Risk-Aware Machine Learning-based Approach for Lightpath QoT Estimation in Optical Networks

Hussein Jammal[1], Omran Ayoub[2], Andrea Bianco[1], Cristina Rottondi[1]
[1]Politecnico di Torino, Torino, Italy [2]University of Applied Sciences of Southern Switzerland, Lugano, Switzerland
hussein.jammal@polito.it, *omran.ayoub@supsi.ch*, *andrea.bianco@polito.it*, *cristina.rottondi@polito.it*

***Abstract*—Machine Learning (ML)-based models for Quality of Transmission (QoT) estimation in optical networks are commonly trained and evaluated using aggregate accuracy metrics such as Mean Squared Error (MSE), which fail to reflect the operational impact of prediction errors. In practice, QoT predictions are compared against feasibility thresholds to make accept/reject decisions, causing errors near such thresholds to entail significantly higher operational risk than equally sized errors far from them.**

To address this asymmetry in terms of operational risk, we introduce a risk-aware ML-based QoT estimation framework based on a Hybrid Ensemble (HE) architecture, which partitions lightpaths by ranges of Signal to Noise Ratio (SNR) margins and allocates higher modeling capacity to operationally critical regions of the feature space. We also propose metrics to quantify expected decision risk and risk disparity across lightpath groups. Experiments on two Space Division Multiplexed (SDM) network datasets show that the proposed approach consistently reduces both aggregate and group-level decision risk. Compared to the best-performing baseline, it achieves up to 7% reduction in prediction error, 8% reduction in expected decision risk, and 7% reduction in risk disparity, while maintaining low system complexity.

***Index Terms*—Quality of Transmission Estimation, Signal-to-Noise Ratio Prediction, Optical Networks;**

I. Introduction

Machine Learning (ML)-based approaches for Quality of Transmission (QoT) estimation have become essential in modern optical networks, allowing operators to make dynamic decisions regarding service provisioning and lightpath routing [1]. When training ML models tasked to predict the Signal-to-Noise Ratio (SNR) of a prospective lightpath, the models' loss function is optimized over the entire training dataset with the objective of minimizing the global average prediction error across all training samples. ML models' performance is then typically evaluated using aggregate metrics such as Mean Absolute Error (MAE) and Mean Square Error (MSE) [2].

However, in practical application scenarios, operators typically compare predicted SNR values provided by ML models against an SNR threshold derived from Bit Error Rate (BER) requirements, to decide whether to accept ("establish") or reject ("block") a prospective lightpath. This process exposes the network to two unwanted outcomes [2, 3]: (i) overestimating SNR for a lightpath whose true value falls below the threshold would result in a connection that cannot ensure acceptable QoT, causing service degradation; (ii) underestimating SNR for a lightpath that is, in fact, feasible, would lead to unnecessary blocking, reducing network utilization and, consequently, revenue. The likelihood of a prediction error leading to such consequences depends on the SNR margin, defined as the difference between the actual SNR value and the SNR threshold required for acceptable BER performance [4]. As a consequence, lightpaths with large margins can absorb prediction errors without violating the threshold, whereas those with small margins are highly vulnerable even to minor estimation errors.

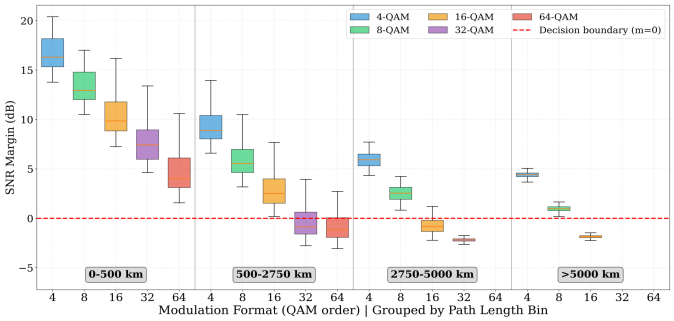


Fig. 1: SNR margin distribution across path length bins and modulation formats for Dataset 1 used later in our experiments.

Fig. 1 illustrates this phenomenon by reporting the distributions of the SNR margin across combinations of path length bins and modulation formats for a Space Division Multiplexing (SDM) network dataset [5]. Lightpaths shorter than 500 km consistently exhibit large positive SNR margins across all modulation formats, indicating that their true SNR values remain well above decision thresholds. For such lightpaths, even substantial prediction errors are unlikely to affect provisioning outcomes. In contrast, lightpaths longer than 500km show progressively smaller margins, with higher-order modulations approaching or falling below zero (specifically, 16-, 32- and 64-QAM for the range 500-2750km, 16- and 32-QAM for the 2750-5000km range, 8- and 16-QAM for lightpaths longer than 5000km). For these cases, prediction accuracy becomes operationally critical. This implies that prediction errors are more likely to lead to incorrect deployment decisions for longer lightpaths than for shorter ones.

This work has been partially supported by the European Union under the Italian National Recovery and Resilience Plan (NRRP) of NextGenerationEU, partnership on "Telecommunications of the Future" (PE00000001 - program "RESTART") and by the Swiss Innovation Agency Innosuisse through the innovation project SUSTAINET (No. 119.588 INT-ICT), carried out within the EUREKA Cluster CELTIC-NEXT SUSTAINET-Advance project.

978-3-903176-78-2 © 2026 IFIP

Based on these observations, ML-based lightpath QoT estimation models should be evaluated not only based on their aggregate prediction accuracy but also on the *operational risk their predictions induce, and how such risk is distributed across lightpaths groups*. For instance, a 0.5 dB prediction error for a lightpath operating far above its required SNR threshold contributes the same to the reported MAE as a 0.5 dB error for a lightpath with a predicted SNR very close to the threshold. However, if deployed, the latter is substantially more likely to result in unacceptable QoT than the former.

Building on this viewpoint, we propose a risk-aware approach to QoT estimation aimed at reducing the operational risk induced by ML models' predictions, and balancing that risk across lightpaths, rather than by equalizing prediction accuracy. To this end, we propose and develop an approach based on splitting lightpaths into specific groups and then utilizing a *Hybrid Ensemble* (HE) architecture, in which heterogeneous learning models are jointly employed. More in detail, we first partition the feature space describing the lightpaths according to their margin width and then train a dedicated model for each partition or subsets of partitions, concentrating modeling capacity, i.e., the ability to capture complex patterns, on operationally critical samples (i.e., lightpaths with smaller margins). We introduce a metric to evaluate the operational risk of errors, and another metric to quantify the disparity of this risk across groups (i.e., partitions) of lightpaths.

We compare our proposed approach against a standard lightpath QoT estimation approach, considering three baseline models of different computational capacity. Results on two SDM network datasets with different fiber types demonstrate that our proposed approach reduces the gap in operational risk across lightpath groups by up to 7%, relative to the best performing baseline approach, while also improving prediction performance in terms of MAE and Root MSE (RMSE).

The remainder of this paper is organized as follows. Sec. II reviews related work. Sec. III presents the problem statement and methodology. Sec. IV describes the experimental setup. Sec. V discusses the results and Sec. VI concludes the paper.

II. Related Work

Lightpath QoT estimation using ML techniques has been extensively studied [1, 6], evolving from binary classification to regression enabling direct SNR prediction, benchmarking such approaches against analytical models [7]. Probabilistic frameworks introduced cost functions penalizing overestimation more heavily than underestimation, recognizing their asymmetric operational consequences [2]. While these approaches provide accurate predictions, their evaluation mainly relies on aggregate metrics such as MAE and RMSE, which average performance across all lightpaths. Margin minimization is a central theme in this evolution, as traditional safety margins lead to resource under-utilization [8], consuming 3–6 dB of SNR budget [4]. ML-based approaches can tighten these margins, improving estimation accuracy [9], with extensions handling multiple uncertainties through probabilistic regressors [10]. However, these approaches optimize aggregate capacity without examining how errors distribute across lightpath categories.

More recently, research has started addressing uncertainty in SNR predictions provided by ML models. Dropout-based inference flags low-confidence predictions [11], Gaussian process regression provides uncertainty estimates [12], calibrated probabilistic regression corrects for overconfidence [13], conformal prediction produces confidence bands with coverage guarantees [14], and margin-driven approaches trigger retraining when conditions drift [15]. These methods apply policies uniformly, without examining whether certain groups experience higher risk. Prior system-level studies have also highlighted that standard accuracy metrics may not correlate with throughput and network-level performance when errors are biased or concentrated in critical operating regions [3]. These observations motivate evaluating QoT estimators by how often their errors can change accept/reject decisions, not only by their global average prediction error.

Therefore, our work focuses on the operational consequences of prediction errors, rather than on the prediction errors themselves, which vary significantly depending on the SNR margin at which a lightpath operates. To this end, our work introduces the problem of risk-aware lightpath QoT estimation and proposes a HE architecture that partitions lightpaths according to their margin width and trains a specialized model for each partition, thereby concentrating modeling capacity on lightpaths with small margins, where even minor estimation inaccuracies are most likely to translate into service disruption or unnecessary blocking. In addition, we introduce novel risk-aware evaluation metrics that quantify both the operational risk induced by estimation errors and the disparity of this risk across lightpath groups.

III. Problem Formulation and Methodology

A. Problem Formulation

We refer to the problem addressed in this work as *risk-aware ML-based lightpath QoT estimation*. While it formally builds upon the classical supervised learning setting for regression, it departs from conventional formulations by explicitly accounting for the operational asymmetry of prediction errors. The goal is not only to achieve high predictive accuracy, but also to mitigate the adverse impact of erroneous QoT estimates on network operation and decision-making.

The underlying ML problem is formulated as follows. Given dataset $D = \{(\mathbf{x}_i, y_i)\}_{i=1}^{N}$, where $\mathbf{x}_i \in \mathbb{R}^f$ is a vector of f features including physical characteristics (path length, number of links, modulation format) and spectral context (adjacent channels' properties) of a lightpath and $y_i \in \mathbb{R}$ denotes the observed SNR in dB, the objective is to learn a function $\mathcal{P} : \mathbb{R}^f \to \mathbb{R}$ that predicts the lightpath's SNR.

While prediction accuracy remains important, operators are concerned with whether predictions lead to correct provisioning decisions. For each lightpath i operating with modulation format m, we define the true SNR y_i, predicted SNR $\hat{y}_i$, and the required SNR threshold τ_m corresponding to the minimum SNR required by a lightpath using modulation format m

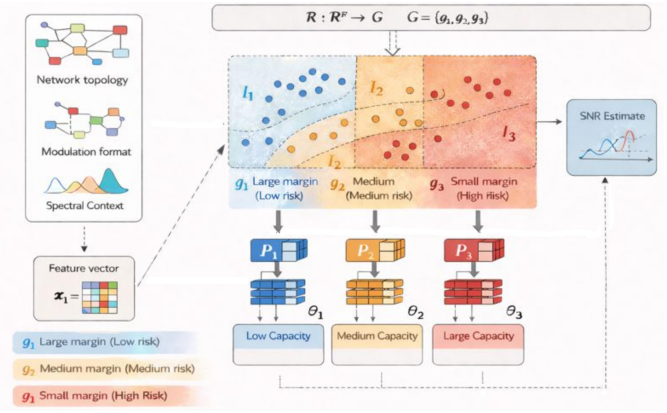


Fig. 2: The Risk-Aware Hybrid Ensemble prediction approach.

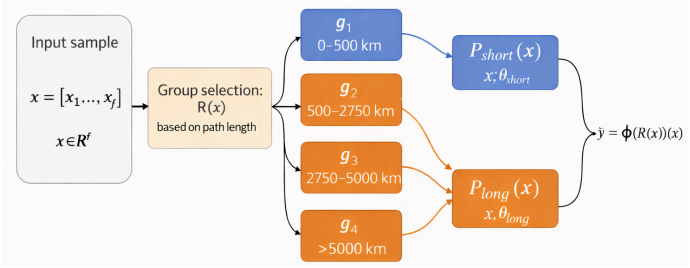


Fig. 3: Partitioning used in experiments.

TABLE I: Model Hyperparameters

Parameter	Short-Path	Long-Path	Baselines
Max Depth	10	12	8
Estimators	1,200	1,500	1,000 / 2,000 / 3,000
Learning Rate	0.05	0.03	0.08

TABLE II: Dataset Characteristics

Property	D1	D2
Fiber Type	Bu-SMF	WC-MCF
Topology	CONUS	CONUS
Samples	860,276	864,105
SNR Range (dB)	9.15 – 29.77	9.59 – 29.77
Path Length (km)	24.21 – $\approx 7,346$	24.21 – $\approx 6,891$

for acceptable BER. The correct and estimated provisioning decisions are, respectively:

$$d_{i,m} = \mathbf{1}(y_i \geq \tau_m), \quad \hat{d}_{i,m} = \mathbf{1}(\hat{y}_i \geq \tau_m) \tag{1}$$

where $d_{i,m} = 1$ indicates a feasible lightpath operating with modulation format m, $d_{i,m} = 0$ indicates an infeasible one, $\hat{d}_{i,m} = 1$ indicates that lightpath i using modulation format m is deemed feasible based on $\hat{y}_i$, and $\hat{d}_{i,m} = 0$ otherwise.

Decision errors carry asymmetric consequences. *False Acceptance* (FA) occurs when $\hat{d}_{i,m} = 1$ but $d_{i,m} = 0$: the lightpath is provisioned but fails QoT requirements, causing service disruption. *False Rejection* (FR) occurs when $\hat{d}_{i,m} = 0$ but $d_{i,m} = 1$: a viable lightpath is unnecessarily blocked, resulting in unserved traffic requests. Since service disruption is substantially more severe than capacity underutilization, FA is typically associated to a higher penalty than FR.

B. Methodology

Our proposed approach, referred to as *risk-aware*, is depicted in Fig. 2, and primarily builds on addressing the unequal distribution of operational risk by partitioning lightpaths into groups based on their margin width. Since the SNR margin depends on the unknown target value y_i, these cannot be defined directly in terms of margins at inference time. Instead, they are approximated through a partition of the feature space that captures the observed correlation between lightpath characteristics and SNR margin distributions. Building on this partitioning, we employ a HE architecture that assigns a specialized model to each group, dedicating modeling capacity to operationally critical regions where lightpaths operate closer to feasibility thresholds and prediction errors are more likely to result in false acceptances or false rejections.

We define a set of groups $G = \{g_1, g_2, \ldots, g_K\}$ by means of a function $\mathcal{R} : \mathbb{R}^f \rightarrow G$, which assigns each lightpath i, characterized by feature vector $\mathbf{x}_i$, to one of the groups according to:

$$\mathcal{R}(\mathbf{x}_i) = g_k \quad \text{if } \mathbf{x}_i \in I_k, \tag{2}$$

where $\{I_k\}_{k=1}^{K}$ defines a partition of $\mathbb{R}^f$ into non-overlapping regions, induced by (a subset of) feature dimensions. The choice and granularity of this partitioning can be adapted to the operational scenario and informed by empirical observations of how lightpath features relate to SNR margin distributions (e.g., groups may be defined based on specific combinations of path-length ranges and modulation formats, which are known to influence SNR margins). Similarly, the size of the ensemble is a configurable parameter, allowing the architecture to be tailored to different computational requirements.

At inference time, a generic lightpath with feature vector $\mathbf{x}$ is first assigned to its corresponding group via $\mathcal{R}(\mathbf{x})$ and then evaluated using the regressor $\mathcal{P}_{\mathcal{R}(\mathbf{x})}$ specialized for that group. Specialization is achieved by associating each group $g_k \in G$ with a distinct set of model hyperparameters $\boldsymbol{\theta}_k$, which control the model's capacity (e.g., depth, number of estimators, learning rate) and are selected to reflect the operational risk of the corresponding group. In particular, groups characterized by smaller expected SNR margins are assigned hyperparameter configurations that provide increased computational capacity, enabling the model to capture the more complex relationships present in decision-critical operating conditions. The resulting prediction is given by $\hat{y} = \mathcal{P}_{\mathcal{R}(\mathbf{x})}(\mathbf{x}; \boldsymbol{\theta}_{\mathcal{R}(\mathbf{x})})$.

We remark that the proposed HE approach is agnostic to the choice of the partitioning criterion and supports arbitrary feature-based group definitions.

C. Evaluation Metrics

We introduce three evaluation metrics: Margin-Weighted MAE (MW-MAE), which captures prediction accuracy near decision thresholds, and two risk-aware metrics, Expected Decision Risk (EDR) and Risk Disparity (RD), which quantify the operational consequences of prediction errors.

MW-MAE: it aims to shed light on how prediction errors are distributed across decision thresholds. Consider the SNR margin as $\mu_i = y_i - \tau_m$, representing the difference between true SNR and the required SNR threshold. Standard MAE weights all errors equally, but errors near the threshold (small $|\mu_i|$) are operationally critical while errors far from the threshold (large $|\mu_i|$) rarely affect provisioning. To capture this, we model MW-MAE as follows:

$$\text{MW-MAE} = \frac{1}{N}\sum_{i=1}^{N} w_i \cdot |y_i - \hat{y}_i| \tag{3}$$

where $w_i = \exp(-|\mu_i|/\sigma)$ assigns higher weight to samples operating near the decision threshold, and σ controls the decay rate. Following low-margin design considerations in [8], we set $\sigma = 3\ dB$, emphasizing errors near feasibility boundaries while gradually down-weighting lightpaths with large margin.

EDR: Inspired by [16], EDR permits to evaluates predictions by their induced operational consequences:

$$\begin{aligned}\text{EDR} = \frac{1}{N}\sum_{i=1}^{N}\Big[c_{\text{FA}}\,\mathbf{1}(\hat{d}_{i,m}=1,\, d_{i,m}=0) \\ + c_{\text{FR}}\,\mathbf{1}(\hat{d}_{i,m}=0,\, d_{i,m}=1)\Big].\end{aligned} \tag{4}$$

where N represents the total number of samples, and c_{FA} and c_{FR} denote the costs of FA and FR, respectively. Accordingly, EDR differs from standard regression losses by judging predictions by decision harm rather than error magnitude. For instance, a 2 dB error that does not cross the threshold contributes zero risk, while a 0.3 dB error that causes threshold crossing incurs cost. Consequently, a lower EDR is preferred, as it indicates that decision-related risk is effectively minimized.

RD: RD is defined as the difference in EDR across the diverse groups, computed as follows:

$$\text{RD} = \max_{g\in G}\text{EDR}_g - \min_{g\in G}\text{EDR}_g. \tag{5}$$

where EDR_g is the EDR of lightpaths of group g:

$$\begin{aligned}\text{EDR}_g = \frac{1}{|g|}\sum_{i\in g}\Big[c_{\text{FA}}\,\mathbf{1}\{\hat{d}_{i,m}=1,\ d_{i,m}=0\} \\ + c_{\text{FR}}\,\mathbf{1}\{\hat{d}_{i,m}=0,\ d_{i,m}=1\}\Big].\end{aligned} \tag{6}$$

A lower RD indicates that decision risk is evenly distributed across the considered groups, meaning that lightpaths experience comparable risk regardless of group membership. In contrast, a higher RD reflects substantial disparities in risk across groups, which is undesirable. Therefore, a low RD is preferred, as it implies balanced operational risk among lightpath groups. Naturally, achieving a low RD is meaningful only when accompanied by a low overall EDR in the first place. Consequently, an approach characterized by both low EDR and low RD is able to simultaneously minimize operational risk and ensure equitable treatment across groups.

IV. Experimental Setup

This section presents the experimental setup and describes the HE configuration used for evaluation.

TABLE III: Sample distribution across path-length bins.

Path-length bin (km)	D1 #samples	D2 #samples
g_1	34,647	34,677
g_2	496,184	498,046
g_3	273,194	273,460
g_4	56,251	57,922

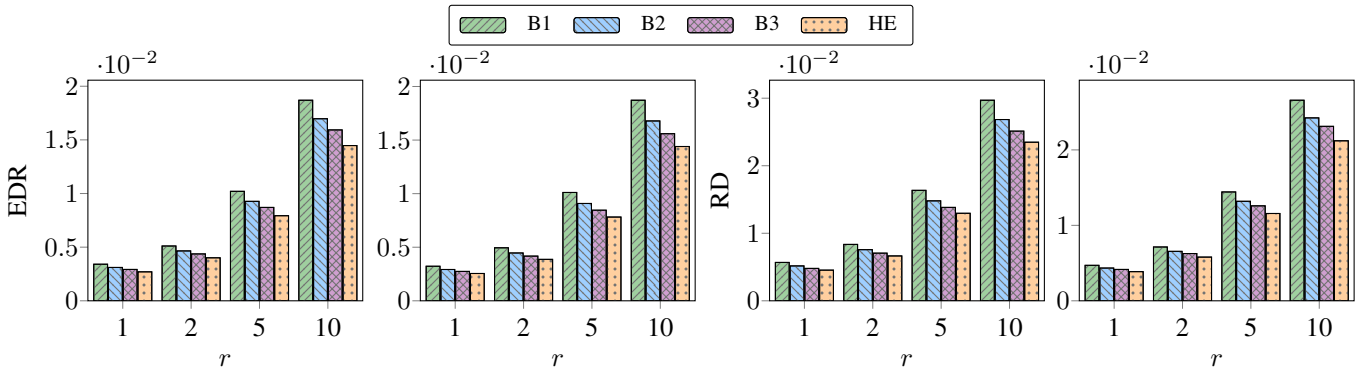
We specialize function $\mathcal{R}$ by defining lightpath groups according to path length, which, based on the empirical evidence reported in Fig. 1, is observed to correlate with SNR margin distributions in the considered datasets. We then define a set of four groups $G = \{g_1, g_2, g_3, g_4\}$ by partitioning the path length range into $K = 4$ disjoint intervals: $I_1 : [0, 500]$ km, $I_2 : (500, 2750]$ km, $I_3 : (2750, 5000]$ km, and $I_4 :> 5000$ km. Each lightpath is assigned to a group via $\mathcal{R}(\mathbf{x})$ based on the path length component of its feature vector. Following the general formulation, model specialization is achieved by associating groups with different predictors. For the sake of conciseness, in this study we limit number of predictors to two, whose hyperparameters are optimized and reported in Tab. I. Specifically, a short-path model is assigned to the lowest-risk group g_1, while a long-path model is shared across the remaining groups $\{g_2, g_3, g_4\}$, which are associated with smaller SNR margins. The short-path model is configured with moderate depth and a higher learning rate, whereas the long-path model employs greater depth, a larger number of estimators, and a lower learning rate, to better capture high-risk operating conditions where lightpaths are closer to feasibility thresholds and prediction accuracy is operationally critical. At inference time, each lightpath is first assigned to a group and then evaluated using the corresponding predictor.

We compare the HE against three baseline approaches, denoted as B1, B2, and B3, corresponding to models with 1,000, 2,000, and 3,000 estimators, respectively. Notably, the highest-capacity B3 contains more estimators than the two predictors composing the HE (2,700 estimators overall).

We compute the SNR thresholds τ_m through analytical BER formulas for each modulation format, and use a BER threshold (BER_{th}) of 2.0×10^{-2}, such as in [5]. We numerically solve for τ_m such that $\text{BER}(\tau_m) = \text{BER}_{\text{th}}$, as per equations in in [17]. The resulting thresholds are 6.34 dB, 9.75 dB, 12.80 dB, 15.70 dB and 18.53 dB for 4-QAM (QPSK), 8-QAM, 16-QAM, 32-QAM, and 64-QAM, respectively.

We evaluate the approach on two SDM datasets from the CONUS topology (Tab. II), generated using the GNRF approximation for non-identical channels with Amplified Spontaneous Emission (ASE), Non-Linear Interference (NLI), and inter-core crosstalk impairments [5]. Dataset D1 employs Bundled Single-Mode Fibers (Bu-SMF), while dataset D2 uses Weakly-Coupled Multi-Core Fibers (WC-MCF), both reporting 26 features including lightpath characteristics (path length, number of links, modulation format) and adjacent channel's properties. Tab. III lists the number of samples per group in D1 and D2.

We adopt XGBoost (XGB) as ML algorithm for all models,



(a) EDR across D1. (b) EDR across D2. (c) RD across D1. (d) RD across D2.

Fig. 4: EDR ((a) and (b)) and RD ((c) and (d)) achieved by *B1, B2, B3, and HE* on D1 and D2.

TABLE IV: Overall prediction performance across D1 and D2, averaged across 5 folds. Best results are reported in bold.

	D1			D2		
Method	MAE	RMSE	MW-MAE	MAE	RMSE	MW-MAE
B1	0.035	0.045	0.017	0.034	0.043	0.016
B2	0.032	0.041	0.015	0.030	0.039	0.015
B3	0.030	0.039	0.014	0.029	0.037	0.014
HE	**0.028**	**0.037**	**0.013**	**0.027**	**0.035**	**0.013**
Improvement	*6.7%*	*5.1%*	*7.1%*	*6.9%*	*5.4%*	*7.1%*

with hyperparameters optimized via grid search. Model performance is evaluated via 5-fold cross-validation with shuffled splits (80% training, 20% testing), and results are averaged over folds. Models are trained using a squared error loss and evaluated using: MAE, RMSE, MW-MAE, EDR and RD. We vary the risk ratio (r) $c_{\text{FA}}/c_{\text{FR}} \in \{1, 2, 5, 10\}$ to capture different operational priorities, ranging from symmetric cost settings (ratio 1) to highly disruptive scenarios (ratio 10), in which provisioning a failing lightpath is significantly more costly than blocking a feasible one.

V. Numerical Results

We now discuss numerical results, first addressing prediction performance and then focusing on risk metrics.

Prediction Performance. Tab. IV reports prediction metrics across both datasets, with the last row showing the HE improvement over the best-performing baseline (B3). The proposed approach consistently improves MAE and RMSE, achieving reductions of 6.7% and 6.9% in MAE and 5.1% and 5.4% in RMSE for D1 and D2, respectively. While such gains are expected given the use of specialized models, improving aggregate accuracy is not the primary objective; rather, the goal is to reduce decision risk, even if MAE and RMSE were to remain unchanged. In principle, this could occur if performance improves for certain groups while degrading for others, yet still results in lower overall risk. Nevertheless, the results demonstrate that our approach achieves both reduced decision risk and improved predictive performance.

The HE approach also improves MW-MAE by 7.1% across both datasets, indicating higher accuracy in the decision-critical regime near the SNR threshold, where small errors can lead to incorrect provisioning decisions. These results indicate that our proposed approach is capable of reducing the average prediction error and concentrating its accuracy where it matters most.

Takeaway 1: Beyond conventional accuracy gains, the HE method improves precision near feasibility thresholds, effectively reducing operational risk.

Risk Evaluation. We now turn to risk analysis. Fig. 4a and Fig. 4b report the EDR across different risk ratios $r \in \{1, 2, 5, 10\}$. Lower EDR indicates improved provisioning reliability by jointly reducing false acceptances and false rejections. Results show that the HE approach achieves an EDR improvement of 8.0% with respect to B3 across D1 and D2, respectively. Although EDR naturally increases with larger values of r for all methods, our approach consistently attains the lowest overall EDR across all considered settings. These results indicate that the proposed method is more effective at mitigating decision-level risk, particularly under increasingly conservative operational scenarios. Importantly, this risk reduction is achieved through a lightweight ensemble strategy that introduces only marginal additional system complexity, making the approach practical for deployment in real-world network control pipelines.

We now focus on RD. Fig. 4c and Fig. 4d report the RD, which quantifies the maximum EDR gap across lightpath groups, across both datasets for different values of r. From an operational perspective, a lower RD indicates a more balanced distribution of risk across the network, ensuring that no particular group of lightpaths disproportionately suffers from prediction-induced decision errors. Results show that our approach achieves RD improvements of 6.0% for D1 and 7.8% for D2 with respect to B3. Similar to EDR, RD increases with larger values of r for all methods, reflecting the increased penalty for FAs, yet, results show that the percentage improvement is maintained consistent. These results further confirm that the proposed HE architecture not only reduces overall decision risk but also distributes it more equitably across lightpath groups. This is particularly relevant for network operators, as it provides assurance that the QoT estimation model exhibits a consistent risk profile across different lightpath groups, thereby enhancing the overall

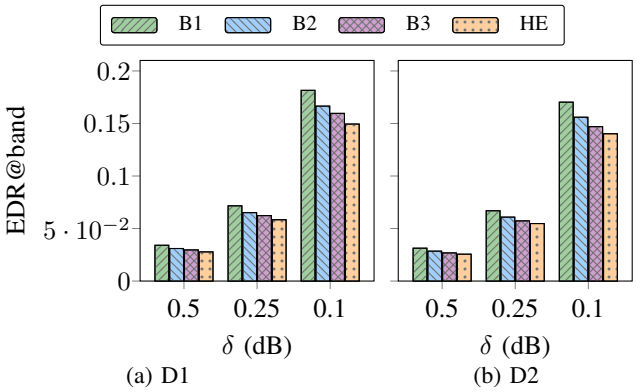


Fig. 5: EDR evaluated within a decision-critical band of width $\pm\delta$ around the SNR threshold, at risk ratio $r = 1$ across D1 and D2.

reliability of provisioning decisions.

Takeaway 2: By jointly reducing EDR and RD across all risk ratios, the proposed approach mitigates both aggregate and group-level decision risk, enabling safer and more balanced QoT-driven provisioning with a consistent risk profile across lightpath groups.

To provide further insight into the risk reduction achieved in decision-critical operating conditions, Fig. 5 reports the EDR computed on samples falling within a band of with $\pm\delta$ around their respective SNR thresholds. As δ decreases, the evaluation isolates progressively smaller subsets of samples operating closest to the feasibility boundary, where even minor prediction errors can flip the provisioning decisions. Results show that EDR increases as the band narrows for all methods, yet the proposed approach consistently achieves the lowers EDR across all band widths ($\delta = \{0.5, 0.25, 0.1\}$ dB). Relative to the strongest baseline (B3), HE reduced EDR by approximately 6% and 4.5%, for D1 and D2, respectively, across all tested values of δ, with the improvement increasing at the tightest setting ($\delta = 0.1$ dB). This confirms that the HE approach improves reliability specifically for lightpaths operating with minimal margin, where decision errors are most consequential.

Takeaway 3: The proposed approach provides consistent risk reduction across all levels of operational criticality.

Computational Overhead. Training time increases with model capacity for the baseline model ranging from 8.3 s (B1) to 15.2 s (B2) and 22.2 s (B3). The HE requires 71.3 s (about $3.2\times$ slower than B3), since two specialized models are trained sequentially. All runtimes were measured and obtained on a machine with a NVIDIA GPU 40 and 26GB RAM. Inference speed follows a similar trend: B1 is the fastest model, with an inference time of 0.37μs per sample, followed by B2 at 0.5μs and B3 at 0.66μs. The HE required an inference time of 1.28μs per sample ($1.9\times$ relative to B3) due to the predictor selection step. Despite this increase, the HE does not incur significantly higher time for practical deployment.

VI. Conclusion

This paper presented a hybrid ensemble approach for lightpath QoT estimation that addresses the unequal distribution of operational risk across network segments. Compared to the baseline models, the proposed method achieves $6 \sim 7\%$ reduction in MAE and $6 \sim 7\%$ reduction in risk disparity across lightpath groups by partitioning lightpaths according to path length and allocating higher model capacity to operationally critical groups. The approach requires no modifications to the training objective or post-hoc calibration; the improvement arises from allocating model capacity according to risk exposure of different lightpath groups, rather than their representation in the training data. Future work may explore adaptive threshold selection and extension to multi-attribute partitioning schemes.

References

[1] Y. Pointurier, "Machine learning techniques for quality of transmission estimation in optical networks," *J. Opt. Commun. Netw.*, vol. 13, no. 4, pp. B60–B71, 2021.

[2] M. Ibrahimi *et al.*, "Machine learning regression for QoT estimation of unestablished lightpaths," *J. Opt. Commun. Netw.*, vol. 13, no. 4, pp. B92–B101, 2021.

[3] M. Lonardi *et al.*, "Machine learning for quality of transmission: a picture of the benefits fairness when planning wdm networks," *J. Opt. Commun. Netw.*, vol. 13, no. 12, pp. 331–346, 2021.

[4] Y. Pointurier, "Design of low-margin optical networks," *J. Opt. Commun. Netw.*, vol. 9, no. 1, pp. A9–A17, 2017.

[5] H. Akbari *et al.*, "Datasets for qot estimation in sdm networks," *J. Opt. Commun. Netw.*, vol. 17, no. 6, pp. 514–525, 2025.

[6] C. Rottondi *et al.*, "Machine-learning method for quality of transmission prediction of unestablished lightpaths," *J. Opt. Commun. Netw.*, vol. 10, no. 2, pp. A286–A297, 2018.

[7] R. M. Morais and J. Pedro, "Machine learning models for estimating quality of transmission in DWDM networks," *J. Opt. Commun. Netw.*, vol. 10, no. 10, pp. D84–D99, 2018.

[8] S. J. Savory *et al.*, "Design considerations for low-margin elastic optical networks in the nonlinear regime," *J. Opt. Commun. Netw.*, vol. 11, no. 10, pp. C76–C85, 2019.

[9] E. Seve *et al.*, "Learning process for reducing uncertainties on network parameters and design margins," *J. Opt. Commun. Netw.*, vol. 10, no. 2, pp. A298–A306, 2018.

[10] O. Karandin *et al.*, "Probabilistic low-margin optical-network design with multiple physical-layer parameter uncertainties," *J. Opt. Commun. Netw.*, vol. 15, no. 7, pp. C129–C137, 2023.

[11] T. Tanimura *et al.*, "OSNR estimation providing self-confidence level as auxiliary output from neural networks," *J. Lightw. Technol.*, vol. 37, no. 7, pp. 1717–1723, 2019.

[12] F. Meng *et al.*, "Self-learning monitoring on-demand strategy for optical networks," *J. Opt. Commun. Netw.*, vol. 11, no. 2, pp. A144–A154, 2018.

[13] N. Di Cicco, F. Musumeci, and M. Tornatore, "Calibrated regression for quality of transmission estimation in optical networks," in *Proc. BalkanCom*, 2022, pp. 21–25.

[14] K. Rezaei, O. Ayoub, P. Monti, and C. Natalino, "Policy-driven conformal prediction for trustworthy QoT estimation," in *Proc. OFC*, 2026.

[15] P. Lechowicz, C. Natalino, and P. Monti, "QoT estimation with margin-driven transfer learning in time-varying optical networks," in *Proc. OFC*, 2025.

[16] V. Komisarenko and M. Kull, "Cost-sensitive classification with cost uncertainty: Do we need surrogate losses?" *Mach. Learn.*, vol. 114, no. 4, 2025.

[17] J. G. Proakis and M. Salehi, *Digital Communications*, 5th ed. McGraw-Hill, 2008.