Performance Analysis of Transfer-learning Approaches for QoT Estimation of Network Operating with 400ZR

Performance Analysis of Transfer-learning Approaches for QoT Estimation of Network Operating with 400ZR

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 Umar Masood
Politecnico di Torino, IT
muhammad.masood@polito.it

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*—In the last decade, internet traffic has increased exponentially due to the expansion of bandwidth-intensive applications and the evolution of the concept of the internet of things. To sustain this growth in internet traffic, network operators insist on maximizing the utilization of already deployed network infrastructure to its maximum capacity to maximize the CAPEX. In this context, an accurate and earlier calculation of the Quality of transmission (QoT) of the lightpaths (LPs) is essential for minimizing the required margins that result from the uncertainty of the working point of network elements. This article presents a novel QoT-Estimation (QoT-E) framework assisted by Transfer-learning (TL). The main focus of this study is to present a detailed analysis of two major TL approaches, i.e., the Transfer-learning feature extraction (TLFE) approach and the Transfer-learning fine-tuning (TLFT) method, and demonstrate their effectiveness in minimizing the uncertainties in QoT-E in comparison with standard baseline models like Artificial neural network (ANN) and Convolutional-neural network (CNN). The Generalized signal-to-noise ratio (GSNR) is considered a characterizing parameter for the QoT of LP. The dataset utilized in this analysis is generated synthetically using the GNPy platform. Promising results are achieved by reducing the overall required margin and extracting the residual network capacity.**

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

I. Introduction

In recent years, there has been an enormous growth in internet traffic driven by the adoption of modern technologies such as 5G, cloud computing, and the Internet of Things. As a result, the core optical network is under pressure to meet the high capacity demands. Installing new infrastructure or enhancing the capacity of existing installed optical networks are typically the two options that can be used to address this problem. The first option requires a substantial CAPEX investment and is unfeasible from the operator's perspective. The alternate solution, however, is more practical because it may potentially boost the returns on already installed network infrastructure.

Currently, a major part of the existing optical transport networks employs Wavelength-division multiplexing (WDM) around a spectral window of $\approx$ 4 THz in the C-band. For further capacity enhancement, modern technologies such as Band-division multiplexing (BDM) are being adopted to utilize the residual capacity of already deployed WDM optical transport systems over the complete low-loss spectrum of optical fibers (e.g., $\approx$ 54 THz in ITU G.652.D fiber) [1]. In addition to this, employing modern technologies such as Software-defined networking (SDN) and Elastic optical networks (EONs) allows for the efficient use of this WDM spectrum with the dynamic and adaptive provisioning of network resources. These two technologies facilitate the optical networks to evolve towards fully disaggregated networks. The important step in network disaggregation is to carefully consider the Optical line systems (OLSs) that include fibers and amplifiers to interconnect the network nodes.

The QoT degradation is eventually determined by the OLS controller, which operates in the control plane, and is responsible for deciding the amplifier operating point. The nominal operating point needs to be carefully defined to improve reliance on the QoT degradation. The accurate determination of these parameters eventually leads to deploying a lower margin in LP and allows for better utilization of network resources with higher deployed traffic rates. Therefore, the precise estimation of the QoT of LP before its deployment becomes a key factor for effective resource utilization and high performance of optical networks. Various supervised Machine learning (ML) models have been explored recently as an alternative solution to conventional QoT estimation approaches in order to provide fast and precise QoT estimation [2]–[5]. Most of these ML-based approaches require a fair amount of dataset for training with the assumption that both training and testing datasets possess the same distribution and feature space. In the case of optical networks, each system is deployed with different configuration settings, and the traditional ML model trained on one system will unlikely perform well on another system. Therefore, a new ML model needs to be developed for specific system settings, which can be adaptive and require a fair amount of data samples for retraining. In such scenarios, TL emerged as a promising solution to achieve high performance by utilizing the learned knowledge of an already trained ML model with a reduced number of

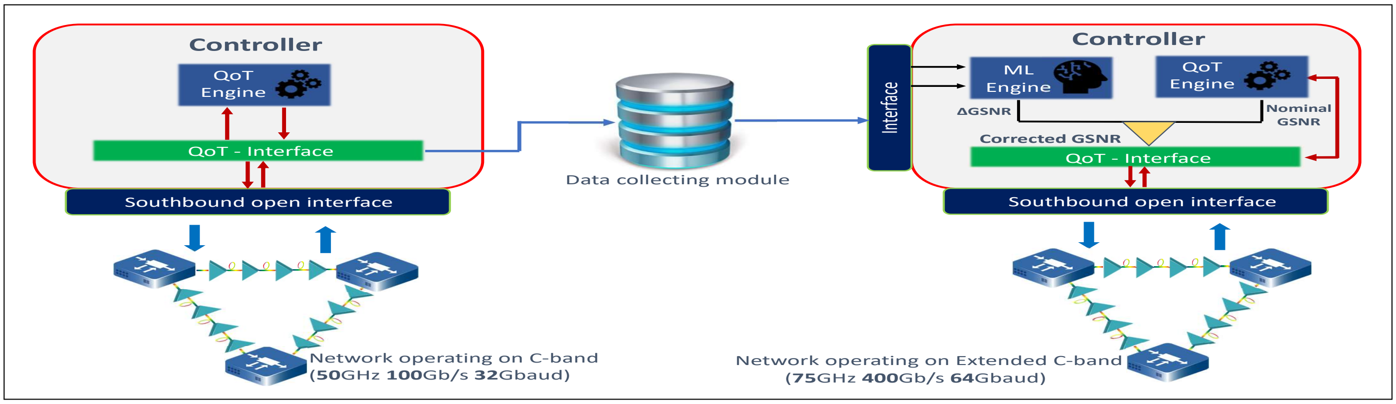

Fig. 1: Transfer-learning schematic utilizing traditional C-band knowledge to assist QoT-E engine of extended C-band network.

samples and training time. The TL applies the concept of generalization; a model trained on one problem can be used for another related problem without starting to learn from scratch. The model begins learning from the already learned knowledge.

In this paper, we extend our previous work done in [6], wherein we proposed a TL-assisted framework for an extended C-band network to reduce the uncertainties in GSNR computation. The GSNR metric is used to assess the QoT estimation in the given scenario well. The GSNR uncertainties are typically introduced by the amplifier's gain ripples, noise figure, and variation of spectral load. In this paper, we try to improve the performance of our already proposed method in [6]; Transfer-learning partial tuning (TLPT). The main focus of this study is to present a detailed analysis of two different TL approaches, i.e. TLFE approach and the TLFT approach and demonstrate their effectiveness in correcting GSNR uncertainties in a newly deployed extended C-band network. Furthermore, the performance of these two new TL approaches is compared with standard baseline models, ANN and CNN, trained from scratch.

II. Simulation Framework and Dataset Analysis

This study considers a software-defined optical network with re-configurable-optical add-drop-multiplexers (ROADMs) as the nodes and OLS as the edges [7]. The OLS under consideration is assumed to operate at their optimal operating point, and the only factor accounting for the perturbation behaviour of the physical layer is the noise figure and amplifier's ripple gain. The gain-ripples oscillate in response to variations in the spectral load. With some degree of working point uncertainty, OLS controllers can thus guarantee that they operate at the nominal operating point. The LPs are transparently deployed on the WDM flexible grid system on the lower layer, linking the transceivers and enabling dual-polarization multilevel modulation formats [8]. Multiple impairments impact the LPs during their transmission, but amplified-spontaneous noise-(ASE) and non-linear impairments are the prominent ones (NLI). Each in-line amplifier (ILA) introduces statistically independent ASE noise, accumulating throughout propagation. The NLI of every span, however, is statistically associated with one another [9].. The following equation gives the total $GSNR$ of each LP propagating across the OLS:

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

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

The simulation setup considers the conventional C-band network (referred to as the EU network) and the extended C-band (USA network). Although the two networks under analysis have different topologies, they employ identical amplifiers and fiber types.

The conventional C-band has a total bandwidth of approximately $\approx 4\,$THz, enabling it to transport 80 channels over a typical 50 GHz grid. In contrast, the extended C-band has a total bandwidth of approximately $\approx 4.8\,$THz, allowing it to carry 64 channels over a 75 GHz grid. Traditional C-band and extended C-band transceivers are configured with a raised-root-cosine filter and run at 32 Gbaud and 64 Gbaud, respectively. The Erbium-doped fiber amplifiers (EDFAs) taken into account for both networks are set up to function in a fixed output power mode with 0 dBm/channel. It is envisaged that both networks' connections will function with standard single-mode fiber (SMF) with an 80 km reach. The ILAs in both networks are assumed to have a random gain ripple with a 1 dB variance and a noise figure for each amplifier in the range of 3.5 to 4.5 dB. In Table I, the specifics of the network simulation parameters are reported.

An open-source GNPy package is used to mimic the described scenario and generate synthetic datasets by providing the abstraction of the physical layer [10]. The GNPy library builds the network models for the physical layer using an end-to-end simulated environment. The datasets are collected for the extended C-band network (USA network) and the C-band network (EU network). The detailed description of both considered network is provided in [6]. The acquired dataset for the conventional C-band network is a subset of 2^{80}, with 80 channels being the maximum number of ways the spectral load might be realized. In contrast, the generated dataset for an expanded C-band network is 2^{64}, with 64 operating channels. The difference in traffic load and overall

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
Channel Spacing (Extended C-Band)	75 GHz
Span Length	80 km
WDM Comb (C-Band)	80
WDM Comb (Extended C-Band)	64
Baud Rate (C-Band)	32 $Gbaud$
Baud Rate (Extended C-Band)	64 $Gbaud$
Amplifier Noise Figure	[3.5 - 4.5] dB [10]
Nominal Amplifier Noise Figure	4 dB
Amplifier Gain Ripple	Variation of 1 dB
Nominal Amplifier Gain Ripple	Flat
Fiber Type	Standard SMF

bandwidth consumption for the two networks varies from 34% to 100%. The schematic diagram of the TL approach is shown in Fig. 1. The proposed approach uses the EU network's learned knowledge to train the TL-agent application program interface (API) that is incorporated with the core engine of the QoT estimator engine running in the network controller of the USA network. The core-QoT estimation engine uses nominal parameters to calculate the LP GSNR. The QoT-estimator engine calculates the GSNR using these nominal parameters provided by the vendors. However, there is uncertainty due to the change in the network elements' working point during its operation. Due to the uncertainty in the GSNR prediction, a margin is required, which lowers the deployable traffic rate and leads to the under-utilization of network resources. In the described framework, the core QoT-estimator engine of another newly installed network (USA-network in this scenario) is assisted by a TL agent trained on the dataset from an already operational network (EU-network). This work mainly aims to reduce the uncertainty caused by noise figures and ripple gain in amplifiers.

III. Transfer-learning Approaches for QoT Estimation

The proposed TL scheme relies on an already well-trained ML model and its weights, which have been trained on a sufficiently large dataset. Exploiting an already well-trained model minimizes the overhead of developing a model from scratch, which needs a fair amount of dataset and time. This study considers three supervised TL approaches and two baseline models. The proposed TL approaches are based on the ANN model and are well-trained on the large dataset obtained from the EU network operating on the C-band. In these approaches, the feature space used to characterize the GSNR of LP consists of power, NLI, ASE noise, and the number of spans. The concept of TL is illustrated in the Fig. 2. The proposed TL approaches and baseline models are developed by utilizing high-level Keras API running on top of the TensorFlow platform. A brief description of these approaches is given below.

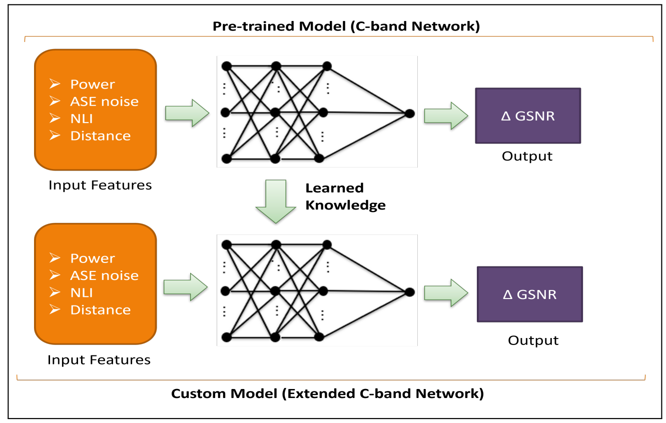

Fig. 2: Transfer Learning Methodology.

A. Transfer-learning Partial Tuning (TLPT)

This approach has already been implemented in our previous work [6]. The model utilized a pre-trained ANN model (trained on a C-band network) and preserved the weights of top layers as the knowledge learned from these layers. Two new layers have been added to this model, and the newly added layers were merged along with the already present layers of the pre-trained model. The newly added layers enable the retraining of the model with a small number of samples obtained from the extended C-band network and thus enable it to work in another ecosystem (extended C-band network). The presented approach is referred to as partial tuning because the retraining is not performed on the entire model; instead, the model is partially tuned to adjust the weights.

B. Transfer-learning Feature Extraction (TLFE)

In this approach, we preserved the weights of the entire pre-trained model developed in [6]. We added two new layers, each with 500 neurons on top of the pre-trained model. In order to avoid the vanishing gradient problem, a ReLU-based activation function is employed in each hidden layer and linear-activation function is used at output layer. A dropout layer with a rate of 0.20 is inserted to discard 20% of the neurons to stop co-adaptive learning in each iteration and prevent overfitting. The last layer of a pre-trained model is more specific to the C-band network prediction task on which the model is trained; we also replaced it with a newly added fully-connected layer. We trained newly added layers from scratch on top of the pre-trained model to leverage the knowledge learned previously on the dataset acquired from the C-band network. The training is performed by employing an Adaptive-moment-estimation (Adam) optimizer, with a learning rate of 0.01 on a minimal number of samples acquired from the extended C-band network. The retraining of the entire model is not required in this approach. The already trained model contains the generic features that are helpful in extracting meaningful insights from the new samples obtained from the extended C-band network.

C. Transfer-learning Fine Tuning (TLFT)

In this approach, the initial pre-trained model developed in the TLFE approach is duplicated, and weights are fine-

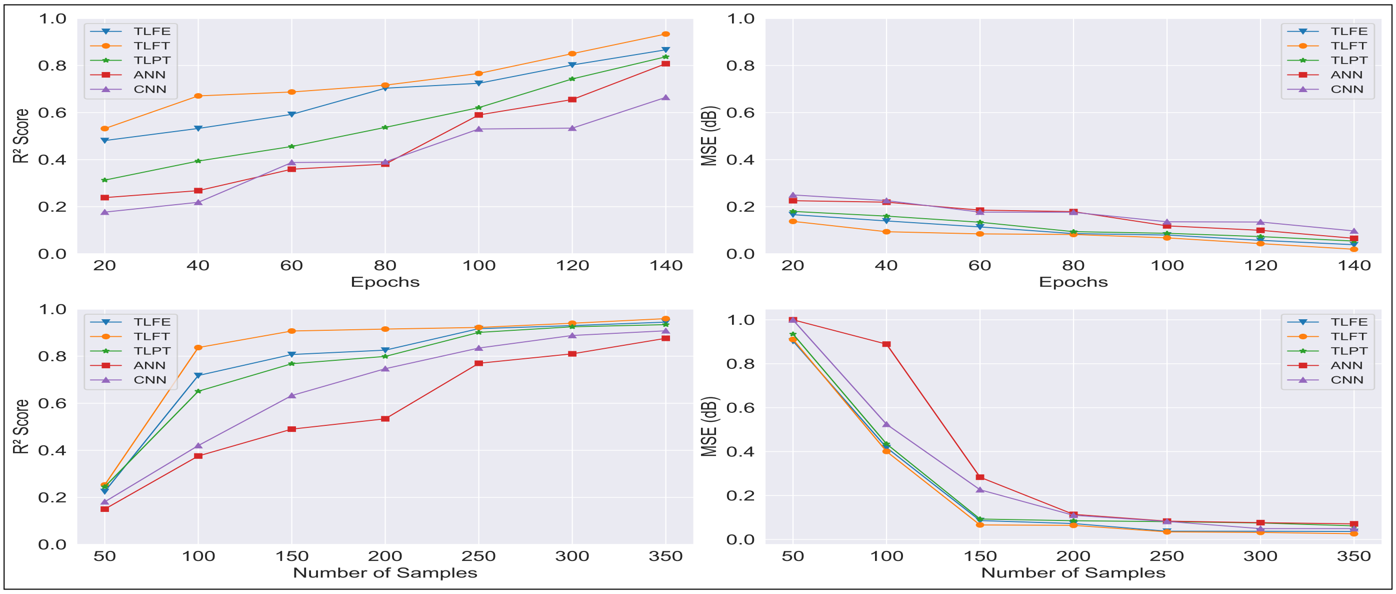

Fig. 3: Comparison of transfer learning schemes with baseline models.

tuned entirely to develop a TL model for an extended C-band network. All the model layers unfreeze and optimize the shared weights to make them more appropriate for performing well on the new network (extended C-band); a low learning rate (0.001) is used to retain some of the cognitive information learned previously. The retraining of the entire model is performed with the adam optimizer on a small dataset of extended C-band network, and weights are adapted from their initial values to reduce the prediction uncertainty on the extended C-band network.

D. Baseline Models

Two baseline models, i.e., CNN and ANN, are selected for comparison to better evaluate the performances of the above TL approaches. Both baseline models are trained from scratch using the Adam optimizer on the dataset acquired from the extended C-band network.

- Convolutional Neural Network (CNN): The proposed CNN model contains the feature-extraction and regression units for end-to-end training. Before activating the feature-extraction unit, z-score normalization is applied to the dataset obtained from an extended C-band network. After that, the normalized dataset is transformed from 2-dimension to 3-dimension to get suitable 3-dimensional data. Next, the normalized data is forwarded to the feature-extraction unit, which has two convolutional layers and an input layer. The convolutional layers are used as a feature extractor to capture the important characteristics of the input data for better prediction. Additionally, each convolutional layer uses the ReLU activation function to speed up training. Furthermore, the average-pooling layers are inserted between subsequent convolutional layers to perform spatial pooling. The output of the feature-extraction unit is transformed into a 1-dimensional array by employing the flattening layer, which is then passed to the regression unit for prediction.
- Artificial Neural Network (ANN): We developed the ANN model consisting of an input layer with 257 neurons, two hidden layers each with 500 neurons, and an output layer with one neuron. An activation function based on ReLU is utilized in each hidden layer, whereas a linear activation function is employed at output layer. In addition, a dropout layer with a 0.20 rate is placed to discard 20% of the random neurons to prevent the over-fitting problem.

IV. Performance Analysis

In this section, we evaluate the performance of the GSNR predictions in extended C-band networks both with and without employing TL approaches (discussed in Section III), using the R^2-score metric and mean squared error (MSE) defined in Eq. 2 and Eq. 3.

$$R^2 = 1 - \frac{\sum_i^n \left(\mathrm{y}_i^{\text{predicted}} - \mathrm{y}_i\right)^2}{\sum_i^n \left(\overline{y}_i - \mathrm{y}_i\right)^2} \tag{2}$$

$$MSE = \frac{1}{n}\sum_i^n \left(\Delta\mathrm{GSNR}_i^{\text{predicted}} - \Delta\mathrm{GSNR}_i^{\text{actual}}\right)^2 \tag{3}$$

where n is the total number of test samples, $\mathrm{y_i}^{\text{predicted}}$ represents the predicted values, $\mathrm{y_i}$ collectively represents the values in the dataset and $\overline{y}_i = 1/n \sum_i^n y_i$. In addition to this, $\Delta\mathrm{GSNR}^{\text{predicted}} = \mathrm{GSNR}^{\text{nominal}} - \mathrm{GSNR}^{\text{predicted}}$, while $\Delta\mathrm{GSNR}^{\text{actual}} = \mathrm{GSNR}^{\text{nominal}} - \mathrm{GSNR}^{\text{actual}}$.

The R^2-score measures the proportion of variance in the ML model's predictions compared to the overall variance in test data. The R^2 metric computes the values in the range of 0 and 1. If the R^2 score value is 1, it indicates that the ML model is trained perfectly, while a 0 value indicates that the model

TABLE II: Detailed performance analysis of transfer learning schemes.

Paths	Transfer learning			Without Transfer learning	
	TLPT MSE (dB)	**TLFE MSE (dB)**	**TLFT MSE (dB)**	**ANN MSE (dB)**	**CNN MSE (dB)**
Birmingham → Bismarck	0.079	0.064	0.041	0.084	0.081
Bismarck → Boston	0.079	0.062	0.053	0.251	0.133
Boston → Buffalo	0.085	0.071	0.063	0.122	0.095
Charlotte → Chicago	0.098	0.074	0.043	0.213	0.171
Cleveland → Columbus	0.083	0.061	0.042	0.173	0.131
Dallas → Denver	0.083	0.041	0.033	0.099	0.091
Detroit → ElPaso	0.083	0.037	0.031	0.098	0.092
ElPaso → Fresno	0.065	0.038	0.032	0.081	0.071
Greensboro → Hartford	0.094	0.062	0.021	0.12	0.099

won't perform well on an unseen dataset. The MSE is used to measure how precise the model predictions are. We consider 12000 testing data samples obtained from an extended C-band network to assess the performance of all the discussed approaches in Section III. The upper two graphs in Fig. 3 plot the R^2 score and MSE (computed on the test dataset) against the number of epochs. Firstly we fixed the number of samples to 175 and then varied the number of epochs from 20 to 140.

As we can see that all TL approaches perform well as compared to baseline models (CNN and ANN) even on a minimal number of epochs. The TLFT approach outperforms all the other approaches with an R^2 score value of 0.93 with 140 epochs. It illustrates that TLFT can capture 93% of the variance in the dataset, while the TLFE and TLPT capture 86% and 83% variance, respectively. In the MSE graph, in the beginning, all TL approaches obtained similar MSE. As the number of epochs increases, the value of MSE also improves for all TL approaches. Again, TLFT obtains the lowest MSE with a value of 0.01 dB compared to TLFE with a value of 0.03 dB and TLPT with 0.09 dB. Both the baseline models do not perform well for GSNR prediction on an extended C-band network. To further assess the performance of these approaches, we plotted the R^2 score and MSE against the number of samples (for 140 epochs), as demonstrated in the lower two graphs of Fig.3. It is shown that all our proposed TL approaches perform very well as compared to baseline models (trained from scratch). With few training epochs and a limited number of samples (175 samples in the given scenario), TL approaches can obtain high accuracy in predicting the GSNR in an extended C-band network compared to conventional re-training approaches.

To compare the performance of TL approaches proposed in this paper with our previous work in [6], we reported both the results of TL approaches and without TL approaches on the nine paths of extended C-band (USA network) as shown in Table II. The two new approaches, TLFE and TLFT significantly reduce the MSE in GSNR estimation compared to TLPT proposed previously, against each path of the extended C-band network. By noticing the MSE values against each path, it is clear that the TL approach outperforms the baseline approaches (Without TL). The TLFT approach is excellent for correcting the GSNR uncertainty in an extended C-band network.

V. Conclusions

In this work, we investigated two TL approaches, TLFE and TLFT, to train the data-driven agent on the dataset acquired from the traditional C-band network to reduce the uncertainty in the GSNR prediction of an extended C-band network operating with 400ZR. By utilizing the previously learned knowledge and slowly adjusting the weights of the entire model in the TLFT approach, we can achieve the best performance with a small number of training samples and a reduced number of epochs. It is indicated in the results that both approaches have the potential to perform well to reduce the uncertainties in QoT estimation of an LP in an extended C-band network, utilizing the learned knowledge from the traditional C-band network.

References

[1] A. Ferrari, A. Napoli, J. K. Fischer, N. Costa, A. D'Amico, J. ao Pedro, W. Forysiak, E. Pincemin, A. Lord, A. Stavdas, J. P. F.-P. Gimenez, G. Roelkens, N. Calabretta, S. Abrate, B. Sommerkorn-Krombholz, and V. Curri, "Assessment on the achievable throughput of multi-band ITU-T G.652.D fiber transmission systems," Journal of Lightwave Technology **38** (2020).

[2] Y. Pointurier, "Machine learning techniques for quality of transmission estimation in optical networks," J. Opt. Commun. Netw. **13**, B60–B71 (2021).

[3] 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).

[4] 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).

[5] 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).

[6] 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.

[7] 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.

[8] https://www.itu.int/rec/T-REC-G.694.1/en.

[9] 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.

[10] 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).