Query Based Iterative Learning Approach for Lightpath Deployment in Optical Networks

Query Based Iterative Learning Approach for Lightpath Deployment in Optical Networks

Fehmida Usmani
National University of Sciences & Technology (NUST), Pakistan
fusmani.dphd18seecs@seecs.edu.pk

Ihtesham Khan
Politecnico di Torino, IT
ihtesham.khan@polito.it

Hafsa Tariq
National University of Sciences & Technology (NUST), Pakistan
htariq.bese20seecs@seecs.edu.pk

Muhammad Shahzad
National University of Sciences & Technology (NUST), Pakistan
muhammad.shehzad@seecs.edu.pk

Arsalan Ahmad
National University of Sciences & Technology (NUST), Pakistan
arsalan.ahmad@seecs.edu.pk

Vittorio Curri
Politecnico di Torino, IT
curri@polito.it

***Abstract*—Predicting the Quality of Transmission (QoT) of a Lightpath (LP) before its actual deployment is important for efficient resource utilization. Conventionally, analytical models using closed-loop formulation estimate QoT, which imposes substantial margins to avoid network outages. Recently, data-driven techniques have been shown as a potential alternative with excellent precision and real-time applicability. However, data-driven techniques require sufficient training data, which might be challenging to acquire during real network operations. In this context, we proposed a novel unsupervised Iterative learning (IL) framework developed on top of the Random forest (RF) classifier for QoT estimation of LP before deployment. We considered the Generalized signal-to-noise ratio (GSNR) as a characterizing parameter for QoT estimation of LP. Our simulation results illustrate that, by employing the proposed iterative learning approach, we can obtain 99% classification accuracy with a reduced number of training samples compared to the traditional supervised learning approach.**

***Index Terms*—Machine learning; Quality of Transmission estimation; Generalized SNR; Iterative learning.**

I. Introduction

The main phase in the design and operation of optical networks is the prediction of the QoT of an LP. The modern alternative to the analytical models, such as the Gaussian Noise (GN) model, which implements conservative approaches to account for model generalizations and to cater to any imperfections, is the use of Machine learning (ML) techniques for QoT estimation [1]–[5]. These data-driven methods for QoT prediction of LPs are generally based on supervised learning techniques, which demand a considerable volume of data of already-deployed LPs to acquire a training dataset from which ML-based models can understand the augmented knowledge to assist the operator in the estimation of the QoT of forthcoming LPs [6].

Nevertheless, in a real operational network, the size of acquired telemetry is limited, and thus the training dataset size of supervised learning algorithms is comprised. Parallel to insufficient telemetry, it is also highly improbable that abnormal LPs will be noticed during regular network operation mainly due to the conservative system-design implementation, which ensures that the network will never turn out of service. This further decreased the size of the training dataset and added more bias to the data, affecting the accuracy of data-driven techniques. For the supervised ML-based models, the training dataset realizations must be sufficiently large, and the dataset must mimic the actual operational scenario in order to obtain a reasonable level of prediction accuracy.

In this context, IL (also known as "active learning" or sometimes "query learning") emerged as a promising solution in many contemporary ML problems, where it is difficult to obtain a decent amount of labeled data for training. IL-based algorithms aim to enhance the classification performance by utilizing the sequentially added useful training samples while lowering the labeling cost [7]. Contrary to IL, in traditional supervised learning, obtaining a large amount of labeled data to train the predictive model is fairly expensive, while obtaining a large volume of unlabeled data is relatively simple. The core concept of IL is that if the freedom is given to the ML model to select the data it learns from, it can achieve greater accuracy with less labeled training samples. The comparison between traditional supervised learning and query-based IL is demonstrated in Fig. 1. In traditional supervised learning, the entire dataset is labeled by an expert and given at once to train the machine learning model. In contrast, the query-based IL approach tries to minimize the overhead of data labeling by querying the label of selective unlabeled samples from the human annotator or oracle in each iteration [8], [9]. In this manner, the training dataset is increased sequentially in each iteration, the ML model is retrained on the newly added samples, and obtain highly accurate results with a few data samples while lowering the cost of acquiring labeled data. The effectiveness of IL approach is demonstrated in variety of applications such as image classification [10], semantic segmentation [11] and text classification [12]. In [13] and [14], active learning based approach is employed for QoT estimation as well.

In this work, we proposed a novel unsupervised query-based IL approach for classifying LP into good or bad QoT before its deployment. We consider an unsupervised learning approach to distinguish our work from the previous works [13] and [14]. In contrast, IL is applied as a supervised learning problem in previous works, where they initialize the model's training with few labeled data samples. In our scenario, we assume that no labeled dataset is available for initial training. The effectiveness of the unsupervised active learning-based approach is also demonstrated in [15] for linear regression problems. The major contributions of our works are as follows:

- An unsupervised query-based IL approach is proposed to classify the LP QoT into good or bad before deployment.
- We propose to use uncertainty sampling with entropy for sample selection.
- We demonstrate the good performance of our proposed scheme compared to the traditional approach.

II. Iterative learning engine for QoT Estimation

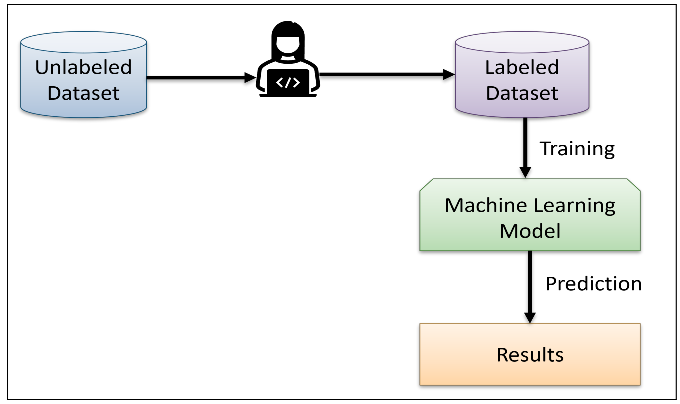
We proposed to develop the pool-based unsupervised IL framework on top of the RF classifier to evaluate the QoT of an unestablished LP in advance. The main goal of the IL approach is to increase the number of training samples iteratively by choosing the most useful samples. In this manner, our RF classifier will be trained on the most informative set of samples and achieve high performance with a reduced number of training samples. The proposed IL framework is developed using a high-level python-based Scikit-learn library [16].

A. Random Forest Classifier

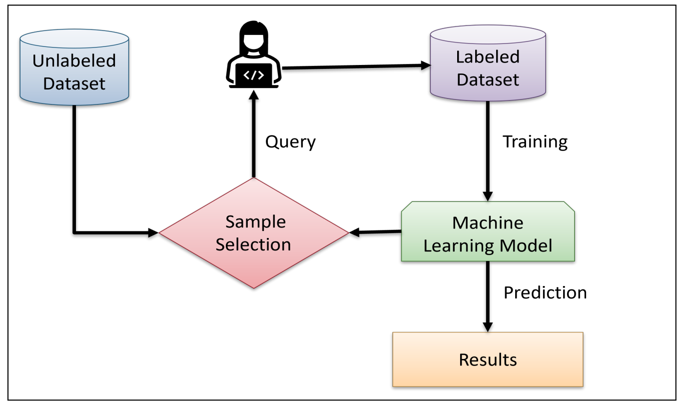
RF classifier is based on the ensemble learning approach that is comprised of several decision trees. It was initially presented to address the classification problems [17]. Each tree generates a prediction for a new testing sample. Consequently, the test sample gets voted for that prediction class. The class with the highest votes serves as the sample's prediction label. We developed the RF classifier using the feature space vector $\mathcal{V}$, containing power, ASE noise, NLI, and the number of spans to classify an LP into good or bad QoT.

B. Query-based iterative learning solution

Considering the available dataset $\mathcal{D}$={$x_1, x_2, \ldots, x_n$}, where x_i represents a sample of dimension vector $\mathcal{V}$. The label y for a sample x_i is $y_i \in$ {0,1}. The selected pool of unlabeled samples is given by $\mathcal{P}$ $\{x_1, x_2, \ldots, x_n\}$, where $\mathcal{P} \in \mathcal{D}$. The IL framework aims to develop the efficient data model by repeatedly querying the labels of most useful samples $\mathcal{U}$ from the expert system while considering the threshold value of a fixed budget $\mathcal{B}$. Let $\mathcal{L}$={$x_1, x_2, \ldots, x_n$} is a dataset in which labeled samples are added. The flowchart of our proposed methodology is illustrated in Fig. 2. In the first step, we randomly select the pool $\mathcal{P}$ of 1000 unlabeled samples (500 samples for each class) from the given dataset $\mathcal{D}$. We initialize the RF classifier with the unlabeled pool of data to extract the underlying distribution of data. Note that, we wrap our RF classifier using the Sklearn Classifier method described in [16] to cope with missing label values. Each labeled sample belongs to a certain class i.e, 0 or 1 in our case, while the labels for the samples in dataset $\mathcal{P}$ are unknown to us. For every unlabeled sample in dataset $\mathcal{P}$, we assume a random variable Y which denotes the membership class. We obtain the probabilities $\hat{\mathcal{P}}$ for membership class estimation of Y from our RF classifier. The degree of uncertainty of Y in class membership is computed in terms of entropy as follows:



(a)



(b)

Fig. 1: (a), Traditional Supervised Learning Approach (b), Iterative Learning Approach

$$E(Y) = -\sum_{y} \left(\hat{\mathcal{P}}(y|x_P) log \hat{\mathcal{P}}(y|x_P) \right) \quad (1)$$

where y denotes the classes, x_P represents a sample from the dataset $\mathcal{P}$. The higher value of E indicates the increased uncertainty of a classifier about the distribution of a sample class. In each round of IL, we calculate the probabilities of class membership for all the samples in dataset $\mathcal{P}$. To select the most useful samples from the set $\mathcal{P}$ for training, we utilized uncertainly sampling with the entropy method. The sample with the highest value of entropy is queried from an expert system to obtain a label. Let (x_L, y_L) be a sample for which label is obtained from the expert system, this sample will be added to the training dataset $\mathcal{L}$={$x_1, x_2, \ldots, x_n$}, and at the same time, the corresponding sample is removed from the dataset $\mathcal{P}$. The RF classifier is retrained on the dataset $\mathcal{L}$ until the budget is exhausted (number of iterations in our case)

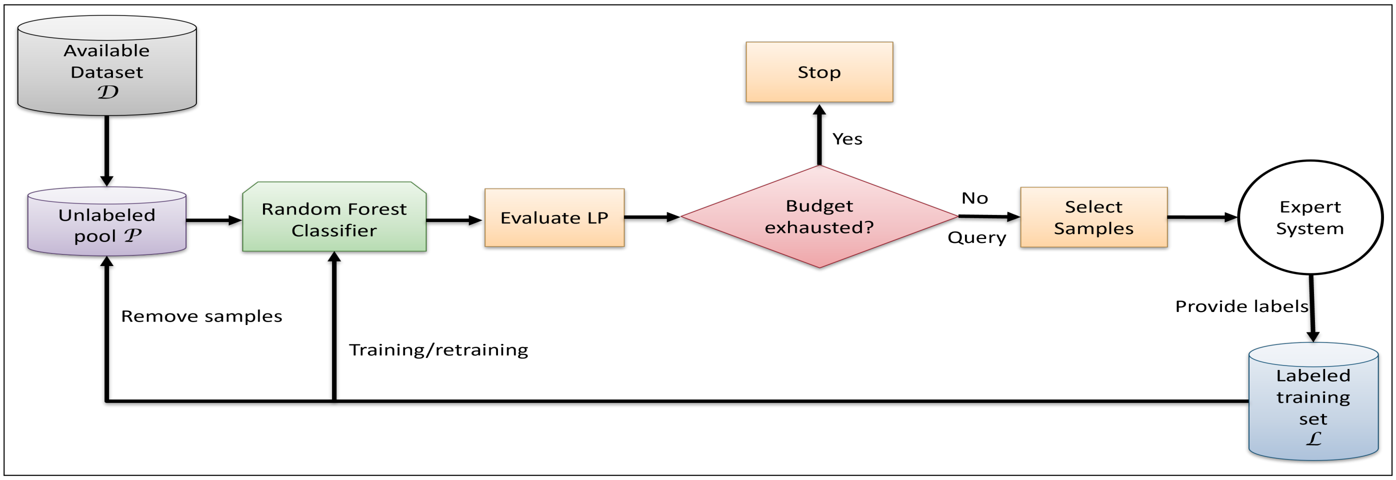


Fig. 2: Proposed Iterative Learning based Framework.

TABLE I: Network simulation parameters.

Simulation Parameters	
Launch Power/ Channel	0 dBm
Dispersion (D)	16.0 $ps/nm/km$
Attenuation coefficient (α)	0.2 dB/km
Channel Spacing (C-Band)	50 GHz
Span Length	80 km
WDM Comb (C-Band)	80
Baud Rate (C-Band)	32 $Gbaud$
Amplifier Noise Figure	[3.5 - 4.5] dB [18]
Amplifier Gain Ripple	Variation of 1 dB
Fiber Type	Standard SMF

III. Simulation environment and dataset Generation

In this work, we consider a software-defined open optical network, with an Optical line system (OLS) serving as an instance of network edges and Reconfigurable optical add-drop multiplexers (ROADMs) operating as an instance of network nodes [19]. The considered OLSs are being operated at their optimal working point, and the perturbed behaviour of the physical layer is described mainly by the ripple gain of the amplifier. These ripples gain fluctuates around when the spectral load variates. Therefore, even with some degree of uncertainty in the operating point, OLS controllers can guarantee that they are running at the nominal operating point. To connect the transceivers and allow for the use of dual-polarization multilevel modulation formats, LPs are deployed transparently on the Wavelength division multiplexing (WDM) flexible grid system at the transmission layer. In the course of transmission, LPs are degraded by a number of impairments, the most significant of which are Amplified spontaneous noise (ASE) and Non-linear impairments (NLI). Each In-line amplifier's (ILA) contribution to the ASE noise in the propagation is statistically independent, yet it all adds up. However, there is a statistically significant correlation between each span's NLI [20]. The overall $GSNR$ of each LP traversing through OLS is given as:

$$\frac{1}{GSNR} = \sum_{n} \frac{1}{GSNR_n} \tag{2}$$

where n is the number of OLSs that the LP passed along a particular path, the ASE and NLI over the specified path are both taken into account by the GSNR metric.

The simulation framework considers the EU network operating over a traditional C-band network. The traditional C-band has a total bandwidth of $\approx 4\,\text{THz}$, which has the potential of carrying 80 channels over a standard 50 GHz grid. The transceivers of the traditional C-band band operate at 32 Gbaud, shaped with a raised-root-cosine filter. The Erbium-doped-fiber amplifiers (EDFAs) considered for both networks are configured to operate in a constant output power mode with 0 dBm/channel. The network connections are assumed to work with standard Single-mode fiber (SMF) with a span of 80 km. The ILAs are considered to have a randomly selected noise figure for each amplifier in the 3.5 to 4.5 dB range, along with a random gain ripple with a 1 dB variation. The details of network simulation parameters are reported in Table I [21].

The scenario is modeled by generating synthetic datasets that abstract the physical layer using the open-source GNPy tool. With the help of the GNPy library, a full stack simulation environment of physical layer network models is simulated [18]. For a traditional C-band network, the generated dataset is a subset of 2^{80}, with 80 channels as the maximum possible realization of the spectral load. The traffic load utilization is 34% to 100% of the total bandwidth utilization.

IV. Performance Analysis and results

In this section, we evaluate the performance of the IL approach for QoT estimation. The QoT estimation metric for any given LP traversed by OLSs from any source to the destination is given by the GSNR metric, which incorporates the effect of ASE from the amplifier and fiber NLI (see Eq. 3).

$$\text{GSNR} = \frac{P_{\text{Rx}}}{P_{\text{ASE}} + P_{\text{NLI}}} = \left(\text{OSNR}^{-1} + \text{SNR}_{\text{NL}}^{-1}\right)^{-1} \tag{3}$$

where $\text{OSNR} = P_{\text{Rx}}/P_{\text{ASE}}$, $\text{SNR}_{\text{NL}} = P_{\text{Rx}}/P_{\text{NLI}}$, P_{Rx} is the signal power of the particular channel at the receiver, P_{ASE} is the power of the ASE noise and P_{NLI} is the power of the NLI. The GSNR accurately calculates the Bit error rate (BER)

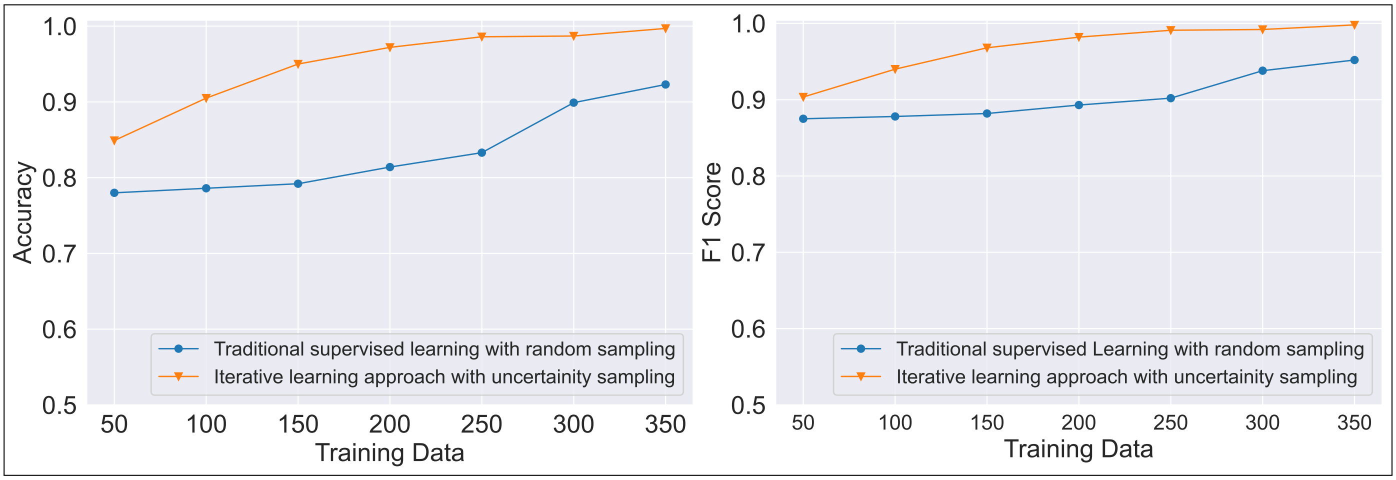


Fig. 3: Comparison of traditional and proposed IL approach in terms of classification accuracy and F1 Score.

by analyzing the transceiver's back-to-back characterization, as the BER has been thoroughly mentioned in several vendor demonstrations using commercial products [22]. The proposed scheme enables the operator to predict the status of upcoming deploying LP in terms of its QoT (GSNR). The proposed IL approach works as a binary classifier and makes the classification based on the GSNR estimation of LP (see Eq. 4). The $OSNR$ sensitivity threshold at the receiver considered in this analysis is based on [23]

$$\text{GSNR} > OSNR_{\text{Rx}} \longrightarrow 1 \ (otherwise\ 0) \tag{4}$$

The performance of the proposed approach is analyzed on 1000 samples of test data. The proposed model is evaluated using the accuracy and F1 score metric. The evaluation of the test dataset is repeated 100 times to obtain meaningful results for both accuracy and F1 score. We compare the performance of our proposed IL approach with the traditional supervised learning approach which adopts the random sampling method for training data selection. In the random sampling approach, data samples are chosen with the same probability, while in our approach samples are selected on the basis of uncertainty sampling based on entropy. The considered supervised learning model is initially trained on 50 labeled data samples. Both schemes are based on an iterative approach and a new data sample is added to the training dataset in each iteration. We run the IL cycles 50 times ($\mathcal{B}$=50). Fig. 3 plots the accuracy against the number of data samples for each scheme. We varied the number of samples from 50 to 350 to analyse the performance. The performance of both approaches is improved as we increase the number of training samples. As we can see that our proposed IL outperforms the traditional supervised learning approach using only the 50 new data samples. The accuracy of 99% is achieved with the IL approach considering 250 data samples, while for the same number of data samples supervised learning approach obtains 83% accuracy. To further assess the performance of our proposed approach, we plot the F1 score obtained against the training samples as shown in the Fig. 3. Our proposed approach is able to achieve a 95% F1 score with 150 data samples, whereas the traditional approach with random sampling obtains a 79% F1 score with the same number of samples. By observing the results, it is evident that IL based approach shows excellent results for the classification of LP into good or bad QoT.

V. Conclusions

In this work, we investigated a novel IL-based framework which utilizes the entropy-based uncertainty-sampling method to determine the most useful samples for training the machine learning model. This framework does not require any initial labeled dataset for training. It is demonstrated in the results that, without using any initially labeled dataset for initial training, with the reduced number of labeled training samples, IL based approach is able to achieve 99% accuracy in classifying the good or bad QoT of LPs before deployment.

References

[1] I. Khan, M. Bilal, and V. Curri, "Assessment of cross-train machine learning techniques for qot-estimation in agnostic optical networks," OSA Continuum **3**, 2690–2706 (2020).

[2] T. Panayiotou, S. P. Chatzis, and G. Ellinas, "Performance analysis of a data-driven quality-of-transmission decision approach on a dynamic multicast-capable metro optical network," J. Opt. Commun. Netw. **9**, 98–108 (2017).

[3] F. Usmani, I. Khan, M. Siddiqui, M. Khan, M. Bilal, M. U. Masood, A. Ahmad, M. Shahzad, and V. Curri, "Cross-feature trained machine learning models for QoT-estimation in optical networks," Optical Engineering **60**, 125106 (2021).

[4] F. Usmani, I. Khan, M. U. Masood, A. Ahmad, M. Shahzad, and V. Curri, "Convolutional neural network for quality of transmission prediction of unestablished lightpaths," Microwave and Optical Technology Letters **63**, 2461–2469 (2021).

[5] F. Usmani, I. Khan, M. U. Masood, A. Ahmad, M. Shahzad, and V. Curri, "Transfer learning aided qot computation in network operating with the 400zr standard," in *2022 International Conference on Optical Network Design and Modeling (ONDM),* (2022), pp. 1–6.

[6] K. Christodoulopoulos, P. Kokkinos, A. Di Giglio, A. Pagano, N. Argyris, C. Spatharakis, S. Dris, H. Avramopoulos, J.-C. Antona, C. Delezoide *et al.*, "Orchestra-optical performance monitoring enabling flexible networking," in *2015 17th International Conference on Transparent Optical Networks (ICTON),* (IEEE, 2015), pp. 1–4.

[7] D. A. Cohn, Z. Ghahramani, and M. I. Jordan, "Active learning with statistical models," (1996).

[8] H. Hino, "Active learning: Problem settings and recent developments," (2020).

[9] P. Ren, Y. Xiao, X. Chang, P.-Y. Huang, Z. Li, B. B. Gupta, X. Chen, and X. Wang, “A survey of deep active learning,” (2020).
[10] O. Sener and S. Savarese, “Active learning for convolutional neural networks: A core-set approach,” (2017).
[11] R. Mackowiak, P. Lenz, O. Ghori, F. Diego, O. Lange, and C. Rother, “Cereals - cost-effective region-based active learning for semantic segmentation,” (2018).
[12] C. Schröder and A. Niekler, “A survey of active learning for text classification using deep neural networks,” (2020).
[13] D. Azzimonti, C. Rottondi, and M. Tornatore, “Reducing probes for quality of transmission estimation in optical networks with active learning,” Journal of Optical Communications and Networking **12**, A38–A48 (2020).
[14] S. Yao, C.-W. Hsu, L. Kong, Q. Zhou, S. Shen, R. Zhang, S.-J. Su, Y. Alfadhli, and G.-K. Chang, “Data efficient estimation for quality of transmission through active learning in fiber-wireless integrated network,” Journal of Lightwave Technology **39**, 5691–5698 (2021).
[15] Z. Liu and D. Wu, “Unsupervised pool-based active learning for linear regression,” (2020).
[16] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg *et al.*, “Scikit-learn: Machine learning in python,” Journal of machine learning research **12**, 2825–2830 (2011).
[17] L. Breiman, “Random forests,” Machine learning **45**, 5–32 (2001).
[18] A. Ferrari, M. Filer, K. Balasubramanian, Y. Yin, E. Le Rouzic, J. Kundrat, G. Grammel, G. Galimberti, and V. Curri, “Gnpy: an open source application for physical layer aware open optical networks,” Journal of Optical Communications and Networking **12** (2020).
[19] A. Ahmad, A. Bianco, E. Bonetto, M. Garrich, and J. R. Oliveira, “Switching node architectures in flexible-grid networks: A performance comparison,” in *2014 International Conference on Optical Network Design and Modeling,* (2014), pp. 49–54.
[20] V. Curri, “Software-defined wdm optical transport in disaggregated open optical networks,” in *2020 22nd International Conference on Transparent Optical Networks (ICTON),* (IEEE, 2020), pp. 1–4.
[21] I. Khan, M. Bilal, M. U. Masood, A. D’Amico, and V. Curri, “Lightpath qot computation in optical networks assisted by transfer learning,” Journal of Optical Communications and Networking **13**, B72–B82 (2021).
[22] M. Filer, M. Cantono, A. Ferrari, G. Grammel, G. Galimberti, and V. Curri, “Multi-Vendor Experimental Validation of an Open Source QoT Estimator for Optical Networks,” JLT **36**, 3073–3082 (2018).
[23] “Cisco transceiver modules - cisco 400g digital coherent optics qsfp-dd optical modules data sheet,” (2021).