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

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], Vittorio Curri[b]
[a]National University of Sciences and Technology (NUST), Islamabad, Pakistan
[b]DET, Politecnico di Torino, Corso Duca degli Abruzzi 24, Turin, 10129 Italy

Abstract. The ever-increasing demand for global internet traffic, together with evolving concepts of Software-defined network (SDN) and Elastic-optical-network (EON), not only demands the total capacity utilization of underlying infrastructure but also demands dynamic, flexible, and transparent optical network. Generally, 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 lightpaths (LP) establishment is crucial for reducing this provisioning margins. This article proposes and compares several data-driven Machine learning (ML) models to make an accurate calculation of QoT before the actual establishment of 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 LP is the Generalized Signal-to-Noise Ratio (GSNR) which accumulates the impact of both Non-Linear Interference (NLI) and Amplified Spontaneous Emission (ASE) noise. The dataset is generated synthetically using well tested GNPy simulation tool. Promising results are achieved, showing that the proposed Neural network (NN) 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.

Keywords: Machine learning, Quality of Transmission estimation, Generalized SNR.

*Arsalan Ahmad, arsalan.ahmad@seecs.edu.pk

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 (FHD) or 4K, cloud computing and the Internet of Things (IoT). This increasing trend of global internet traffic requires the maximum utilization of the remaining capacity of the already working network infrastructure. Around 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 dis-aggregation paves a path for the technologies such as EON and SDN. The distinctive features of SDN and EON offers versatile and dynamic resource provisioning in optical networks for both control and data plane.[2,3] EONs provide flexible spectral assignment in the data plane and boost the network's capacity while lowering network cost. This adaptability results in much more complex 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.

Today's optical networks have started evolving to partial dis-aggregation, with a full dis-aggregation goal eventually. The prime step towards 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 QoT degradation.[4,5] The precise accomplishment of this working point leads to achieve 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 GSNR, which comprises the accumulated effect of NLI and 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 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), where the network operator has the sufficient knowledge about the working point of network nodes and provide 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 where the system administrator does not have the 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 *target domain* to accurately establish the LP with reduced margin. Generally, the controller can acquire an accurate description of the system parameters i.e.,*network status*. The QoT-E exploit several analytical approaches that can measure the GSNR with a very good precision as shown in.[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 analytic approach is not recommended in such an agnostic scenario to estimate the QoT of LP prior to its deployment.

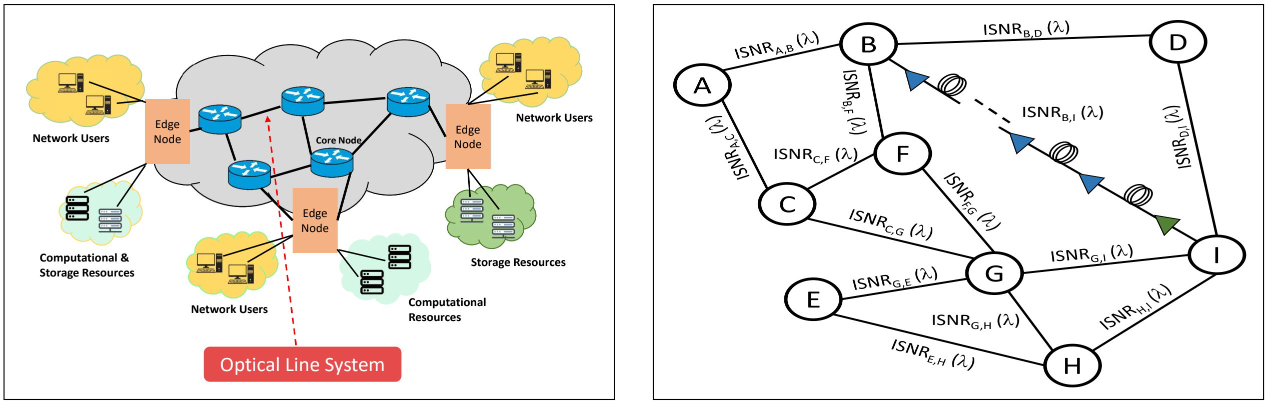
To overcome this challenge, we opted to use data-driven ML approach as an alternative way, which has already been proved very effective in several different contexts of managing optical networks; look at[8–11] for performance monitoring of optical network operations. A comprehensive assessment of ML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

The notable difference between the past literature and the present work is that we proposed 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 by 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 the identical fiber type and communication devices but 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 supposed as *S* network, a type of network where the operator has a complete description of network elements' operational parameters. The other network is considered as *T* network, which is intrinsically a type of network where the operator has only a basic description of the operational parameters of network elements.

The rest of the paper is organized as follows: Section 2, briefly explains the physical layer's abstraction to efficiently implement a multi-layer optimization, simultaneously with the argument that an accurate QoT-E has a fundamental role in minimizing the system margin. Moreover, we also pitched 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 un-established LP. In Section 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 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 EDFA noise figure, ripple gain and insertion losses. In Section 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 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 Section 6, we produced detailed results. Finally, the conclusion and future research work are illustrated in Section 7.

2 Overview of Optical Transport Network

Generally, an optical network consists of Optical Network Elements (ONE) connected through bidirectionally fiber links, where traffic demand is added/dropped or routed, as shown in Fig. 1a. The amplifiers are placed after a specific span length using the Erbium-Doped Fiber Amplifiers



(a) (b)

Fig 1: (a), Architecture of Optical Network (b), Optical Transport Network

(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. Further to this, 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, where 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 (PDM) is exploit to propagate from particular source to its specific destination. Along with the transmission, LP suffers different propagation impairments like amplifier noise added as an ASE, fiber propagation, and filtering penalties applied by ROADM. Also, the fiber propagation has been 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] In opposite to this, 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 source node to destination node is given by the well-acknowledged GSNR measurement, which combines both the aggregated effect of ASE noise and NLI disturbance. Generally, GSNR is defined as:

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

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 non-linear effects, while fiber propagation generates P_{NLI}, which relies on the spectral-load and the power of the distinct channel.[4] In these circumstances, it is pretty much 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 the QoT impairments. Simultaneously, given a specific pair of source and destination encounters the cumulative impairments of previously traversed OLSs along with ROADM effects. Each crossed OLS adds a specific amount of NLI and ASE noise. For the purpose of QoT, the abstraction of OLS can be achieved by a single parameter known as SNR degradation which generally depends upon the frequency ($\mathrm{GSNR}_i(f)$), if the OLS controllers can retain 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 represent ROADM network nodes, while the E represent OLSs having $\mathrm{GSNR}_i(f)$ as weights on the consequent edges, shown in Fig. 1b. Specifically, for a given LP from the 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 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.), is 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 un-reliable 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 rely on the telemetry data achieving from the EDFAs and the Optical Channel Monitor (OCM). This specific technique is feasible for an accurate QoT prediction by utilizing the network's current state's telemetry. In opposite 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, leading to substantial unreliability in the QoT margin.[15]

The final method examines the dataset that obtains the QoT responses against arbitrary spectral-loads of *S* network. As mentioned earlier, the generation of 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 utilizing a dataset comprised of spectral-load samples of an *S* network for training and complies a correct QoT-E for every generated spectral-load section of *T* network. In distinction to the former method, where just telemetry data is explored, this procedure employs the QoT-E centered on the GSNR reaction to particular spectral-load arrangements of *S* network, used for an accurate GSNR prediction of *T* network. Additionally, this arrangement does not require any information about physical parameters of the OLS as compared to the first technique. Therefore, this approach gives an excellent playground to utilize the ML-DA method. In this activity, we focus on the third procedure, which is based upon 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 *T* network.

3 Background on Machine Learning Models

This section briefly explains the ML techniques we have applied for QoT estimation of un-established LP. Generally, ML has a wide range of applications in optical communications and networking.[35] 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 un-established LP, and also 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 un-established 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 is used, which is as follows:

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

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

3.2 K-Nearest Neighbours

K-nearest neighbors (KNN) is a type of supervised non-parametric ML model. KNN attempts to classify the data sample into a particular category by utilizing training dataset. We applied this model to make predictions based on feature similarity by calculating the distance between new data point and training data points. The hyper-parameters 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

It 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 the trees (without giving importance to anyone specific tree) are averaged to give the final output. In Random Forest, 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 where 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 brings in the insights from each of them, and we are left with a better generalization of the output. We also applied 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 Support Vector Machines (SVM). SVR is used to cater regression problems where continuous output is predicted. The following essential parameters are used to configure SVR:

- **Kernel**: It is used to map data from lower dimension to higher dimension at lower computation cost. It is beneficial in finding the best hyper-plane. We applied linear Kernel to our problem.
- **Hyper Plane**: It is a line used to predict the continuous output.
- **Decision Boundary**: Two parallel lines are drawn with ϵ distance from the hyper-plane to define a margin.

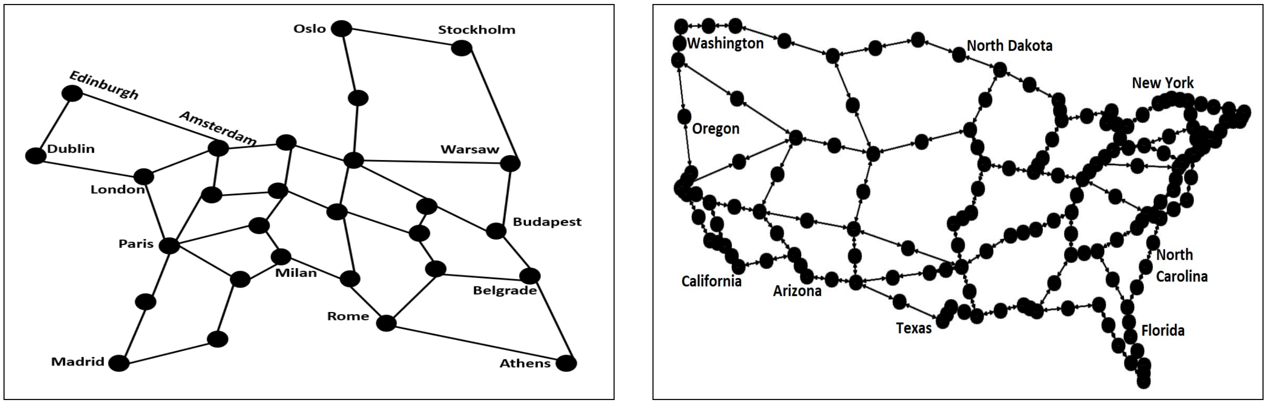
In SVR, we are trying to fit the maximum allowable error within a the tolerable range that is defined by the ϵ value. The goal is to find a function $f(x)$ that deviates by a value not greater than ϵ for each training point from the output prediction. The best fit line is the one with the maximum number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

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

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

$$\epsilon \leq y - xw - b \leq -\epsilon \tag{5}$$

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



(a) European Network (b) USA Network

Fig 2: Networks Topologies

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 where 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 back propagation 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 Model is a parametric ML model which 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 co-efficient of each variable and x is the set of input features. The model estimates the values of intercept (B_0) and the co-efficient (B_1). Linear regression has a different kind of optimization strategy. In our work to estimate QoT, we applied 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 50GHz is considered to have 76 channels on the C-band. Due to computational resources limitation, only 76 channels are considered over the total bandwidth of approximately 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 dB to 6 dB with a ripple gain variation uniformly with 1 dB variation. All the links are operated using a Standard Single-Mode Fiber (SSMF) having a typical span length of approximately 80 km. To these fiber impairments such as fiber attenuation (α) = 0.2 dB/km and dispersion (D) = 16 ps/nm/km are also considered. To create the simulation model realistic, the statistics of insertion losses are determined by an exponential distribution with $\lambda = 4$, as described in the study.[36,37] The paths are computed using the Dijkstra algorithm, with the metrics used is the shortest distance path. For the computation of GSNR, the ASE noise is modeled as Additive White Gaussian Noise (AWGN) with bilateral Power Spectral Density (PSD), including both polarization. The nonlinear impairments are modeled by the analytical perturbation model, such as Generalized Gaussian Noise (GGN) model.[38]

Table 1: **Source-Destination pairs and Number of Spans of 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 USA Network**

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

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 GGN model.[7,38]This simulator has been developed by Open Optical & Packet Transport–Physical Simulation Environment (OOPT–PSE) working group within the Telecom Infra Project (TIP). In,[38] 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 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; European (EU) network and USA network shown in Fig. 2a and Fig. 2b respectively. The EU Network is considered well-deployed and represents the *S* network while 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 6 source-to-destination (s $\rightarrow$ d) pairs of EU network and 11 s $\rightarrow$ d pairs of USA network shown in Table 1 and Table 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

Table 3: Hyperparameters of ML Models

ML Model	Parameter	Value
Decision Tree	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
Random Forest	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	$'Linear'$
Neural Network	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

every s $\rightarrow$ d pair, we considered 3000 realizations of arbitrary traffic flow varying between 34% to 100% of overall operational bandwidth. Thus for 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 investigated the different normalization methods on the prediction performance of our machine learning models. Based on the value of the evaluation metric (mean absolute error), 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 against each input feature is used to normalize the vector of each feature.[39] It helps to reduce the effect of outliers from the data and overcomes the problem of dominant features entirely.[40] It is used as follows:

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

where μ and σ is the mean and standard deviation against each feature, 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 section 3, are simulated in this

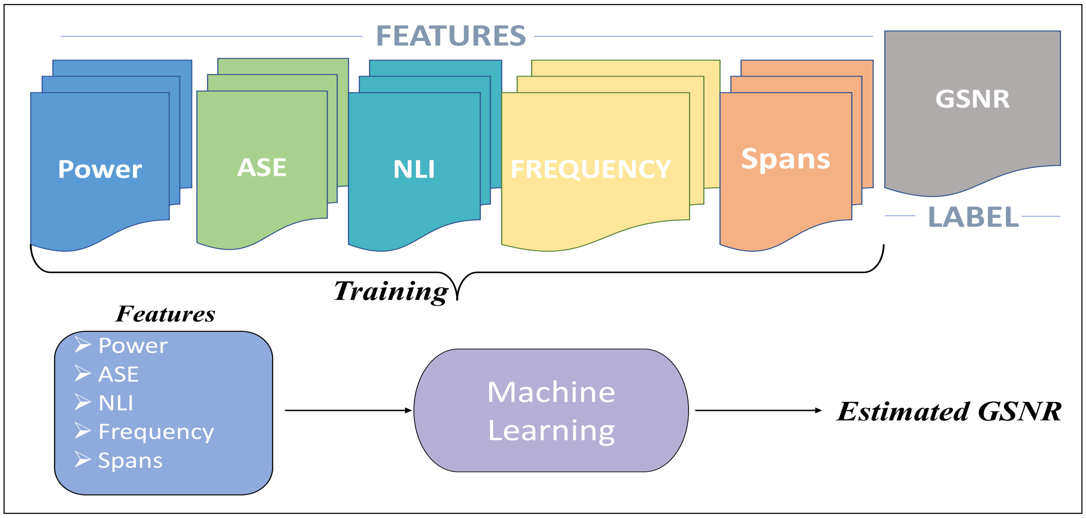


Fig 3: Machine Learning Module

section. The standardized dataset is divided into two sets: *train* and *test* set. The train set consists of four paths (12,000 samples) for the EU network, while 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, shown in Fig. 3. All the proposed models are evaluated using the Mean Absolute Error (MAE) metric to quantify the GSNR predictions of the ML models by taking the mean absolute difference of all the predicted values with the actual values. Moreover, the models described in Section 3 are simulated using MATLAB® platform and are configured using the simulation parameters given in the Table 3.

6 Results & Discussion

In this section, the performance comparison of six ML models in the Same Domain (SA) and 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 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 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 EU Network and then testing it on other paths. This section exploits all 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 depicted in Fig. 4. It shows the results of all the proposed models, i.e., Actual and Predicted GSNR with mean (μ) and standard deviation (σ). Observing the statistics μ and σ in Fig. 4. it is depicted that the NN model trained on cross features shows excellent results in terms of GSNR prediction, whereas the KNN model is showing the worst prediction performance

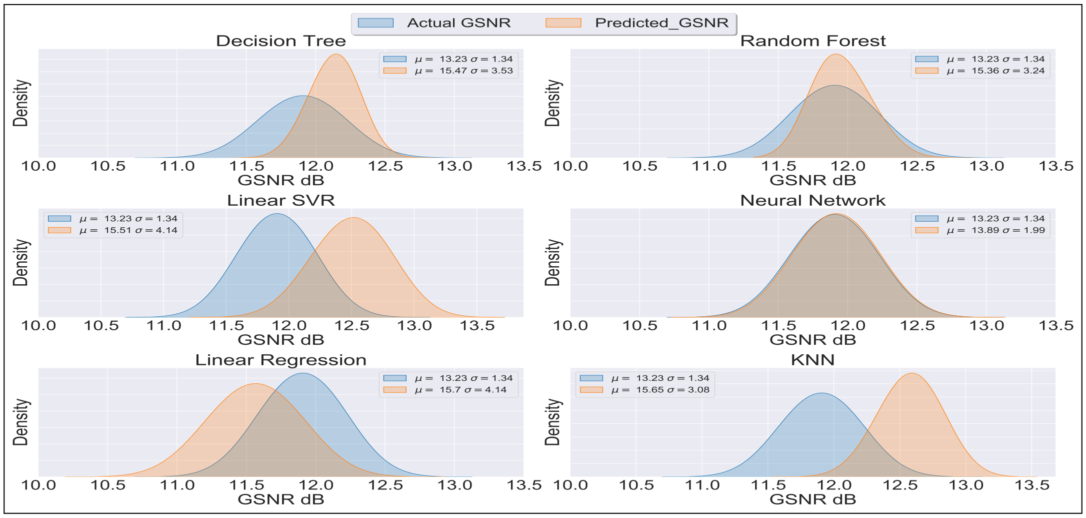


Fig 4: EU network path Paris to Rome: Cross features training.

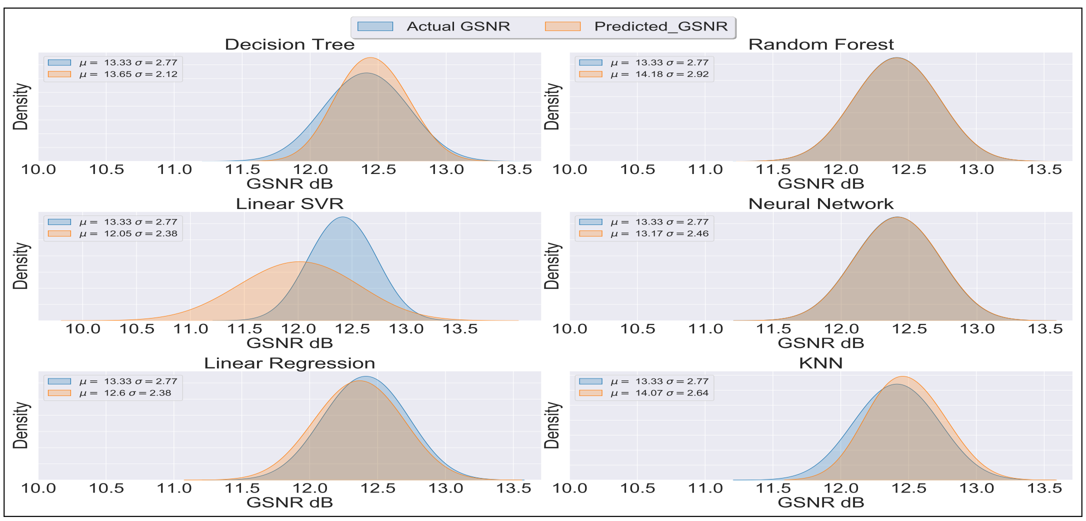


Fig 5: USA network path Kanas to Las-Vegas: Cross features training.

in comparison with all proposed models. NN's performance gets better each time 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.

To evaluate the performance of the proposed ML models using the DA approach, i.e., training on four paths of EU network and testing on two paths of USA network are reported in Table 2. The outcome of the DA approach is shown in Fig. 5 and Fig. 6. It shows the proposed ML models' prediction performance against the two paths, i.e., Kanas City to Las-Vegas and Milwaukee to Minneapolis, of the USA network. Observing the statistics of μ and σ, it is pretty 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 the proposed models. Based on the performance of our NN model, we make the following observations. NN model still performs better in case of DA for the USA network because of its ability to learn complex hidden patterns, leading to better generalization. NN continuously adjusts weights at each input to further optimize results.

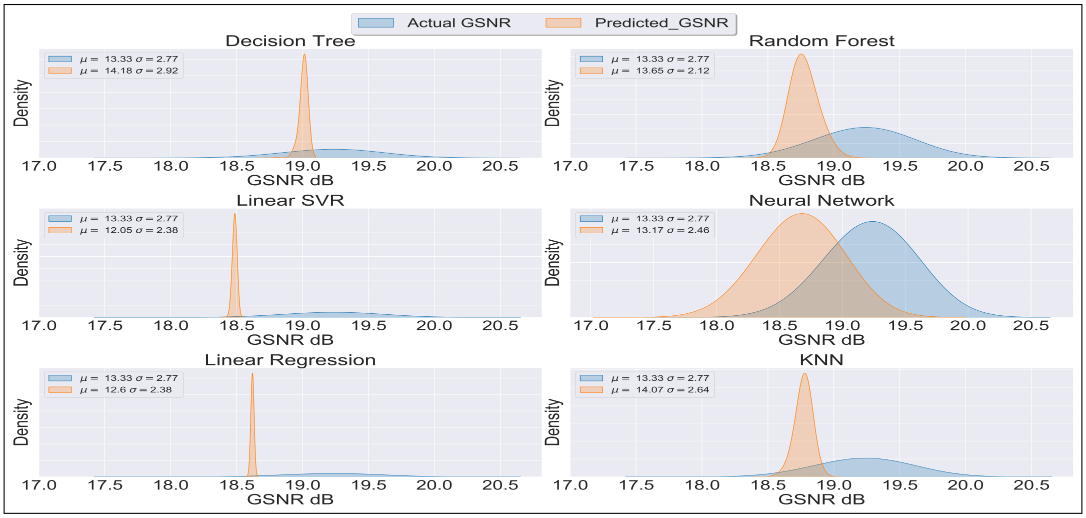


Fig 6: USA network path Milwaukee to Minneapolis: Cross features training.

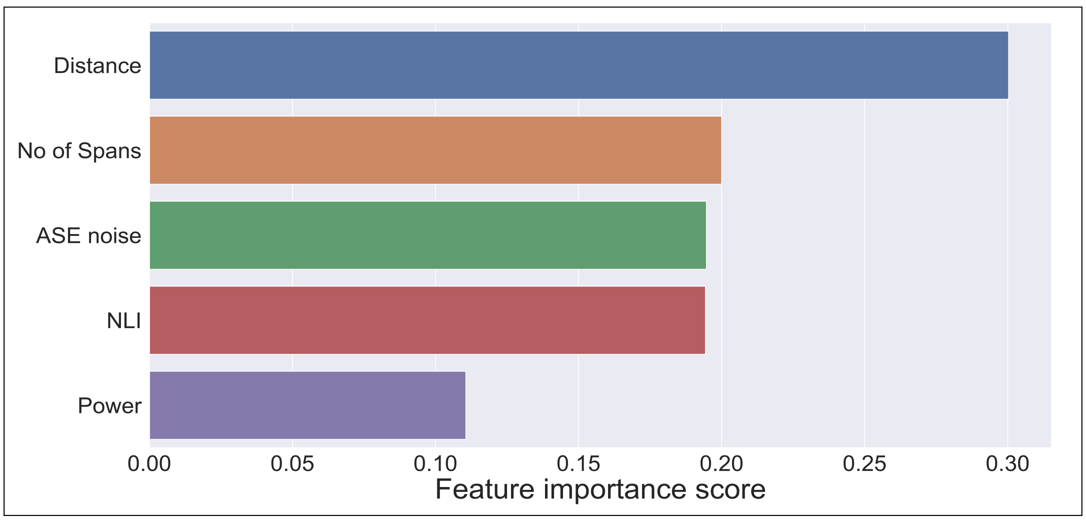


Fig 7: Features importance

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). Firstly, we applied the Random Forest model to determine the feature importance for GSNR prediction. The Random Forest model helps in finding 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. Fig. 7 depicts the importance of features to the label (GSNR). On the y-axis, various used features are shown, while on the x-axis 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 the proposed models are shown in Fig. 8. The statistics of μ and σ demonstrate

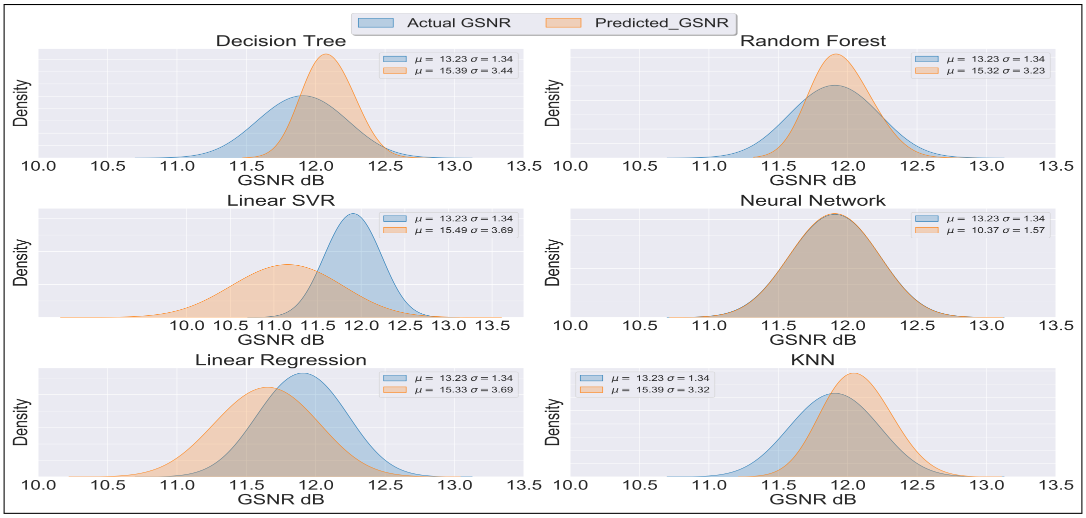


Fig 8: EU network path Paris to Rome: Relevant features training.

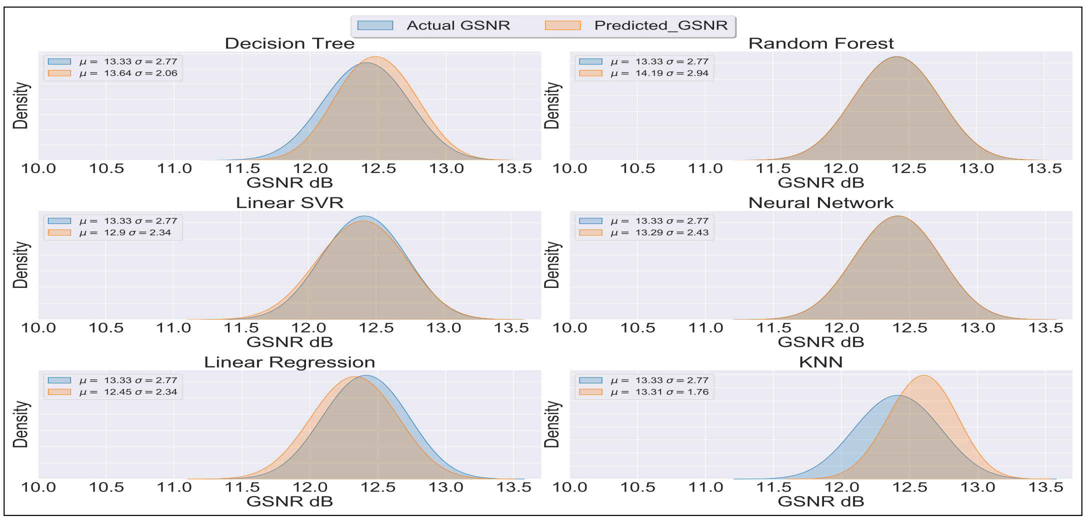


Fig 9: USA network path Kanas to Las-Vegas : Relevant features training

NN's excellent prediction performance against other models. Moreover, for the DA approach, the distribution of predicted vs. actual GSNR for the CUT on the two paths, i.e., Kanas to Las-Vegas and Milwaukee to Minneapolis is shown in Fig. 9 and Fig. 10. The results in both Fig. 9 and Fig. 10 show that NN also performs best in the case of DA as compared to all the proposed models.

6.3 *Cross feature vs. Relevant feature training*

This section compares ML models based on 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 demonstrates that the MAE of Decision Tree is reduced when trained on relevant features for both networks, but it does not perform well on USA network when trained on cross features due to its poor learning of underlying associations in the dataset. On the other hand, Random Forest leverages several decision trees for feature selection, hence its overall performance is better than the decision tree for both networks when trained on cross features and relevant

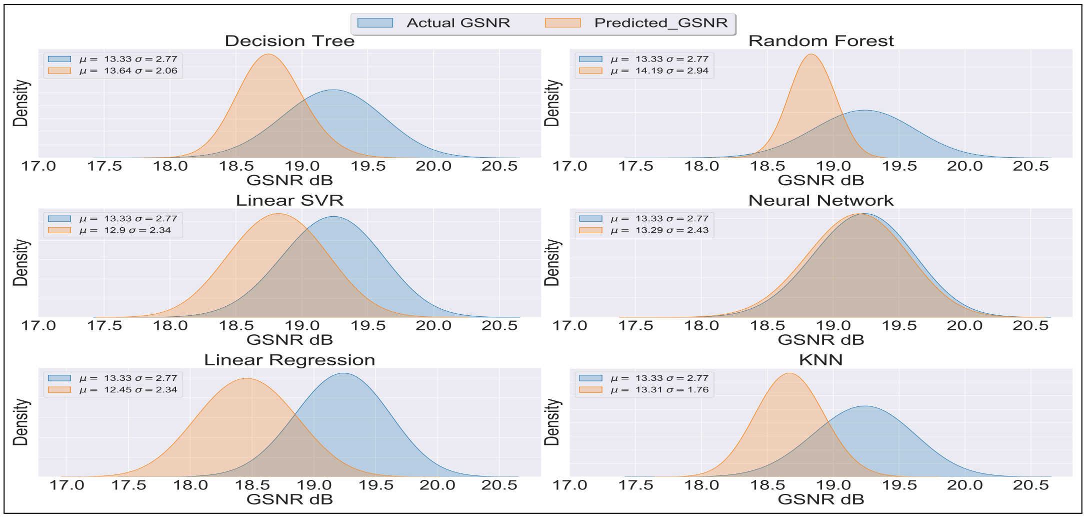


Fig 10: USA network path Milwaukee to Minneapolis :Relevant features training.

features. The 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 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 to work on feature similarity. In the case of KNN model, the performance is degraded when trained on cross features and it failed to fit the underlying relationship in the dataset and depicts the worst generalization on the USA network as compared to 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 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 prediction error each time for both the training and testing data in the same as well as in the domain adaptation scenario. These simulations are performed using our best performing model, the neural network model trained on relevant features whose results, including the mean and standard deviation of ΔGSNR distribution, are reported in Fig. 8 and Fig. 9. The average mean absolute error of the prediction error in the training and testing dataset about the EU network is around 0.00127 dB and 0.00132 dB, respectively. Similarly, the standard deviation (confidence level) in the respective training and the testing dataset is around 0.0000114 dB and 0.000018 dB. Moreover, to check the robustness of the model in the domain adaptation scenario, we also tested the trained model on the different network topology (i.e., USA network). The mean absolute error in prediction on the USA network is around 0.0053 dB, while the standard deviation (confidence level) is found to be 0.00015 dB. Overall, what we analyzed is that when relevant features are considered for training ML models, the performance gets 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 are taking the features of all 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 a lightpath request; it can estimate the QoT of a lightpath 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 Elastic Optical Networks (EONs), where its output will be used by Routing and Spectrum Assignment (RSA) decision tools to make the final decision about the lightpath deployment

Table 4: Comparison of MAE of EU and USA network

European Network			USA Network	
ML Model	**cross feature MAE(dB)**	**relevant feature MAE(dB)**	**cross feature MAE(dB)**	**relevant feature MAE(dB)**
Decision Tree	0.0745	0.0471	0.2277	0.1874
Random Forest	0.0089	0.0089	0.0597	0.0477
Linear SVR	0.0777	0.0532	0.3481	0.3103
Neural Network	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

7 Conclusion

We investigated different ML techniques to predict the QoT of LP of an unseen network before its deployment. The prior prediction of the QoT of LP in an un-seen network is an essential step for the optimal design of the network and reliable LP deployment with a low margin. The GSNR of LP is used as a QoT metric which 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 USA network when trained on relevant features and 0.007 dB for European network and 0.008 dB for 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 is improved. The presented results clearly show that ML-based techniques, especially NN, significantly reduce the provisioning GSNR margin in both SA and DA scenarios. For future perspective, additional work is required considering broad range of system configurations to prove the effectiveness of this approach for real world applications.

References

1 Cisco, "Cisco Visual Networking Index: Forecast and Trends, 2017–2022," tech. rep., Cisco (2017).

2 O. Gerstel, M. Jinno, A. Lord, *et al.*, "Elastic optical networking: A new dawn for the optical layer?," *IEEE Communications Magazine* **50**(2), s12–s20 (2012).

3 G. Zhang, M. De Leenheer, A. Morea, *et al.*, "A survey on ofdm-based elastic core optical networking," *IEEE Communications Surveys & Tutorials* **15**(1), 65–87 (2012).

4 V. Curri, A. Carena, A. Arduino, *et al.*, "Design strategies and merit of system parameters for uniform uncompensated links supporting nyquist-WDM transmission," *JLT* **33**, 3921–3932 (2015).

5 R. Pastorelli, "Network optimization strategies and control plane impacts," in *OFC*, OSA (2015).

6 M. Filer, M. Cantono, A. Ferrari, *et al.*, "Multi-Vendor Experimental Validation of an Open Source QoT Estimator for Optical Networks," *JLT* **36**, 3073–3082 (2018).

7 A. Ferrari, M. Filer, K. Balasubramanian, *et al.*, "Gnpy: an open source application for physical layer aware open optical networks," *JOCN* **12**(6), C31–C40 (2020).

8 M. Freire, S. Mansfeld, D. Amar, *et al.*, "Predicting optical power excursions in erbium doped fiber amplifiers using neural networks," in *2018 (ACP)*, 1–3, IEEE (2018).

9 J. Thrane, J. Wass, M. Piels, *et al.*, "Machine learning techniques for optical performance monitoring from directly detected pdm-qam signals," *JLT* **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 *2018 (OFC)*, 1–3, IEEE (2018).

11 L. Barletta, A. Giusti, C. Rottondi, *et al.*, "Qot estimation for unestablished lighpaths using machine learning," in *OFC*, OSA (2017).

12 J. Mata, I. de Miguel, R. J. Duran, *et al.*, "Artificial intelligence (ai) methods in optical networks: A comprehensive survey," *OSN* **28**, 43–57 (2018).

13 T. Jiménez, J. C. Aguado, I. de Miguel, *et al.*, "A cognitive quality of transmission estimator for core optical networks," *JLT* **31**(6), 942–951 (2013).

14 A. Caballero, J. C. Aguado, R. Borkowski, *et al.*, "Experimental demonstration of a cognitive quality of transmission estimator for optical communication systems," *Optics express* **20**(26), B64–B70 (2012).

15 A. D'Amico, S. Straullu, A. Nespola, *et al.*, "Using machine learning in an open optical line system controller," *JOCN* **12**(6), C1–C11 (2020).

16 E. Seve, J. Pesic, C. Delezoide, *et al.*, "Learning process for reducing uncertainties on network parameters and design margins," *Journal of Optical Communications and 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," *Journal of Optical Communications and Networking* **9**(1), 98–108 (2017).

18 W. Mo, Y.-K. Huang, S. Zhang, *et al.*, "Ann-based transfer learning for qot prediction in real-time mixed line-rate systems," in *2018 (OFC)*, 1–3, IEEE (2018).

19 R. Proietti, X. Chen, A. Castro, *et al.*, "Experimental demonstration of cognitive provisioning and alien wavelength monitoring in multi-domain eon," in *Optical Fiber Communication Conference*, W4F–7, Optical Society of America (2018).

20 I. Khan, M. Bilal, M. Siddiqui, *et al.*, "Qot estimation for light-path provisioning in un-seen optical networks using machine learning," in *ICTON*, 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 *ICTON*, IEEE (2020).

22 C. Rottondi, L. Barletta, A. Giusti, *et al.*, "Machine-learning method for quality of transmission prediction of unestablished lightpaths," *Journal of Optical Communications and Networking* **10**(2), A286–A297 (2018).

23 S. Aladin and C. Tremblay, "Cognitive tool for estimating the qot of new lightpaths," in *OFC*, OSA (2018).

24 I. Khan, M. Chalony, E. Ghillino, *et al.*, "Machine learning assisted abstraction of photonic integratedcircuits in fully disaggregated transparent optical networks," in *ICTON*, IEEE (2020).

25 R. Di Marino, C. Rottondi, A. Giusti, *et al.*, "Assessment of domain adaptation approaches for qot estimation in optical networks," in *2020 Optical Fiber Communications Conference and Exhibition (OFC)*, 1–3, IEEE (2020).

26 I. Khan, M. Bilal, M. Umar Masood, *et al.*, "Lightpath qot computation in optical networks assisted by transfer learning," *IEEE/OSA Journal of Optical Communications and Networking* **13**(4), B72–B82 (2021).

27 F. Usmani, I. Khan, M. U. Masood, *et al.*, "Convolutional neural network for quality of transmission prediction of unestablished lightpaths," *Microwave and Optical Technology Letters* **63**(10), 2461–2469 (2021).

28 D. Azzimonti, C. Rottondi, A. Giusti, *et al.*, "Active vs transfer learning approaches for qot estimation with small training datasets," in *Optical Fiber Communication Conference*, M4E–1, Optical Society of America (2020).

29 https://www.itu.int/rec/T-REC-G.694.1/en.

30 A. Ahmad, A. Bianco, H. Chouman, *et al.*, "Exploring the effects of physical layer parameters in wdm based fixed and flexible-grid networks," in *2015 International Conference on Optical Network Design and Modeling (ONDM)*, 128–133 (2015).

31 A. Ahmad, A. Bianco, H. Chouman, *et al.*, "A transmission layer aware network design for fixed and flexible grid optical networks," in *2015 17th International Conference on Transparent Optical Networks (ICTON)*, 1–4, IEEE (2015).

32 D. J. Elson, G. Saavedra, K. Shi, *et al.*, "Investigation of bandwidth loading in optical fibre transmission using amplified spontaneous emission noise," *Optics express* **25**(16), 19529–19537 (2017).

33 A. Nespola, S. Straullu, A. Carena, *et al.*, "Gn-model validation over seven fiber types in uncompensated pm-16qam nyquist-wdm links," *IEEE Photonics Technology Letters* **26**(2), 206–209 (2013).

34 D. Pilori, F. Forghieri, and G. Bosco, "Residual non-linear phase noise in probabilistically shaped 64-qam optical links," in *OFC*, (2018).

35 F. N. Khan, Q. Fan, J. Lu, *et al.*, "Applications of machine learning in optical communications and networks," in *Optical Fiber Communication Conference*, M1G–5, Optical Society of America (2020).

36 Y. Ando, "Statistical analysis of insertion-loss improvement for optical connectors using the orientation method for fiber-core offset," *IEEE Photonics Technology Letters* **3**(10), 939–941 (1991).

37 M. Filer, J. Gaudette, Y. Yin, *et al.*, "Low-margin optical networking at cloud scale," *J. Opt. Commun. Netw.* **11**, C94–C108 (2019).

38 M. Cantono, D. Pilori, A. Ferrari, *et al.*, "On the Interplay of Nonlinear Interference Generation with Stimulated Raman Scattering for QoT Estimation," *JLT* **PP**(99), 1–1 (2018).

39 S. Kotsiantis, D. Kanellopoulos, and P. Pintelas, "Data preprocessing for supervised leaning," *World Academy of Science, Engineering and Technology, International Journal of Computer, Electrical, Automation, Control and Information Engineering* **1**, 4104–4109 (2007).

40 D. Singh and B. Singh, "Investigating the impact of data normalization on classification performance," *Appl. Soft Comput.* **97**, 105524 (2020).

List of Figures

List of Tables