Cross-feature trained machine learning models for QoT-estimation in optical networks

Fehmida Usmani,[a] Ihtesham Khan,[b] Mehek Siddiqui,[a] Mahnoor Khan,[a] Muhammad Bilal,[b] M. Umar Masood,[b] Arsalan Ahmad,[a,*] Muhammad Shahzad,[a] and Vittorio Curri[b]

[a]National University of Sciences and Technology, Islamabad, Pakistan
[b]DET, Politecnico di Torino, Corso Duca degli Abruzzi, Turin, Italy

Abstract. The ever-increasing demand for global internet traffic, together with evolving concepts of software-defined networks and elastic-optical-networks, demand not only the total capacity utilization of underlying infrastructure but also a dynamic, flexible, and transparent optical network. In general, worst-case assumptions are utilized to calculate the quality of transmission (QoT) with provisioning of high-margin requirements. Thus, precise estimation of the QoT for the lightpath (LP) establishment is crucial for reducing the provisioning margins. We propose and compare several data-driven machine learning (ML) models to make an accurate calculation of the QoT before the actual establishment of the LP in an unseen network. The proposed models are trained on the data acquired from an already established LP of a completely different network. The metric considered to evaluate the QoT of the LP is the generalized signal-to-noise ratio (GSNR), which accumulates the impact of both nonlinear interference and amplified spontaneous emission noise. The dataset is generated synthetically using a well-tested GNPy simulation tool. Promising results are achieved, showing that the proposed neural network considerably minimizes the GSNR uncertainty and, consequently, the provisioning margin. Furthermore, we also analyze the impact of cross-features and relevant features training on the proposed ML models' performance. © *2021 Society of Photo-Optical Instrumentation Engineers (SPIE)* [DOI: 10.1117/1.OE.60.12.125106]

Keywords: machine learning; quality of transmission estimation; generalized SNR.

Paper 20210444 received May 3, 2021; accepted for publication Oct. 18, 2021; published online Dec. 3, 2021.

1 Introduction

In the last decades, optical transmission systems revolutionized the telecommunication industry to meet the rapidly increasing global internet traffic demands. In the last few years, internet traffic has been growing continuously[1] with the evolution of new technologies and bandwidth-intensive applications, such as video-on-demand, full high definition or 4K, cloud computing, and the Internet of Things. This increasing trend of global internet traffic requires the maximum utilization of the remaining capacity of the already working network infrastructure. In this direction, the fundamental key enabler technologies are coherent technology for optical transmission and DWDM for spectral usage of fiber propagation. In addition to these, the network disaggregation paves a path for the technologies such as elastic optical networks (EON) and software-defined network (SDN). The distinctive features of SDN and EON offer versatile and dynamic resource provisioning in optical networks for both the control and data planes.[2,3] EONs provide flexible spectral assignment in the data plane and boost the network's capacity while lowering the network cost. This adaptability results in much more complex lightpath (LP) provisioning than conventional fixed-grid wavelength-division multiplexing (WDM) networks. On the other hand, the SDN controller in the control plane handles the operating points of different network components independently, which consequently enables the customized network management.

*Address all correspondence to Arsalan Ahmad, arsalan.ahmad@seecs.edu.pk

Today's optical networks have started evolving to partial dis-aggregation, with a goal of full dis-aggregation eventually. The prime step toward network dis-aggregation is the consideration of optical-line systems (OLSs), which link the network nodes. In the current reference frame, the capacity of OLS controllers to function at the optimum working point determines the quality of transmission (QoT) degradation.[4,5] The precise accomplishment of this working point leads to the achievement of a lower margin and higher traffic rates for traffic deployment. Therefore, it is compulsory to utilize the QoT-E for accurate estimation of LP performance—the path computation—before its deployment. In the current context, QoT is effectively evaluated by the generalized signal-to-noise ratio (GSNR), which comprises the accumulated effect of nonlinear interference (NLI) and amplified spontaneous emission (ASE) noise.[6] Exploiting the transceiver characteristics, the GSNR describes the feasibility of the path as well as deployable rate. Typically, the network elements suffer a variation on their working point (insertion losses, noise and gain figure, ripples in amplifiers, etc.). This implies uncertainty in the QoT-E that needs a system margin to avoid out-of-services.

In the present investigation, we suppose a domain adaption (DA) approach. The DA approach uses only the available data from source domain "S" (e.g., well-deployed in-service network), in which the network operator has sufficient knowledge about the working point of network nodes and provides useful information related to the QoT prediction in the target domain "T" (e.g., a newly deployed or unseen network), i.e., the kind of a network in which the system administrator does not have sufficient information of the operating point of network components. This work aims to minimize the margin in the GSNR estimation of the target domain. This decrease in the GSNR uncertainty enables the network controller in the target domain to accurately establish the LP with a reduced margin. In general, the controller can acquire an accurate description of the system parameters, i.e., network status. The QoT-E exploits several analytical approaches that can measure the GSNR with a very good precision as shown in Ref. 7. The use of an analytical approach is not reliable without the exact knowledge of system parameters, as it is essential to acquire system parameters in the current context of DA. The present work-frame regarding DA concludes that the use of an analytic approach is not recommended in such an agnostic scenario to estimate the QoT of the LP prior to its deployment.

To overcome this challenge, we opted to use the data-driven ML approach as an alternative way, which has already been proved to be very effective in several different contexts of managing optical networks; see Refs. 8–11 for performance monitoring of optical network operations. A comprehensive assessment of ML practiced in optical networks was described in Ref. 12. In particular, moving toward a distinct interest of this investigation, i.e., estimation of the QoT-E of the LP before its actual deployment, a few very effective ML-based approaches, for instance, the cognitive case-based reasoning method, were presented in Ref. 13. The experimental results related to Ref. 13 obtained with real field data were described in Ref. 14. In Ref. 15, the ML-based approach was presented to control OLS in an open environment. An approach based on random forest (RF) was presented to utilize the already accumulated database in Ref. 16 to decrease unreliability in design margins and network parameters. A neural network (NN) model was trained to measure the Q-factor for a multicast communication scenario in Refs. 17–19. Numerous ML-based approaches were proposed in Refs. 20 and 21 for QoT-estimation of the LP. In Ref. 22, a binary classifier based on RF was proposed to estimate the bit-error-ratio (BER) of LPs before their establishment. The authors in Ref. 23 trained three classifiers, i.e., RF, support vector machine (SVM), and K-nearest neighbor (KNN) for QoT estimation. The performance of all of these approaches is evaluated in terms of accuracy. Furthermore, the investigation presented in Ref. 23 concluded that the SVM shows good accuracy but performs worst in terms of computation time. The authors in Ref. 24 used NN to characterize the integrated circuits consequently used for their full and accurate softwarization. In Ref. 25, the authors evaluated the performance of two DA approaches for ML-assisted QoT-E of an optical LP for a fixed/variable number of available training samples from the source/target domain. The authors studied two networks characterized by different topologies but utilizing identical fiber type and communication devices and assessed the performance of two DA approaches depending on the number of available train realizations from the target domain. The results in Ref. 25 stated that the DA-based approach worked better than standard ML techniques. The authors in Ref. 26 not only proposed ML for QoT-E but also reported the statistical closed-form approach to the QoT margin setting. In Ref. 27, a CNN-based QoT estimator was proposed for unestablished LPs for the DA scenario. Finally, the authors in Ref. 28 analyzed the QoT-E accuracy delivered by a few active learning (AL) and DA methods on two different network topologies. The results presented in Ref. 28 announced significant improvements using an AL approach with some extra samples acquired from the target domain.

The notable difference between the literature and the present work is that we propose several ML techniques considering the DA approach for the system margin minimization of the T network using the mimicked data of GSNRs response to specific traffic configurations of LPs of the S network in an open environment. We also evaluate and compare ML models' performance using cross-feature training and relevant features training approaches for GSNR prediction. The generation of the dataset is achieved synthetically by perturbing the nominal working point of network components. In the present practice, the synthetic dataset is created against two different networks characterized by different topologies utilizing an identical fiber type and communication devices, but they are different in terms of the most delicate parameters of amplifiers, i.e., noise figure, amplifier ripple gain, and fiber insertion losses. Regarding these two companion networks, the first one is the S network, a type of network in which the operator has a complete description of the network elements' operational parameters. The other network is the T network, which is intrinsically a type of network in which the operator has only a basic description of the operational parameters of the network elements.

The rest of the paper is organized as follows: Sec. 2 briefly explains the physical layer's abstraction to efficiently implement a multilayer optimization, simultaneously with the argument that an accurate QoT-E has a fundamental role in minimizing the system margin. Further, we also propose several potential methods to obtain information about OLS attributes, each providing a diverse decline of the GSNR uncertainty. Section 3 explains the background of ML techniques used for QoT estimation of the unestablished LP. In Sec. 4, the simulation conducted to model an open OLS composed of cascaded amplifiers and fibers is described. The data generation and the technique used for refining the dataset before applying to the ML models are reported. The dataset is generated synthetically against two different networks using the open-source GNPy simulator. The two mimicked datasets are perturbed by varying erbium-doped fiber amplifiers (EDFA) noise figure, ripple gain, and insertion losses. In Sec. 5, we reported the configuration parameters for the proposed ML techniques, which are used in the context of the DA approach by exploiting the dataset of the already well-deployed S network. The proposed ML techniques predict the GSNRs of the LPs of the T network before its actual deployment with significant accuracy. Moreover, we also define the characterization of features and labels of the proposed ML models and the metric used to evaluate them. Later, in Sec. 6, we produce detailed results. Finally, the conclusion and future research work are discussed in Sec. 7.

2 Overview of Optical Transport Network

In general, an optical network consists of optical network elements (ONE) connected through bidirectionally fiber links, in which traffic demand is added/dropped or routed, as shown in Fig. 1(a). The amplifiers are placed after a specific span length using the EDFAs technique/Raman amplification or optionally used in combination. In the present state-of-the-art optical network, ONE connected through fibers are generally expressed as an OLS and a particular controller with the specific characteristic feature to configure the operating point of every amplifier traversing through the link, and the spectral load is provided at the input of every fiber span. In addition, the transport layer adding/dropping or routing services is delivered using reconfigurable optical add/drop multiplexers (ROADM). The DWDM, spectral usage technology according to the ITU-T recommendations,[29] can exploit either the fixed or flexible spectral grid that characterizes the spectral slots for both grid architectures.[30,31] Utilizing either grid architecture, LPs are deployed; LPs are the logical abstraction of feasible links between node-to-node given the traffic demands. Furthermore, against every deployed LP, a polarization-division-multiplexing is exploit to propagate from a particular source to its specific destination. Along with the transmission, the LP suffers different propagation impairments such as amplifier noise added as an ASE, fiber propagation, and filtering penalties applied by ROADM. Also, it has been

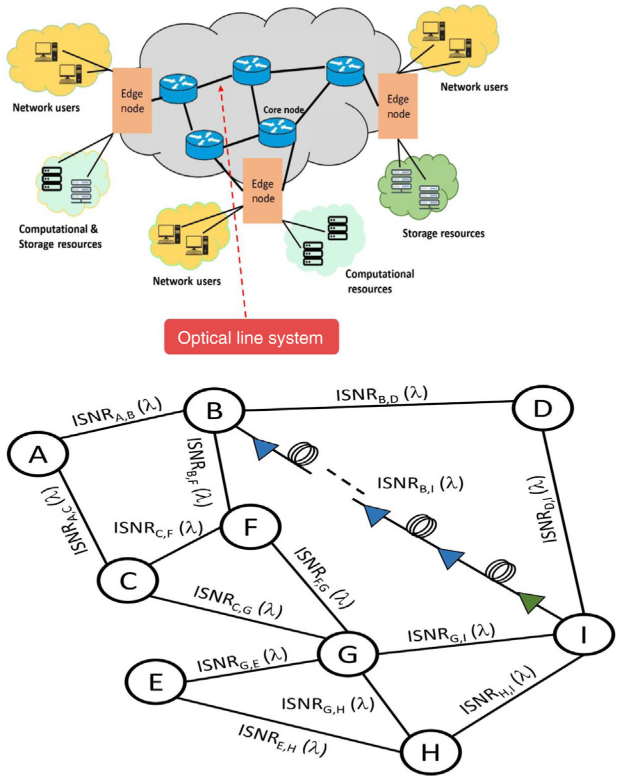

Fig. 1 (a) Architecture of optical network and (b) optical transport network.

extensively demonstrated that the fiber propagation on an uncompensated optical coherent transmission system impairs the QoT of deployed LPs by introducing amplitude and phase noise.[4,32–34] This introduced phase noise is efficiently counterbalanced by the receiver's DSP module, using a carrier phase estimator algorithm. This particular set of noise can only be considered for very high symbol rate communication designed for short distance.[34] By contrast, the amplitude noise, typically described as the NLI, always impairs the performance. It is a Gaussian disturbance that accumulates with the receiver's ASE noise. Finally, the ROADMs filtering penalty also decreases the QoT level, which is generally estimated as an extra loss.

2.1 *QoT-Estimation Metric*

The QoT-E metric for a particular LP routed by definite OLSs from the source node to the destination node is given by the well-acknowledged GSNR measurement, which combines both the aggregated effect of ASE noise and NLI disturbance. In general, GSNR is defined as

$$\mathrm{GSNR} = \frac{P_{\mathrm{Rx}}}{P_{\mathrm{ASE}} + P_{\mathrm{NLI}}} = (\mathrm{OSNR}^{-1} + \mathrm{SNR}_{\mathrm{NL}}^{-1})^{-1}, \tag{1}$$

where $\mathrm{OSNR} = P_{\mathrm{Rx}}/P_{\mathrm{ASE}}$, $\mathrm{SNR}_{\mathrm{NL}} = P_{\mathrm{Rx}}/P_{\mathrm{NLI}}$, 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. Analyzing the transceiver's back-to-back characterization, the GSNR accurately provides the BER, as BER has been extensively stated in different vendor demonstrations with the use of industrial products.[6] The nonlinear effects P_{NLI} are generated during the fiber propagation, which relies on the spectral-load and the power of the distinct channel.[4] In these circumstances, it is fairly clear that there is an optimal spectral load for each specific OLS that maximizes the GSNR.[5] Examining the LP propagation effects against a specific pair of source and destination, we provide an abstract view of the operation as a combined impact of every single ONE that adds up to the QoT impairments. For any given LP request between a specific pair of source and destination, the total impairments are cumulative of the previously traversed OLSs and

ROADM losses.[35] Each crossed OLS adds a specific amount of NLI and ASE noise. For the purpose of the QoT, the abstraction of OLS can be achieved by a single parameter known as SNR degradation, which generally depends on the frequency ($\mathrm{GSNR}_i(f)$), if the OLS controllers can keep the OLS running at the ideal operation point. Therefore, an optical network can be generally abstracted as a weighted graph (W), where W = (vertices (V), edges (E)) corresponds to the specific networking topology. The V represents ROADM network nodes, and the E represents OLSs having $\mathrm{GSNR}_i(f)$ as weights on the consequent edges, shown in Fig. 1(b). Specifically, for a given LP from source node I to destination node F that passes through intermediary nodes B, the QoT is

$$\mathrm{GSNR}_{\mathrm{IF}}^{-1}(f) = \mathrm{GSNR}_{\mathrm{IB}}^{-1}(f) + \mathrm{GSNR}_{\mathrm{BF}}^{-1}(f). \tag{2}$$

Following network level abstraction, LPs deployment can be feasible for a specific source node to the destination node with the reduced margin, which relies on the GSNR of a particular source to the destination path.

2.2 *Methods for QoT Estimation*

This section describes the various possible approaches for acquiring knowledge about the characteristics of an OLS, with each allowing the distinct GSNR measurement. In the initial approach, the data obtained from ONE, for example, static description of elements (e.g., connector loss, noise figure amplifier gain, etc.), are utilized to achieve precise QoT-E vendor-specific systems. Concerning this specific method, several analytical approaches are available to assess the GSNR and characterize the OLS elements. Nevertheless, this strategy, based on static data, may not be reliable as the ONE experiences continuous performance degradation owing to the aging effect, heading to gradually unreliable QoT-E after a specific period.

The next method is utilizing the telemetry data to examine the network status instantly. Assuming an agnostic operating of OLS in an open environment, the controller of OLS largely relies on the telemetry data from the EDFAs and the optical channel monitor. This specific technique is feasible for an accurate QoT prediction by utilizing the network's current state's telemetry. In contrast to the former method, this technique does not depend on the static parameters of ONE. Thus, it eliminates the unreliability in the QoT-E precision introduced because of device aging factor as discussed in the earlier technique. However, this unique technique dilemma is that the response of GSNR, particularly the OSNR part, significantly relying on the configuration of spectral-load, leads to substantial unreliability in the QoT margin.[15]

The final method examines the dataset that obtains the QoT responses against arbitrary spectral-loads of the S network. As mentioned earlier, the generation of the dataset is performed during the operating period of the S network by estimating the OLS response with regard to GSNR for numerous spectral-load arrangements. This specific case comes up with a perfect playing field to employ ML. An ML technique utilizes a dataset of spectral-load samples from a S network for training to produces an accurate QoT-E for every generated spectral load region of a T network. Distinct from the former method, in which just telemetry data are explored, this procedure employs the QoT-E centered on the GSNR reaction to particular spectral-load arrangements of the S network, used for an accurate GSNR prediction of T network. In addition, this arrangement does not require any information about physical parameters of the OLS as compared with the first technique. Therefore, this approach gives an excellent playground to utilize the ML-DA method. For this, we focus on the third procedure, which is based on the ML method. This approach uses the GSNR related to the individual spectral-load configurations of the previously established S network for its training and predicts the QoT of the T network.

3 Background on Machine Learning Models

This section briefly explains the ML techniques that we applied for QoT estimation of an unestablished LP. In general, ML has a wide range of applications in optical communications and networking.[36] The ML model learns from previous knowledge of the network and then uses that learned knowledge to make predictions. Recently, QoT prediction of an unestablished LP with ML models has gained a lot of attention.[12,15,16] In this work, six ML models are employed to estimate the QoT of an unestablished LP, and domain adaptation (DA) capability of these models is assessed. In the following, we briefly present a short overview of these employed ML models.

3.1 *Decision Trees*

We propose using the decision tree (DT) model to assess the feasibility of an unestablished LP in the DA scenario, i.e., transferring the source data distribution learned from a known network to another related target network with a different distribution. DT constructs a tree based on the decisions made by exploring dataset features in different aspects. It has three essential parameters: the maximum number of splits, minimum leaf size, and minimum parent size. We applied a greedy approach to data to minimize the cost function and obtained the optimum values for these parameters. A standard regression cost function representing the mean absolute error (MAE) is used as follows:

$$E = \frac{1}{N}\sum_{i=1}^{N}(y - y'), \tag{3}$$

where y denotes the ground truth (actual value) and y' represents the predictive value, i.e., GSNR in our case. N represents the total number of samples. We sum over all of the samples in our dataset to get the total error; then we keep splitting the tree until an optimal value is reached.

3.2 *K-Nearest Neighbors*

KNN is a type of supervised nonparametric ML model. KNN attempts to classify the data sample into a particular category by utilizing the training dataset. We apply this model to make predictions based on feature similarity by calculating the distance between new data point and training data points. The hyperparameters required for KNN are

- Number of nearest-neighbors (K)
- Distance metric

In our simulation environment, we kept $K = 5$ and used Euclidean distance to calculate the distance between the test point and training samples.

3.3 *Random Forest*

RF is a type of ML model that uses ensemble learning, which is based on the bagging tree technique. In this technique, each tree runs independently, and at the end, the results of all of the trees (without giving importance to any one specific tree) are averaged to give the final output. In RF, each tree has a random set of training observations and a random subset of features used to form a tree. If we only rely on the decision of a single tree, the scope of the output information is limited. However, if we construct multiple such trees and average their output, our net information from the output is much greater. We applied the bagging technique in which n different ensembles are created to give different profound knowledge about the dataset because their outputs are not correlated. When we average these ensembles, it effectively includes the insights from each of them, and we are left with a better generalization of the output. We also apply this technique to figure out the important set of features to predict our label better.

3.4 *Linear Support Vector Regression*

Linear support vector regression (SVR) is a type of supervised ML model that works on the same idea as SVMs. SVR is used to cater regression problems in which a continuous output is predicted. The following essential parameters are used to configure SVR:

- **Kernel**: It is used to map data from the lower dimension to the higher dimension at lower computation cost. It is beneficial for finding the best hyperplane. We applied linear kernel to our problem.
- **Hyperplane**: It is a line used to predict the continuous output.
- **Decision boundary**: Two parallel lines are drawn with a distance ϵ from the hyperplane to define a margin.

In SVR, we are trying to fit the maximum allowable error within the tolerable range that is defined by the ϵ value. The goal is to find a function $f(x)$ that deviates by a value not greater than ϵ for each training point from the output prediction. The best fit line is the one with the maximum number of data points:

$$f(x) = xw + b. \tag{4}$$

For the linear hyper plane, the equation that satisfies our support vector regressor to predict the QoT of the LP is

$$\epsilon \le y - xw - b \le -\epsilon. \tag{5}$$

To fit the maximum allowable error within a tolerable range, we define the value of $\epsilon = 3$ in our simulation environment.

3.5 *Neural Networks*

NN is an ML model inspired by the human nervous system to process information. It comprises the input layer, hidden layers, and output layer, where the layers are sets of neurons. NN typically learns with a feedback process in which the predicted output is compared with the actual output. The difference between them is then calculated. The error gradient is computed for every preceding layer using a backpropagation algorithm to adjust the weights using a stochastic gradient descent algorithm. For QoT estimation, we applied the NN model with several tuned parameters to get an efficient model providing high accuracy.

3.6 *Linear Regression*

Linear regression is a parametric ML model that uses a statistical technique to find the linear relationship between the input feature (x) and the output label (y). The mathematical representation for the linear regression model is as follows:

$$y = B_0 + B_1 x, \tag{6}$$

where y is the output variable, B_0 is the intercept, B_1 is the coefficient of each variable, and x is the set of input features. The model estimates the values of the intercept (B_0) and the coefficient (B_1). Linear regression has a different kind of optimization strategy. In our work to estimate the QoT, we apply the ordinary least square method that takes more than one input feature and requires no weighting function.

The hyperparameters for our employed ML models are given in Table 3.

4 Simulation Model and Synthetic Data Generation

This section describes the simulation model and the considered network topologies, the library used for physical layer abstraction and data generation, and the technique used for refining the dataset before applying it to ML models.

The proposed work simulates an open OLS composed of cascaded amplifiers and fibers. In the simulation setup, the grid size of 50 GHz is considered to have 76 channels on the C-band. Due to limitations on computational resources, only 76 channels are considered over the total bandwidth of ∼4 THz. The transmitter generates signals at 32 GBaud, shaped with a root-raised-cosine filter. The signal's launch power is set to 0 dBm, which is kept constant by EDFA, operating at a constant output power mode of 0 dBm per channel. The noise figure of EDFA is varied uniformly, in the range of 4.5 to 6 dB with a ripple gain variation uniformly with 1 dB variation. All of the links are operated using a standard single-mode fiber having a typical span length of ∼80 km. Fiber impairments such as fiber attenuation $(\alpha) = 0.2$ dB/km and dispersion $(D) = 16$ ps/nm/km are also considered. To create a realistic simulation model, the insertion losses are determined by an exponential distribution with $\lambda = 4$, as described in another study.[37,38] The paths are computed using the Dijkstra algorithm, with the metrics used being the shortest distance path. For the computation of GSNR, the ASE noise is modeled as additive white Gaussian noise with bilateral power spectral density, including both polarizations. The nonlinear impairments are modeled by the analytical perturbation model, such as the generalized Gaussian noise (GGN) model.[39]

The dataset is generated synthetically, mimicking the receiver's signal power, NLI generation during the signal propagation against two different networks, and ASE-noise accumulation using the GNPy simulator. The GNPy is an open-source optimization library that is spectrally resolved and is formulated on the GGN model.[7,39] This simulator has been developed by open optical and packet transport–physical simulation environment (OOPT–PSE) working group within the telecom infra project. In Ref. 39, GNPy is validated on a real network for QoT estimation of the LPs. It exhibits outstanding accuracy for GSNR prediction. It provides an end-to-end simulation environment to develop the network model on the physical layer. This library defines route planning in mesh optical networks and can include customized network elements in the network. The synthetic dataset is generated against two different network topologies: the European (EU) network and the USA network, shown in Figs. 2(a) and 2(b), respectively. The EU network is

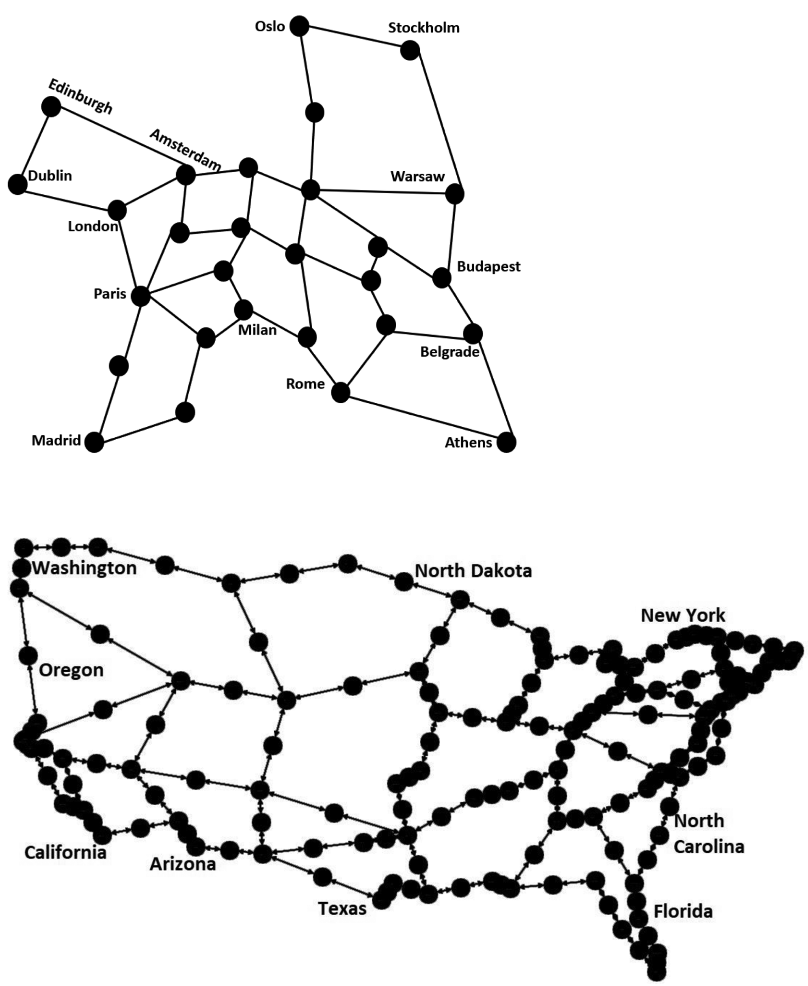

Fig. 2 Networks topologies.

Table 1 Source–destination pairs and number of spans of the European network.

Source	Destination	Number of spans
Amsterdam	Berlin	8
Brussels	Bucharest	30
Frankfurt	Istanbul	34
Vienna	Warsaw	7
Paris	Rome	34

Table 2 Source–destination pairs and number of spans of the USA network.

Source	Destination	Number of spans
Kansas City	Las Vegas	30
Milwaukee	Minneapolis	6

considered to be well deployed and represents the S network, and the USA network represents the T network. The two considered networks are the same in terms of fiber and ONE. However, they are different concerning the amplifier's delicate parameters (noise figure and amplifier ripples gain) and fiber insertion losses. The dataset used in this work consists of six source-to-destination ($s \rightarrow d$) pairs of the EU network and 11 $s \rightarrow d$ pairs of the USA network, shown in Tables 1 and 2. The spectral load realization against each simulated link of a dataset is a subset of 2^{76}. In the considered spectral load realization for every $s \rightarrow d$ pair, we considered 3000 realizations of arbitrary traffic flow varying between 34% and 100% of overall operational bandwidth. Thus for the EU network topology, 18,000 realizations are generated, and for the USA network topology, 33,000 realizations are generated. The considered dataset is then normalized to scale the values. We investigate the different normalization methods on the prediction performance of our ML models. Based on the value of the evaluation metric (MAE), we believe that z scale normalization seems to be a good choice for our case. In the z score, the mean and standard deviation of each input feature are used to normalize the vector of each feature.[40] This helps to reduce the effect of outliers from the data and overcomes the problem of dominant features entirely.[41] It is used as follows:

$$Z = \frac{X - \mu}{\sigma}, \tag{7}$$

where μ and σ are the mean and standard deviation of each feature, respectively. The considered Z-score normalization is applied to both the train and the test data.

5 Machine Learning Models Orchestration

This section describes the characterized features and labels of ML models and the metric used to evaluate the ML models. Furthermore, the models, depicted in Sec. 3, are simulated in this section. The standardized dataset is divided into two sets: the train set and the test set. The train set consists of four paths (12,000 samples) for the EU network, whereas the test set consists of the last one path (3000 samples) of the EU network and two paths of the USA network (6000 samples). The parameters utilized to describe ML models' features include ASE, NLI, received signal power, span length, total distance, and channel frequency of 76 channels, as shown in Fig. 3. All of the proposed models are evaluated using the MAE metric to quantify the GSNR predictions of the ML models by taking the mean absolute difference of all of the predicted values with the actual values. Moreover, the models described in Sec. 3 are simulated using MATLAB® platform and are configured using the simulation parameters given in Table 3.

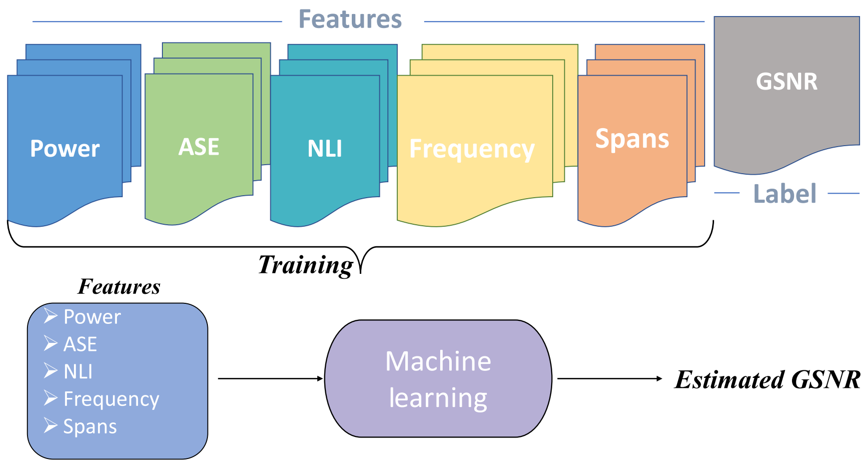

Fig. 3 ML module.

Table 3 Hyperparameters of the ML models.

ML model	Parameter	Value
DT	Max no. of splits	obs-1 = 5407
	Min leaf size	4
	Min parent size	10
	Split criteria	"*mse*"
	Purne	"On"
	Purne criteria	"mse"
KNN	k	5
	Distance metric	Euclidean
RF	Method	"Bag"
	Min leaf size	4
	No. of cycles	50
	No. of Var to sample	1/3 of max splits
Linear SVR	ϵ	0.3
	kernal	"Linear"
NN	No. of hidden layers	3
	No. of units	3
	Activation function	"ReLU"
	—	"Linear"
	Learning rate	0.01
	No. of epochs	1000
Linear regression	Equation	Linear
	Method	Ordinary least squares

6 Results and Discussion

In this section, the performance comparison of six ML models in the same domain (SA) and the DA scenario is reported. In the SA scenario, we trained the ML models on the EU network and tested it on other EU network paths. In the DA scenario, ML models exploit the knowledge of the EU network to estimate the output label (i.e., GSNR of the LP) of the USA network. Moreover, we also perform feature engineering to find out the importance of features for GSNR prediction. Furthermore, the evaluation and comparison of models are also performed using the cross and relevant features model training approaches for the specific label (i.e., GSNR of the LP in our case).

6.1 *ML Models Trained on Cross Features*

We first investigated the MAE using the SA approach, i.e., training an ML model on some paths of the EU Network and then testing it on other paths. This section exploits all of the available features of 76 channels to perform cross-feature training of ML models to estimate the GSNR of channel 1. Using the paths reported in Table 1, the first four paths of the EU network are used to train the ML models, and the last path is used for testing the models. The result of the test path, from Paris to Rome, is shown in Fig. 4. It shows the results of all of the proposed models, i.e., actual and predicted GSNR with mean (μ) and standard deviation (σ). As shown by the μ and σ in Fig. 4, the NN model trained on cross features shows excellent results in terms of GSNR prediction, whereas the KNN model shows the worst prediction performance in comparison with all proposed models. NN's performance gets better each time that we train it on training data because of the iterative learning approach, whereas in KNN the training dataset is given once for it to work.

The performances of the proposed ML models using the DA approach, i.e., training on four paths of the EU network and testing on two paths of the USA network, are reported in Table 2. The outcome of the DA approach is shown in Figs. 5 and 6, which show the proposed ML models' prediction performance against the two paths, i.e., Kansas City to Las Vegas and Milwaukee to Minneapolis, of the USA network. AS shown by the of μ and σ, it is fairly clear that the predictions with the NN model seem to follow the same distribution as the actual values and it outperforms all other proposed ML models, whereas KNN is again performing worst among all of the proposed models. Based on the performance of our NN model, we make the

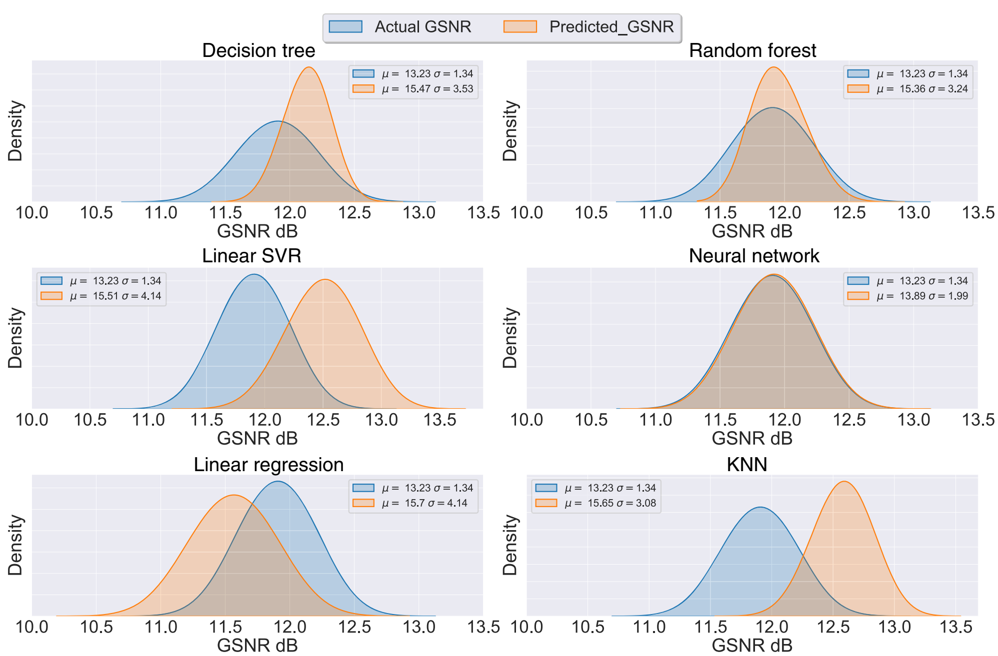

Fig. 4 EU network path Paris to Rome: cross features training.

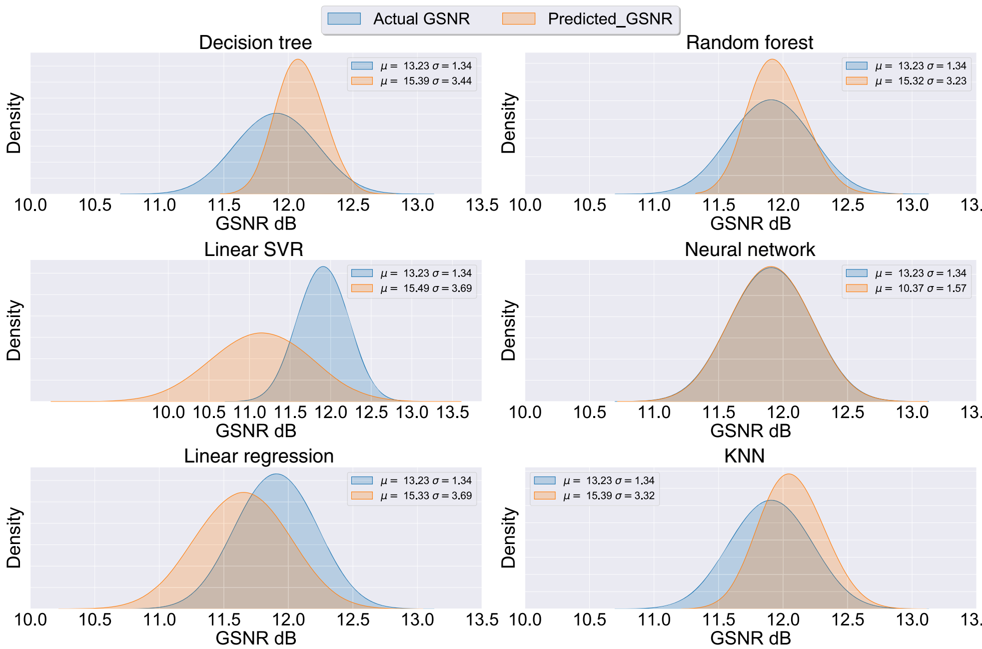

Fig. 5 USA network path Kansas to Las Vegas: cross features training.

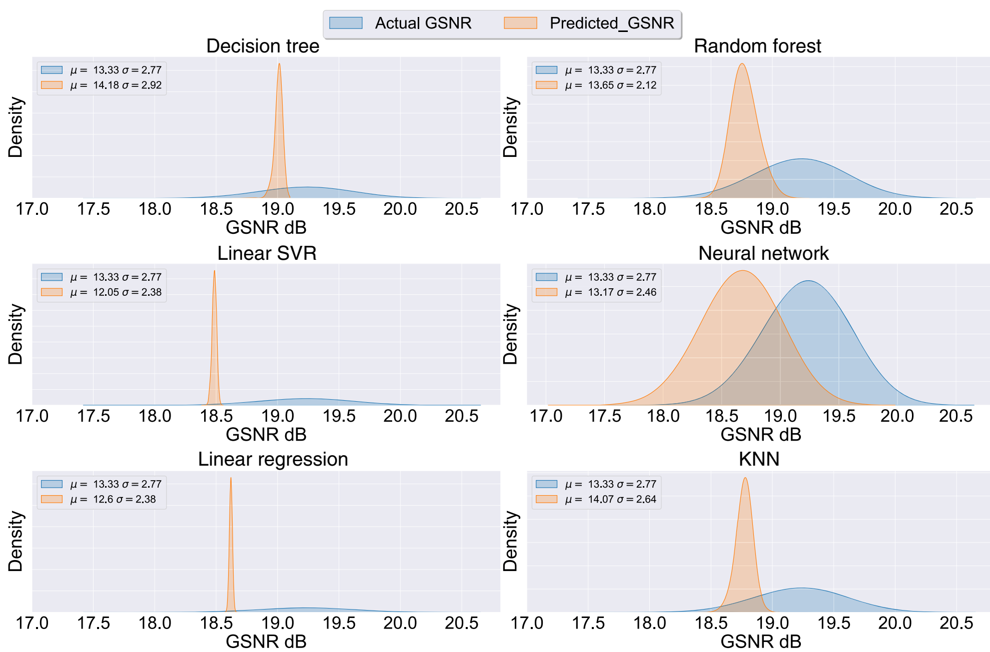

Fig. 6 USA network path Milwaukee to Minneapolis: cross features training.

following observations. First, the NN model still performs better in the case of DA for the USA network because of its ability to learn complex hidden patterns, leading to better generalization. Second, NN continuously adjusts weights at each input to further optimize results.

6.2 *ML Models Trained on Relevant Features*

This section analyzes the impact of relevant features in predicting the GSNR of channel 1, which is supposed to be channel-under-test (CUT). First, we applied the RF model to determine the

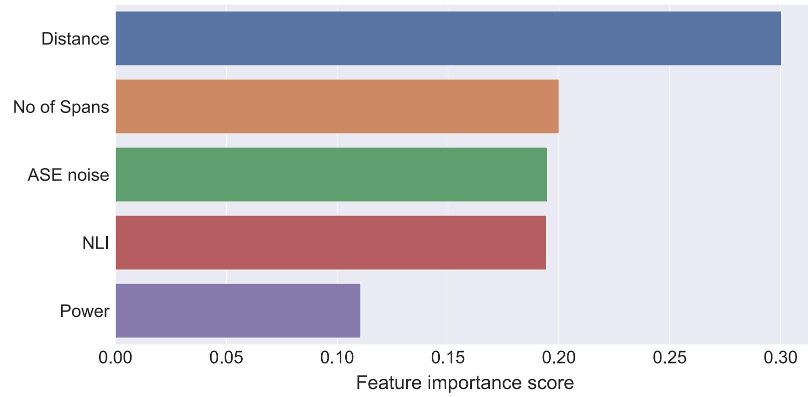

Fig. 7 Features importance.

feature importance for GSNR prediction. The RF model helps to find the features that have more impact on the prediction of the target label (i.e., the GSNR in our case). It performs well due to its property of randomly sampling the features and the data points. Figure 7 shows the importance of features to the label (GSNR). On the y axis, various used features are shown, and on the x axis, the importance score is plotted. The results shown in Fig. 7 represent that the distance between source and destination is the most important feature followed by the number of spans, ASE, NLI, and power of the LP in predicting the GSNR of LP. We trained the ML models on all of these relevant CUT features and evaluated CUT's predictions against one test path of the SA (European) network and two DA (USA) network paths. First, we see the SA network results with a test path, i.e., Paris to Rome. The results of the actual and predicted distribution of the SA test path against all of the proposed models are shown in Fig. 8. The μ and σ demonstrate NN's excellent prediction performance against other models. Moreover, for the DA approach, the distribution of predicted versus actual GSNR for the CUT on the two paths, i.e., Kansas to Las Vegas and Milwaukee to Minneapolis is shown in Figs. 9 and 10. The results in both Figs. 9 and 10 show that NN also performs best in the case of DA as compared with all of the proposed models.

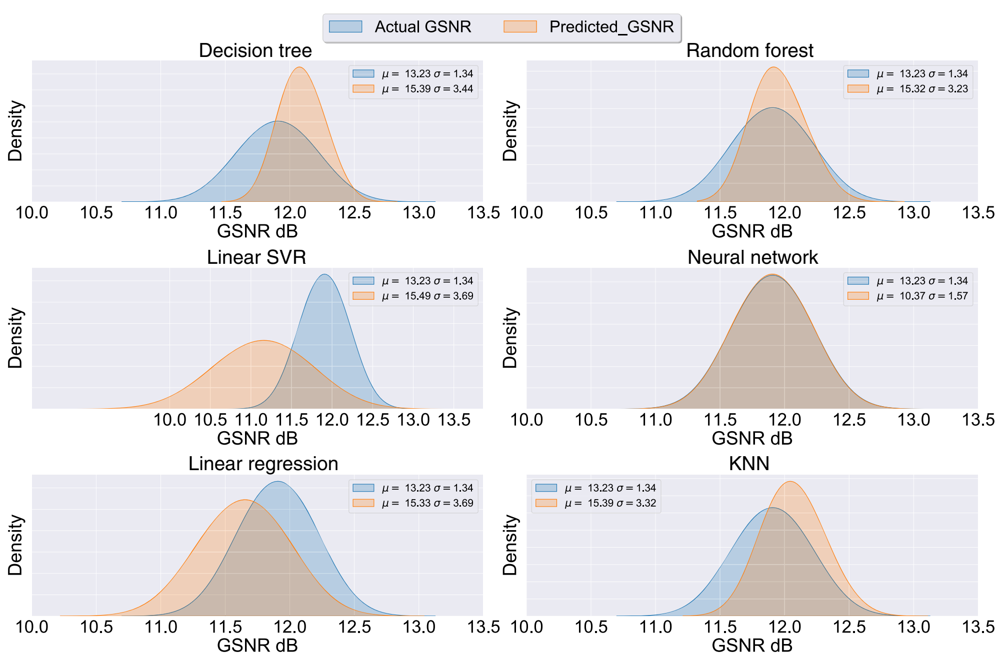

Fig. 8 EU network path Paris to Rome: relevant features training.

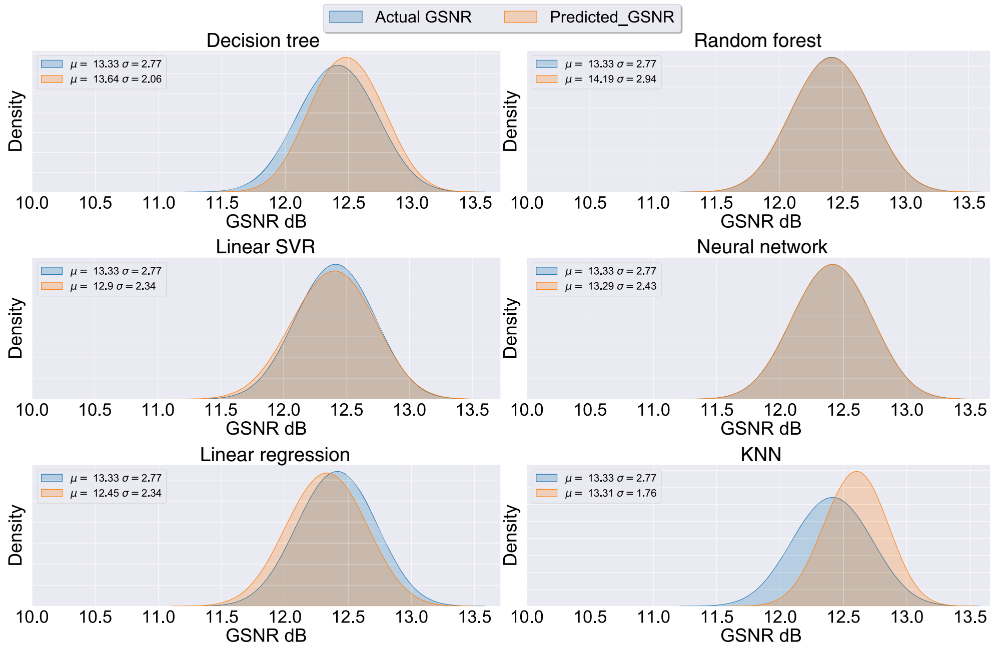

Fig. 9 USA network path Kansas to Las Vegas: relevant features training.

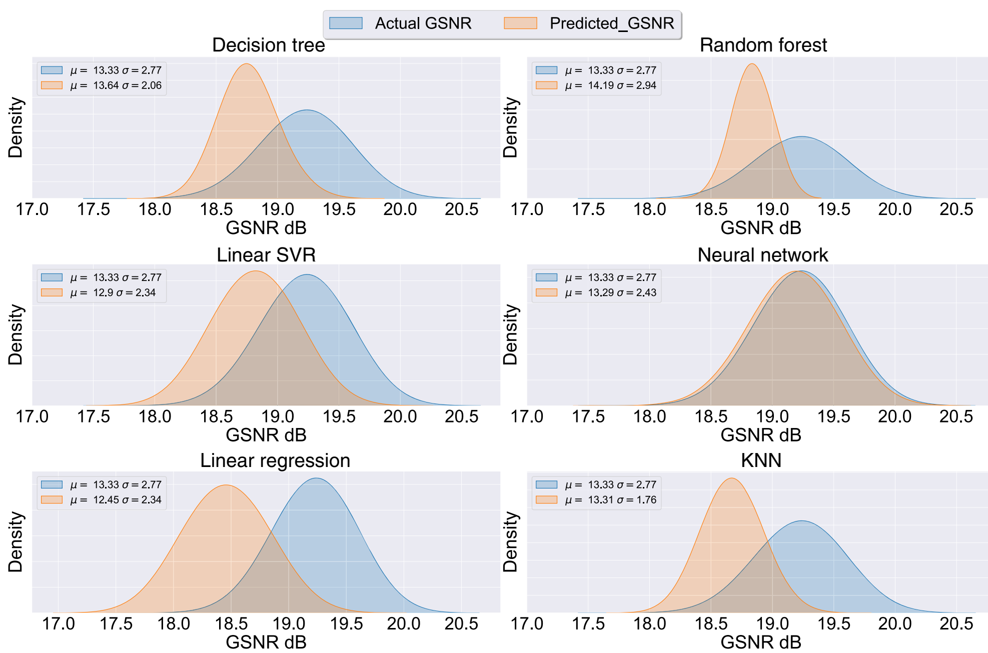

Fig. 10 USA network path Milwaukee to Minneapolis: relevant features training.

6.3 *Cross Feature versus Relevant Feature Training*

This section compares ML models based on the MAE when trained on cross features and relevant features. Table 4 illustrates the MAE comparison of the EU and USA networks. For the given simulation scenario, Fig. 4 shows that the MAE of DT is reduced when trained on relevant features for both networks, but it does not perform well on the USA network when trained on cross features due to its poor learning of underlying associations in the dataset. On the other hand, RF leverages several DTs for feature selection; hence its overall performance is better than the DT for both networks when trained on cross features and relevant features. The

Table 4 Comparison of the MAE of the EU and USA networks.

	European network		USA network	
ml model	CROSS feature MAE (dB)	relevant feature MAE (dB)	Cross feature MAE (dB)	Relevant feature MAE (dB)
DT	0.0745	0.0471	0.2277	0.1874
RF	0.0089	0.0089	0.0597	0.0477
Linear SVR	0.0777	0.0532	0.3481	0.3103
NN	0.0072	0.0013	0.008	0.0054
Linear regression	0.08919	0.0612	0.3912	0.3606
KNN	0.1653	0.0758	0.6759	0.3886

performance of linear SVR and linear regression is almost similar, and it gets better when relevant features are considered for training in both scenarios. On the other hand, the performance of NN for the EU network is outstanding when trained on cross features and relevant features because of its cognitional ability to learn complex and hidden patterns very well. To further analyze the NN model's performance trained with cross features and relevant features, we tested it on the USA network, including more test data. When NN is tested on the USA network path with relevant feature training, it generalizes very well and gives an excellent performance for both networks. For the given scenario, KNN performs well when trained on relevant features of CUT because of its good capability in working on feature similarity. In the case of the KNN model, the performance is degraded when trained on cross features, it fails to fit the underlying relationship in the dataset, and it depicts the worst generalization on the USA network as compared with all other models. The MAE of the KNN model is increased when trained on cross features particularly for the USA network because it does not properly approximate the relation between the input and output of a dataset. Observing these results, we conclude that NN-based models have the potential to generalize well on an unseen network with good performance when trained on relevant features of CUT. To take the confidence level into consideration along with prediction error, we executed the simulation 10 times with 1000 epochs and computed the prediction error each time for both the training and testing data in the same and the DA scenarios. These simulations are performed using our best performing model, the NN model trained on relevant features, the results of which, including the mean and standard deviation of ΔGSNR distribution, are shown in Figs. 8 and 9. The average MAE of the prediction error in the training and testing datasets of the EU network is around 0.00127 and 0.00132 dB, respectively. Similarly, the standard deviation (confidence level) in the respective training and the testing datasets is around 0.0000114 and 0.000018 dB. Moreover, to check the robustness of the model in the DA scenario, we also tested the trained model on a different network topology (i.e., USA network). The MAE in prediction on the USA network is around 0.0053 dB, and the standard deviation (confidence level) is found to be 0.00015 dB. Overall, the analysis shows that, when relevant features are considered for training ML models, the performance is enhanced.

Our best performing NN model provides a viable solution for practical implementation into SDN-based optical networks for real-time QoT estimation of LPs. It is executed on a system with an Intel®Core™i7 8550U 1.80 GHz CPU workstation equipped with 8 GB of RAM. The model's computation time is 0.2 ms when trained on relevant features. With the use of cross-features, we take the features of all of the relevant channels to make an appropriate decision about the lightpath deployment. It can be considered an important component for online network operating tools for QoT estimation in real time. It can improve network efficiency as on the arrival of an LP request; it can estimate the QoT of an LP in real time. In contrast, the typical techniques require extensive computational effort when applied to real-time scenarios. Its other application is in the design of EONs, in which its output will be used by routing and spectrum assignment decision tools to make the final decision about the LP deployment.

7 Conclusion

We investigated different ML techniques to predict the QoT of the LP of an unseen network before its deployment. The prior prediction of the QoT of the LP in an unseen network is an essential step for the optimal design of the network and reliable LP deployment with a low margin. The GSNR of the LP is used as a QoT metric that comprises the effect of both NLI and ASE noise accumulation. Our simulation results show that NN performs best with an MAE of 0.001 dB for the European network and 0.005 dB for the USA network when trained on relevant features and 0.007 dB for the European network and 0.008 dB for the USA network when trained on cross features.

We performed feature engineering and observed that, when the models are trained only on relevant features, the prediction performance improves. The presented results clearly show that ML-based techniques, especially NN, significantly reduce the provisioning GSNR margin in both SA and DA scenarios. For the future, additional work considering a broad range of system configurations is required to prove the effectiveness of this approach for real world applications.

References

1. Cisco, "Cisco visual networking index: forecast and trends, 2017–2022," Technical Report, Cisco (2017).
2. O. Gerstel et al., "Elastic optical networking: a new dawn for the optical layer?" *IEEE Commun. Mag.* **50**(2), s12–s20 (2012).
3. G. Zhang et al., "A survey on OFDM-based elastic core optical networking," *IEEE Commun. Surv. Tutorials* **15**(1), 65–87 (2013).
4. V. Curri et al., "Design strategies and merit of system parameters for uniform uncompensated links supporting nyquist-WDM transmission," *J. Lightwave Technol.* **33**, 3921–3932 (2015).
5. R. Pastorelli, "Network optimization strategies and control plane impacts," in *Opt. Fiber Commun. Conf.*, OSA (2015).
6. M. Filer et al., "Multi-vendor experimental validation of an open source QoT estimator for optical networks," *J. Lightwave Technol.* **36**, 3073–3082 (2018).
7. A. Ferrari et al., "Gnpy: an open source application for physical layer aware open optical networks," *J. Opt. Commun. Networking* **12**(6), C31–C40 (2020).
8. M. Freire et al., "Predicting optical power excursions in erbium doped fiber amplifiers using neural networks," in *Asia Commun. and Photonics Conf.*, IEEE, pp. 1–3 (2018).
9. J. Thrane et al., "Machine learning techniques for optical performance monitoring from directly detected PDM-QAM signals," *J. Lightwave Technol.* **35**(4), 868–875 (2017).
10. F. N. Khan, C. Lu, and A. P. T. Lau, "Optical performance monitoring in fiber-optic networks enabled by machine learning techniques," in *Opt. Fiber Commun. Conf. and Expos.*, OSA, pp. 1–3 (2018).
11. L. Barletta et al., "QOT estimation for unestablished lighpaths using machine learning," in *Opt. Fiber Commun. Conf. and Exhib.*, OSA (2017).
12. J. Mata et al., "Artificial intelligence (AI) methods in optical networks: a comprehensive survey," *Opt. Switching Networking* **28**, 43–57 (2018).
13. T. Jiménez et al., "A cognitive quality of transmission estimator for core optical networks," *J. Lightwave Technol.* **31**(6), 942–951 (2013).
14. A. Caballero et al., "Experimental demonstration of a cognitive quality of transmission estimator for optical communication systems," *Opt. Express* **20**(26), B64–B70 (2012).
15. A. D'Amico et al., "Using machine learning in an open optical line system controller," *J. Opt. Commun. Networking* **12**(6), C1–C11 (2020).
16. E. Seve et al., "Learning process for reducing uncertainties on network parameters and design margins," *J. Opt. Commun. Networking* **10**(2), A298–A306 (2018).
17. 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. Networking* **9**(1), 98–108 (2017).

18. W. Mo et al., "Ann-based transfer learning for qot prediction in real-time mixed line-rate systems," in *Opt. Fiber Commun. Conf.*, Optical Society of America, pp. 1–3 (2018).
19. R. Proietti et al., "Experimental demonstration of cognitive provisioning and alien wavelength monitoring in multi-domain EON," in *Opt. Fiber Commun. Conf.*, Optical Society of America, W4F–7 (2018).
20. I. Khan et al., "QOT estimation for light-path provisioning in un-seen optical networks using machine learning," in *22nd Int. Conf. Transparent Opt. Networks*, IEEE (2020).
21. I. Khan, M. Bilal, and V. Curri, "Advanced formulation of QOT-estimation for un-established lightpaths using cross-train machine learning methods," in *22nd Int. Conf. Transparent Opt. Networks*, IEEE (2020).
22. C. Rottondi et al., "Machine-learning method for quality of transmission prediction of unestablished lightpaths," *J. Opt. Commun. Networking* **10**(2), A286–A297 (2018).
23. S. Aladin and C. Tremblay, "Cognitive tool for estimating the qot of new lightpaths," in *Opt. Fiber Commun. Conf. and Expos.*, OSA (2018).
24. I. Khan et al., "Machine learning assisted abstraction of photonic integrated circuits in fully disaggregated transparent optical networks," in *22nd Int. Conf. Transparent Opt. Networks*, IEEE (2020).
25. R. Di Marino et al., "Assessment of domain adaptation approaches for QOT estimation in optical networks," in *Opt. Fiber Commun. Conf. and Exhib.*, IEEE, pp. 1–3 (2020).
26. I. Khan et al., "Lightpath qot computation in optical networks assisted by transfer learning," *IEEE/OSA J. Opt. Commun. Networking* **13**(4), B72–B82 (2021).
27. F. Usmani et al., "Convolutional neural network for quality of transmission prediction of unestablished lightpaths," *Microwave Opt. Technol. Lett.* **63**(10), 2461–2469 (2021).
28. D. Azzimonti et al., "Active vs. transfer learning approaches for QOT estimation with small training datasets," in *Opt. Fiber Commun. Conf.*, Optical Society of America, M4E–1 (2020).
29. https://www.itu.int/rec/T-REC-G.694.1/en.
30. A. Ahmad et al., "Exploring the effects of physical layer parameters in WDM based fixed and flexible-grid networks," in *Int. Conf. Opt. Network Des. and Model.*, pp. 128–133 (2015).
31. A. Ahmad et al., "A transmission layer aware network design for fixed and flexible grid optical networks," in *17th Int. Conf. Transparent Opt. Networks*, IEEE, pp. 1–4 (2015).
32. D. J. Elson et al., "Investigation of bandwidth loading in optical fibre transmission using amplified spontaneous emission noise," *Opt. Express* **25**(16), 19529–19537 (2017).
33. A. Nespola et al., "GN-model validation over seven fiber types in uncompensated PM-16QAM Nyquist-WDM links," *IEEE Photonics Technol. Lett.* **26**(2), 206–209 (2014).
34. D. Pilori, F. Forghieri, and G. Bosco, "Residual non-linear phase noise in probabilistically shaped 64-QAM optical links," in *Opt. Fiber Commun. Conf.* (2018).
35. A. Ahmad et al., "Switching node architectures in flexible-grid networks: a performance comparison," in *Int. Conf. Optical Network Design and Modeling*, pp. 49–54 (2014).
36. F. N. Khan et al., "Applications of machine learning in optical communications and networks," in *Opt. Fiber Commun. Conf.*, Optical Society of America, M1G–5 (2020).
37. Y. Ando, "Statistical analysis of insertion-loss improvement for optical connectors using the orientation method for fiber-core offset," *IEEE Photonics Technol. Lett.* **3**(10), 939–941 (1991).
38. M. Filer et al., "Low-margin optical networking at cloud scale," *J. Opt. Commun. Network* **11**, C94–C108 (2019).
39. M. Cantono et al., "On the interplay of nonlinear interference generation with stimulated raman scattering for QoT estimation," *J. Lightwave Technol.* **36**(15), 3131–3141 (2018).
40. S. Kotsiantis, D. Kanellopoulos, and P. Pintelas, "Data preprocessing for supervised leaning," *Int. J. Comput. Electr. Autom. Control Inf. Eng.* **1**, 4104–4109 (2007).
41. D. Singh and B. Singh, "Investigating the impact of data normalization on classification performance," *Appl. Soft Comput.* **97**, 105524 (2020).

Biographies of the authors are not available.