADM card. If any of these two decisions is wrong, the misclassification will lead to an unnecessary replacement of a device, with consequent waste of resources.

For this reason, it is extremely important to allow ML models to provide *informed* decisions, e.g., by associating to each prediction a numerical score that quantifies the confidence of the prediction itself. A widely adopted instrument to evaluate this confidence is by *uncertainty quantification* of a predicted class for a given data instance [87], which is based on the concept of probabilistic classification [88]. The information brought by prediction uncertainty can be exploited by decision maker to, e.g., decide whether to trust the output provided by the ML model automatically, or if it is more convenient to explore more accurately the data instance, e.g., by involving a human domain expert and/or collect more information to fine-tune the decision. Another approach to provide confident and informed decisions is *conformal prediction*, based on ML models providing classification (similarly, regression) outputs in the form of *prediction sets* (respectively, prediction intervals) which globally provide a predefined level of formal guarantees of correctness [89].

An example of uncertainty quantification in ML-based ONFM is available in [43], where the authors leverage quantile regression and Bayesian uncertainty for QoT forecasting models relying on multi-layer perceptrons and Long-Short-Term-Memory-based Artificial Neural Networks (LSTM ANNs). Moreover, probabilistic classification is used also in [35] to evaluate model uncertainty of failure detection.

#### F. How do we explain model decisions?

Besides providing a quantitative metric to assess prediction uncertainty, it is often desirable to obtain *qualitative* explanations of model behavior, e.g., what features had a major impact in the classification of a specific data instance, or what are the features that globally provide more information for a model. *Explainable Artificial Intelligence (XAI)* includes a set of mathematical tools that provide quantifiable explanations of models' behavior. While there already exist simple algorithms, such as linear or rule-based (e.g., decision trees) models, that are inherently interpretable, i.e., they can provide an interpretation of model decisions which can be understood by humans (e.g., the set of thresholds on the considered features that are evaluated in a

tree-based model), the power of XAI is that it is mostly model agnostic, and can be adopted regardless the ML model used to perform training and inference, hence including more complex non-linear models such as ANNs.

Several XAI frameworks exist, including the well-known Local Interpretable Model-Agnostic Explanations (LIME) [90], however, the most frequently adopted in the context of ONFM is SHAP. SHAP approach takes its origin from game theory and the concept of *Shapley value*, that aims at fair payouts assignment to players based on their contribution to the total payout in a game. In the context of XAI, the game corresponds to the process of model inference by the ML model, features play the role of players of the game, and the payout of the game is represented by the output of the ML model, e.g., a certain class for an ONFM classification application.

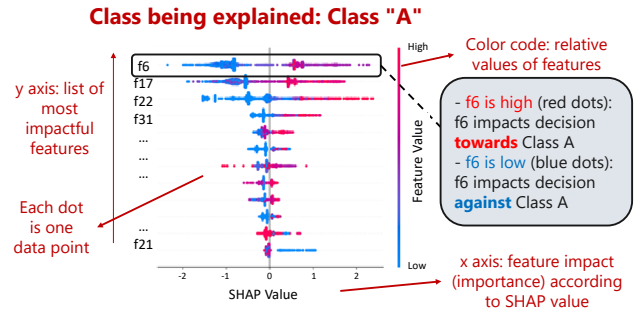
The Shapley (or SHAP) value is computed as the *marginal contribution* brought by a given feature towards a certain class, averaged across all possible subsets of features in the problem. More formally, SHAP measures the individual contribution of each feature for each data point by evaluating the difference in model output with and without using the feature, as detailed in Eq. 1.

$$\phi_i(f, x) = \sum_{z' \subseteq x} \frac{|z'|!(M - |z'| - 1)!}{M!} [f(z') - f(z'/i)] \quad (1)$$

Here the SHAP value  $\phi_i(f, x)$  is calculated for feature  $i$  and for a given data point  $x$  and a ML model  $f$ . The marginal contribution  $\phi_i(f, x)$  is calculated over all possible coalitions  $z'$ , i.e., the subsets of the overall features set  $x$ . Term  $\frac{|z'|!(M - |z'| - 1)!}{M!}$ , represents a weight given to each coalition, where  $|z'|$  is the cardinality of the set of features in  $z'$  and  $M$  is the total number of features. This weight is high for small and large coalitions while it is low for medium-sized coalitions. For each feature  $i$  under evaluation and for each subset of feature  $z'$ ,  $[f_x(z') - f_x(z'/i)]$  is the contribution (or importance) of feature  $i$  to the coalition, evaluated as the difference between  $f_x(z')$ , the output of the model when feature  $i$  is considered, and  $f_x(z'/i)$ , the output of the model without feature  $i$ . Note that same values of a feature in different data points can contribute differently to the output, depending on values of other features for each of the data points, which makes the importance of a single feature valuable, in particular, concerning the coalition with other features.

One of the most widely adopted ways to leverage the informativeness brought by the SHAP value, is to visualize *summary plots*, as the one in Fig. 7, that shows an example of summary plot to explain the behavior of a ML classifier with respect to a generic class "A". Details of the elements of the graph are shown in the figure. This type of graph is useful to understand features importance (the most important are shown in the top part of the figure as they provide the highest absolute range of SHAP values across all data points), and to explain, for each feature and for each data point, how the relative value of that feature has contributed to a decision towards (high positive SHAP) or against (high negative SHAP) the class being explained.

SHAP has been used in [42] to explain the predictions of a ML model estimating laser lifetime (corresponding to the mean time to failure, hence equivalent to failure prediction), by quantifying the relevance of the input features on the individual predictions. Moreover, in ref. [?] the authors include domain experts knowledge as an additional feature used in a classical supervised learning model. Here, the new features encompassing domain



**Fig. 7.** Use of summary plot with SHAP values to explain the behavior of a model with respect to a generic class "A".

expertise are automatically obtained by a decision tree trained on historical data labeled by human experts. To confirm the intuition and the validity of the decision-tree-based expertise-enhancement, the authors use SHAP values to estimate feature importance, showing that the expertise-related features always reveal to be the most important across several different ML models. Additionally, SHAP-based XAI has been adopted in [65] to evaluate the impact of different features collected at different monitoring points within a lightpath in localizing failures, in [91] to evaluate overall feature importance on failure detection, and in [10] to explain misclassifications in QoT estimation.

## 6. CONCLUSION AND FUTURE DIRECTIONS

Although a relevant amount of work has been already produced to address failure management in optical networks by leveraging ML, significant attention is yet to be put in obtaining high-quality data to train ML models and to evaluate and explain model decisions. In this paper, we discussed in detail specific research challenges in these two domains, also identifying available ML frameworks that are useful to address them, including active learning, data augmentation, digital twinning, transfer learning, continual learning, federated learning, uncertainty quantification and XAI. For each of these techniques, we overviewed the main benefits and, where available, some of the papers that have already leveraged them to address the aforementioned challenges. Moreover, we discussed the characteristics of publicly available datasets suitable for various ONFM applications.

Besides these, there are also other promising investigation areas that it is worth mentioning and that are still unexplored. Among them, we believe the most important are in 1) the use of LLMs, transformers and generative-AI to support ONFM in both data preparation and output interpretation/explanation, that are already being investigated in very recent literature [92, 93], 2) the standardization processes required to define common workflows between data collection, data transfer, data processing and model development within large-scale networks, 3) the development of in-network inference, enabled by data plane programmability, to allow implementation of ML models directly in network equipment and at line rate, and 4) the definition of practical deployments of storage and computing resources dedicated to ML-assisted ONFM, considering the presence of multiple geographically-distributed data sources and the possible need for distributed training and inference.

## ACKNOWLEDGMENTS

This work was supported by the Italian Ministry of University and Research under the Italian PRIN Project ZetON and PRIN PNRR project GASTON.

## REFERENCES

1. F. Musumeci, C. Rottondi, G. Corani, S. Shahkarami, F. Cugini, and M. Tornatore, "A Tutorial on Machine Learning for Failure Management in Optical Networks," *J. Light. Technol.* **37**, 4125–4139 (2019).
2. R. Gu, Z. Yang, and Y. Ji, "Machine learning for intelligent optical networks: A comprehensive survey," *Journal of Network and Computer Applications*. **157**, 102576 (2020).
3. G. Villa, C. Tipantúa, D. S. Guamán, G. V. Arévalo, and B. Arguero, "Machine Learning Techniques in Optical Networks: A Systematic Mapping Study," *IEEE Access* **11**, 98714–98750 (2023).
4. S. Cruzes, "Failure Management Overview in Optical Networks," *IEEE Access* **12**, 169170–169193 (2024).
5. D. Wang, C. Zhang, W. Chen, H. Yang, M. Zhang, and A. P. T. Lau, "A review of machine learning-based failure management in optical networks," *Sci. China Inf. Sci.* **65**, 1–19 (2022).
6. M. Nouioua, P. Fournier-Viger, G. He, F. Nouioua, and Z. Min, *A Survey of Machine Learning for Network Fault Management* (Springer International Publishing, Cham, 2021), pp. 1–27.
7. F. Musumeci, "Machine-Learning-Assisted Optical Network Failure Management: Challenges and Pitfalls," in *50th European Conference on Optical Communication (ECOC 2024)*, (2024).
8. X. Yang, T. Eldahrawy, C. Sun, A. Lorences-Riesgo, M. Tornatore, G. Charlet, and Y. Pointurier, "Integrating PPE and Inputs Refinement for Enhanced QoT Estimation and Optimization in an Optical Mesh Network," in *50th European Conference on Optical Communications (ECOC 2024)*, (2024), pp. 491–494.
9. L. E. Kruse and S. Pachnicke, "Joint QoT Estimation and Soft-Failure Localization using Variational Autoencoder," in *2023 International Conference on Optical Network Design and Modeling (ONDM)*, (2023), pp. 1–3.
10. O. Ayoub, S. Troia, D. Andreoletti, A. Bianco, M. Tornatore, S. Giordano, and C. Rottondi, "Towards explainable artificial intelligence in optical networks: the use case of lightpath QoT estimation," *J. Opt. Commun. Netw.* **15**, A26–A38 (2023).
11. S. Allogba, S. Aladin, and C. Tremblay, "Machine-Learning-Based Lightpath QoT Estimation and Forecasting," *J. Light. Technol.* **40**, 3115–3127 (2022).
12. D. Azzimonti, C. Rottondi, A. Giusti, M. Tornatore, and A. Bianco, "Comparison of domain adaptation and active learning techniques for quality of transmission estimation with small-sized training datasets [Invited]," *J. Opt. Commun. Netw.* **13**, A56–A66 (2021).
13. D. Azzimonti, C. Rottondi, and M. Tornatore, "Reducing probes for quality of transmission estimation in optical networks with active learning," *J. Opt. Commun. Netw.* **12**, A38–A48 (2020).
14. 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 J.* **15**, 1–9 (2023).
15. K. Abdelli, M. Lonardi, J. Gripp, S. Olsson, F. Boitier, and P. Layec, "Unsupervised Anomaly Detection and Localization with Generative Adversarial Networks," in *50th European Conference on Optical Communications (ECOC 2024)*, (2024), pp. 998–1001.
16. A. Raj, Z. Wang, F. Slyné, T. Chen, D. Kilper, and M. Ruffini, "Interference Identification in Multi-User Optical Spectrum as a Service using Convolutional Neural Networks," in *50th European Conference on Optical Communications (ECOC 2024)*, (2024), pp. 314–317.
17. Q. Lin, X. Chen, Z. Ouyang, H. Gao, X. Chen, and Z. Li, "Scaling Optical Network Fault Management with Decentralized Graph Learning," in *2024 Optical Fiber Communications Conference and Exhibition (OFC)*, (2024), pp. 1–3.
18. Z. Sun, C. Zhang, M. Zhang, B. Ye, and D. Wang, "Semi-supervised learning model synergistically utilizing labeled and unlabeled data for failure detection in optical networks," *J. Opt. Commun. Netw.* **16**, 541–552 (2024).
19. K. Abdelli, M. Lonardi, J. Gripp, D. Correa, S. Olsson, F. Boitier, and P. Layec, "Anomaly Detection and Localization in Optical Networks Using Vision Transformer and SOP Monitoring," in *2024 Optical Fiber Communications Conference and Exhibition (OFC)*, (2024), pp. 1–3.
20. R. F. Sales, A. Ribeiro, M. F. Silva, F. R. Lobato, A. Sgambelluri, L. Valcarenghi, and J. C. W. Costa, "Disaggregated Confidentiality-Preserving Scheme for Fault Detection in Optical Networks," in *2024 Optical Fiber Communications Conference and Exhibition (OFC)*, (2024), pp. 1–3.
21. S. Aladin, L. Wosinska, and C. Tremblay, "Detecting Anomalies in the Optical Layer Using Unsupervised Machine Learning," in *2024 Optical Fiber Communications Conference and Exhibition (OFC)*, (2024), pp. 1–3.
22. A. N. Ribeiro, R. F. Sales, F. R. Lobato, J. Costa, M. F. Silva, A. Sgambelluri, L. Valcarenghi, and L. Wosinska, "PCA-Assisted Fuzzy Clustering Approach for Soft-Failure Detection in Optical Networks," in *2024 International Conference on Optical Network Design and Modeling (ONDM)*, (2024), pp. 1–5.
23. M. Devigili, M. Ruiz, N. Costa, C. Castro, A. Napoli, J. Pedro, and L. Velasco, "Applications of the OCATA time domain digital twin: from QoT estimation to failure management," *J. Opt. Commun. Netw.* **16**, 221–232 (2024).
24. K. Abdelli, M. Lonardi, J. Gripp, S. Olsson, F. Boitier, and P. Layec, "Risky event classification leveraging transfer learning for very limited datasets in optical networks," *J. Opt. Commun. Netw.* **16**, C51–C68 (2024).
25. W. Yang, C. Zhang, D. Wang, H. Zhu, X. Xu, D. Shi, and M. Zhang, "Data Labeling Using Unsupervised Cascaded Pre-training with Fused Multi-port Data for Optical Failure Management," in *2024 Optical Fiber Communications Conference and Exhibition (OFC)*, (2024), pp. 1–3.
26. L. E. Kruse, S. Kühl, A. Dochhan, and S. Pachnicke, "Experimental demonstration of soft-failure management using variational autoencoder and GAN on optical spectrum," in *49th European Conference on Optical Communications (ECOC 2023)*, (2023), pp. 420–423.
27. M. F. Silva, A. Sgambelluri, A. Pacini, F. Paolucci, A. Green, D. Mascarenas, and L. Valcarenghi, "Confidential Detection of Multiple Failures in Optical Networks: an Experimental Evaluation," in *2023 Optical Fiber Communications Conference and Exhibition (OFC)*, (2023), pp. 1–3.
28. H. Lun, M. Fu, Y. Zhang, H. Jiang, L. Yi, W. Hu, and Q. Zhuge, "A GAN Based Soft Failure Detection and Identification Framework for Long-Haul Coherent Optical Communication Systems," *J. Light. Technol.* **41**, 2312–2322 (2023).
29. L. Z. Khan, J. Pedro, N. Costa, L. De Marinis, A. Napoli, and N. Sambo, "Data augmentation to improve performance of neural networks for failure management in optical networks," *J. Opt. Commun. Netw.* **15**, 57–67 (2023).
30. F. Musumeci, G. G. Marchionni, and M. Tornatore, "Cross-Task and Cross-Lightpath Failure Detection and Localization in Optical Networks Using Transfer Learning," in *ICC 2023 - IEEE International Conference on Communications*, (2023), pp. 435–440.
31. F. Musumeci, V. G. Venkata, Y. Hirota, Y. Awaji, S. Xu, M. Shiraiwa, B. Mukherjee, and M. Tornatore, "Domain adaptation and transfer learning for failure detection and failure-cause identification in optical networks across different lightpaths [Invited]," *J. Opt. Commun. Netw.* **14**, A91–A100 (2022).
32. S. Behera, T. Panayiotou, and G. Ellinas, "Machine learning framework for timely soft-failure detection and localization in elastic optical networks," *J. Opt. Commun. Netw.* **15**, E74–E85 (2023).
33. M. F. Silva, A. Sgambelluri, A. Pacini, F. Paolucci, A. Green, D. Mascarenas, and L. Valcarenghi, "Confidentiality-preserving machine learning algorithms for soft-failure detection in optical communication networks," *J. Opt. Commun. Netw.* **15**, C212–C222 (2023).
34. K. Sun, Z. Yu, L. Shu, Z. Wan, H. Huang, Y. Lei, and K. Xu, "Digital Residual Spectrum-Based Generalized Soft Failure Detection and Identification in Optical Networks," *IEEE Transactions on Commun.* **71**, 324–338 (2023).

35. X. Chen, C.-Y. Liu, R. Proietti, Z. Li, and S. J. B. Yoo, "Automating Optical Network Fault Management with Machine Learning," *IEEE Commun. Mag.* **60**, 88–94 (2022).
36. K. Sun, Z. Yu, H. Huang, J. Zhang, and K. Xu, "Autonomous and Generalized Soft Failure Detection based on Digital Residual Spectrum in Optical Networks," in *2022 Optical Fiber Communications Conference and Exhibition (OFC)*, (2022), pp. 1–3.
37. X. Chen, C.-Y. Liu, R. Proietti, J. Yin, Z. Li, and S. J. B. Yoo, "On Co-operative Fault Management in Multi-Domain Optical Networks Using Hybrid Learning," *IEEE J. Sel. Top. Quantum Electron.* **28**, 1–9 (2022).
38. M. F. Silva, A. Pacini, A. Sgambelluri, and L. Valcarengi, "Learning Long- and Short-Term Temporal Patterns for ML-Driven Fault Management in Optical Communication Networks," *IEEE Transactions on Netw. Serv. Manag.* **19**, 2195–2206 (2022).
39. M. Ibrahim, H. Abdollahi, C. Rottondi, A. Giusti, A. Ferrari, V. Curri, and M. Tornatore, "Machine learning regression for QoT estimation of unestablished lightpaths," *J. Opt. Commun. Netw.* **13**, B92–B101 (2021).
40. S. Liu, D. Wang, C. Zhang, L. Wang, and M. Zhang, "Semi-supervised Anomaly Detection with Imbalanced Data for Failure Detection in Optical Networks," in *2021 Optical Fiber Communications Conference and Exhibition (OFC)*, (2021), pp. 1–3.
41. C. Xing, C. Zhang, Y. Wang, Z. Duan, W. Song, M. Zhang, and D. Wang, "Spatio-Temporal Failure Prediction Using LSTGM for Optical Networks," in *2024 Optical Fiber Communications Conference and Exhibition (OFC)*, (2024), pp. 1–3.
42. K. Abdelli, H. Griebner, and S. Pachnicke, "An Interpretable Machine Learning Approach for Laser Lifetime Prediction," *J. Light. Technol.* **42**, 2094–2102 (2024).
43. S. Yousefi, H. Chouman, P. Djukic, F. Golaghadzadeh, C. Tremblay, and C. Desrosiers, "Forecasting Lightpath Quality of Transmission and Implementing Uncertainty in the Forecast Models," *J. Light. Technol.* **41**, 4871–4881 (2023).
44. S. Behera, T. Panayiotou, and G. Ellinas, "Modeling Soft-Failure Evolution for Triggering Timely Repair with Low QoT Margins," in *GLOBECOM 2022 - 2022 IEEE Global Communications Conference*, (2022), pp. 2140–2145.
45. W. Yang, C. Zhang, X. Jiang, Y. Fan, Z. Bi, J. Xiong, X. Xiao, M. Zhang, and D. Wang, "Spatio-Temporal Graph Attention Networks for Alarm Root Cause Recognition in Optical Transport Network," in *50th European Conference on Optical Communications (ECOC 2024)*, (2024), pp. 1002–1005.
46. L. E. Kruse, S. Kühl, A. Dochhan, and S. Pachnicke, "Experimental investigation of machine-learning-based soft-failure management using the optical spectrum," *J. Opt. Commun. Netw.* **16**, 94–103 (2024).
47. L. E. Kruse, S. Kühl, A. Dochhan, and S. Pachnicke, "Monitoring Data Augmentation of Spectral Information using VAE and GAN for Soft-Failure Identification," in *2024 Optical Fiber Communications Conference and Exhibition (OFC)*, (2024), pp. 1–3.
48. L. Z. Khan, J. Pedro, N. Costa, A. Sgambelluri, A. Napoli, and N. Sambo, "Model and data-centric machine learning algorithms to address data scarcity for failure identification," *J. Opt. Commun. Netw.* **16**, 369–381 (2024).
49. L. Z. Khan, J. Pedro, O. Ayoub, N. Costa, A. Sgambelluri, L. De Marinis, A. Napoli, and N. Sambo, "Optimizing Deep Learning-Based Failure Management in Optical Networks by Monitoring Relative Neural Activity," in *2024 International Conference on Optical Network Design and Modeling (ONDM)*, (2024), pp. 1–3.
50. L. Z. Khan, J. Pedro, N. Costa, A. Napoli, and N. Sambo, "Model-centric versus data-centric machine learning for soft-failure cause identification in optical networks," in *49th European Conference on Optical Communications (ECOC 2023)*, (2023), pp. 1027–1030.
51. C. Tremblay, A. Mahmoudialami, P. A. Ngani Sigué, Y. Pei, D. Côté, D. Boertjes, D. Demeter, and C. Desrosiers, "Detection and root cause analysis of performance degradation in optical networks using machine learning," in *49th European Conference on Optical Communications (ECOC 2023)*, (2023), pp. 1278–1281.
52. L. Z. Khan, P. J. Freire, J. Pedro, N. Costa, A. Napoli, and N. Sambo, "Data Augmentation to Reduce Computational Complexity of Neural-Network-Based Soft-Failure Cause Identifier," in *2023 Optical Fiber Communications Conference and Exhibition (OFC)*, (2023), pp. 1–3.
53. C. Zhang, D. Wang, J. Jia, L. Wang, K. Chen, L. Guan, Z. Liu, Z. Zhang, X. Chen, and M. Zhang, "Potential failure cause identification for optical networks using deep learning with an attention mechanism," *J. Opt. Commun. Netw.* **14**, A122–A133 (2022).
54. M. Ibrahim, F. Temiz, F. Musumeci, and M. Tornatore, "Vertical Federated Learning for Failure Localization in Partially Disaggregated Optical Networks," in *2024 IEEE 25th International Conference on High Performance Switching and Routing (HPSR)*, (2024), pp. 1–6.
55. M. Cai, X. Liu, Y. Zhang, L. Yi, W. Hu, and Q. Zhuge, "A Physics-Based Learning Approach for ROADM-Induced Anomaly Localization and Estimation," *J. Light. Technol.* **42**, 4433–4443 (2024).
56. R. Wang, J. Zhang, Z. Gu, M. Ibrahim, B. Zhang, F. Musumeci, M. Tornatore, and Y. Ji, "Digital-twin-assisted meta learning for soft-failure localization in ROADM-based optical networks," *J. Opt. Commun. Netw.* **16**, C11–C19 (2024).
57. Y. Jiao, P.-H. Ho, X. Lu, J. Tapolcai, and L. Peng, "A Novel Framework for Optical Layer Device Board Failure Localization in Optical Transport Network," *IEEE Transactions on Netw. Serv. Manag.* **21**, 5374–5383 (2024).
58. Y. Jiao, P.-H. Ho, X. Lu, J. Tapolcai, and L. Peng, "A Novel Framework of Failure Localization in Optical Transport Network," *IEEE Commun. Mag.* **61**, 142–147 (2023).
59. R. Wang, J. Zhang, F. Musumeci, B. Zhang, Z. Gu, M. Tornatore, and Y. Ji, "Meta-learning-based failure localization with digital-twin-enabled multi-mirror models in optical networks," in *49th European Conference on Optical Communications (ECOC 2023)*, (2023), pp. 898–901.
60. C. Delezoide, P. Ramantanis, and P. Layec, "Streamlined Failure Localization Method and Application to Network Health Monitoring," *J. Light. Technol.* **41**, 6119–6125 (2023).
61. M. Sena, P. Hazarika, C. Santos, B. Correia, R. Emmerich, B. Shariati, A. Napoli, V. Curri, W. Forysiak, C. Schubert, J. K. Fischer, and R. Freund, "Advanced DSP-Based Monitoring for Spatially Resolved and Wavelength-Dependent Amplifier Gain Estimation and Fault Location in C+L-Band Systems," *J. Light. Technol.* **41**, 989–998 (2023).
62. K. Abdelli, C. Trotschug, H. Griebner, and S. Pachnicke, "Faulty branch identification in passive optical networks using machine learning," *J. Opt. Commun. Netw.* **15**, 187–196 (2023).
63. R. Wang, J. Zhang, S. Yan, C. Zeng, H. Yu, Z. Gu, B. Zhang, T. Taleb, and Y. Ji, "Suspect fault screen assisted graph aggregation network for intra-/inter-node failure localization in ROADM-based optical networks," *J. Opt. Commun. Netw.* **15**, C88–C99 (2023).
64. C. Zeng, J. Zhang, R. Wang, B. Zhang, and Y. Ji, "Multiple attention mechanisms-driven component fault location in optical networks with network-wide monitoring data," *J. Opt. Commun. Netw.* **15**, C9–C19 (2023).
65. O. Karandin, O. Ayoub, F. Musumeci, Y. Hirota, Y. Awaji, and M. Tornatore, "If Not Here, There. Explaining Machine Learning Models for Fault Localization in Optical Networks," in *2022 International Conference on Optical Network Design and Modeling (ONDM)*, (2022), pp. 1–3.
66. K. S. Mayer, R. P. Pinto, J. A. Soares, D. S. Arantes, C. E. Rothenberg, V. Cavalcante, L. L. Santos, F. D. Moraes, and D. A. A. Mello, "Demonstration of ML-Assisted Soft-Failure Localization Based on Network Digital Twins," *J. Light. Technol.* **40**, 4514–4520 (2022).
67. J. Babbar, A. Triki, R. Ayassi, and M. Laye, "Machine learning models for alarm classification and failure localization in optical transport networks," *J. Opt. Commun. Netw.* **14**, 621–628 (2022).
68. H. Yang, X. Zhao, Q. Yao, A. Yu, J. Zhang, and Y. Ji, "Accurate Fault Location using Deep Neural Evolution Network in Cloud Data Center Interconnection," *IEEE Transactions on Cloud Comput.* **10**, 1402–1412 (2022).
69. P. Poggiolini, G. Bosco, A. Carena, V. Curri, Y. Jiang, and F. Forghieri, "The GN-Model of Fiber Non-Linear Propagation and its Applications," *J. Light. Technol.* **32**, 694–721 (2014).
70. I. F. de Jauregui Ruiz, A. Ghazisaeidi, T. Zami, S. Louis, and B. Lavigne, "An accurate model for system performance analysis of optical fibre

- networks with in-line filtering," in *45th European Conference on Optical Communication (ECOC 2019)*, (2019), pp. 1–4.
71. X. Yang, C. Sun, G. Charlet, M. Tornatore, and Y. Pointurier, "Digital-twin-based active input refinement for insertion loss estimation and QoT optimization in C and C + L networks," *J. Opt. Commun. Netw.* **16**, 1261–1274 (2024).
  72. "Automated training dataset collection," <https://github.com/BrandonLJN/automated-training-dataset-collection> (2021).
  73. "Optical Failure Dataset," <https://github.com/Network-And-Services/optical-failure-dataset> (2022).
  74. C. Santos, A. Moawad, B. Shariati, R. Emmerich, P. Safari, C. Schubert, and J. K. Fischer, "Experimental dataset for developing and testing ML models in optical communication systems," *J. Opt. Commun. Netw.* **16**, G1–G10 (2024).
  75. K. Abdelli, F. Azendorf, C. Tropschug, H. Griesser, S. Pachnicke, and J. Choo, "Dataset for Optical fiber faults," <https://dx.doi.org/10.21227/pdnp-1b78> (2022).
  76. k. Abdelli, H. Griesser, P. Ehrle, C. Tropschug, and S. Pachnicke, "An OTDR dataset for optical fiber monitoring," <https://dx.doi.org/10.21227/y0ek-1s16> (2022).
  77. K. Abdelli, C. Tropschug, H. Griesser, and S. Pachnicke, "An OTDR event dataset," <https://dx.doi.org/10.21227/pw43-9p78> (2022).
  78. "Optical Fiber Link Test Bench," <https://dx.doi.org/10.21227/H2N66K> (2018).
  79. N. Di Cicco, M. Ibrahimi, F. Musumeci, F. Bruschetta, M. Milano, C. Passera, and M. Tornatore, "Machine Learning for Failure Management in Microwave Networks: A Data-Centric Approach," *IEEE Transactions on Netw. Serv. Manag.* **21**, 5420–5431 (2024).
  80. Q. Zhuge, X. Liu, Y. Zhang, M. Cai, Y. Liu, Q. Qiu, X. Zhong, J. Wu, R. Gao, L. Yi, and W. Hu, "Building a digital twin for intelligent optical networks [Invited Tutorial]," *J. Opt. Commun. Netw.* **15**, C242–C262 (2023).
  81. M. R. Sena, R. Emmerich, B. Shariati, R.-F. Braun, M. Geitz, J. K. Fischer, and R. Freund, "Link Tomography: A Tool for Monitoring Optical Network and Designing Digital Twins," in *50th European Conference on Optical Communications (ECOC 2024)*, (2024), pp. 176–179.
  82. Y. Qu, H. Ma, Y. Jiang, L. Wang, and J. Yu, "A Network Data Reinforcement Method Based on the Multiclass Variational Autoencoder," *Secur. Commun. Networks* **2022**, 1–10 (2022).
  83. H. Navidan, P. F. Moshiri, M. Nabati, R. Shahbazian, S. A. Ghorashi, V. Shah-Mansouri, and D. Windridge, "Generative Adversarial Networks (GANs) in networking: A comprehensive survey & evaluation," *Comput. Networks* **194**, 108149 (2021).
  84. I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Nets," in *Advances in Neural Information Processing Systems*, vol. 27 Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K. Weinberger, eds. (Curran Associates, Inc., 2014).
  85. Q. Wang, Z. Cai, and F. N. Khan, "Lifelong QoT prediction: an adaptation to real-world optical networks," *J. Opt. Commun. Netw.* **16**, 1159–1169 (2024).
  86. L. Yang, D. Chai, J. Zhang, Y. Jin, L. Wang, H. Liu, H. Tian, Q. Xu, and K. Chen, "A Survey on Vertical Federated Learning: From a Layered Perspective," (2023).
  87. A. Malinin, L. Prokhorenkova, and A. Ustimenko, "Uncertainty in Gradient Boosting via Ensembles," in *International Conference on Learning Representations*, (2021).
  88. V. Vovk, I. Petej, and V. Fedorova, "Large-scale probabilistic predictors with and without guarantees of validity," in *Advances in Neural Information Processing Systems*, vol. 28 (2015).
  89. V. Vovk, A. Gammerman, and G. Shafer, *Algorithmic Learning in a Random World* (Springer-Verlag, Berlin, Heidelberg, 2005).
  90. M. T. Ribeiro, S. Singh, and C. Guestrin, ""Why should I trust you?" Explaining the predictions of any classifier," in *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, (2016), pp. 1135–1144.
  91. C. Zhang, D. Wang, L. Wang, L. Guan, H. Yang, Z. Zhang, X. Chen, and M. Zhang, "Cause-aware failure detection using an interpretable XGBoost for optical networks," *Opt. Express* **29**, 31974–31992 (2021).
  92. 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* **13** (2024).
  93. C. Sun, X. Yang, N. Di Cicco, R. Ayassi, V. V. Garbhapu, P. A. Stavrou, M. Tornatore, G. Charlet, and Y. Pointurier, "First Experimental Demonstration of Full Lifecycle Automation of Optical Network Through Fine-Tuned LLM And Digital Twin," in *50th European Conference on Optical Communications (ECOC 2024)*, (2024), pp. 1002–1005.