# Combining QoT Regression and Modulation Format Classification for Reliable Lightpath Provisioning in Elastic Optical Networks



Citation for the original published paper (version of record):
Natalino Da Silva, C., Arpanaei, F., Lechowicz, P. et al (2024). Combining QoT Regression and Modulation Format Classification for Reliable Lightpath
Provisioning in Elastic Optical Networks. XLII Brazilian Symposium on Telecommunications and Signal Processing



(article starts on next page)

# Combining ML Regression and Classification for Reliable QoT-Aware Lightpath Provisioning in Elastic Optical Networks

Carlos Natalino, Piotr Lechowicz, Farhad Arpanaei, and Paolo Monti

***Abstract*— The dynamic provisioning of lightpaths in elastic optical networks (EONs) requires the decision of which modulation format (MF) to be used by the lightpath. This involves estimating the quality of transmission (QoT) of the unestablished lightpath. Machine learning (ML) has been used as an effective QoT estimator in the presence of uncertain physical layer parameters. However, minor inaccuracies in the estimation may lead to the incorrect modulation format selection, which degrades the reliability of decisions during lightpath provisioning. In this paper, we analyze this issue and propose the use of two ML models, i.e., regression for generalized signal-to-noise ratio (GSNR) estimation, and multi-class modulation format classification. Combined, the models reduce the incorrect modulation format selection compared to the cases using a single model.**

***Keywords*— Machine learning, quality of transmission, unestablished lightpaths.**

## I. Introduction

Dynamic lightpath provisioning in elastic optical networks (EONs) consists of estimating the quality of transmission (QoT) of a new, unestablished, lightpath and solving the routing, modulation format, and spectrum assignment (RMSA) problem considering the existing lightpaths in the network. The decision of which modulation format (MF) depends mainly on the estimated QoT. In the literature, numerous analytical models can be used to estimate the QoT of lightpaths [1], [2]. However, such models require a precise input of the physical layer parameters of the network, which may not be available.

Machine learning (ML) has been investigated as a solution to overcome the uncertainty of physical layer parameters [3] in the QoT estimation of unestablished lightpaths. The QoT estimation problem has been formulated as binary classification [4], [5], [6] or regression [7], [8]. In binary classification, the model is executed for each MF, and the one classified as working and with the highest MF order is selected. In regression, the model outputs the generalized signal-to-noise ratio (GSNR), which is compared to known thresholds for each MF, selecting the one with the highest order that requires at most the obtained GSNR. To the best of our knowledge, the direct multi-class MF classification of unestablished lightpaths has not yet been addressed in the literature. However, the online modulation format identification (MFI) of established lightpaths has already been explored in the literature but represents an entirely different problem [9], [10].

Once a QoT estimation approach has been adopted, the dynamic lightpath provisioning requires solving the RMSA problem. In this problem, a route, MF, and sufficient spectrum need to be allocated. A particular requirement is that the QoT achieved by the lightpath needs to fulfill the minimum requirements of the selected MF. A reliable provisioning will always select a suitable MF. Failure to do so may lead to two issues. If the MF selection process makes an aggressive selection, i.e., selecting an MF that requires higher QoT than achieved, an immediate re-provisioning is needed. On the other hand, a conservative selection may lead to lower spectral efficiency but can be mitigated at a later stage, e.g., during a maintenance window.

In this work, we propose a new approach to the ML-based MF selection that aims at enhancing the reliability of the lightpath provisioning process by reducing the number of aggressive MF selections. We adopt two ML models, one for QoT regression, and another for MF classification, which are used for the same unestablished lightpath. An alternative approach that adopts regression and a margin is also investigated. Then, we propose a new algorithm that, based on the output of the two ML models, defines which MF should be adopted by the unestablished lightpath. Results show that the proposed approach reduces the number of aggressive MF selections by 36% and 44% when compared to traditional approaches. Moreover, the total number of wrong MF selections is lower than the one achieved by regression with or without margins. The results indicate that the evaluation of accuracy metrics of ML-based QoT estimation strategies falls short of fully illustrating their impact on network operation and that deeper analyses are necessary.

C. Natalino, P. Lechowicz, and P. Monti are with the Department of Electrical Engineering, Chalmers University of Technology, 412 96 Gothenburg, Sweden, e-mail: {carlos.natalino, piotr.lechowicz, mpaolo}@chalmers.se. F. Arpanaei is with the Department of Telematic Engineering, Universidad Carlos III de Madrid (UC3M), Leganés, Madrid, 28911 Spain, e-mail: farhad.arpanaei@uc3m.es.

Farhad Arpanaei acknowledges support from the CONEX-Plus programme funded by Universidad Carlos III de Madrid and the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No. 801538.

## II. ML-based Modulation Format Selection in Elastic Optical Networks (EONs)

In this work, we investigate the use of ML models to aid the MF selection during the dynamic provisioning of lightpaths in EONs. In this scenario, a lightpath request contains a source-destination node pair, and the bit rate. Upon the arrival of a

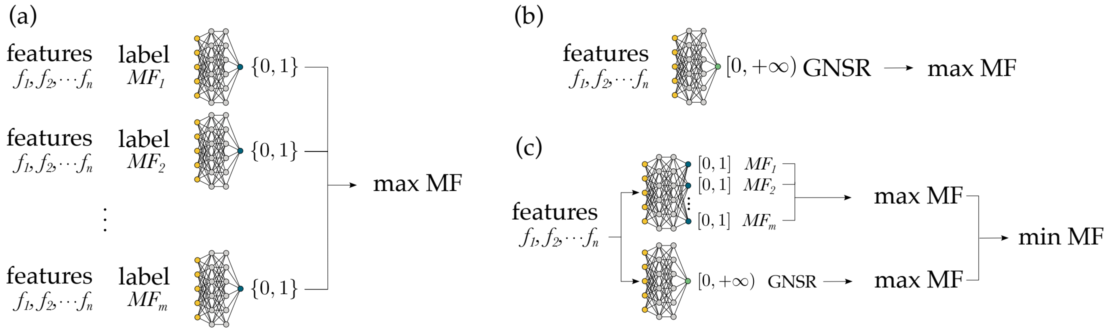



Fig. 1: The two traditional ways to select the modulation format based on ML models: (a) binary classification and (b) GSNR regression. The proposed approach (c) uses multi-class classification in combination with GSNR regression.

lightpath request, the RMSA problem needs to be solved. This entails the selection of a sequence of links (route), the MF to be used by the lightpath, and a set of continuous and contiguous frequency slots (spectrum) sufficient to accommodate the requested bit rate. The decision of the route and spectrum has been the topic of numerous research activities [11] and is out of the scope of this paper.

In this paper, we focus on the ML-based MF selection, which mainly depends on the estimated QoT of the lightpath, once established. This is achieved by training the ML model with the result of previous lightpath provisioning instances. The MF selection depends mainly on the QoT to be achieved by the lightpath once established. The QoT of lightpaths is usually modeled as the GSNR, and each MF establishes a minimum required GSNR which guarantees an acceptable bit error rate (BER). In the literature, the ML-based QoT estimation has been modeled as two kinds of problem: binary classification [4], [5], [6] and regression [7], [8]. These modeling alternatives share a common property: the ML model receives as input a set of features $F = \{f_1, f_2, f_3, ..., f_n\}$ that represent the state of the network and potential lightpath configuration.

Fig. 1(a) illustrates how binary classification can be used for MF selection. An ML model receives a set of features $F$ and the potential MF to be used. The model gives a binary output indicating whether or not that configuration and MF will work. In this case, one inference is needed for each considered modulation format. After collecting the results, the most efficient ML that works can be selected.

Fig. 1(b) illustrates how regression can be used for MF selection. An ML model receives a set of features $F$. The model gives a numerical output that represents the estimated lightpath GSNR. Then, the MF can be performed by selecting the most efficient MF possible for the estimated GSNR. Often a margin is considered to account for inaccuracies in the GSNR estimation, among other effects [12].

Two issues may arise from the incorrect MF selection: conservative or aggressive selection. In a conservative selection, the selected MF has a lower efficiency than the one that could be achieved. This leads to a higher spectrum usage than necessary, impacting the overall spectral efficiency of the network. In an aggressive selection, the selected MF requires a higher GSNR than the one achieved. This leads to a higher BER than allowed, resulting in degraded channel quality with excessive data transmission errors.

Upon provisioning the lightpath and detecting one of these cases, the solution is to re-provision the lightpath with the correct MF. However, in this paper, we consider aggressive selection as the most detrimental one, given that it requires immediate re-provisioning. Moreover, since a less efficient MF will be needed, it might happen that the current path and spectrum selection do not have sufficient free spectrum, which will trigger a complete re-computation of the RMSA solution. Meanwhile, the re-provisioning of a conservative selection can be delayed to a service window, and since the new MF will be more efficient, there are enough spectral resources.

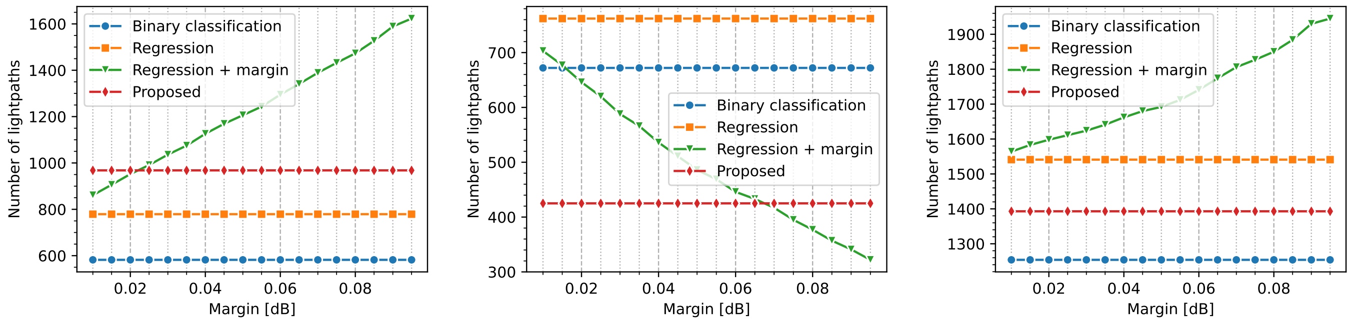
Fig. 1(c) shows the approach proposed in this paper. The first model is a multi-class classifier, unlike the binary classifiers previously used in the literature, responsible for directly outputting the most suitable MF for the lightpath. The second model is a traditional regressor that outputs the estimated QoT (e.g., the GSNR) of the lightpath and can be processed in the same way as in Fig. 1(b) for selecting the best possible MF. The intuition is that the two models will output, each, the best MF achievable according to their training. Given that we assume that an aggressive selection is more detrimental than a conservative one, the output of the two models is combined, and the most conservative decision is adopted.

## III. Performance Assessment

In this section, we evaluate the performance of MF selection using the three ML-based approaches discussed in the previous section. Firstly, we discuss the generation of the dataset using an accurate analytical model. Then, we present and discuss the results. The results are evaluated in terms of the number of lightpaths affected by the wrong MF selection. We evaluate the three strategies discussed in the paper, and how the introduction of a margin to the regression output affects the number of wrong selections, both in terms of aggressive and conservative MF selections. Then, the confusion matrices show the true and predicted MF to be used by the lightpath, and it is possible to identify the aggressive and conservative estimations.

### A. Dataset Generation

The ground truth dataset used in this work was generated by adopting the enhanced Gaussian noise (EGN) analytical model




(a) Conservative (b) Aggressive (c) Total

Fig. 2: Number of lightpaths with wrong MF selection – in terms of (a) conservative, (b) aggressive, and (c) total selections – considering a strategy where a margin is introduced to the regression output.

described in [2]. In this model, the GSNR of a lightpath can be calculated as a function of the set of spans $S$ along the path:

$$GSNR^{-1} = \sum_{s \in S} \left( \frac{P^s}{P^s_{\text{ASE}} + P^s_{\text{NLI}}} \right)^{-1} \quad (1)$$

where, for each span $s \in S$, $P^s$ is the launch power, $P^s_{\text{ASE}}$ is the noise incurred by amplifier spontaneous emissions (ASE), and $P^s_{\text{NLI}}$ is the noise incurred by non-linearity. We use a simulation of the dynamic provisioning of lightpaths. The arrivals follow a Poisson process, where source and destination nodes are uniformly distributed over the nodes in the network topology. The bit rate of the request is uniformly chosen from {10, 40, 100, 400} Gbps. The RMSA problem is solved upon the arrival of each lightpath request. We pre-compute 5 shortest paths for each node pair in the network. The path is randomly selected. We consider six modulation formats: binary phase-shift keying (BPSK), quadrature phase-shift keying (QPSK), 8-, 16-, 32-, and 64-quadrature amplitude modulation (QAM), identified with labels 1–6, with their GSNR threshold set to {3.71, 6.72, 10.84, 13.24, 16.16, and 19.01} dB, respectively. The best modulation format is selected based on the ground truth model, followed by first-fit spectrum allocation.

The following results are generated from the European network topology with 28 nodes and 41 links. Spans within a link have equal length, with a maximum of 80 km. We assume a launch power of 0 dBm across all spans, with a 0.2 dB/km fiber attenuation, and a 4.5 dB noise figure for each amplifier. We simulate the arrival of 100,000 requests, over which 94,109 requests were successfully provisioned.

The following 17 features compose the input to the dataset, divided into lightpath, node, and path features. Lightpath features are the bit rate and center frequency. Node features are the source and destination nodes, and source and destination node degrees. Path features are the total length, number of hops, number of spans, mean, minimum, and maximum link length, mean, minimum, and maximum link usage, and standard deviation of the link usage. These features are normalized, and the target features are set according to the task. Then, the dataset is divided into balanced training and testing datasets in a 50/50 manner. The training set is used for training three ML models: binary classification, regression, and multi-class classification. The testing set is used to obtain the results discussed in the following.

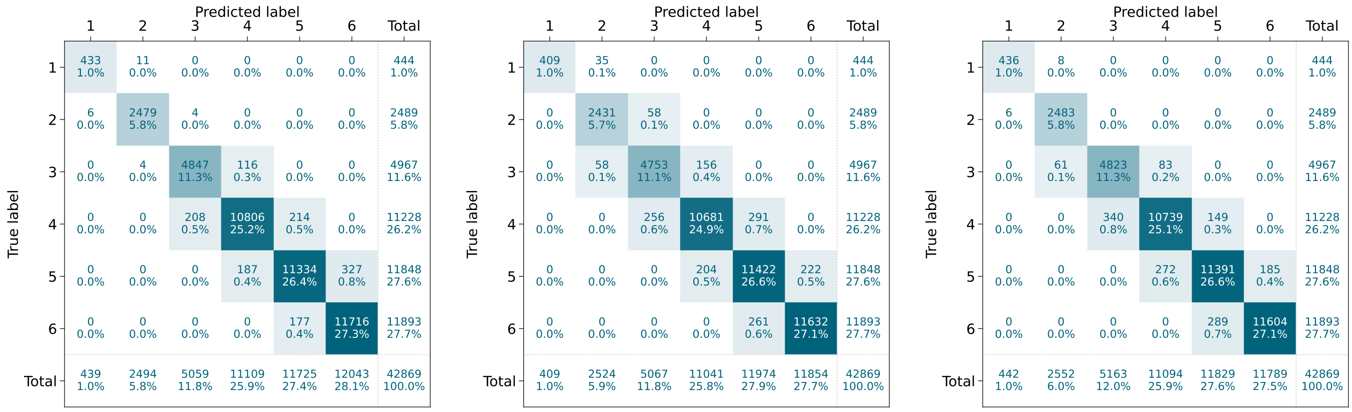
### B. *Results and Discussion*

Fig. 2 shows the number of wrong MF selections for the 3 investigated approaches, in addition to a regression approach that considers a variable margin. Fig. 2(a) shows the number of conservative selections. We can see that the proposed approach leads to the highest number of conservative selections, followed by the regression model, and finally by the binary classification. Intuitively, the number of conservative MF selections increases with the margin, due to the margin requiring a higher GSNR for selecting a given MF. With a very small margin of less than 0.03 dB, the regression model already incurs a higher number of conservative selections than the proposed approach.

Fig. 2(b) shows the number of aggressive MF selections. The proposed approach reduces the number of aggressive selections by 36% when compared to the binary classification. When compared to the regression, the number of aggressive selections is reduced by 44%. For the regression with margin, a margin of at least 0.07 dB is necessary for it to outperform the proposed approach. However, a 0.07 dB margin leads to a very high number of conservative selections, as shown in Fig. 2(a).

Fig. 2(c) shows the total number of wrong MF selections. While the binary classification has the lowest total number, we know from Fig. 2(b) that nearly half of the number is related to aggressive MF selections, which incur the highest overhead. Moreover, regression with and without margin incurs the highest number of wrong selections. Finally, the proposed approach achieves performance between the two traditional approaches, but with the benefit of not having as many aggressive MF selections as the regression model.

Fig. 3 shows the confusion matrices of the MF selection based on the binary classification, regression, and the proposed approach. These results detail the precise selections in the numbers previously shown in Fig. 2. All three strategies provide appropriate accuracy in the range of 96-97%, with




(a) Binary classification (b) Regression (c) Proposed

Fig. 3: Confusion matrices of the MF selection using the three ML-based strategies over the test set.

a maximum of 0.6% difference among them. However, the main changes concern the aggressive and conservative MF selection. The region above the main diagonal represents the cases of aggressive MF selection where the selected MF has a higher order than the possible one. The region below the main diagonal represents the cases of conservative MF selection. The binary classifier has better accuracy on MFs 1–3 and 6, while the regression has higher accuracy on MFs 4 and 5. We can also note that for higher-order MFs, the regression model leads to more aggressive MF selections. Finally, as expected, the proposed approach shows a higher number of conservative selections, in favor of having a lower number of aggressive ones.

In summary, we can observe that by using the output of two ML models, MF selections output by the proposed approach are more reliable than the ones using traditional strategies. In particular, the number of aggressive MF selections that lead to the need for immediately re-provisioning the lightpath is reduced by at least 36%. Meanwhile, the increase in number of conservative selections still makes it achieve a lower total number of wrong selections than the one using regression.

## IV. Conclusions

This work investigated ML-based strategies to select the MF of unestablished lightpaths. We raise awareness for the two issues that may arise from the wrong selection of an MF. Two traditional ML models (binary classification and regression) were investigated. The regression model considering a margin was also included in the analysis. Then, we proposed a new approach that combines multi-class classification and regression models to improve the reliability of the MF selection. Results showed that the proposed approach reduce the number of the most disruptive kind of error. Moreover, the total number of errors of the proposed approach is lower than the one achieved by regression.

This work showcases additional ways that ML-based QoT estimation approaches can be evaluated. In future works, the assessment of how the wrong MF selections impact the overall network spectrum efficiency is required. Moreover, investigating how to improve the accuracy of the models, and potentially combining other types of models, is also an interesting direction.

## References

[1] P. Poggiolini, G. Bosco, A. Carena, V. Curri, Y. Jiang, and F. Forghieri, "The GN-model of fiber non-linear propagation and its applications," *J. Lightwave Technol.*, vol. 32, no. 4, pp. 694–721, Feb.

[2] M. Ranjbar Zefreh, F. Forghieri, S. Piciaccia, and P. Poggiolini, "Accurate closed-form real-time EGN model formula leveraging machine-learning over 8500 thoroughly randomized full C-band systems," *Journal of Lightwave Technology*, vol. 38, no. 18, pp. 4987–4999, 2020.

[3] M. Lonardi, J. Pesic, T. Zami, and N. Rossi, "The perks of using machine learning for QoT estimation with uncertain network parameters," in *OSA Advanced Photonics Congress (AP) (IPR, NP, NOMA, Networks, PVLED, PSC, SPPCom, SOF)*, 2020, p. NeM3B.2.

[4] 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*, vol. 9, no. 1, pp. 98–108, 2017.

[5] R. M. Morais and J. Pedro, "Machine learning models for estimating quality of transmission in DWDM networks," *Journal of Optical Communications and Networking*, vol. 10, no. 10, pp. D84–D99, 2018.

[6] 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*, vol. 10, no. 2, pp. A286–A297, 2018.

[7] M. Ibrahimi, H. Abdollahi, C. Rottondi, A. Giusti, A. Ferrari, V. Curri, and M. Tornatore, "Machine learning regression for QoT estimation of unestablished lightpaths," *Journal of Optical Communications and Networking*, vol. 13, no. 4, pp. B92–B101, 2021.

[8] M. Lonardi, J. Pesic, T. Zami, E. Seve, and N. Rossi, "Machine learning for quality of transmission: a picture of the benefits fairness when planning wdm networks," *Journal of Optical Communications and Networking*, vol. 13, no. 12, pp. 331–346, 2021.

[9] F. N. Khan, K. Zhong, W. H. Al-Arashi, C. Yu, C. Lu, and A. P. T. Lau, "Modulation format identification in coherent receivers using deep machine learning," *IEEE Photonics Technology Letters*, vol. 28, no. 17, pp. 1886–1889, 2016.

[10] Z. Huang, X. Xin, Q. Zhang, H. Yao, F. Tian, and F. Wang, "Modulation format identification method based on multi-feature input hybrid neural network," *IEEE Photonics Journal*, pp. 1–7, 2024.

[11] B. C. Chatterjee, N. Sarma, and E. Oki, "Routing and spectrum allocation in elastic optical networks: A tutorial," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 3, pp. 1776–1800, 2015.

[12] Y. Pointurier, "Design of low-margin optical networks," *Journal of Optical Communications and Networking*, vol. 9, no. 1, pp. A9–A17, 2017.