Experimental Validation of Machine Learning-Based Joint Failure Management and Quality of Transmission Estimation

Lars E. Kruse, Sebastian Kühl, Annika Dochhan, *Member, IEEE*, and Stephan Pachnicke, *Senior Member, IEEE*

***Abstract*—The exponentially growing demand for high-speed data necessitates more complex and versatile networks. Optimization and reliability assurance of such high-complexity networks is getting increasingly important. In this article, we experimentally validate our a machine learning-based framework that combines quality of transmission (QoT) estimation with soft-failure detection, identification, and localization based on the same latent space of a variational autoencoder running on optical spectra obtained by optical spectrum analyzers at high priority nodes in the network. We further investigate the advantages of a variational autoencoder-based soft-failure detection mechanism over a QoT metric-based approach. We use data acquired from optical transmission experiments involving different modulation formats and channel configurations. The results demonstrate that the proposed framework achieves reliable QoT estimation in real world scenarios. Additionally, it effectively detects soft-failures, identifies specific failure types and accurately localizes the occurrence of failures.**

***Index Terms*—Quality of transmission estimation, optical performance monitoring, soft-failures, variational autoencoder, recurrent neural networks.**

I. Introduction

THE issue of growing bandwidth demand has brought significant changes to optical networks. New flexible add-drop multiplexers play a crucial role in enabling more versatile network operations and the implementation of flexible frequency grids. As a result, networks have become more sophisticated, configurable, and adaptable. With the increasing complexity of networks, monitoring and optimizing performance have become increasingly important. In current optical networks, ensuring service level agreements and promised capacity is achieved by incorporating large operating margins that include unallocated system and design margins [1]. However, this conservative approach has the drawback of wasting capacity due to the presence of large margins. In this context, accurate quality of transmission (QoT) estimation becomes vital as it allows for maximizing capacity and potentially enables full self-management of networks in the future by ensuring low-margin optical networking [2]. Accurately estimating QoT in multi-vendor optical networks poses a challenge because precise equipment parameters are often treated as confidential or not precisely known. Consequently, such a multi-vendor network can be referred to as an exact component parameter agnostic network scenario. We have shown in our previous work, that using the optical spectrum obtained by optical spectrum analyzers (OSAs) at high priority nodes in the network as an input for a machine learning framework increases the accuracy of a QoT estimator in such a networking scenario [3]. We assumed a node to be of high priority in a network, if a high amount of transmission paths cross it and it has a high rank.

Data losses and fines due to unmet service-level agreements (SLAs) are the consequences of disruptions in optical connections. Traditionally, optical networks ensure failure security by implementing conservative design solutions with guaranteed redundancies [4]. However, in today's optical networks, network assurance needs to be enhanced through automated and dynamic techniques, especially in such multi-vendor networks. Instead of relying on threshold-based failure detection or probabilistic approaches, the utilization of machine-learning algorithms shows great promise in enabling future networks to have self-management capabilities [5].

Different QoT estimation approaches have been proposed, aiming at evaluating nonlinear impairments in links. Accuracy and speed are always a trade-off in these techniques. The most accurate method is a full-fiber propagation simulation using the split-step Fourier method (SSFM) [6], but it is computationally complex and unsuitable for real-time implementation. Analytical tools like the Gaussian noise (GN) model [7] offer acceptable accuracy with low computation time, being less accurate than SSFM. GN model extensions like the incoherent GN model and close-form methods approximate the nonlinear interference to reduce the computation time. However, both simulative and analytical approaches require precise knowledge of link parameters [8], [9], leading to less accurate QoT estimation [10] with parameter deviations. Furthermore, both models cannot adapt to the presence of soft-failures in the network. Machine learning (ML)-based estimation combines high accuracy and fast

Manuscript received 7 November 2023; accepted 13 November 2023. Date of publication 16 November 2023; date of current version 6 December 2023. This work was supported in part by the CELTIC-NEXT Project AI-NET-PROTECT under Project ID C2019/3-4, and in part by the German Federal Ministry of Education and Research under Grant 16KIS1284. *(Corresponding author: Lars E. Kruse.)*

The authors are with the Chair of Communications, Kiel University, 24143 Kiel, Germany (e-mail: lars.kruse@tf.uni-kiel.de; sebastian.kuehl@tf.uni-kiel.de; annika.dochhan@tf.uni-kiel.de; stephan.pachnicke@tf.uni-kiel.de).

Digital Object Identifier 10.1109/JPHOT.2023.3333420

TABLE I
SELECTED ML-BASED SOFT-FAILURE MANAGEMENT AND QOT ESTIMATION LITERATURE OVERVIEW

Literature	Sim.	Exp.	OPM Data	(ML-)Algorithm	SFD Acc.	SFI Acc.	SFL Acc.	QoT MAE
Vela et al. [11]	√		Rx power, BER	Analytical model	99.06%	SFI: 99.55%		
Furdek et al. [12]	√		BER, block errors, etc.	DBSCAN, SVM	96.2%			
Shariati et al. [13]	√		Optical spectrum (1 Ch.)	SVM	up to 100%	SFI: up to 100%		
Lun et al. [14]		√	PSD	CNN	up to 100%	up to 100%		
Sun et al. [15]	√		Residual spectrum	AE+SVM		97.61%		
Zhang et al. [16]		√	OTN data	BiLSTM		98.73%		
Mayer et al. [17]	√		Tx Power, OSNR	ANN			up to 100%	
Wang et al. [18]		√	ROADM data	SFS-GRN			up to 100%	
Sun et al. [19]		√	Residual spectrum	AE+SVM		96.20%		
Aladin et al. [20]	√		Transmission parameters	GRU, LSTM				0.285 dB
D'Amico et al. [9]		√	Transmission parameters	GNPy				0.6±0.9 dB
Kruse et al. [21]	√		Optical spectrum	LSTM				0.18 dB
Kruse et al. [3]	√	√	Optical spectrum	VAE+LSTM				0.2 dB
Kruse et al. [22]	√		Optical spectrum	VAE+SVM	98.9%	99.6%	90.08%	0.17 dB
Here		√	Optical spectrum	VAE+SVM	99.96%	99.05%	99.13%	0.29 dB

computation by training an estimator on input features correlated to the target metric. Training can take minutes to days, but once trained, estimation is quick, suitable for real-time environments.

Various approaches, including analytical [7], machine learning-based [20], [23], and hybrid methods [24], [25], have recently been explored to evaluate lightpath performance in optical networks using different metrics. The choice of the performance metric is crucial for proactive response to degradation or failures. The key metric of interest for network designers is the lightpath bit error rate (BER), typically expressed as pre-forward error correction (FEC) BER, which determines performance acceptability [2]. OSAs facilitate proactive maintenance and margin optimization by monitoring the optical system's health without demodulation [26]. Measuring the optical signal-to-noise ratio (OSNR), in the reference bandwidth of 0.1 nm, with an OSA allows validation of expected performance during network provisioning without demodulation, as BER or Q-factor depend on the OSNR. However, measuring wavelength division multiplexing (WDM) channel OSNR without service interruption requires dedicated hardware and algorithms. OSNR also includes impairments from transceivers, making it impossible to distinguish different parts of the noise term. This leads to the generalized-OSNR (GOSNR) concept, defined as the OSNR value at which the same BER is achieved in back-to-back transmission (back-trace method) [27]. GOSNR captures only optical impairments from the link, including noise and nonlinear interference, and is estimated for the destination node due to difficulties in obtaining OSNR within a dense WDM signal.

In recent years, significant efforts have focused on exploring machine learning algorithms for managing soft failures, including detection, identification, and localization. Soft-failures arise from degrading and aging of components in the networks and can potentially evolve into hard-failures, i.e. failures that disrupt the service. In [11], signal overlap, tight filtering, gradual drift, and cyclic shift of filters are considered as soft-failures, achieving fault detection using an adaptive threshold mechanism for BER changes and identification using a pattern recognition algorithm. Shariati et al. [13] addressed soft failures caused by filters, using features extracted from the optical spectrum and a support vector machine (SVM) for detection and identification. Lun et al. [14] utilized a one-dimensional convolutional neural network (CNN) on power spectrum density (PSD) from a coherent receiver to identify variations in the PSD, considering soft failures like filter shift, filter tightening, ASE noise increase, and Kerr nonlinear effects from launch power increase. Mayer et al. [17] employed an artificial neural network (ANN) to localize soft failures with limited telemetry data, classifying based on OSNR and transmitter power features to detect amplifier gain degradation, transponder power degradation, and additional fiber losses. A brief overview of the literature on failure management, i.e., soft-failure detection (SFD), soft-failure identification (SFI), and soft-failure localization (SFL), as well as QoT estimation based on different optical performance monitoring (OPM) data is shown in Table I.

It can be seen that with regard to QoT estimation and failure assurance, only one aspect was focused on, although the two issues are related. Since degrading QoT is an indicator for a failure, we will investigate its use as a failure detection mechanism as well.

In this article, we experimentally validate our work from [22] by using the proposed framework on experimental data obtained through a lab setup showing possible benefits of using OSAs in future optical networks to improve its reliability. Furthermore, we set the topic into context while also providing more information on the underlying machine learning algorithms and performance metrics used. The remainder of this paper is organized as follows: First, a brief overview of the theory behind long-short term memory networks, variational autoencoder, and performance metrics is given in Section II. In Section III, the proposed machine learning framework for simultaneous QoT estimation and failure detection, identification and localization is described. Furthermore, the experimental setup, experimental dataset and results are shown in Section IV. A conclusion will be drawn in Section V.

II. BACKGROUND

In this section, we give a short overview of long-short term memory (LSTM) networks, variational autoencoder and

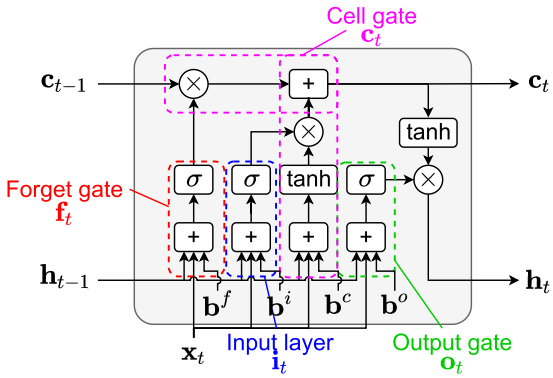

Fig. 1. Basic structure of an LSTM cell; $\mathbf{x}_t$: input at time t, $\mathbf{h}_{t-1}$: output of LSTM unit (last hidden state) at time $t-1$, $\mathbf{c}_{t-1}$: cell state at time $t-1$, $\mathbf{c}_t$: updated cell state, $\mathbf{h}_t$: output (new hidden state). Each gate has a certain bias value $\mathbf{b}$.

performance metrics for machine learning algorithms. LSTMs are designed to overcome the vanishing gradient problem of which recurrent neural networks (RNNs) suffer. However, RNNs overcome the limitations of traditional feed-forward neural networks (FF-NN) which allow only data flow from the input layer through hidden layers to the output, i.e. in forward direction only. RNNs add another recurrent path to the information flow by individually inputting symbols of a sequence and storing internal state between steps [28]. The backpropagation through time (BPTT) algorithm can be used to train RNNs by unfolding the network in time, enabling training of recurrent connections.

Autoencoders are artificial neural networks with a bottleneck training to reconstruct their input data. They can be used for feature engineering in the preprocessing of data for machine learning algorithms. Feature selection and dimensionality reduction of the input data can be achieved with autoencoders. Variational autoencoders as an extension to conventional autoencoders enable the usage of a probabilistic encoder and decoder for generation capabilities and a higher generalization arising from the fact that the probability distribution of the latent space is trained to be Gaussian distributed.

For a deeper understanding of the evaluation for the investigated machine learning algorithms and their tasks, we briefly introduce common evaluation metrics.

A. Long-Short Term Memory

A Long-short term memory is a special kind of RNN designed to overcome the vanishing gradient problem. The vanishing gradient problem can occur during training when the gradients of the loss function with respect to the model parameters become exceedingly small as they propagate backwards through the layers of the network. When gradients become very small, it can be challenging for the model to effectively update its weights, resulting in training slowing down significantly or completely stalling. LSTM networks were initially proposed by Hochreiter and Schmidhuber [29]. Since their introduction, LSTMs have become one of the most popular and effective methods for modeling sequential dependencies in input features. The basic structure of an LSTM is depicted in Fig. 1. In comparison to traditional RNNs, each layer of an LSTM is expanded to include memory cells, which are controlled by gates. These gates manage the flow of information and retain information from previous time steps [29].

An LSTM cell consists of input, forget, and output gates, as well as a cell activation component [28]. These gates regulate the information flow between memory cells based on past inputs to the network. The presence of the forget gate, along with the additive property of the cell state gradients, enables the network to update the parameter in such a way that the different sub gradients do not necessarily agree and are less likely to converge to zero, i.e., the gradients will not vanish.

B. Variational Autoencoder

Autoencoders (AEs) are a type of artificial neural network architecture comprising an encoder $E : X \to Z$ and a decoder network $D : Z \to X$, where $Z \in \mathbb{R}^n, n \in \mathbb{N}^+$. These are jointly trained to reconstruct unlabeled data $X \in \mathbb{R}^m$. By selecting a lower dimension $n < m$, represented by the multivariate latent vector $z = E(x)$ with $x \in X$, the encoder E learns to encode the input data X in a way that enables reconstruction with the decoder $\hat{x} = D(z)$, where $z \in Z$. Trained autoenconders enable applications such as dimensionality reduction by using z instead of x, denoising by utilizing $\hat{x}$, and anomaly detection by measuring the discrepancy between x and $\hat{x}$. This training, to represent the input data with a significantly lower amount of features, is called bottleneck training. Variational Autoencoders (VAEs), introduced by Kingma and Welling [30], are an expansion of autoencoders (AEs). While VAEs share a similar architecture with AEs, they differ in their objective. Instead of directly reconstructing the data, VAEs aim to approximate the unknown distribution P of the data by utilizing a prior distribution p parameterized by θ. The latent vector $\mathbf{z}$ in VAEs is typically assumed to follow a multivariate Gaussian distribution, which enables additional capabilities beyond conventional AEs. For instance, VAEs can generate new data by decoding samples from the latent distribution using the probabilistic decoder. These properties contribute to improved generalization.

In VAEs, the true posterior $p_\theta(\mathbf{z}|\mathbf{x})$ is often intractable, necessitating approximation. This is achieved through a function $q_\phi(\mathbf{z}|\mathbf{x}) \approx p_\theta(\mathbf{z}|\mathbf{x})$, which is parameterized by the probabilistic encoder $E_\phi(\mathbf{x})$. The multivariate latent vector $\mathbf{z}$ is calculated using a reparametrization trick to ensure the model's trainability:

$$\mathbf{z} = \mu + \sigma \odot \varepsilon. \tag{1}$$

Here, μ represents the mean value, σ denotes the standard deviation, and ε is a sample drawn from the multivariate normal distribution $\mathbb{N}(0, I)$. Fig. 2 illustrates the basic structure of a VAE. During training, the objective is to find optimal parameters θ and ϕ that minimize the reconstruction error at the decoder output while preserving the probability distribution in the latent space. By utilizing the encoder of a well-trained autoencoder, VAEs can effectively reduce the input dimension m to n ($n < m$) with minimal information loss. As a result, the latent space represents a meaningful set of features for describing the input

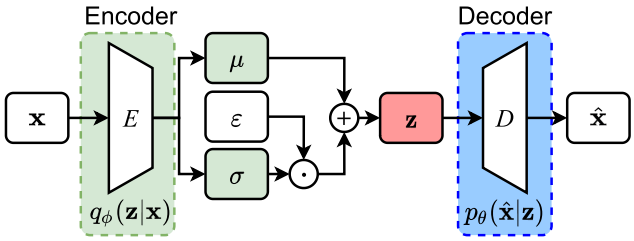

Fig. 2. Basic structure of a VAE; $\mathbf{x}$: input vector, $\mathbf{z}$: multivariate latent vector, $\hat{\mathbf{x}}$: reconstructed input vector; μ: mean value, σ: standard deviation, and ε: sample drawn from multivariate normal distribution.

data, reducing the need for manual feature selection in other machine learning algorithms.

C. Performance Metrics

1) Regression Metrics: Univariate regression models are trained to map input variables to a single output variable. Thus, the output variable is a scalar, which can be, for example, amounts or sizes.

Commonly, the quality of the regression is evaluated using the mean absolute error (MAE) defined as

$$\text{MAE} = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n}, \tag{2}$$

where n is the number of samples, y_i is the prediction, x_i is the true value, and $|y_i - x_i|$, thus, is the absolute error of the i^{th} sample.

While the MAE is a metric in the same scale as the input data, the R^2-score, also known as the coefficient of determination, is a more intuitive performance metric with its output being between 0 and 1, where a score of 1 is indicating a perfect prediction [31]. The R^2-score is defined as

$$R^2(y, x) = 1 - \frac{\sum_{i=1}^{n}(x_i - y_i)^2}{\sum_{i=1}^{n}(x_i - \overline{y})^2}, \tag{3}$$

where $\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ is the mean over all predictions. It provides an indication of how well unseen samples are likely to be predicted by the model.

2) Classification Metrics: In binary classification tasks the classifier maps each input to one of the two classes "true" or "false, leading to

- true positives (TP), i.e. the number of samples classified as "true" while being labeled "true",
- true negatives (TN), i.e. the number of samples classified as "false" while being labeled "false",
- false positives (FP), i.e. the number of samples classified as "true" while being labeled "false", and
- false negatives (FN), i.e. the number of samples classified as "false" while being labeled "true".

The overall effectiveness of a classifier can be evaluated using different metrics arising from the numbers of TP, TN, FP and FN [32]:

- Accuracy is the percentage of correct predictions:

$$\text{A} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FN} + \text{FP} + \text{TN}}. \tag{4}$$

- Precision is the percentage of positive predictions that are actually positive:

$$\text{P} = \frac{\text{TP}}{\text{TP} + \text{FP}}. \tag{5}$$

- Recall is the percentage of actual positives that are predicted positive:

$$\text{R} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \tag{6}$$

- F1-score is the harmonic mean of the precision and recall defined as:

$$\text{F1} = \frac{2\text{TP} + \text{TN}}{2\text{TP} + \text{FN} + \text{FP}}. \tag{7}$$

For imbalanced datasets containing an unequal number of positive and negative samples the F1-score is better suited than the accuracy to evaluate the performance of a classifier. Class imbalance is mostly present in failure datasets, since more normal data can be retrieved than faulty data.

III. Fault Management and QoT Estimation Framework

Autoencoders (AE) have already been used for semi-supervised anomaly detection based on its reconstruction error in previous work (e.g. [15]). In this work, we are using an extension of conventional autoencoders which is a variational autoencoder (VAE), as described in Section II. However, we are not only using the VAE for failure detection but also for feature reduction. The proposed framework was extensively optimized using a grid search with 80,000 configurations. The VAE encoder was configured with an input layer of size 501 in order match the number of spectrum points obtained by the OSA, one hidden layer of size 25, batch normalization, and an output layer of size 12. Consequently, the latent space size is also 12, which matches the input layer size of the VAE decoder. The decoder structure mirrors that of the encoder. The ReLU function serves as the activation function for all layers. In the following the three stage failure management framework as well as the QoT estimator will be explained in more detail.

A. Quality of Transmission Estimation

A QoT estimation framework based on LSTM and FF-NN layers is trained using an experimentally generated dataset presented in the Section IV-A. The dataset contains samples for various link lengths, modulation formats, channel configurations and launch powers per channel. It is assumed that the distances between the starting node N_1 and current node N_S, i.e., the node where the spectrum is obtained, are known. However, the span lengths in the link are not needed to be known exactly as shown in [3].

The framework's overall structure is depicted in Fig. 3. Feature vectors are inserted through an input layer. Those feature vectors include the link length, modulation format per channel, the channel configuration, the launch powers per channel, the distances between the starting node and the current node, and the latent space of the VAE with the current optical spectrum as

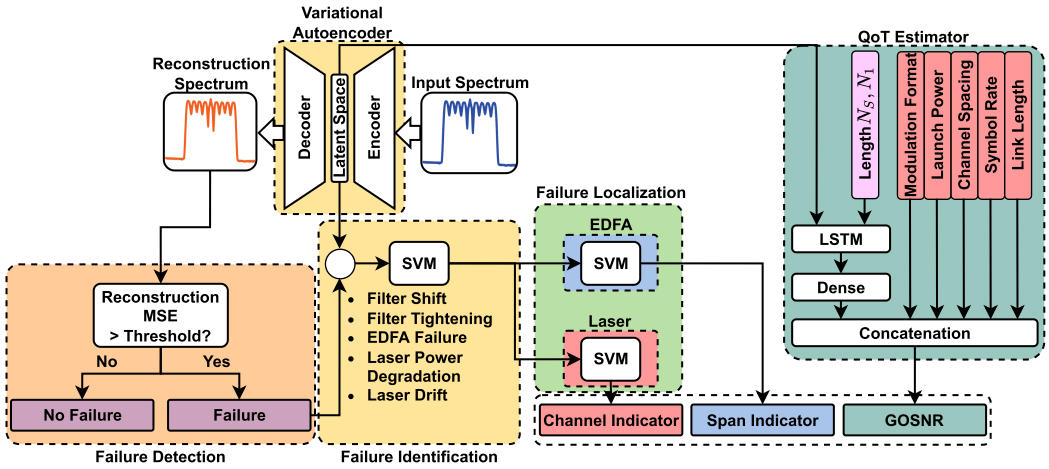

Fig. 3. Fault management and QoT estimation framework based on optical spectrum interpreted by a variational autoencoder. MAE: mean absolute error, SVM: support vector machine, EDFA: Erbium-doped fiber amplifier, LSTM: long-short term memory, N_S: current node number, N_1: starting node.

the input. LSTM layers and a dense layer handle the dimension-changing vectors, i.e. the spectra, while the feature vectors with a fixed size are directly passed to the concatenation layer. The concatenation layer combines the RNN branch's output with the inputs from the transmission features, allowing interpretation by feed-forward NN layers. The FF-NN calculates the output, resulting in a GOSNR value for each node in the link.

The dataset is split into 60% training, 20% validation, and 20% test data before training the framework. Training is performed over 800 epochs using the Adam optimizer [33]. Choosing the optimal layer size is crucial for high accuracy and fast estimation. Oversizing the framework can lead to overfitting and increased computational effort. Thus, different sizes of hidden layers for LSTM RNNs were investigated for optimal MAE. Results show that increasing structure size does not necessarily significantly improve MAE performance. Consequently, the QoT estimation framework utilizes LSTM recurrent layers of sizes 24 and 12.

B. Failure Detection, Identification and Localization

Soft-failure detection is established using the semi-supervised nature of the VAE. During training, the VAE is optimized to minimize the reconstruction error of non-faulty data. If a fault is present, the reconstruction mean squared error (MSE) will be larger than the non-fault MSE. Due to this, we can achieve failure detection by using a threshold-based approach. The threshold is optimized regarding the reconstruction MSE. If a failure is detected, the second stage of the framework is activated. Here, an ML-based classifier is used to identify the soft-failure cause with the latent space of the VAE as the input features. If a certain failure, for example an Erbium-doped fiber amplifier (EDFA) noise figure increase or a channel power drop, is identified, then the failure localization is activated next. The localization of the EDFA, which causes the noise figure increase, can be physically explained as follows: An increase in noise figure and change in gain spectrum due to pump laser degradation leads to different non-linearities in the following span affecting the signal. This leads to an individual distortion on the signal which is also visible in the spectrum obtained by the OSA and used by the ML algorithms. The next EDFA in the link would need to pump more which leads to an additional increase in the ASE noise. A classifier is trained to localize the failure on a per-span or per-channel level. Thus the output of the localization stages are either the span number or the channel number in which the failure is present. In the following section, a support vector machine classifier with a RBF-Kernel (Radial Basis Function) will be used as the classifier for both tasks. The algorithms is optimized using an exhaustive grid search for optimizing the F1-score.

IV. Experimental Validation

In [22], simulations were done to propose a joint QoT estimation and soft-failure management framework. Although simulations can cover a wider variation of configurations of network sizes and components, experiments are indispensable for deployments of such frameworks as well as showing its capabilities in a real-world environment. The process of the experimental validation including setup description and results will be shown in the following.

A. Experimental Setup

The experimental setup serves the purpose of generating experimental data [3]. Offline execution of the DSP is accomplished through MATLAB routines. To create the channel of interest (COI) at the transmitter side, a pseudo-random multi-level sequence (PRMS) with a length of $2^{17} - 1$ is generated. This sequence is then mapped to dual polarization (DP) QPSK, DP 8-QAM, or DP 16-QAM symbols, followed by the addition of training symbols for equalization and synchronization. Prior to up-sampling from the symbol rate of 32 Gbaud to the DAC sampling rate (88 GSa/s), the signal undergoes predistortion to account for the characteristics of the electrical amplifier and digital-to-analog converter (DAC). Subsequently, the signal is shaped using a root-raised cosine filter with a roll-off factor

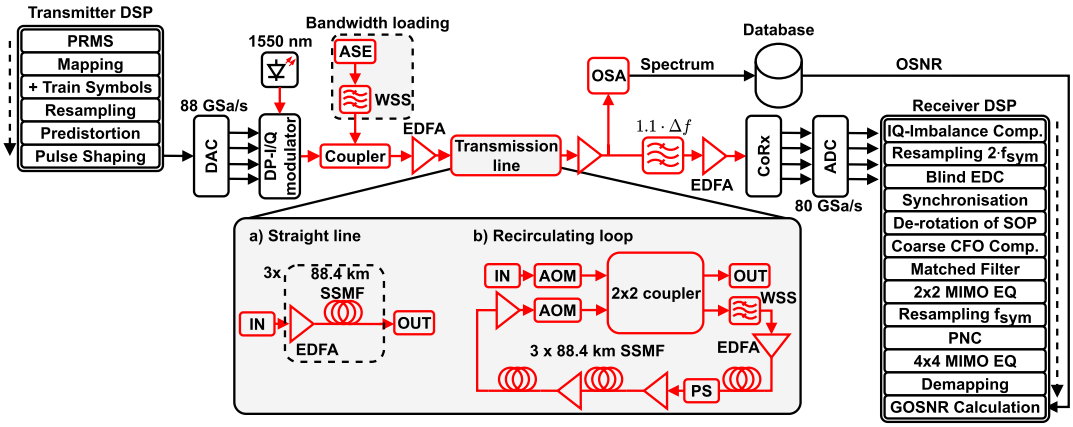

Fig. 4. Experimental transmission setup with a) a straight line and b) a recirculating loop. PRMS: pseudo-random multilevel sequence, DAC: digital-to-analogue converter, ASE: amplified-spontaneous emission, WSS: wavelength selective switch, EDFA: Erbium-doped fiber amplifier, PS: polarization scrambler, EDC: electrical dispersion compensation, SOP: state of polarization, CFO: carrier frequency offset, PNC: phase-noise compensation.

of 0.2, resulting in an almost rectangular spectrum. Digital-to-analog conversion is achieved by an arbitrary waveform generator (AWG) operating at 88 GSa/s, with an effective number of bits (ENOB) of 5.5 bits.

The COI is generated by an external laser at a wavelength of 1550.004 nm, coupled with a DP-IQ modulator driven by the DAC via four driver amplifiers. For the generation of the other WDM channels (loaders), i.e., one, three or five in total, a programmable wavelength-shaping filter (II-VI WS4000A) (WSS) is employed, utilizing an amplified spontaneous emission (ASE) noise source as input. This process results in shaped ASE noise that represents the WDM channels proximate to the COI. The wavelength-shaping filter has a periodically repeating filter bandwidth aligned with the channel spacing and is configured to equalize all channels at the output. Opting for noise-loading offers advantages over traditional channel generation as it reduces complexity on the transmitter side, requiring only one modulator, one laser, and one DAC. Comparatively, the characteristics of a noise-loaded signal closely resemble those of a conventional WDM signal [34]. The COI and loaders are combined using a 3 dB-coupler and subsequently amplified using an EDFA. After amplification, the COI and loaders are ready for the transmission over the fiber. The experimental setup is depicted as a high-level model in Fig. 4. It provides a comprehensive framework for generating and manipulating signals, allowing for the study of various aspects related to channel transmission and performance evaluation.

For generating non-faulty data, the EDFA output is fed into the recirculating loop. The loop is composed of another waveshaper (Finisar WS4000S) being used as a gain-flattening filter followed by three spans and a polarization scrambler (Fig. 4). Each span consists of an EDFA running at a constant output power of 10.5 dBm and a variable optical attenuator (VOA) after the EDFA to get the desired launch power for the following 88.4 km SSMF. After the first span, the polarization scrambler is deployed to randomize the polarization shift effects from the fibers.

At the receiver side, the signal is amplified using an additional EDFA before being filtered to isolate the COI. Subsequently, the COI is detected using a coherent receiver. Analog-to-digital conversion (ADC) is performed by an oscilloscope operating at 80 GSa/s. The received signal is subject to various impairments, which can be classified as either uncompensated (primarily noise and nonlinearities) or compensated. Compensated disturbances include IQ-skews and IQ-imbalances originating from both the transmitter and receiver, laser phase noise from the transmitter and receiver, chromatic dispersion, polarization mode dispersion (PMD), rotation of the state of polarization (SOP), carrier frequency offset, and receiver laser phase noise. Offline receiver digital signal processing (DSP) is carried out using standard algorithms specifically designed for coherent DP-WDM systems [3], [35]. Towards the end of the DSP chain, the GOSNR is computed using pre-measured lookup tables that establish the relationship between OSNR and Q-factor for the given configurations, employing the back-trace method. To obtain the spectrum, an OSA with an optical resolution of 10 pm, such as the Advantest Q8384, is utilized. The experimental parameters for the investigations are summarized in Table II. In total, a non-faulty dataset with 21,600 spectra is obtained and used for the QoT estimation as well as partly for the training of the VAE.

TABLE II
EXPERIMENTAL SYSTEM PARAMETERS

Parameter	Value
Modulation format	DP-QPSK, DP-8-QAM, DP-16-QAM
Symbol rate	32 GBaud
Channel spacing	37.5 GHz
Number of channels	1, 3, 5
Launch power	-3, -2, -1, 0 dBm
Center wavelength	1550.004 nm
Loop length	265.2 km
Loop iterations	1 to 6
OSA resolution	10 pm

1) Soft-Failure Emulation: We consider five different types of soft-failures that we emulate in our lab:

- An increase in EDFA noise figure (NF),

TABLE III
EXPERIMENTALLY EMULATED SOFT-FAILURES

Soft-failure	Range	Steps
EDFA noise figure (NF) increase	0.2 to 2 dB	0.2 dB
Transmit laser drift	-2.5 to 2.5 GHz	0.5 GHz
Transmit laser power drop	-2.5 to 2.5 dB	0.5 dB
Filter tightening	1 to 5 GHz	1 GHz
Filter shift	-2 to 2 GHz	1 GHz

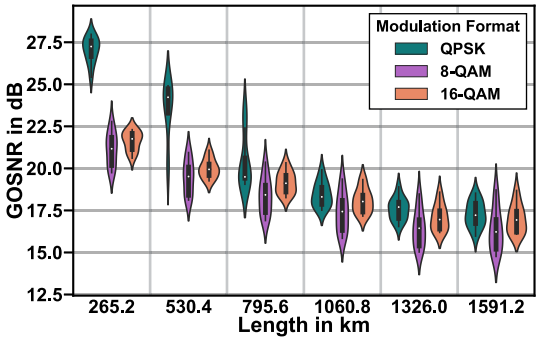

Fig. 5. GOSNR distribution over length for QPSK, 8-QAM, and 16-QAM.

- a shift in transmit laser frequency,
- a drop in transmit laser power,
- tightening of the filter characteristic,
- and a shift in the filter's center frequency.

To generate fault data, a straight-line experiment is conducted using three spans of 88.4 km SSMF, where various failure scenarios are emulated. The soft failure cases we examine are summarized in Table III . To emulate an increase in EDFA noise due to pump laser failure, a VOA is placed at the midstage access of the inline EDFAs. The attenuation of the VOA is adjusted from 0.2 to 2 dB in steps of 0.2 dB. To simulate a laser drift, the transmit laser frequency for the center channel is varied from its center frequency by −2.5 to 2.5 GHz in steps of 0.5 GHz. A power drop in a laser is emulated by varying the laser power between −2.5 to 2.5 dB in steps of 0.5 dB. The same procedure is followed for different randomly selected channels in the waveshaper responsible for noise shaping the loaders. For filter tightening, the waveshaper that generates the loaders is used to narrow the channels by 1 to 5 GHz in steps of 1 GHz. By shifting the center frequency of the waveshaper from −2 to 2 GHz in steps of 1 GHz, a filter shift is achieved. Sweeping through the experimental parameters from Table II and the emulated soft-failure parameters from Table III results in approximately 800 spectra per failure type and a total failure dataset of 5,600 spectra.

2) Quality of Transmission Measurements: To obtain the GOSNR, the experimentally obtained Q-factors are used for a back-trace to a back-to-back (B2B) configuration [3]. The distribution of GOSNR over different lengths is depicted in Fig. 5 due to the varying parameters from Table II. Overall, it can be observed that the curves characterized by the medians decrease with higher transmission distances and flatten out above 1500 km. This flattening is due to the Q-factor-OSNR curves becoming steeper below a specific low OSNR value and leveling off at high OSNR values. Furthermore, it is noticeable that the modulation formats QPSK and 16-QAM exhibit smaller violins compared to 8-QAM, and in general, the medians of the 8-QAM values are lower than those of 16-QAM. This distinction arises from 8-QAM being more susceptible to nonlinearities caused by unequal symbol distances in multi-channel transmissions. It can be seen that almost all violins have the greatest curvature in the lower GOSNR range. This indicates that the occurrence of lower GOSNR values is greater than that of the high values.

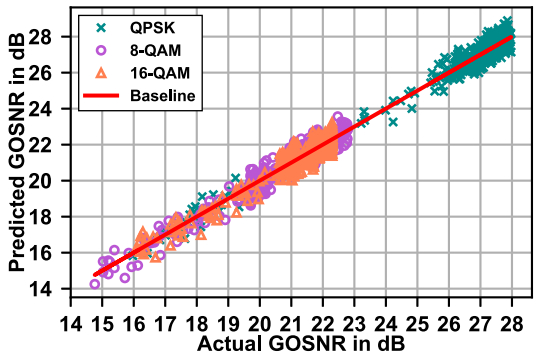

Fig. 6. Predicted GOSNR by the estimator over actual GOSNR for all configurations in the dataset; R^2-score: 0.9846, Mean absolute error: 0.29 dB.

B. Experimental Results

1) Quality of Transmission Estimation: To test the capabilities of the framework with regards to QoT estimation, the dataset is split into 60% training, 20% validation, and 20% test data. The dataset contains non-faulty data from the recirculating loop experiment and faulty data from the straight line experiment. After the training, the QoT estimator is tested on the test data as well as the failure data. The result of the prediction is depicted in Fig. 6. It can be seen that only small deviations from the baseline occur which is represented in the R^2-score of 0.9846 and the MAE of 0.29 dB. Furthermore, a high density of GOSNR values in the range from 16 to 23 dB and between 25.5 and 28 dB can be seen. This is due to the different modulation formats used, i.e. the high values arise from the QPSK modulated samples and the lower regime is populated by the samples modulated as 8-QAM and 16-QAM symbols. The datapoints in the plots outside of these regimes are originating from soft-failures induced in the failure measurements. A more in-depth analysis of soft-failure detection with the help of the QoT metric will be given in the following section. The generalization capabilities of the QoT estimation block in this framework have been evaluated in [36]. We showed the benefits of using the optical spectrum as input features for the QoT estimation tasks and the advantage in generalization when feature changes are present.

2) Soft-Failure Detection: First, we assume soft-failure detection using a QoT metric based threshold mechanism. We have to define an acceptable range of GOSNR fluctuation for the straight line experiment with a length 265.2 km to find a suitable threshold for every modulation format. If we look at the histogram of the GOSNR in Fig. 7, we can see a peak for QPSK around 26.2 dB which indicates, that some of the failures are located in this area. For 8-QAM this would be

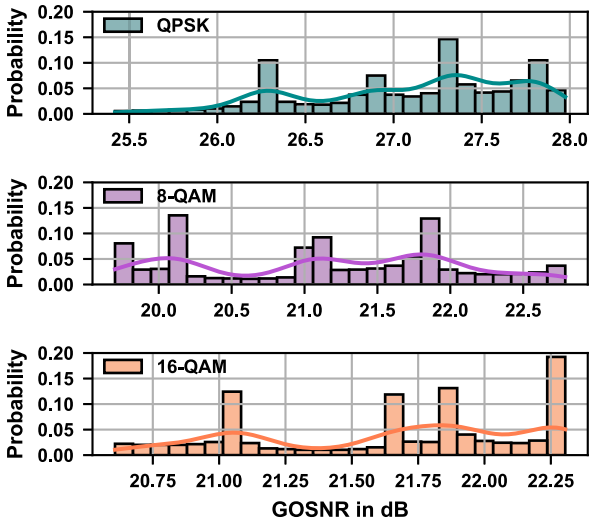

Fig. 7. Histogram of the occurrence of GOSNR values in the dataset; solid lines: smoothed probability distribution from the kernel density estimate.

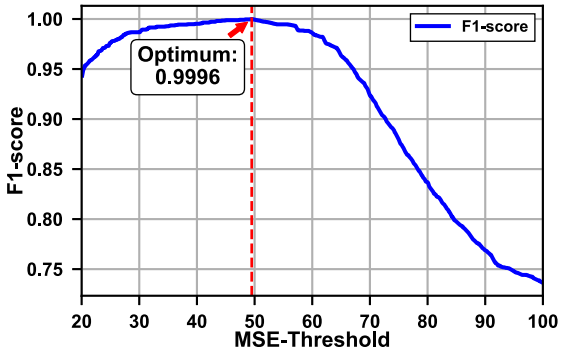

Fig. 8. F1-score over MSE-Threshold with indicated optimal threshold.

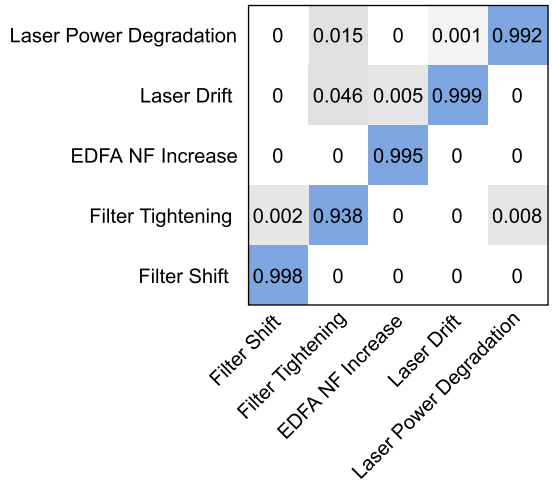

Fig. 9. Confusion matrix for failure identification by the SVM-based classifier with an overall F1-score of 0.9905.

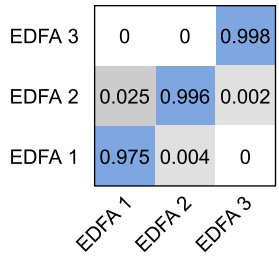

Fig. 10. Confusion matrix for EDFA soft-failure localization by the SVM-based classifier with an overall F1-score of 0.9913.

around 20.3 dB and for 16-QAM this would be at 21.1 dB. If we assume these values as the failure threshold for each modulation format, we achieve an F1-score of 0.7135. This is due to the emulated soft-failures not necessarily leading to a major decrease in GOSNR.

For the second mechanism, we use the VAE-based failure detection. For this, the MSE for normal data and fault data is computed. Afterwards, we sweep over the MSE values to find the optimal threshold with regards to the F1-score. This sweep is depicted in Fig. 8. With the optimum threshold we achieve an F1-score for the failure detection of 0.9996. The overall shape of the MSE-threshold against F1-score curve shows other trends: If the threshold is Set too low, we might as well get a high F1-score, however, this is due to some failures having a low MSE from the normal spectra. A threshold which is set too high will lead to bad classification performances, since the MSE would be outside of the majority of the samples.

3) Soft-Failure Identification: For evaluation of the SVM-based classifier for identification, the soft-failure dataset from the straight line experiment is divided into 60% training, 20% validation, and 20% test data. After the before mentioned grid search, the SVM reaches an F1-score of 0.9905. This highly accurate classification can also be seen in Fig. 9. Here, the confusion matrix for the identification is depicted. It can be derived that the overall accuracy is high. A misclassification of the filter shift as filter tightening can only be seen in 0.2% of the tested failures. A higher amount of wrongly classified samples can be seen in the case of filter tightening. These discrepancies are due to the dataset containing also small deviations from the normal state.

4) Soft-Failure Localization: For the localization purposes of the framework, the soft-failures regarding EDFA noise figure increase are considered, since these are the only failures, which can be localized on a per-span level in our lab setup. The localization of an increase in EDFA noise figure becomes more challenging when considering individual spans. This difficulty is evident in the confusion matrix shown in Fig. 10, which represents the performance of a support vector classifier. In this matrix, it can be observed that the first EDFA's localization accuracy is lower compared to the last EDFA in the link. This discrepancy is attributed to the SVM with RBF kernel's inability to differentiate between an increase in the noise figure of an EDFA and a variation in the noise figure itself. Additionally, even a slight attenuation of 0.2 dB in the midstage access of an EDFA only leads to a minor increase in ASE (amplified spontaneous emission) noise. However, the overall accuracy of the SVM-based localization is high. This is also shown in the F1-score of 0.9913.

TABLE IV
OVERALL FRAMEWORK PERFORMANCE

Framework stage	Max. R2/F1-score
QoT Estimation	0.9846
Soft-Failure Detection	0.9996
Soft-Failure Identification	0.9905
Soft-Failure Localization	0.9913

V. CONCLUSION

In this paper, we experimentally validated our proposed machine learning-based framework from [22] for joint QoT estimation and soft-failure detection, identification and localization based on optical spectra obtained by sparsely deployed OSAs. Furthermore, we investigated the advantages of a VAE-based soft-failure detection mechanism over a QoT metric based approach. The framework was validated on experimental data acquired for 32 GBaud DP-QPSK, DP-8-QAM, and DP-16-QAM with up to 5 channels with 37.5 GHz spacing. For the generation of the QoT estimation data, a recirculating loop with up to 6 loop iterations of three spans of 88.4 km SSMF with three EDFAs has been used. The soft-failure data is acquired over a three span setup. The emulated soft-failures include EDFA noise figure increase, transmit laser frequency drift, transmit laser power drop, filter tightening, and filter shift. The results show, that the same VAE can be used for reliable QoT estimation, shown by an R2-score of 0.9846 and an MAE of 0.29 dB, and soft-failure detection (F1-score: 0.9996), identification (F1-score: 0.9905) and localization (F1-score: 0.9913). The overall performance of the framework is summarized in Table IV. We show that leveraging the optical spectrum for simultaneous QoT estimation and soft-failure management is beneficial and can be used in practical transmission scenarios.

REFERENCES

[1] Y. Pointurier, "Design of low-margin optical networks," *J. Opt. Commun. Netw.*, vol. 9, no. 1, pp. A9–A17, Jan. 2017.

[2] Y. Pointurier, "Machine learning techniques for quality of transmission estimation in optical networks," *J. Opt. Commun. Netw.*, vol. 13, no. 4, pp. B60–B71, Apr. 2021.

[3] L. E. Kruse, S. Kü"hl, A. Dochhan, and S. Pachnicke, "Experimental investigation of spectral data enhanced QoT estimation," *J. Lightw. Technol.*, vol. 41, no. 18, pp. 5885–5894, Set. 2023.

[4] D. Rafique, T. Szyrkowiec, H. Grießer, A. Autenrieth, and J.-P. Elbers, "Cognitive assurance architecture for optical network fault management," *J. Lightw. Technol.*, vol. 36, no. 7, pp. 1443–1450, 2017.

[5] 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. Lightw. Technol.*, vol. 37, no. 16, pp. 4125–4139, Aug. 2019.

[6] J. Shao, X. Liang, and S. Kumar, "Comparison of split-step Fourier schemes for simulating fiber optic communication systems," *IEEE Photon. J.*, vol. 6, no. 4, pp. 1–15, Aug. 2014.

[7] P. Poggiolini, "The GN model of non-linear propagation in uncompensated coherent optical systems," *J. Lightw. Technol.*, vol. 30, no. 24, pp. 3857–3879, Dec. 2012.

[8] V. Curri, "GNPy model of the physical layer for open and disaggregated optical networking," *J. Opt. Commun. Netw.*, vol. 14, no. 6, pp. C92–C104, Jun. 2022.

[9] A. D'Amico et al., "Experimental validation of GNPy in a multi-vendor flex-grid flex-rate WDM optical transport scenario," *J. Opt. Commun. Netw.*, vol. 14, no. 3, pp. 79–88, Mar. 2022.

[10] Q. Zhuge et al., "Application of machine learning in fiber nonlinearity modeling and monitoring for elastic optical networks," *J. Lightw. Technol.*, vol. 37, no. 13, pp. 3055–3063, Jul. 2019.

[11] A. P. Vela et al., "BER degradation detection and failure identification in elastic optical networks," *J. Lightw. Technol.*, vol. 35, no. 21, pp. 4595–4604, Nov. 2017.

[12] M. Furdek, C. Natalino, A. Di Giglio, and M. Schiano, "Optical network security management: Requirements, architecture, and efficient machine learning models for detection of evolving threats," *J. Opt. Commun. Netw.*, vol. 13, no. 2, pp. A144–A155, 2021.

[13] B. Shariati, M. Ruiz, J. Comellas, and L. Velasco, "Learning from the optical spectrum: Failure detection and identification," *J. Lightw. Technol.*, vol. 37, no. 2, pp. 433–440, Jan. 2019.

[14] H. Lun et al., "Soft failure identification for long-haul optical communication systems based on one-dimensional convolutional neural network," *J. Lightw. Technol.*, vol. 38, no. 11, pp. 2992–2999, Jun. 2020.

[15] K. Sun, Z. Yu, L. Shu, Z. Wan, and K. Xu, "Generalized soft failure identification enabled by digital residual spectrum and autoencoder," in *Proc. IEEE Opt. Fiber Commun. Conf. Exhib.*, 2021, pp. 1–3.

[16] C. Zhang et al., "Potential failure cause identification for optical networks using deep learning with an attention mechanism," *J. Opt. Commun. Netw.*, vol. 14, no. 2, pp. A122–A133, Feb. 2022.

[17] K. S. Mayer, J. A. Soares, R. P. Pinto, C. E. Rothenberg, D. S. Arantes, and D. A. A. Mello, "Machine-learning-based soft-failure localization with partial software-defined networking telemetry," *J. Opt. Commun. Netw.*, vol. 13, no. 10, pp. E122–E131, Oct. 2021.

[18] R. Wang et al., "Suspect fault screen assisted graph aggregation network for intra-/inter-node failure localization in ROADM-based optical networks," *J. Opt. Commun. Netw.*, vol. 15, no. 7, pp. C88–C99, Jul. 2023.

[19] K. Sun, Z. Yu, L. Shu, Z. Wan, H. Huang, and K. Xu, "Experimental demonstration of soft failure identification based on digital residual spectrum and machine learning," in *Proc. IEEE Int. Conf. U.K.-China Emerg. Technol.*, 2021, pp. 81–84.

[20] S. Aladin, A. V. S. Tran, S. Allogba, and C. Tremblay, "Quality of transmission estimation and short-term performance forecast of lightpaths," *J. Lightw. Technol.*, vol. 38, no. 10, pp. 2807–2814, May 2020.

[21] L. E. Kruse, S. Kü"hl, and S. Pachnicke, "Exact component parameter agnostic QoT estimation using spectral data-driven LSTM in optical networks," in *Proc. IEEE Opt. Fiber Commun. Conf.*, 2022, pp. 1–3.

[22] L. E. Kruse and S. Pachnicke, "Joint QoT estimation and soft-failure localization using variational autoencoder," in *Proc. IEEE Int. Conf. Opt. Netw. Des. Model.*, 2023, pp. 1–3.

[23] C. Rottondi, L. Barletta, A. Giusti, and M. Tornatore, "Machine-learning method for quality of transmission prediction of unestablished lightpaths," *J. Opt. Commun. Netw.*, vol. 10, no. 2, pp. A286–A297, Feb. 2018.

[24] I. Sartzetakis, K. K. Christodoulopoulos, and E. M. Varvarigos, "Accurate quality of transmission estimation with machine learning," *J. Opt. Commun. Netw.*, vol. 11, no. 3, pp. 140–150, Mar. 2019.

[25] E. Seve, J. Pesic, and Y. Pointurier, "Associating machine-learning and analytical models for quality of transmission estimation: Combining the best of both worlds," *J. Opt. Commun. Netw.*, vol. 13, no. 6, pp. C21–C30, Jun. 2021.

[26] I. Khan, M. Bilal, M. U. Masood, A. D'Amico, and V. Curri, "Lightpath QoT computation in optical networks assisted by transfer learning," *J. Opt. Commun. Netw.*, vol. 13, no. 4, pp. B72–B82, Apr. 2021.

[27] H. J. Cho et al., "Constellation-based identification of linear and nonlinear OSNR using machine learning: A study of link-agnostic performance," *Opt. Exp.*, vol. 30, no. 2, pp. 2693–2710, 2022.

[28] H. Salehinejad, S. Sankar, J. Barfett, E. Colak, and S. Valaee, "Recent advances in recurrent neural networks," *CoRR*, 2018. [Online]. Available: http://arxiv.org/abs/1801.01078

[29] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[30] D. P. Kingma et al., "An introduction to variational autoencoders," *Found. Trends Mach. Learn.*, vol. 12, no. 4, pp. 307–392, 2019.

[31] D. C. Montgomery, E. A. Peck, and G. G. Vining, *Introduction to Linear Regression Analysis*. Hoboken, NJ, USA: Wiley, 2021.

[32] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Inf. Process. Manage.*, vol. 45, no. 4, pp. 427–437, 2009.

[33] D. P. Kingma, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Representations*, 2015.

[34] T. Richter, J. Pan, and S. Tibuleac, "Comparison of WDM bandwidth loading using individual transponders, shaped, and flat ASE noise," in *Proc. Opt. IEEE Fiber Commun. Conf. Expo.*, 2018, pp. 1–3.

[35] S. Ohlendorf, T. Wettlin, S. Pachnicke, and W. Rosenkranz, "Optimized flexible mappings with multidimensional modulation for coherent optical transport," in *Proc. IEEE 45th Eur. Conf. Opt. Commun.*, 2019, pp. 1–4.

[36] L. E. Kruse, S. Kühl, and S. Pachnicke, "Generalizable QoT estimation based on spectral data driven LSTM in exact component parameter agnostic networks," in *Proc. IEEE Eur. Conf. Opt. Commun.*, 2022, pp. 1–4.