Evaluating Cross- feature Trained Machine Learning Models for Estimating QoT of Unestablished Lightpaths

Evaluating Cross-feature Trained Machine Learning Models for Estimating QoT of Unestablished Lightpaths

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

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

Mehek Siddiqui
National University of Sciences & Technology (NUST), Pakistan
meheksiddiqui@gmail.com

Mahnoor Khan
National University of Sciences & Technology (NUST), Pakistan
mahnoor.khan.au@gmail.com

Muhamamd Bilal
DET, Politecnico di Torino, Italy
muhammad.bilal@polito.it

M Umar Masood
DET, Politecnico di Torino, Italy
muhammad.masood@polito.it

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

Muhammad Shahzad
National University of Sciences & Technology (NUST), Pakistan
mshehzaad@hotmail.com

Vittorio Curri
DET, Politecnico di Torino, Italy
curri@polito.it

***Abstract*—The rapid increase in bandwidth-driven applications has resulted in exponential internet traffic growth, especially in the backbone networks. To address this growth of internet traffic, operators always demand the total capacity utilization of underlying infrastructure. In this perspective, precise estimation of the quality of transmission (QoT) of the lightpaths (LPs) is vital for reducing the margins provisioned by uncertainty in network equipment's working point. This article proposes and compares several data-driven Machine learning (ML) based models to estimate QoT of unestablished LP before its deployment in the future deploying network. The proposed models are cross-trained on the data acquired from an already established LP of an entirely different in-service network. The metric considered to evaluate the QoT of LP is the Generalized Signal-to-Noise Ratio (GSNR). The dataset is generated synthetically using well tested GNPy simulation tool. Promising results are achieved to reduce the GSNR uncertainty and, consequently, the provisioning margin.**

***Index Terms*—Machine learning, Quality of Transmission estimation, Generalized SNR**

I. Introduction

In the past few years, the telecommunication industry transformed to meet the vastly growing internet traffic requirements using optical-transmission systems. In recent times, a myriad of bandwidth-intensive applications are being deployed, consequently increasing the internet traffic [1]. The full utilization of the already deployed infrastructure's residual-capacity is needed to meet these ever-increasing internet traffic requirements. In this direction, the cornerstone technologies are; coherent-technology and DWDM, particularly optical transmission and fiber usage. Furthermore, the optical network's goal towards network dis-aggregation also paves a path for SDN and EON technologies. SDN's notable features are to provide adaptive and dynamic utilization of network resources both in control and data plane [2], [3]. Moreover, EONs enable flexible spectrum assignment, enhances the network capacity at a meager network cost. Due to this resilience, the LP-provisioning problem becomes more crucial than conventional fixed grid wavelength-division-multiplexing (WDM) networks.

At present, optical networks have started progressing towards network dis-aggregation goal. For the network dis-aggregation, the first requirement is to inspect the optical-line-systems (OLSs) that associate the network nodes. Currently, the degradation of QoT is based on an OLS-controller's proficiency to run at the best possible working point [4], [5]. The more precisely this working point is achieved, the more significant the traffic rate for deployed traffic is achieved, and consequently, it leads to a minimum margin for traffic deployment. Hence, it is imperative to estimate an accurate QoT of LP before its actual deployment with a reduced margin. Generally, QoT is well appraised by the GSNR, which incorporates the cumulative effect of NLI and ASE noise [6]. The GSNR reports the path viability along with the deployable rate by utilizing the attributes of the transceiver.

In the current work, we assume a Domain-Adaption (DA) approach. In this technique, the data is acquired from the already deployed in-service network (source domain "S") to learn the valuable knowledge about the network, and later on, this learned knowledge is utilized to estimate the QoT in an un-used or newly deployed network (target domain "T"). This investigation's primary focus is to minimize the uncertainty in GSNR assessment of the *T*. This leads to reliable LP deployment in the *T* network with minimum margin. Typically the QoT-E uses several analytical techniques to measure the GSNR precisely with the required knowledge of system parameters as shown in [7]. The analytical approach cannot be implemented in the absence of the required description of system-parameters. Therefore, it is impossible to implement it in the present DA's scenario as we do not know the target do-

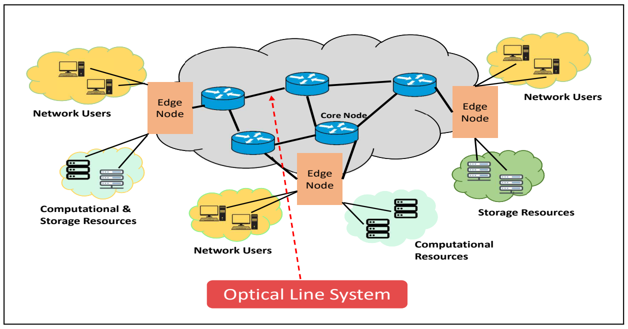

Fig. 1: Architecture of Optical Network

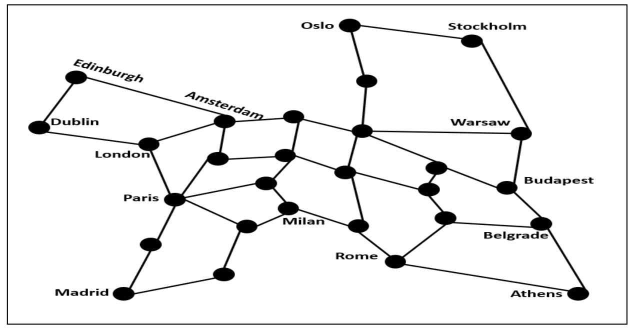

Fig. 2: European Network

main T network's exact physical characteristics. Consequently, the current investigation corresponding to DA reports that the analytic techniques are not working on getting provisioning of QoT of LP before its actual deployment in such an agnostic scenario. To overcome this, we choose to use an alternative ML paradigm that has already been proved to be very useful in various scenarios of managing optical networks; look at [8]–[11]. An extensive survey of ML practices in optical-networks is given in [12]. Specifically, coming towards QoT-E of LP before its establishment, an approach based on random forests (RF) is suggested to employ the stored-database in [13] to reduce the unreliability concerning network design margins and parameters. Ample ML techniques are proposed in [14], [15] for estimating QoT-E of LP prior to its establishment. An approach based on RF for binary classification is proposed in [16] to estimate the bit-error-ratio (BER) of un-established LPs. The performance evaluation of two DA techniques is carried out in [17]. Finally, in [18] the authors examined the QoT-E accuracy furnished by a DA approach and active learning (AL) technique on two dissimilar network-topologies.

The significant difference between the past investigation and the present investigation is that we propose to employ various ML-based approaches considering the DA technique for the reduction of system-margin of the T network utilizing the mimicked dataset of GSNRs response against particular traffic settings of LPs of the S network in an open environment. We also compare the performance of employed ML models considering cross-feature training and relevant-feature training techniques for GSNR estimation.

II. Networks Model & Data Generation

Typically an optical-transport-network generally consists of a set of Optical-Network-Elements (ONE) linked via bidirectional optical fiber as shown in Fig. 1. After a certain span distance, the amplifiers are installed using the Raman amplification method or the Erbium-Doped Fiber Amplifiers (EDFAs) method, or the hybrid of both. ONE linked via fibers is generally demonstrated as an optical-line system (OLS) and a particular controller to perform configurations in an advanced network. In our proposed network model, we simulate an open OLS consists of several cascaded-amplifiers and optical fibers. In our simulation framework, the C-band with a 50GHz grid size and 76 channels are considered. Owing to resource limitation, we have considered only 76 channels over the total bandwidth of around 4 THz. The transmitter produces signals at 32 GBaud. The launch-power for a signal is kept to 0 dBm and fixed by employing the EDFA approach. The noise value for EDFA varies uniformly between 4.5 dB and 6 dB with a 1 dB uniform variation of ripple gain. Standard-Single-Mode Fiber (SSMF) is considered for all the links with a total distance of 80 km. Fiber-impairments particularly dispersion (D) = 16 ps/nm/km and attenuation (α) = 0.2 are also considered. To make simulation framework more realistic, insertion losses are also calculated with $\lambda = 4$, as reported in [19], [20]. We make use of the Dijkstra algorithm to compute the shortest paths. For the GSNR computation, we modeled ASE noise and other nonlinear impairments by utilizing Additive-White-Gaussian Noise (AWGN) and Generalized Gaussian-Noise (GGN) models, respectively [21]. GNPy tool is utilized to mimic the signal characteristics i.e., received power, NLI, ASE noise, etc., during its propagation against two distinct networks. The GNPy is an open-source library based on GGN model and used for optimization [7], [21]. It simulates an end-to-end environment for the network physical layer abstraction.

TABLE I: European Network Source-Destination pairs

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

TABLE II: USA Network Source-Destination pairs

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

The synthetic-dataset is generated against two distinct network-topologies; European (EU) network and USA network as illustrated in Fig. 2 and Fig. 3 respectively. Here we considered the EU Network as a well-deployed source S network, whereas the USA Network acts as T network. Both networks are identical to fiber type and ONE. Nevertheless, they are distinct in terms of fiber insertion-losses and amplifier specifications such as ripple gain and noise figure. The dataset employed for this investigation is comprised 5 paths for EU-network and 2 paths for USA-network given in Table I and

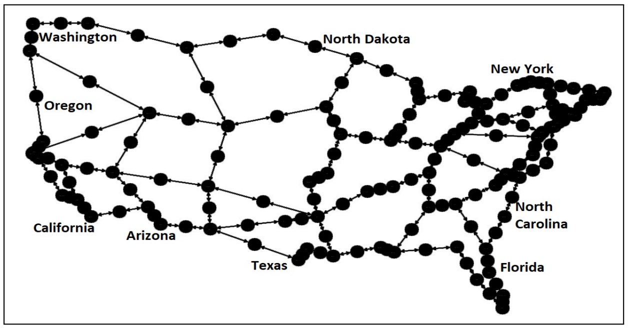

Fig. 3: USA Network

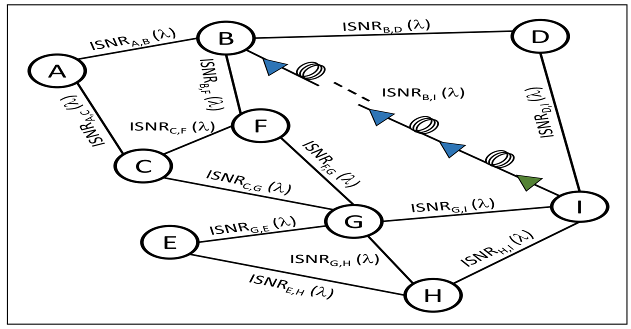

Fig. 4: Optical Transport Network

Table II. The spectral load considered for each simulated-link in a dataset is a subset of 2^{76}. We considered arbitrary traffic in the range of 34% to 100% of total utilized bandwidth for each source-to-destination (s $\rightarrow$ d) pair. Furthermore, normalization is performed on a generated dataset to get scaled values using the Z-score normalization.

III. GSNR as QoT-Estimation Metric

Generally, the well-acknowledged metric utilized for QoT-E of a specific LP routed by particular OLSs from source-node to destination-node is specified GSNR computation, which combines the accumulated effect of NLI perturbation and ASE-noise. Generally, GSNR is defined as Eq. 1, where OSNR $= P_{\mathrm{Rx}}/P_{\mathrm{ASE}}$, $\mathrm{SNR}_{\mathrm{NL}} = P_{\mathrm{Rx}}/P_{\mathrm{NLI}}$, P_{Rx} represents the received signal power of a specific channel, P_{ASE} defines the ASE-noise power, whereas P_{NLI} represents the NLI power.

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

Exploring the sequential characteristics of a transceiver, the GSNR seems to provide more precise BER measurements as indicated in [6]. The P_{NLI} is generated through fiber-propagation, and it depends on the spectral-load and power of the particular channel [4]. In this scenario, it is explicit that there is an ideal spectral-load for each particular OLS that enhances GSNR [5]. Inspecting LP propagation's consequences for a particular source and destination pair, we provide an abstracted operation for every ONE as a cascading effect that adds up QoT deterioration. Simultaneously, a given LP from a particular source-node to destination-node also encounters various impairments of all the OLSs traversed earlier. Each crossed-OLS introduce a certain quantity of NLI and ASE-noise. However, in addition to the ROADMs effects, every LP encounters the aggregated impairments of all the OLSs traversed earlier. For QoT, we can abstract the OLS with a single variable which is marked as SNR degradation and generally depends on the frequency ($\mathrm{GSNR}_i(f)$). Hence, we can abstract a network as a weighted-graph (W), where W = (V, E) corresponds to the specific network topology concerning the above scenario. The vertices (V) are used to represent the nodes of the ROADM network, whereas the edges (E) are used as the OLSs having $\mathrm{GSNR}_i(f)$ degradation illustrated as weights on the corresponding edges, demonstrated in Fig. 4. Specifically, for an LP routed from source-node I to destination-node F that passes along intermediate nodes B, the QoT is given as follows:

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

Following the availability of network abstraction, LPs' provisioning with a reduced margin for a particular source to destination is possible based on the *GSNR* of a particular source to destination path.

IV. Machine Learning Models

In general, ML has a broad variety of applications in both optical communications and networking [22]. ML model implements previously learned knowledge to make predictions. In this work, three ML models are utilized to evaluate the QoT of an un-deployed LP. In the following, we shortly give an overview of these applied ML models.

A. Random Forest

It is an ML model that utilizes an ensemble-learning strategy that depends on the bagging tree method. In this approach, each tree executes separately, and eventually, the mean of the outcome of all the trees is taken to get the final result. In RF, an arbitrary subset of features and training samples are used to form a tree. We implemented the Bagging technique, where n separate ensembles are designed to provide intellectual information about the dataset, the output of these ensembles is not correlated. When we get the mean of these ensembles, it usefully brings in the insights from each of them, and we achieve a perfect generalization of the end result.

B. Neural Networks

Neural Networks (NN) is an ML model inspired by the human brain to process information. It generally includes the input layer, hidden layers, and output layer, where each layer comprises the set of neurons. NN usually learns with a feedback approach where the predicted label is compared with the actual label, then the difference between them is calculated. The error is assessed for each former layer using a back-propagation mechanism to adapt the weights with a stochastic gradient descent approach. For QoT estimation, we executed the NN model with various tuned parameters to obtain an effective model providing good accuracy.

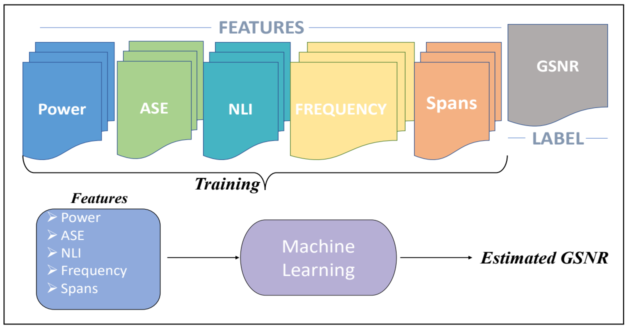

Fig. 5: Machine Learning Module

TABLE III: Hyperparameters of ML Models

ML Model	Parameter	Value
Random Forest	Method	$'Bag'$
	Min leaf size	4
	Cycles	50
	No of Var to sample	1/3 of max splits
Neural Network	Hidden layers	3
	No of units	3
	Activation Function	$'ReLU'$ $'Linear'$
	Learning rate	0.01
	Epochs	1000
KNN	k	5
	Distance metric	Euclidean

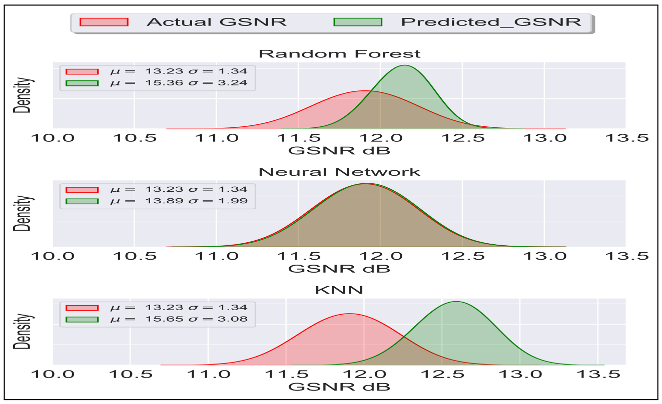

Fig. 6: Paris to Rome: Cross features training.

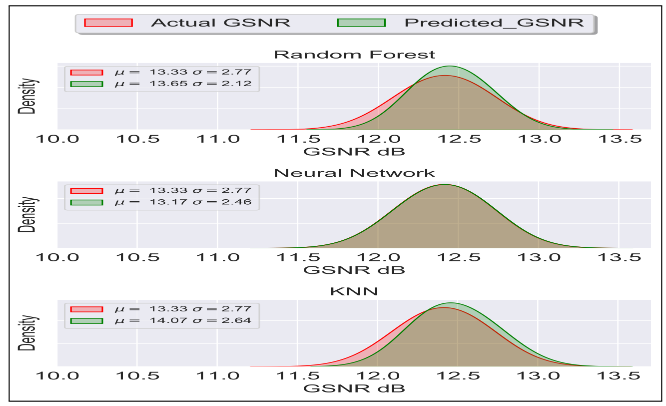

Fig. 7: Kanas to Las-Vegas: Cross features training.

C. K-Nearest Neighbours

K-nearest neighbors (KNN) is a kind of non-parametric supervised ML model. We implemented this model to make estimations by calculating the distance between new instances and training data points and then classify the data point based on feature similarity. The foremost hyper-parameters needed for KNN are number of nearest neighbors (k) and distance metric.

V. Results & Discussion

This section describes the definition of features and labels used for ML models and the performance comparison of three proposed ML models in both the Same Domain (SA) and DA framework. The normalized dataset is split-ted into *train* and *test* set. The train set incorporates 12,000 samples (four paths) of the EU network, while the test set includes 3000 samples (one path) from the EU network and 6000 samples (two-path) from the USA network. The set of ML models' features are received signal power, frequency of channel NLI, span-length, ASE noise, and the total distance for all the 76 channels, illustrated in Fig. 5. The Mean Absolute Error (MAE) metric is used to evaluate all the employed ML models; it quantifies the GSNR estimations by taking the difference between the actual value and the predicted value. Furthermore, the proposed models are simulated using MATLAB® platform and are configured applying the specifications/hyperparameters given in the Tab. III.

We compare the performance of three employed ML models in the Same Domain (SA) and DA scenarios. In the SA scenario, both the training and testing are performed on the EU network paths. In contrast, for the DA scenario, ML models utilized the learned knowledge from the EU network to estimate the GSNR of LP in the USA network. Furthermore, we also implemented a feature engineering approach to obtain important features in GSNR prediction. Additionally, the models are trained on the cross and relevant features against the particular label (i.e., GSNR of LP) for performance assessment and comparison. Initially, we explored the MAE for the SA mechanism, where the ML model is trained and tested on the EU network paths. For this approach, we utilized all the features of 76 channels for performing cross-feature model training to evaluate the GSNR of channel 1. We reported the paths in Table I, where the first four paths of the EU Network are utilized for training the ML models, and the last path is utilized for assessing the performance of models. Fig. 6 demonstrates the performance of all the models with a mean (μ) and standard deviation (σ) on the test path (i.e., From Paris to Rome). It clearly illustrates that the NN model exhibits excellent results concerning GSNR prediction. However, the KNN model shows the worst performance than all other employed ML models. The performance of NN's becomes more good owing to the iterative learning mechanism.

For the further performance assessment of ML models, the DA mechanism is employed where the models are trained on

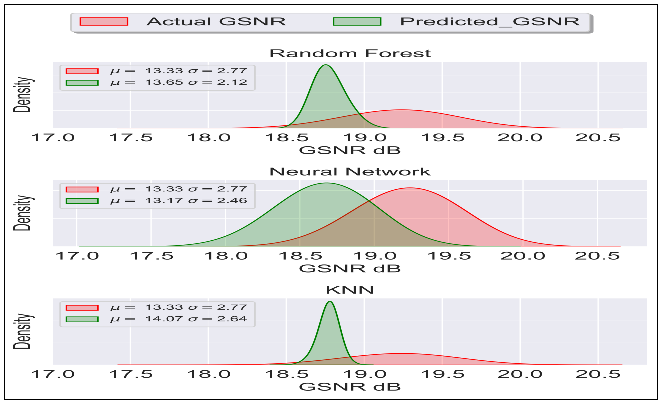

Fig. 8: Milwaukee to Minneapolis: Cross features training.

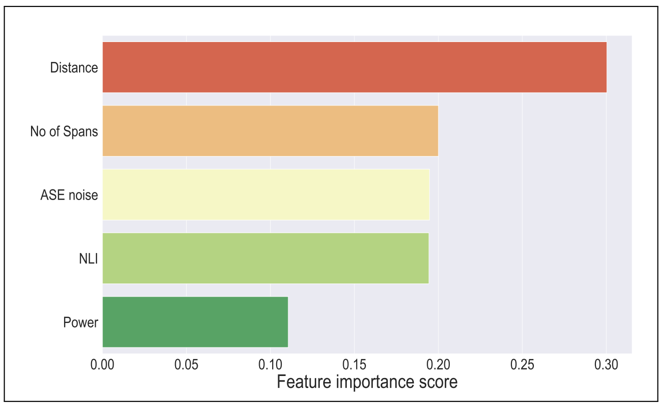

Fig. 9: Features importance

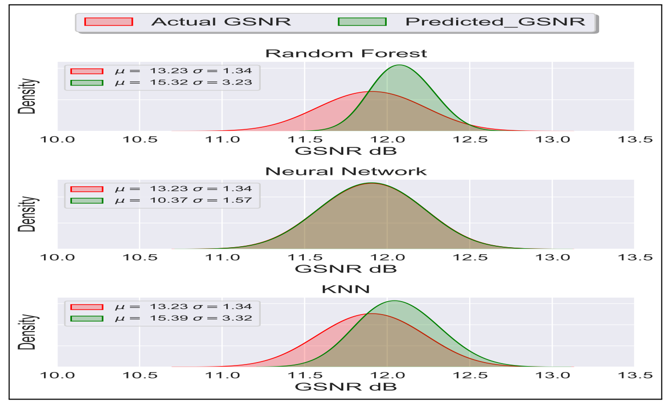

Fig. 10: Paris to Rome: Relevant features training.

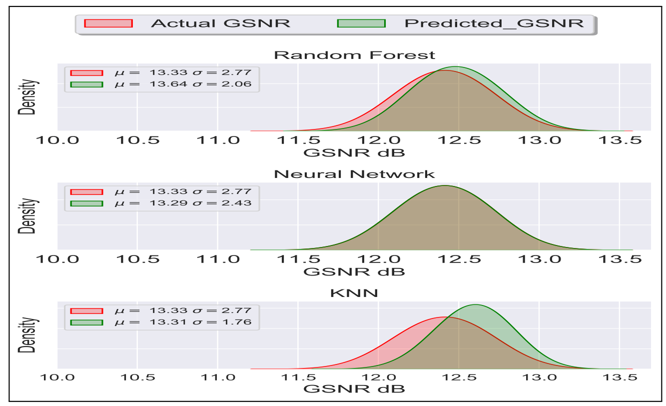

Fig. 11: Kanas to Las-Vegas: Relevant features training

EU network and tested on USA network paths given in Table II. The performance of the DA mechanism for all the ML models on two test path of the USA network (i.e., From Kanas City to Las-Vegas and Milwaukee to Minneapolis) is depicted in Fig. 7 and Fig. 8. In the results, it is explicitly demonstrated that the outcome of the predictions of the NN model follows the same trend as the actual values, and it performs better than all the other employed ML models. However, the prediction performance of KNN is again the worst among all the models. Based on these results, we build the following statement that the NN can generalize effectively in DA's case due to its potential to learn more complex patterns very well. It gets optimum results by continual adjustment of weights at each input layer.

To perform feature engineering, we implemented the Random Forest ML model to find out a set of features that are most important for the prediction of GSNR. In Fig. 9, several employed features (on the y-axis) and their importance score (on the x-axis) is illustrated concerning our label (i.e, GSNR in our case). It is demonstrated in Fig 9. that the total-distance is the most striking feature followed by the number of spans, ASE, NLI, and power for the estimation of GSNR of LP. After analyzing a set of important features, we trained the ML models on the relevant important features of channel 1 which is assumed as a Channel-Under-Test (CUT), and assessed the estimation of CUT against one test path of the EU network (SA scenario) and two test paths of USA network (DA scenario).

Firstly, we evaluated all the ML models' performance on a test path, i.e., Paris to Rome in the SA scenario. The result of the actual and estimated distribution of all the employed models is depicted in Fig.10. From the values of σ and μ, it is clear that NN's performance is excellent compared to other implemented ML models. Furthermore, in the case of the DA scenario, the distribution of actual against predicted GSNR for the CUT on the two test paths, i.e., Kanas to Las-Vegas and Milwaukee to Minneapolis, is illustrated in Fig. 11 and Fig. 12. The results in both Fig. 11 and Fig. 12 show that NN also performs best in the case of DA, whereas the performance of KNN is worst as compared to all the other employed ML models.

The comparison of ML models for the whole EU and USA network based on MAE is illustrated in Fig.13, when trained on cross features vs. relevant features. In the case of the EU network, Fig. 13 shows that all the three models' performance is improved with training on relevant features. When these models are tested on the USA network to examine NN's performance further considering both the cross feature and relevant feature training, the results in terms of MAE are demonstrated in Fig. 13. From the results, we observed that NN trained on the relevant feature generalizes well on

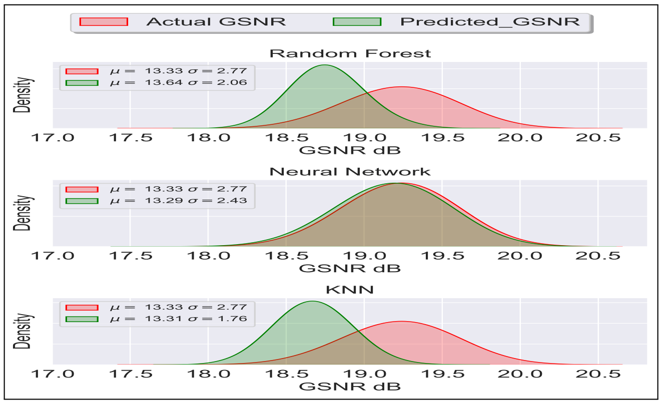

Fig. 12: USA network path Milwaukee to Minneapolis: Relevant features training.

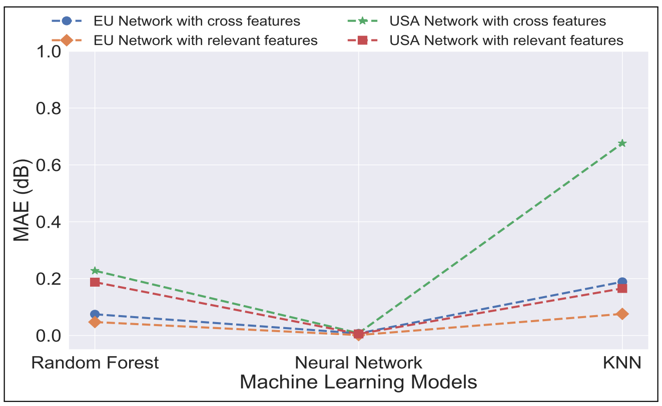

Fig. 13: Comparison of MAE of EU and USA Network

USA network paths and gives an outstanding performance. Moreover, the performance of KNN and Random Forest is also enhanced by employing relevant feature training.

From these observations, we conclude that the NN model trained on CUT's relevant-features is capable of generalizing effectively on an unseen-network with an outstanding performance. Generally, we observed that when ML models are trained with relevant features, their performance gets boosted.

VI. Conclusion

We explored various ML approaches to estimate the QoT of LP of an unseen network before its deployment. We utilized the GSNR of LP as a QoT metric. From our simulation results, we conclude that NN produces the best results for the EU and USA network with an MAE of 0.001 dB and 0.005 dB are achieved considering relevant feature training. For cross-features-based training with the EU network, we achieved MAE of 0.007 dB and 0.008 dB for the USA network. Overall, we observed that all the employed ML models' performance gets enhanced when relevant features are considered for training. The results clearly illustrate that ML-based approaches, particularly NN, notably diminish the provisioning GSNR-margin in both SA and DA scenarios.

References

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

[2] O. Gerstel, M. Jinno, A. Lord, and S. B. Yoo, "Elastic optical networking: A new dawn for the optical layer?" IEEE Communications Magazine **50**, s12–s20 (2012).

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

[4] V. Curri, A. Carena, A. Arduino, G. Bosco, P. Poggiolini, A. Nespola, and F. Forghieri, "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, G. Grammel, G. Galimberti, and V. Curri, "Multi-Vendor Experimental Validation of an Open Source QoT Estimator for Optical Networks," JLT **36**, 3073–3082 (2018).

[7] A. Ferrari, M. Filer, K. Balasubramanian, Y. Yin, E. Le Rouzic, J. Kundrát, G. Grammel, G. Galimberti, and V. Curri, "Gnpy: an open source application for physical layer aware open optical networks," JOCN **12**, C31–C40 (2020).

[8] M. Freire, S. Mansfeld, D. Amar, F. Gillet, A. Lavignotte, and C. Lepers, "Predicting optical power excursions in erbium doped fiber amplifiers using neural networks," in *2018 (ACP),* (IEEE, 2018), pp. 1–3.

[9] J. Thrane, J. Wass, M. Piels, J. C. Diniz, R. Jones, and D. Zibar, "Machine learning techniques for optical performance monitoring from directly detected pdm-qam signals," JLT **35**, 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),* (IEEE, 2018), pp. 1–3.

[11] L. Barletta, A. Giusti, C. Rottondi, and M. Tornatore, "Qot estimation for unestablished lighpaths using machine learning," in *OFC,* (OSA, 2017).

[12] J. Mata, I. de Miguel, R. J. Duran, N. Merayo, S. K. Singh, A. Jukan, and M. Chamania, "Artificial intelligence (ai) methods in optical networks: A comprehensive survey," OSN **28**, 43–57 (2018).

[13] E. Seve, J. Pesic, C. Delezoide, S. Bigo, and Y. Pointurier, "Learning process for reducing uncertainties on network parameters and design margins," Journal of Optical Communications and Networking **10**, A298–A306 (2018).

[14] I. Khan, M. Bilal, M. Siddiqui, M. Khan, A. Ahmad, M. Shahzad, and V. Curri, "Qot estimation for light-path provisioning in un-seen optical networks using machine learning," in *ICTON,* (IEEE, 2020).

[15] 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).

[16] C. Rottondi, L. Barletta, A. Giusti, and M. Tornatore, "Machine-learning method for quality of transmission prediction of unestablished lightpaths," Journal of Optical Communications and Networking **10**, A286–A297 (2018).

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

[18] D. Azzimonti, C. Rottondi, A. Giusti, M. Tornatore, and A. Bianco, "Active vs transfer learning approaches for qot estimation with small training datasets," in *Optical Fiber Communication Conference,* (Optical Society of America, 2020), pp. M4E–1.

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

[20] M. Filer, J. Gaudette, Y. Yin, D. Billor, Z. Bakhtiari, and J. L. Cox, "Low-margin optical networking at cloud scale," J. Opt. Commun. Netw. **11**, C94–C108 (2019).

[21] M. Cantono, D. Pilori, A. Ferrari, C. Catanese, J. Thouras, J.-L. Augé, and V. Curri, "On the Interplay of Nonlinear Interference Generation with Stimulated Raman Scattering for QoT Estimation," JLT **PP**, 1–1 (2018).

[22] F. N. Khan, Q. Fan, J. Lu, G. Zhou, C. Lu, and P. T. Lau, "Applications of machine learning in optical communications and networks," in *Optical Fiber Communication Conference,* (Optical Society of America, 2020), pp. M1G–5.