Convolutional neural network for quality of transmission prediction of unestablished lightpaths

Received <day> <Month>, <year>; | Revised <day> <Month>, <year>; | Accepted <day> <Month>, <year>

DOI: xxx/xxxx

ARTICLE TYPE

Convolutional Neural Network for Quality of Transmission Prediction of Unestablished Lightpaths

Fehmida Usmani[1,2] | Ihtesham Khan*[3] | Muhammad Umar Masood[3] | Arsalan Ahmad[1] | Muhammad Shahzad[1] | Vittorio Curri[3]

[1]School of Electrical Engineering & Computer Science, National University of Sciences & Technology, Islamabad, Pakistan

[2]Department of Information Technology, Khwaja Fareed University of Engineering & Information Technology , Rahim Yar Khan, Pakistan

[3]Department of Electronics & Telecommunications, Politecnico di Torino, Torino, Italy

Correspondence
*Ihtesham Khan, Corso Duca degli Abruzzi, 24, 10129 Torino, Italy.. Email: ihtesham.khan@polito.it

Present Address
Corso Duca degli Abruzzi, 24, 10129 Torino, Italy.

Abstract

Advance estimation of the Quality of transmission (QoT) of a Lightpath (LP) before its actual deployment is of capital importance for the network's operator in optical networks' optimized design. Due to the continuous advancement in evolving concepts of Software-defined networks (SDN) and Elastic-optical-network (EON), the number of design parameters is growing dramatically, making the LP deployment more complex. Typically, worst-case assumptions are utilized to calculate the QoT with provisioning of high-margin requirements. To this aim, precise and advanced estimation of the QoT of the LP is important for reducing this provisioning margin. In this investigation, we present Convolutional neural networks (CNN) based architecture to accurately calculate QoT before the actual deployment 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). The synthetic dataset is generated by utilizing well appraised GNPy simulation tool. Promising results are achieved, showing that the proposed CNN model considerably minimizes the GSNR uncertainty and, consequently, the provisioning margin.

KEYWORDS:
Machine Learning, Quality of Transmission, Convolutional Neural Network, Generalized SNR, Domain Adaption.

1 | INTRODUCTION

In the last decades, optical networks' design and management are continuously evolving to handle the rapidly increasing global internet traffic demands. The internet traffic has been growing continuously[1] with the advancement of new technologies and bandwidth-rigorous applications, such as video-on-demand, Full high definition (FHD) or 4K, and the Internet of things (IoT). This continuous increase in global internet traffic requires the full capacity exploitation of the already deployed network infrastructure. In this context, the underlying key-enabler technologies are coherent technology for optical transmission and Dense wavelength division multiplexing (DWDM) employed for spectral usage of fiber propagation. EON and SDN pave a path for the open and dis-aggregated optical network in addition to these technologies. The distinctive characteristics of EON and SDN provide flexible and dynamic resource provisioning in optical networks for both control and data planes[2,3]. EON introduces flexibility in the spectral assignment in the data plane and uplifts the network's capacity while minimizing network

[0]**Abbreviations:** CNN, Convolutional Neural Network; QoT, Quality of Transmission

cost. This resilience leads towards much more intricate LP-provisioning than typical fixed-grid WDM networks. However, the SDN-controller manages the operating-points of distinct network elements independently in the control plane, while enabling personalized network management. Today's optical networks have begun to evolve towards a partial dis-aggregation, with an eventual target of full dis-aggregation. The key step towards network dis-aggregation is examining the Optical-Line Systems (OLSs), which associate the network elements. At present, the proficiency of OLS controllers to handle the optimal working-point determines the deterioration in the QoT[4,5]. This working-point's precise accomplishment leads to a lower margin and higher traffic deployment rates; therefore, it is obligatory to employ the (Quality of transmission estimation) QoT-E for precise estimation of LP performance – the *path computation* – before its actual establishment. In this direction, QoT is effectively assessed by the GSNR, which incorporates the collective effect of both Non-linear interference (NLI) and Amplified spontaneous emission (ASE) noise respectively[6]. Utilizing the features of the transceiver, the GSNR reports the suitability of the path and the deploy-able rate. Conventionally, the network nodes tolerate a variation on their working point (ripples in amplifiers, insertion losses, noise, and gain figure, etc.). This leads to develop an uncertainty in QoT-E that needs a system-margin to avoid network out-of-services or network outage.

In the current study, we suppose a Domain-Adaption (DA) approach. The DA-approach make use of the available data from *source domain "S"* (e.g., well-deployed in-service network), where the network operator has the exact knowledge about the working-point of network elements and exploit the useful information to estimate the QoT of LP in the *target domain "T"* (e.g., a recently deployed or unseen-network), i.e., the network where the system administrator does not have the enough information of the operating point of network elements. The goal of this investigation is to diminish the margin in the GSNR-measurement of the *target domain*. This reduction in the GSNR unreliability permits the network-controller in *target domain* to precisely set up the LP with a low margin. Commonly, the controller can get an exact representation of the system-specifications i.e., *network status*. The QoT-E makes use of different analytical methods that can compute the GSNR with a very well accuracy as indicated in[7]. The application of an analytical method is not valid without the *exact* knowledge of system-specifications, as it is a pre-requisite to get system-specifications in the present scenario of DA. The latest work-frame about DA infers that the utilization of the analytic method is not recommended for predicting the QoT of LP before its set up in such an agnostic scenario. [8,9,10,11] To overcome this challenge, we choose to apply a data-driven approach as an alternative way, which has already been proved very efficacious for managing optical networks; as demonstrated[12,13,14]. A thorough analysis of various applications of ML in optical networks is described in[15]. The authors in[16] employed Neural network (NN) to distinguish the integrated-circuits and for their complete and precise softwarization. In specific to a distinct focus of this work, i.e., estimation of QoT of LP before its actual set up, some very efficacious ML-based techniques, for instance, the methodology based on Cognitive-case-reasoning (CBR) is described in[17]. In[18], a data-driven ML-based technique is demonstrated to handle OLS in an open environment. Different ML-based techniques are investigated in[19,8,9] for QoT-prediction of LP. Authors in[20] employed CNN for performance monitoring of optical-transport-network. In[21] 1-dimensional CNN model is proposed to estimate multi-step performance in operational optical-networks by utilizing bit-error-rate of data. In[10], the authors assessed the performance of two DA-based mechanisms for ML assisted QoT-E of an optical LP. The authors in[22], presented an ML-based technique for QoT-E along with the statistical closed-form method for QoT margin-setting. The authors in[23] presented a transfer-learning-based deep-neural-network architecture for OSNR monitoring. Finally, the authors in[11] studied the QoT-E accuracy given by a few Active-Learning (AL) and DA approaches on two distinct network topologies.

The remarkable distinction of the present study is that we proposed to employ CNN for the system-margin minimization of the *T* network considering the mimicked data of GSNRs response-to-specific traffic configurations of LPs of the *S* network in an open environment.There are a couple of motivations for employing CNN for this particular work. Most of the ML-based approaches discussed in the literature are required to perform feature extraction manually prior to their learning process. But CNN is capable of learning domain-specific features automatically. On the other hand, the conventional fully connected neural networks result in complex networks and are not capable of minimizing the spatial dependencies. whereas CNN has sparse connectivity with a reduced number of trainable parameters that leads to reduced computation complexity and memory requirements. The dataset-generation is carried out *synthetically* by employing the GNpy simulation tool against two distinct networks specified by dissimilar topologies employing the identical fiber-type and communication devices but are distinguished regarding the most exquisite specifications of amplifiers and fiber losses. Our simulation results show that CNN performs very well with an MAE of 0.18 dB for the *S* network and on average 0.2 dB for the *T* network.

The rest of the paper is structured as follows: In Section 2, the simulation performed to imitate an open-OLS and data-generation are reported. Section 3, shortly describes the argument that precise QoT-E in terms of GSNR has a prime role in the

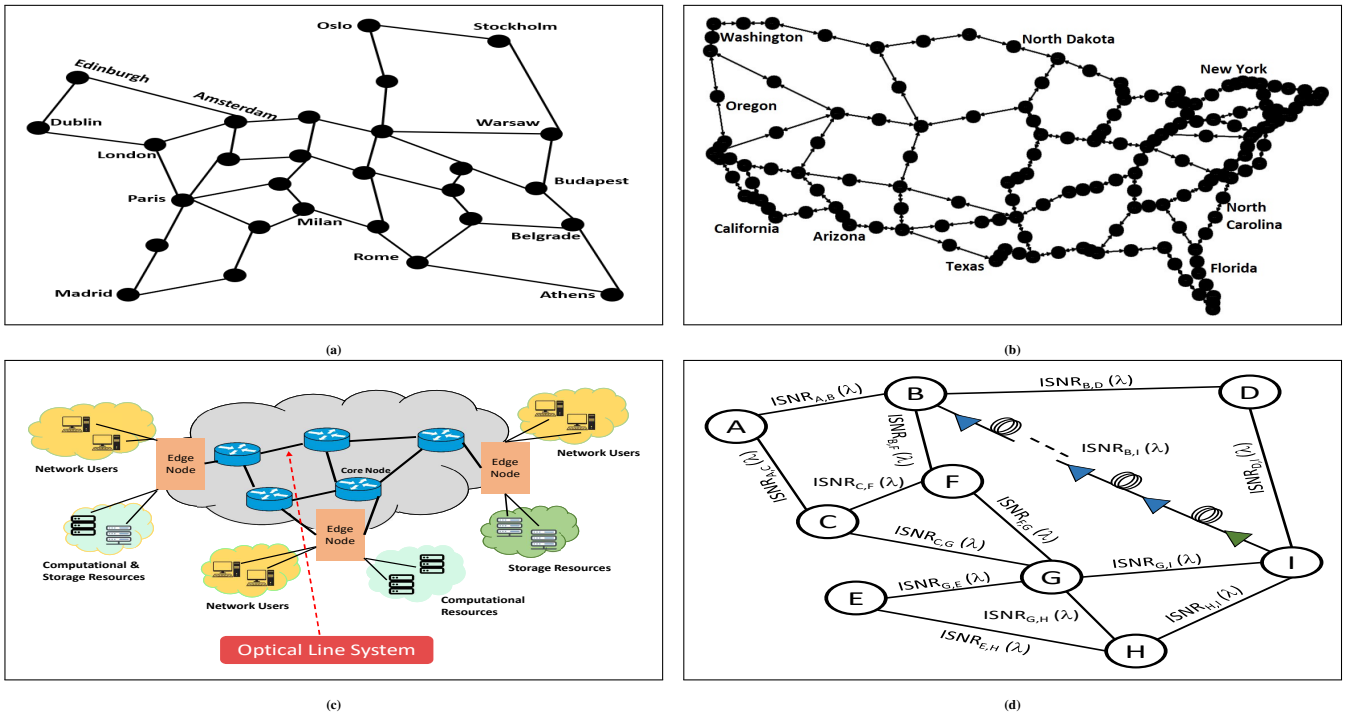

FIGURE 1 (a) European Network (b) USA Network (c) Architecture of Optical Network (d) Optical Transport Network

reduction of the system margin. In Section 4, we described the proposed CNN architecture, which is used in the context of the DA-based approach. Later, in Section 5, we presented detailed results. Finally, the conclusion is given in Section 6.

2 | SYSTEM MODEL & DATA GENERATION

The proposed work simulates an open OLS which incorporates cascaded-amplifiers and fibers. For the simulation setup, the grid size of 50GHz is assumed with 76 channels on the C-band. Only 76 channels over the total bandwidth of approximately 4 THz are examined because of limited computational-resources. 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 Erbium-doped fiber amplifier (EDFA) (EDFA), operating at a constant output power mode of 0 dBm per channel. The noise-figure of EDFA changes uniformly, between the range of 4.5 dB to 6 dB with a ripple-gain variation uniformly with 1 dB variation. Standard single-mode fiber (SSMF) is assumed for all the links with a total distance of approximately 80 km. In addition, 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[24,25]. The paths are computed using the Dijkstra algorithm, with the metric 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 polarizations. The nonlinear impairments are modeled by the analytical perturbation model, such as Generalized Gaussian Noise (GGN) model[26,27].

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 formulated on Gaussian Noise (GN) model[27,7]. 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; European (EU) network and USA-network demonstrated in Fig. 1 a and Fig. 1 b 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 Optical network elements (ONE). However, they are different concerning the amplifier's delicate parameters (noise and ripples gain) and fiber insertion losses. The dataset used in this work consists of 5 source-to-destination (s → d) pairs of EU network and 2 s → d pairs of USA-network presented in Tab. 1 . The spectral load realization against each simulated link

of a dataset is a subset of 2^{76}, where 76 represents the total number of channels. We considered 3000 realizations of arbitrary traffic flow varying between 34% to 100% of overall operational bandwidth for every s → d pair. Thus for EU network topology, 15,000 realizations are generated, and for the USA network topology, 6,000 realizations are generated. The considered dataset is then *normalized* to scale the values. For normalization, Z-score normalization is used: $z = \frac{X-\mu}{\sigma}$ 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.

TABLE 1 Dataset Description

Source-Destination pairs and Number of Spans of European and USA Network					
European Network			USA Network		
Source	**Destination**	**Spans**	**Source**	**Destination**	**Spans**
Amsterdam	Berlin	8	Orlando	Philadelphia	30
Brussels	Bucharest	30	Houston	Jacksonville	6
Frankfurt	Istanbul	34			
Vienna	Warsaw	7			
London	Madrid	34			

3 | GSNR AS QOT-ESTIMATION METRIC

In general, an optical network composed of Optical-Networks-Elements (ONE) coupled via two-way fiber links, where traffic is routed, add on/dropped as illustrated in Fig. 1 c. The amplifiers are laid down following a certain span distance considering the EDFAs/Raman amplification or hybrid of both. In the advance optical network, ONE coupled via fibers is typically demonstrated with a discrete controller and an OLS with the particular specifications to place the operational-point for every single amplifier passing through the link. Moreover, the transport-layer services are implemented utilizing Re-configurable Optical Add/Drop Multiplexers (ROADM). As stated in recommendations of International telecommunication union (ITU-T)[28], the spectral-usage of DWDM technology is capable of exploiting both fixed or variable spectral-grid that distinguish the spectral-slots for both grid architectures[29]. Exploiting either architecture of grid, LPs can be established, where each LP represents the logical abstraction of suitable links between node-to-node concerning traffic demands. Further to this, a Polarization-Division-Multiplexing (PDM) is utilized over each established LP to propagate from a specified source to destination. In addition to the transmission, LP tolerates various impairments for instance ASE-noise, fiber propagation, and filtering retributions implemented by ROADM. Also, it has been widely reported in the literature that during fiber propagation, the QoT of the lightpath is affected by the amplitude and phase noise.[30,31,4,32]. This incorporated phase-noise is effectually counter-weighted by the receiver's DSP-module, applying a carrier-phase estimator. These specific noises are particularly examined for a very small distance along with a high symbol-rate communication model[32].On the contrary, the amplitude noise, commonly known as NLI consistently degrades the performance. Finally, due to the penalty of ROADMs-filtering the level of QoT level decreases, which is commonly measured as an additional loss.

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

The QoT-E metric for a particular LP routed by definite OLSs from source to destination is given by the well-acknowledged GSNR measurement, which combines both the aggregated effect of NLI disturbance and ASE noise. Generally, GSNR is determined by Eq. 1, where $\mathrm{OSNR} = P_{\mathrm{Rx}}/P_{\mathrm{ASE}}$, $\mathrm{SNR}_{\mathrm{NL}} = P_{\mathrm{Rx}}/P_{\mathrm{NLI}}$, P_{Rx} is the power of the signal at certain channel at the receiver, P_{ASE} denotes the ASE noise power and P_{NLI} denotes the NLI power. Analyzing the specifications of the transceiver, the GSNR accurately provides the BER, as BER is a common terminology stated by different vendors during the demonstration of industrial products[6]. The non-linear effects P_{NLI} generated during fiber propagation 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 QoT, the abstraction of OLS is represented as a single variable 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 operating point. Therefore, an optical network is typically abstracted as a weighted graph (W),where W is represented by vertices (V) and edges (E). The V of W are actually ROADM nodes, whereas

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

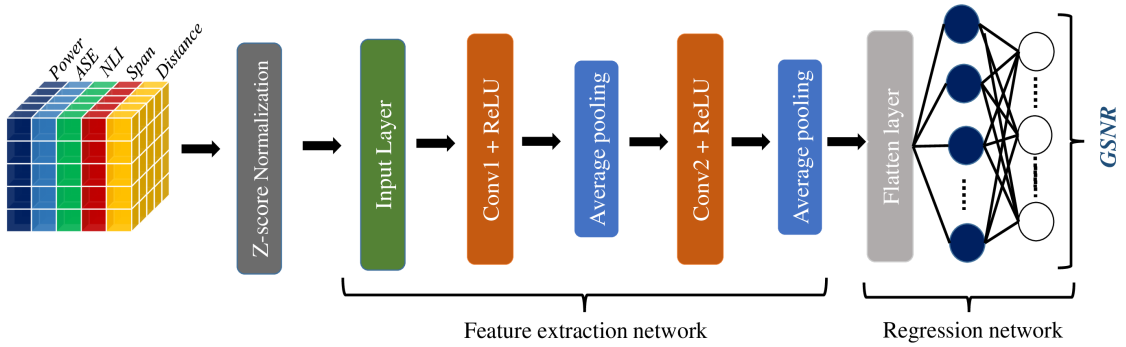

FIGURE 2 Schematic of CNN architecture for QoT-E of LPs using Power, ASE, NLI, Span-length and Distance as features.

the E represent OLSs with $\mathrm{GSNR}_i(f)$ represented as the weights on a particular edge, shown in Fig. 1 d. Specifically, for a given LP from the source node I to destination node F that passes through intermediary nodes B, the QoT defines as Eq. 2. Following abstraction at network-level, LPs deployment could be feasible for a specific source node to destination node with the reduced margin, which relies on the *GSNR* of a specific source to the destination path.

4 | CNN ARCHITECTURE FOR GSNR ESTIMATION

The ML paradigm, particularly the CNN which is a subset of deep learning, provides striking attributes that cannot be precisely measured using analytical models. Typically, ML models achieve cognitive ability by exploiting different perceptive sets of rules to obtain the training data's inherent information. The trained model explores the abstraction of inherent knowledge to execute logical decisions during the testing phase. Commonly CNN is a well-acknowledged model to perform best with image data. The proposed CNN-based model works with numerical data to explore its effectiveness for GSNR estimation of unestablished LP.

In the present work, the dimensionality of the problem makes it more complex to apply the frameworks of the fully connected NN; as for each channel $\mathcal{N}_{CH}$, we consider five distinct feature $\mathcal{N}_F$ so the total number of mapping features are $\mathcal{N}_F \times \mathcal{N}_{CH}$ (i.e, ($\mathcal{N}_F$=4 $\times$ $\mathcal{N}_{CH}$=76) + span=1). The mapping between such a large number of input features and the respective GSNR configuration needs a system with many trainable factors, which increases the training time and will be easily susceptible to problems such as over-fitting and local minimum. Also, using the conventional approach of a fully connected network would create unnecessary complexity as it would not be able to get benefit from the hidden correlation between the input data. Regarding these two key problems, we concluded CNN as more remarkable, as they have been designed to process data coming in multiple arrays, as images. Moreover, they can effectively encapsulate the spatial and temporal dependencies in two-dimensional form data by applying relevant filters[33] and weight sharing.

TABLE 2 Structural and Optimization Statistics of CNN

(a) Specifications of Proposed CNN Architecture

Layer	Type	Number of neurons	Kernel-size	Stride
1	Input	305	-	-
2	Convolutional	64	3	1
3	Avg-pooling	64	2	2
4	Convolutional	64	3	1
5	Avg-pooling	64	2	2
6	Flatten	-	-	-
7	Fully Connected	80	-	-
8	Output	1	-	-

(b) Performance Comparison of Different Activation Functions

Activation Function	MAE (dB)
Linear	0.40
Tanh	0.25
ReLU	0.18
Sigmoid	0.31
-	-
-	-
-	-
-	-

For the CNN-based model studied in this investigation, we consider a set of features and number of samples as a 2-dimensional input to the network intending to estimate the GSNR of an unestablished LP illustrated in Fig 2 . The proposed framework comprises two network stages for end-to-end training, i.e., feature-extraction network and regression-network. First, the pre-processing of the input data is performed using z-score normalization. In this work, we utilize a 2-dimensional dataset consist of 12,000 rows (number of samples) and 305 columns (set of features), respectively. In order to obtain compatible 3-dimensional

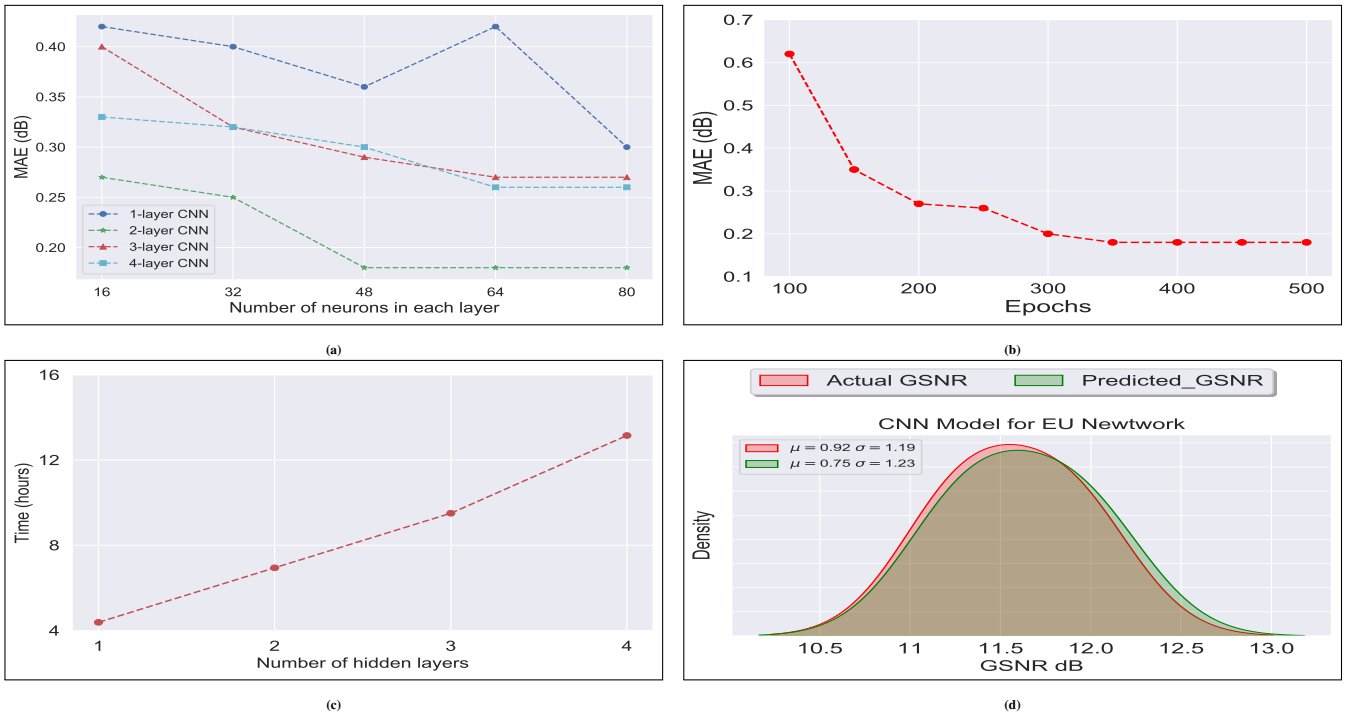

FIGURE 3 (a) MAE against different layer-configurations (b) MAE against Epochs (c) Training time against number of layers (d) EU network path London to Madrid

data, we reshaped our data dimensions from 2-dimension (12000 x 305) to 3-dimension (12000 x 305 x 1). After that, the 3-dimensional normalized data is directed to the feature-extraction network which consists of a total of three layers an input layer and two convolution layers (Conv-layers). Conv-layers are used as a feature detector to extract significant features from the input data for better prediction. In addition to this, each conv-layer is using *ReLU* activation function to accelerate the training process. Moreover, the average-pooling layers are placed between succeeding Conv-layers to carry out spatial-pooling. The main purpose of pooling layers is to lower the spatial size of input feature-maps, leading to the reduced number of parameters and computational-complexity. It is worth mentioning that the reduction in spatial size is equivalent to the kernel size of the pooling layer that is 2×2 in our case and it reduces the spatial size of input features by a factor of 2. Thus, each layer produces a compact and informative description of input features. The output of the feature-extraction network is a 3-dimension representation of input data. Moreover, the flattening layer is used to convert 3-dimensional representation into a 1-dimensional array of feature vectors and then this feature vector is finally passed to the regression network. The purpose of this network is to map extracted features to the GSNR of LP. This network includes one fully connected layer and an output layer with 1 neuron to output the estimated GSNR of LP. The detailed parameters of our proposed CNN architecture are given in Tab. 2 a.

5 | RESULTS & DISCUSSION

In this section, we first assess the performance of our proposed CNN model for the Same Domain (SD) scenario and explore the effect of different CNN-layer configurations on the model performance. Then, we further evaluate the performance of our proposed model in the DA scenario.

Initially, we train the CNN model on the EU network ("S" network), considering 2 hidden layers with 32 neurons in each layer for the SD scenario. We utilized 12,000 samples of the EU network for training and 3000 samples from the EU network for testing.The utilized set of features to describe CNN models' features include power, ASE-noise, span-length, NLI, and the total distance of the path. The proposed CNN model's performance is assessed using the Mean-absolute-error (MAE) metric; it quantifies the GSNR-predictions of the CNN model by getting the mean-absolute difference of all the estimated values with the actual values. The initial configuration of our proposed architecture illustrated in Fig. 2 provides an MAE of 0.32 dB. Also, to obtain a more accurate prototype, various configurations of the proposed CNN layout are investigated. In the prosecution of this

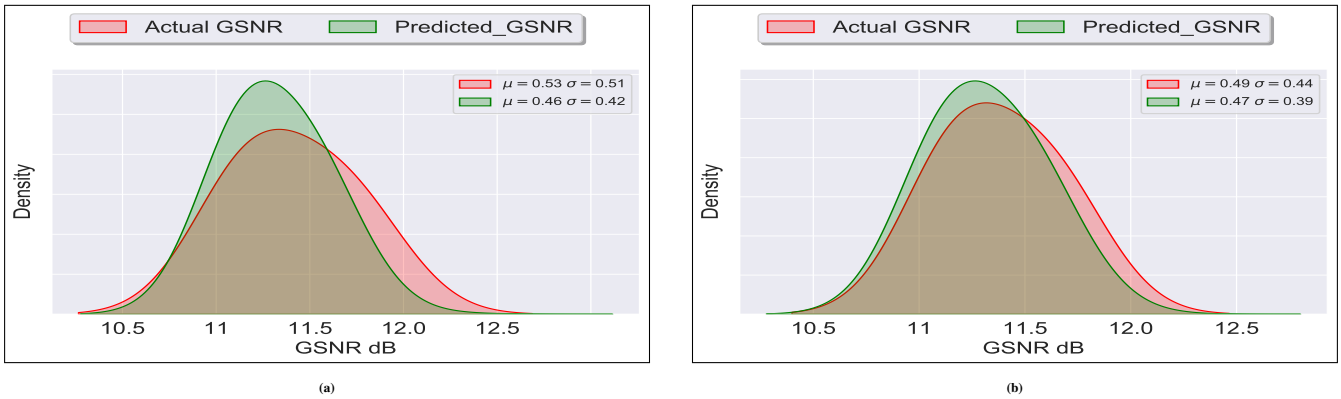

FIGURE 4 (a) USA network path Houston to Jacksonville (b) USA network path Orlando to Philadelphia

objective, Fig. 3 a demonstrates the MAE of the predicted GSNR against different CNN architectural designs with an increasing number of neurons for each layer. Observing Fig. 3 a, the architecture with 2-layers produces better results in terms of MAE than the more dense architecture. The addition of extra hidden layers leads to a more extensive and complex network layout. In addition to this, Fig. 3 a also reveals that with the increase in the number of neurons for each layer ranging from 32 to 64, the model's performance enhances. Moreover, when we further increase the number of neurons, the CNN-model performance does not get better any further with the given number of 12,000 training samples. Therefore, we choose to analyze the performance of CNN architecture with 2 hidden-layers and 64 neurons for evaluating both the SD and DA scenarios.

The proposed model is developed by utilizing the python-based higher-level Application program interfaces (APIs) of the open-source TensorFlow© platform, particularly Keras library. Furthermore, we consider four activation functions for this investigation to assess their impact on model performance (see Tab. 2 b). In Tab. 2 b the comparison of four considered activation functions exhibits that the model with the ReLU as an activation function performs better, having lowest MAEs compared to other models[34]. Moreover, we utilized the default Adaptive moment estimation (ADAM) optimizer for this study. To avoid over-fitting the model, we set the epochs as the stopping factor. The training phase is accomplished by executing several epochs until there is no further improvement in the model's performance. The performance concerning the training step is depicted in Fig. 3 b, which demonstrates that, at the start of the training procedure, MAE is very high (i.e., 0.62 dB), and it gradually decreases with the increase in the number of epochs till specific limit (350 epochs in this case), further incrementation in epochs do not improve the system performance. Finally, we found that the proposed model converged at 350 epochs in the present simulation scenario. The considered CNN model is executed on a system with an Intel® Core™ i7 8550U 1.80 GHz CPU workstation equipped with 8 GB of RAM. The CNN model's training time against several layers is demonstrated in Fig. 3 c. The proposed CNN model takes $\approx$ 7 hours to train with 2 hidden layers, 64 neurons for 350 epochs.

The proposed CNN model's performance is evaluated by using the MAE metric for SD approach, where the CNN model is trained on some EU-Network paths and tested on another path of EU-network. The first four paths of the EU Network are used to train the CNN models, and the last path is used for testing the models reported in Tab. 1 . We utilized all the given features of the 76 channels for training the CNN model to estimate the GSNR of single channel under test (channel-1 is considered in this case). The result of the test path, from Paris to Rome, is depicted in Fig. 3 d. Fig. 3 d shows the Predicted GSNR value against Actual GSNR with a mean (μ) and standard deviation (σ). Observing the statistics μ and σ, it is pretty much clear that the CNN model performs very well in terms of GSNR prediction. The CNN exploits its ability of dimensionality reduction to perform automatic learning of the different correlations among a large set of features. In CNN, the dimensionality reduction is achieved by synergic use of convolution and pooling operations. This reduction in dimension of a massive set of input features substantially minimizes the risk of over-fitting and computational complexity.

We further evaluate the CNN model's performance in DA scenario, where the model is trained on four paths of EU-network and then tested on two paths of USA-network given in Tab. 1 . The result of the DA approach is illustrated in Fig. 4 a and Fig. 4 b. These results demonstrate the proposed CNN models' prediction performance against the two paths, i.e., Houston to Jacksonville and Orlando to Philadelphia, of the USA network. Observing the result statistics shows that the predicted-GSNR values with the CNN model seem to follow the same distribution as the real GSNR-values.The error in predicting GSNR (ΔGSNR) is defined as : $\Delta\text{GSNR} = \text{GSNR}_{Predicted} - \text{GSNR}_{Actual}$. The ΔGSNR μ and σ of the worst predicted path (Houston to Jacksonville) of USA network is -0.1375 dB and 0.124 dB. The maximum error in predicting GSNR using the proposed CNN is ($\Delta\text{GSNR}^{max} = 0.372$ dB), and is estimated by considering 3σ of the worst predicted path of the USA network. From the obtained

results, we concluded that the CNN model also performs very well in DA scenario due to its capability to learn complex hidden patterns, resulting in better generalization.

6 | CONCLUSION

The advanced prediction of the QoT of LP before its actual deployment has techno-economic importance for the network's operator during the design and operating phase of optical networks. In this context, a CNN-based framework is proposed in the SD and DA scenarios for precise LP estimation before the network's actual deployment. The proposed CNN architecture consists of two networks: 1) feature-extraction with input and 2 Conv-layers to extract useful features and 2) a regression network to estimate GSNR of LP before its actual provisioning in a network. Our simulation results show that the proposed framework performs very well in predicting GSNR for both the SD and DA scenarios scenario.

Acknowledgements

This publication has been produced with co-funding of the European Union for the Asi@Connect Project under Grant contract ACA 2016-376-562.

References

[1] Cisco . Cisco Visual Networking Index: Forecast and Trends, 2017–2022. tech. rep., Cisco; 2017.

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

[3] Zhang G, De Leenheer M, Morea A, Mukherjee B. A survey on OFDM-based elastic core optical networking. *IEEE Communications Surveys & Tutorials* 2012; 15(1): 65–87.

[4] Curri V, Carena A, Arduino A, et al. Design Strategies and Merit of System Parameters for Uniform Uncompensated Links Supporting Nyquist-WDM Transmission. *JLT* 2015; 33(18): 3921–3932. doi: 10.1109/jlt.2015.2447151

[5] Pastorelli R. Network Optimization Strategies and Control Plane Impacts. In: OSA; 2015

[6] Filer M, Cantono M, Ferrari A, Grammel G, Galimberti G, Curri V. Multi-Vendor Experimental Validation of an Open Source QoT Estimator for Optical Networks. *JLT* 2018; 36(15): 3073–3082. doi: 10.1109/jlt.2018.2818406

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

[8] Khan I, Bilal M, Curri V. Advanced formulation of QoT-estimation for un-established lightpaths using cross-train machine learning methods. In: IEEE. ; 2020.

[9] Khan I, Bilal M, Curri V. Assessment of cross-train machine learning techniques for QoT-estimation in agnostic optical networks. *OSA Continuum* 2020; 3(10): 2690–2706.

[10] Di Marino R, Rottondi C, Giusti A, Bianco A. Assessment of Domain Adaptation Approaches for QoT Estimation in Optical Networks. In: IEEE. ; 2020: 1–3.

[11] Azzimonti D, Rottondi C, Giusti A, Tornatore M, Bianco A. Active vs transfer learning approaches for QoT estimation with small training datasets. In: Optical Society of America. ; 2020: M4E–1.

[12] Freire M, Mansfeld S, Amar D, Gillet F, Lavignotte A, Lepers C. Predicting Optical Power Excursions in Erbium Doped Fiber Amplifiers using Neural Networks. In: IEEE. ; 2018: 1–3.

[13] Thrane J, Wass J, Piels M, Diniz JC, Jones R, Zibar D. Machine learning techniques for optical performance monitoring from directly detected PDM-QAM signals. *JLT* 2017; 35(4): 868–875.

[14] Khan FN, Lu C, Lau APT. Optical performance monitoring in fiber-optic networks enabled by machine learning techniques. In: IEEE. ; 2018: 1–3.

[15] Mata J, Miguel dI, Duran RJ, et al. Artificial intelligence (AI) methods in optical networks: A comprehensive survey. *OSN* 2018; 28: 43–57.

[16] Khan I, Chalony M, Ghillino E, et al. Machine learning assisted abstraction of photonic integratedcircuits in fully disaggregated transparent optical networks. In: IEEE. ; 2020.

[17] Jiménez T, Aguado JC, Miguel dI, et al. A cognitive quality of transmission estimator for core optical networks. *JLT* 2013; 31(6): 942–951.

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

[19] Khan I, Bilal M, Siddiqui M, et al. QoT estimation for light-path provisioning in un-seen optical networks using machine learning. In: IEEE. ; 2020.

[20] Tanimura T, Hoshida T, Kato T, Watanabe S, Morikawa H. Convolutional neural network-based optical performance monitoring for optical transport networks. *IEEE/OSA Journal of Optical Communications and Networking* 2019; 11(1): A52-A59.

[21] Mezni A, Charlton DW, Tremblay C, Desrosiers C. Deep Learning for Multi-Step Performance Prediction in Operational Optical Networks. In: ; 2020: 1-2.

[22] Khan I, Bilal M, Umar Masood M, D'Amico A, Curri V. Lightpath QoT computation in optical networks assisted by transfer learning. *IEEE/OSA Journal of Optical Communications and Networking* 2021; 13(4): B72-B82. doi: 10.1364/JOCN.409538

[23] Xia L, Zhang J, Hu S, Zhu M, Song Y, Qiu K. Transfer learning assisted deep neural network for OSNR estimation.. *Optics express* 2019; 27 14: 19398-19406.

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

[25] Filer M, Gaudette J, Yin Y, Billor D, Bakhtiari Z, Cox JL. Low-margin optical networking at cloud scale. *J. Opt. Commun. Netw.* 2019; 11(10): C94–C108. doi: 10.1364/JOCN.11.000C94

[26] Cantono M, Pilori D, Ferrari A, et al. On the Interplay of Nonlinear Interference Generation with Stimulated Raman Scattering for QoT Estimation. *JLT* 2018; PP(99): 1-1. doi: 10.1109/JLT.2018.2814840

[27] Poggiolini P. The GN Model of Non-Linear Propagation in Uncompensated Coherent Optical Systems. *J. Lightwave Technol.* 2012; 30(24): 3857–3879.

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

[29] Ahmad A, Bianco A, Chouman H, Curri V, Marchetto G, Tahir S. Exploring the effects of physical layer parameters in WDM based fixed and flexible-grid networks. In: ; 2015: 128-133

[30] Elson DJ, Saavedra G, Shi K, et al. Investigation of bandwidth loading in optical fibre transmission using amplified spontaneous emission noise. *Optics express* 2017; 25(16): 19529–19537.

[31] Nespola A, Straullu S, Carena A, et al. GN-model validation over seven fiber types in uncompensated PM-16QAM Nyquist-WDM links. *IEEE Photonics Technology Letters* 2013; 26(2): 206–209.

[32] Pilori D, Forghieri F, Bosco G. Residual Non-Linear Phase Noise in Probabilistically Shaped 64-QAM Optical Links. In: ; 2018.

[33] LeCun Y, Bengio Y, Hinton G. Deep learning. *nature* 2015; 521(7553): 436–444.

[34] Nwankpa C, Ijomah W, Gachagan A, Marshall S. Activation functions: Comparison of trends in practice and research for deep learning. *arXiv preprint arXiv:1811.03378* 2018.