XAI-Guided Optimization of a Multilayer Network Regression Model

Katarzyna Duszyńska[1], Paweł Polski[1], Michał Włosek[1],
Aleksandra Knapińska[1], Piotr Lechowicz[1,2], Krzysztof Walkowiak[1]
[1]Department of Systems and Computer Networks, Wrocław University of Science and Technology, Wrocław, Poland
[2]Department of Electrical Engineering, Chalmers University of Technology, Gothenburg, Sweden
aleksandra.knapinska@pwr.edu.pl

***Abstract*—The recent technological advances create increased network capacity demand, highlighting the need for new network optimization methods. However, the proposed solutions require broad testing with numerous time-consuming simulations. Thus, estimation methods based on Machine Learning (ML) are developed to improve this process. In this work, we create a regression network model to predict four resource utilization metrics using the input set of connection requests. Using eXplainable Artificial Intelligence (XAI) tools, we optimize the proposed model for faster inference without a decrease in prediction quality.**

***Index Terms*—multilayer network, resource allocation, machine learning, explainable artificial intelligence.**

I. Introduction

The growing demand for network capacity, accelerated by the technology development, amplifies the need for new solutions to fit more traffic within the backbone infrastructure. New, data-driven methods based on Artificial Intelligence (AI) are designed to cope with the growing demands and better utilize the backbone networks [1], [2]. Elastic Optical Networks (EONs) are one of the breakthroughs for the physical layer, enabling drastically more efficient resource utilization [3]. Combined into a multilayer design with virtual topologies and quality of service (QoS) considerations, they are a powerful solution for the current needs [4].

The fundamental optimization problem in EONs is routing and spectrum assignment (RSA), where for each demand, a routing path and suitable optical corridor are assigned with respect to spectrum contiguity and continuity constraints. Typically, the optimization objective is to reduce spectrum usage by allocating connections using the lowest possible wavelength indices. The problem is complex but grows particularly challenging for large topologies and request sets. Thus, sophisticated methods are required for setting upper and lower bounds and improving solution finding [5]. The optimization problem becomes even more difficult in multilayer networks where many layers are jointly considered [6]. However, for a thorough evaluation of a proposed algorithm solving the problem, a broad assessment of multiple test cases is required.

Considering the significant time and resources required for solving numerous instances for thorough evaluation, recent research efforts focus on various estimation methods based on Machine Learning (ML). Specifically, broad testing is required to thoroughly evaluate the effectiveness of the proposed solutions in various traffic conditions. Furthermore, instant estimation of chosen performance metrics is beneficial for an operator to set expectations of how the network should operate under the chosen algorithm considering various traffic conditions. To this end, simulations using a network digital twin enable a broad "what-if" analysis [7]. However, considering the non-negligible simulation time, such an assessment is computationally heavy and time-consuming. Therefore, a regression model of a network trained using the already performed simulations is a promising approach for improving this process. Recently, models for predicting various metrics, including bandwidth blocking [8], [9] or latency [10], are proposed. In this regard, in our previous work [11], we showed how it is possible to build a regression model of a network using only the input set of connection requests for successful prediction of various performance metrics regarding resource utilization, which yielded very promising results.

However, as the employed ML models operate in a black-box manner, their internal operation and reasoning remain unknown. Because the primary considered measure is usually the prediction quality, the models may grow into unnecessarily large or complex structures with numerous parameters extracted based on domain knowledge and available data. Consequently, extensive datasets are essential for their successful training and operation. Next to the large data requirements, the lack of interpretability of the proposed black box models is another significant factor holding out the network operators from using ML approaches in real-world applications. Because the models do not provide any insights into their internal operation or reasoning behind their decisions, they often seem untrustworthy and, thus, less likely to be used despite promising results.

The identified issues can be, however, addressed by employing the emerging explainable AI (XAI) techniques [12], which is the focus of this work. Considering the estimation of the complex problem of RSA in multilayer networks, we analyze various ML models using connection request sets as inputs and investigate the contribution of individual features to the final model output. We explore how the predictor operation changes depending on the forecasted performance metric and ML algorithm to identify trends and dependencies. Finally, we use the gained knowledge for thoughtful feature selection and performance optimization without notable quality degradation.

The remainder of this work is organized as follows. In Section II, we overview the related works. In Section III, we provide details about the network and traffic model with the multilayer RSA algorithm. In Section IV, we build a regression model for estimating four different network performance met-

ISBN 978-3-903176-63-8 © 2024 IFIP

rics and analyze their operation with XAI. Finally, we conduct a numerical evaluation of the developed solutions in Section V and conclude the work in Section VI.

II. Related work

In this Section, we give an overview of the works related to the main aspects of this work, including network performance estimation and XAI in optical networks.

Estimating various metrics to improve or substitute time-consuming simulations is gaining popularity in various domains. Effective methods for traffic prediction [13] or quality of transmission (QoT) estimation [14] are proven to significantly benefit new AI-based routing algorithms. At the same time, the estimation of the overall network performance, expressed as bandwidth blocking or resource utilization enables bottleneck prediction [8], better modulation format selection [9] and fragmentation management [15] and broader algorithm evaluation [11], [16]. Finally, first attempts to substitute the heuristic-based RSA with ML-based models are also developed [17]. However, the prospective research direction seems to be ML-based optimization of complex ILP algorithms for finding an optimal solution through estimating lower or upper bounds [5], [18].

XAI techniques have been investigated in few problems concerning optical networks, including failure localization [19], cause identification [20], or traffic identification [21] and prediction [22]. First efforts to explain the reinforcement-learning-based RSA are also made [23]. However, the prime example is the QoT estimation, where various use cases have been investigated, including model optimization through feature selection and uncertainty quantification [24].

To the best of our knowledge, this is the first work to apply explainability to the problem of multilayer network operation estimation.

III. Network model and allocation algorithm

In this Section, we first provide details about the multilayer network model and allocation algorithm and then describe the conducted simulations.

We consider a two-layer network model with a physical optical (EON) topology at the bottom and a virtual packet (IP) layer at the top. The virtual topology comprises lightpaths set up in the physical network, as illustrated in Fig. 1. The layers are optimized jointly, exchanging information about the free and used bandwidth, enabling traffic grooming. The connection requests to be provisioned are characterized by their source and destination nodes and bitrate.

In this work, we aim to estimate the operation of the multilayer network optimization algorithm proposed in [25], outlined in Alg. 1. In a nutshell, it starts by sorting the connection requests by bitrate (line 1). For each request, it first checks if a direct lightpath from its source to its destination exists and has enough spare bandwidth (line 3). If so, the request is groomed into it without changing the topology (line 4). Otherwise, a new lightpath is requested in the optical layer to accommodate the request's bitrate (lines 6 and 7). The lightpaths are allocated considering ten shortest paths sorted by the index of their highest occupied frequency slot. The spectral efficiency of each candidate path

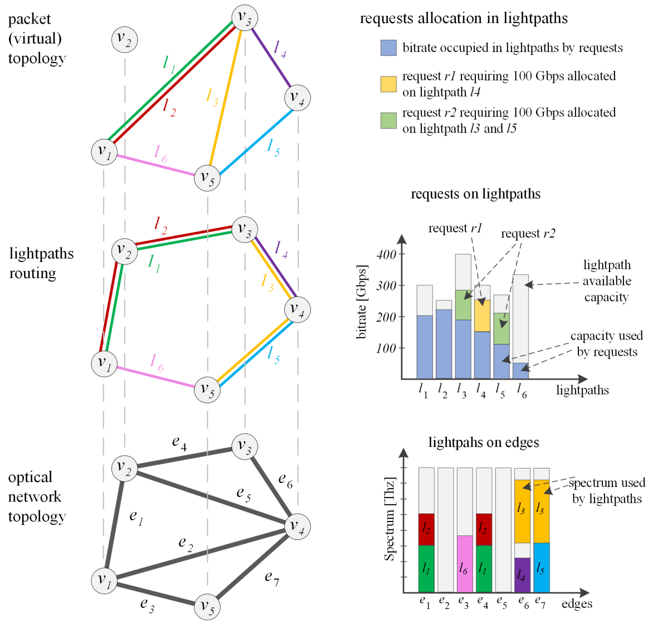

Fig. 1: Overview of the network model and traffic grooming.

Algorithm 1 RSA for multilayer IP-over-EON networks

1: Sort requests by bitrate
2: **for** each request **do**
3:     **if** a direct lightpath from its source to its destination exists and has enough free space **then**
4:         groom the request into this lightpath
5:     **else**
6:         set up a new lightpath in EON layer
7:         allocate the request into the new lightpath
8:     **end if**
9: **end for**

is calculated using the distance-adaptive transmission rule [26] for the chosen transceiver model. In this work, we assume the Ciena Wavelogic 5 Extreme commercial transceiver model, with the transmission reach, supported bitrate and the number of frequency slots (FSs) of its available modulation formats (MFs) given in Tab. I.

TABLE I: MFs – transmission reach and supported bitrate based on [27].

MF	reach	bitrate	# FSs
QPSK	no limit	200G	6
8QAM	no limit	400G	9
16QAM	800 km	400G	6
16QAM	1600 km	600G	9
32QAM	200 km	800G	9

The algorithm's performance for a given set of connection requests can be characterized by various metrics. In this work, we consider four of them: the *highest occupied slot* (*highestSlot*) – a general metric assessing the overall spectrum occupancy, the *average highest occupied slot* (*avgHighestSlot*) – giving a broader idea of the network saturation considering all links, the *sum of occupied slots* (*sumOfSlots*) – describing the utilization of network links considering spectrum fragmentation, and the *number of active transceivers* (*transceivers*), which can represent the network's energy usage and operational cost [28].

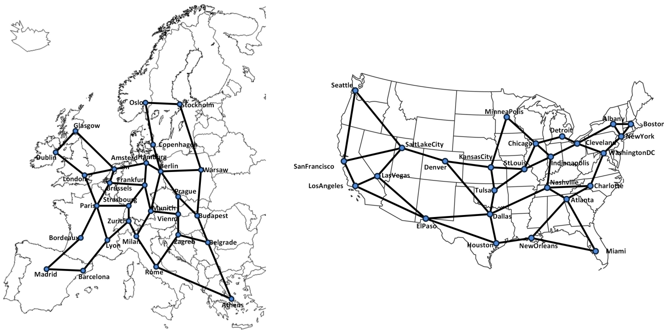

Fig. 2: Considered network topologies: Euro28 with 28 nodes and 82 links (left) and US26 with 26 nodes and 84 links (right).

For a thorough evaluation of the network performance using the employed algorithm, we perform simulations with various traffic conditions. In our experiments, we consider two large topologies, illustrated in Fig. 2. For each topology, we generate 100 connection request sets uniformly distributed between node pairs with bitrate in the 50-150 Gbps range. An example connection request set for each topology is plotted in Fig. 3. From each simulation, we save the values of the four aforementioned performance metrics. In turn, each datapoint consists of an input connection request set and the corresponding metric values. On a machine with the Intel Core i5-1038NG7 processor and 16 GB of RAM, the average time of one simulation (implemented in Java) was 95 ms for Euro28 and 109 ms for US26.

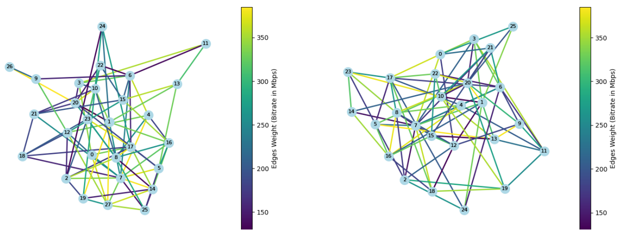

Fig. 3: Example datasets for Euro28 (left) and US26 (right).

IV. Network regression model and its optimization

In this Section, we first build ML models for estimating four network performance metrics using based on an input set of requests and then analyze their operation using XAI techniques.

In Section III, we described our multilayer network optimization algorithm and the performed simulations. From our experiments, we now build a dataset for each topology for further evaluation. That is, each set of connection requests is associated with the resulting values of four network performance metrics (*highestSlot*, *avgHighestSlot*, *sumOfSlots*, and *transceivers*). The proposed black-box regression model of a network outputs its estimation of a chosen network performance metric, taking a set of connection requests as inputs. In particular, similarly to [29], the same dataset can be used for predicting a different target according to current needs. For 100 requests, each characterized by the source node, destination node, and bitrate – the ML model then uses 300 inputs. According to our previous experiments, such a setting enables successful forecasting of each of the metrics.

However, with a significant number of inputs for large sets of requests, it is critical to analyze the importance of each of them. In particular, although contemporary deep learning models use thousands or even millions of parameters, they are not always the preferred choice. Training such models requires extensive datasets and considerable computing resources. Thus, simpler models are desirable if only they are able to deliver acceptable performance, having the advantage of smaller data and power requirements [30]. For that reason, analyzing the importance of each model input feature is desirable to achieve further reduction in model size.

To this end, we first analyze the feature importance of the gathered datasets using decision trees. These simple regressors contain information about the importance of each feature, expressed as its frequency of being the split criterion. The rankings of the top 10 most important features for each topology are plotted in Fig. 4. In both cases, seven out of ten parameters describe the bitrate of various connections, and only three describe the source or destination nodes. The trends found for the remaining metrics, omitted here, are equivalent, clearly showing the highly informative value of the requests' demanded bandwidth. Although this result might seem somewhat surprising, it is very sensible, considering the models are trained on the results from a number of conducted simulations. Even though the connection requests originate and terminate at various network nodes, their routing paths cross similar network regions, and thus, common trends can be extracted by the ML models. Furthermore, such a behavior is observed for each of the predicted network performance metrics.

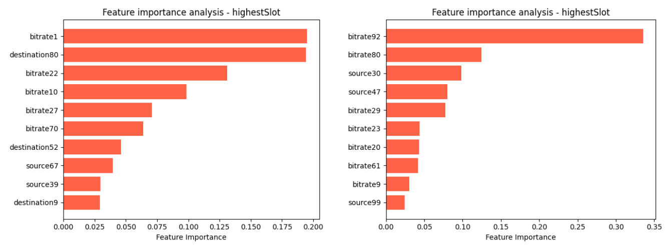

Fig. 4: Feature importance in decision trees for the *highestSlot* metric for Euro28 (left) and US26 (right).

In the second stage, we build multiple regression models for each topology and analyze the feature contribution in each of them using the SHapley Additive exPlanations (SHAP) [31]. This game-theoretic approach estimates the contribution of each feature to the final output of any ML model in a *post-hoc* manner, i.e., after it is trained on a specific dataset. The local explanations (for each individual instance) are combined into a global summary. The results can be easily interpreted from the summary plots (see the examples in Fig. 5 and Fig. 6). This visual representation contains multiple pieces of information.

Each plot includes the ranking of the ten most contributing features to the model output. The points in each row represent the feature contribution in individual problem instances (data points). The colors denote the feature value (red for high and blue for low). The position on the x-axis presents the SHAP value, which is the impact on the model output, i.e., in which direction and by how much the value of a specific feature moves the final model decision.

In this part of the experiment, we use five diverse regressors to analyze feature contributions in various settings. To this end, we chose a single tree model – decision tree (CART), an ensemble method – random forest (RF), a nearest neighbor model – k nearest neighbors with $k = 5$ (KNN), a simple regressor – linear regression (LR), and a support vector machine with a poly kernel (SVR). The regressor selection is strictly connected with the high data acquisition cost (each sample in a dataset requires performing a full simulation). Although deep learning models often yield exceptional prediction quality, they require enormous amounts of training data and vast computing resources. Following the green networking paradigm, we choose the smallest models offering the expected performance in preliminary analysis.

In Fig. 5 and Fig. 6, we present example analysis results. The primary trend observed for all the estimated metrics and all investigated regressors is that the main drivers for the model decision are the information about requests' bitrates. In particular, all of the features visible in Fig. 5 and half of the features visible in Fig. 6 are bitrate information. In the remaining cases, which are not shown due to space constraints, the majority of the most-contributing inputs also describe the requested bandwidth of the connections. However, there are also evident differences between the discussed plots. For the SVR, the contribution of each of the top features is similar, but its direction depends highly on the individual instance. On the other hand, the RF has more apparent trends, which are visible as a higher concentration of data points in groups. Moreover, the ranking of the most important connections is vastly different. Individual connection requests might play a distinct role in the decisions of various regression models trained on the same data. That shows how various ML models operate in very different manners and create their internal functions in their own ways.

From the conducted investigation, we can, however, identify a common trend found in the analysis of the decision-tree-based feature importance and the SHAP-based feature contribution. That is, the models most often use the information about the bitrate. As discussed, different regressors extract common trends from the traffic load at various network regions using the requested bandwidth despite the exact connection origin and destination. Thus, let us use this knowledge for feature selection. In particular, previous studies showed the benefits of SHAP values for feature selection [32], [33]. However, considering the common trends between both feature analysis methods, for versatility and efficiency, our approach selects the most important features according to their importance in decision trees. In particular, the decision-tree-based analysis is orders of magnitude faster than the SHAP-based one for the same dataset. According to our research reported above, the found trends are highly similar. In our proposed XAI-based feature selection method, presented in Alg. 2, the algorithm selects only the features carrying any information. In our testing, the vast majority of the selected inputs are the bitrate features.

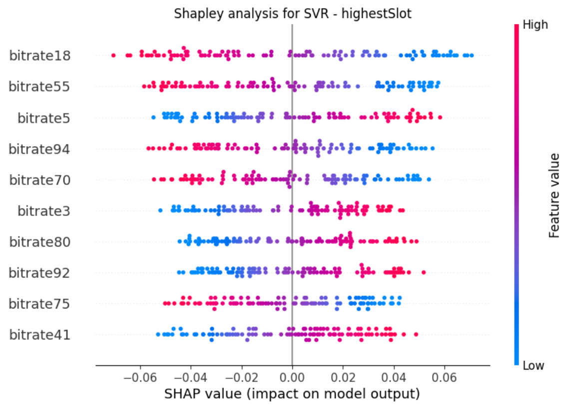

Fig. 5: SHAP summary plot for the SVR model and *highestSlot* metric for Euro28.

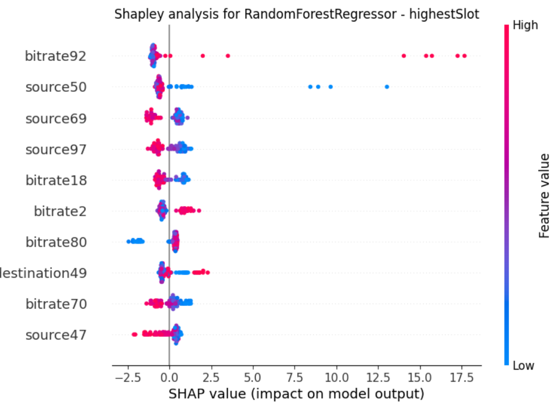

Fig. 6: SHAP summary plot for the RF model and *highestSlot* metric for Euro28.

Algorithm 2 XAI-based feature selection

Input: D – Full dataset
Output: D_t – Transformed dataset with selected features

Symbols:
FI – Feature importance obtained from the decision tree
IC – Table of indexes of features that carry information
c – Features derived from the dataset

function FEATURESELECTION(D)
  $FI \leftarrow DecisionTreeRegressor(D)$
  $IC \leftarrow \oslash$
  for $c \leftarrow D$ **do**
    if $FI[c] > 0$ **then**
      $IC \leftarrow IC \cup \{c\}$
    end if
  end for
end function

TABLE II: MAPE with the results of the t-student test for the predicted metrics for Euro28.

Metric	CART 1	SVR 2	KNN 3	RF 4	LR 5	CART-FS 6	SVR-FS 7	KNN-FS 8	RF-FS 9	LR-FS 10
highestSlot	0.125	0.094	0.093	0.092	0.146	0.131	0.091	0.094	0.099	0.109
	—	1, 5, 6, 10	1, 5, 6	1, 5, 6, 10	—	—	1, 5, 6, 10	1, 5, 6	1, 5, 6	5
avgHighestSlot	0.104	0.067	0.074	0.074	0.100	0.105	0.071	0.077	0.075	0.186
	10	1, 5, 6, 10	1, 5, 6, 10	1, 5, 6, 10	10	10	1, 5, 6, 10	1, 5, 6, 10	1, 5, 6, 10	—
sumOfSlots	0.072	0.051	0.054	0.053	0.068	0.072	0.051	0.057	0.055	0.117
	—	1, 5, 6, 10	1, 5, 6, 10	1, 5, 6, 10	10	—	1, 5, 6, 10	1, 5, 6, 10	1, 5, 6, 10	—
transceivers	0.032	0.020	0.025	0.023	0.031	0.032	0.023	0.026	0.024	0.031
	—	1, 3, 5, 6, 8, 9, 10	1, 5, 6, 10	1, 5, 6, 10	—	—	1, 5, 6, 10	1, 5, 6, 10	1, 5, 6, 10	—

TABLE III: MAPE with the results of the t-student test for the predicted metrics for US26.

Metric	CART 1	SVR 2	KNN 3	RF 4	LR 5	CART-FS 6	SVR-FS 7	KNN-FS 8	RF-FS 9	LR-FS 10
highestSlot	0.197	0.132	0.145	0.146	0.249	0.192	0.135	0.150	0.147	0.207
	5	1, 5, 6, 10	1, 5, 6, 10	1, 5, 6, 10	—	5	1, 5, 6, 10	1, 5, 6, 10	1, 5, 6, 10	5
avgHighestSlot	0.133	0.085	0.090	0.089	0.130	0.133	0.086	0.094	0.093	0.270
	10	1, 5, 6, 10	1, 5, 6, 10	1, 5, 6, 10	10	10	1, 5, 6, 10	1, 5, 6, 10	1, 5, 6, 10	—
sumOfSlots	0.074	0.054	0.054	0.055	0.084	0.075	0.055	0.058	0.057	0.119
	5, 10	1, 5, 6, 10	1, 5, 6, 10	1, 5, 6, 10	10	10	1, 5, 6, 10	1, 5, 6, 10	1, 5, 6, 10	—
transceivers	0.033	0.022	0.023	0.023	0.036	0.032	0.024	0.027	0.024	0.035
	5	1, 5, 6, 7, 8, 10	1, 5, 6, 8, 10	1, 5, 6, 8, 10	—	5	1, 5, 6, 10	1, 5, 6, 10	1, 5, 6, 10	—

V. RESULTS

In this Section, we analyze the performance of the proposed XAI-based feature selection method for all the considered network metrics and regressors for both topologies in terms of prediction quality and runtime. The chosen quality measure is the mean absolute percentage error (MAPE) for a direct comparison between parameters differing vastly in volume. In particular, the average value of *highestSlot* is 96.57 for Euro28 and 147.87 for US26; the average value of *avgHighestSlot* is 65.57 for Euro28 and 66.85 for US26; the average value of *sumOfSlots* is 3677.25 for Euro28 and 3134.25 for US26; and the average value of *transceivers* is 201.32 for Euro28 and 200.42 for US26. To ensure the reliability of the obtained results, in our experiments, we use the 5x2 cross-validation. In the next step, we perform the t-student statistical test with an importance level of 0.05 to determine the significance of the differences between various methods.

The results of our experiments considering the prediction quality are presented in Tab. II for Euro28 and Tab. III for US26. The models denoted with "-FS" were trained using feature selection. For each metric and model, two pieces of information are available: the MAPE value at the top, and the indexes (left to right) of models from which the considered model is significantly better at the bottom. Additionally, the yellow color indicates no statistically significant difference between the same model using all features and after feature selection. For both topologies, the first observation is that using the XAI-based feature selection on various models and metrics does not negatively impact the predictive performance. In most of the cases, the MAPE is at a very similar level, with no statistically significant difference. There are even instances where using fewer features improved the prediction quality. Interestingly, some of the metrics appeared more difficult to predict than others. In particular, the *highestSlot* posed the biggest challenge, while the *transceivers* appeared not at all challenging. Regardless, the lowest overall errors were noted for the SVR regressor in all of the forecasted metrics.

Finally, let us discuss the algorithm runtime. As it does not differ significantly between metrics, below we discuss the averaged measurements. Similarly to the network simulations, the experiments were performed on a machine with the the Intel Core i5-1038NG7 processor and 16 GB of RAM.

In Tab. IV, we present the training and inference times for the considered algorithms, metrics, and topologies, with the models using the complete feature sets and after feature selection. To recall, the average time of one simulation is approximately 100 ms. Thus, obtaining the estimation of a selected network performance metric for a given set of connection requests is orders of magnitude faster than performing a complete simulation. This observation highlights the importance of the performed analysis. Considering the high prediction quality obtained from models trained using data from 100 simulations, estimating the network operation for other connection request sets can be done almost instantly.

Furthermore, it is easily noticeable that applying the proposed XAI-based feature selection mechanism allowed a further significant runtime reduction. It is visible in all models, especially tree-based ones (CART and RF). In particular, their training is four and six times faster, respectively. For the model providing the highest-quality predictions, the SVR, the prediction time is shortened by as much as 15% after the applied XAI-based feature selection. Combined with the

TABLE IV: Algorithm training and prediction time.

Topology	CART	SVR	KNN	RF	LR	CART-FS	SVR-FS	KNN-FS	RF-FS	LR-FS
Training time [ms]										
Euro28	12.44	1.20	0.27	683.94	3.56	2.96	1.02	0.31	215.76	1.08
US26	12.39	1.14	0.27	694.00	3.26	3.06	1.01	0.32	220.55	1.09
Prediction time [ms]										
Euro28	0.16	0.27	2.03	2.55	0.13	0.13	0.23	0.62	2.64	0.10
US26	0.15	0.26	1.76	2.59	0.13	0.13	0.23	0.63	2.63	0.09

discussed no quality decrease, it proves the usefulness of the conducted analysis. The achieved benefits are twofold. First, the considerable model size and, thus, runtime decrease enable a broader "what-if" analysis of the network operation amid various traffic conditions compared to traditional simulation. Second, when an optimal solution is required, the estimation obtained from the regression model combined with some assumed uncertainty can provide boundaries for an ILP to get the solution much faster.

VI. Conclusions

In this paper, we considered the problem of optimizing a regression model of a multilayer network to estimate its performance for an input set of connection requests using XAI techniques. We analyzed five different regression models' operation and predictive performance to estimate four network performance metrics for two large topologies. We conducted an analysis of decision-tree-based feature importance and SHAP-based feature contribution. We showed how the bitrate information is more informative to the models than the data regarding connection source and destination nodes. Finally, we proposed a XAI-based feature selection method for the regressors. Our experimental evaluation revealed the effectiveness of the proposed methodology regarding the prediction quality and runtime. We demonstrated that with no significant quality degradation, the predictions can be obtained significantly faster. Additionally, using a smaller model supports the green networking paradigm.

In the future, we plan to use the created regression model of the heuristic with its uncertainty for creating upper and lower bounds for more effective ILP implementation to find an optimal solution faster.

Acknowledgment

This work was supported by the National Science Center, Poland under Grant 2019/35/B/ST7/04272.

References

[1] T. Panayiotou *et al.*, "Survey on machine learning for traffic-driven service provisioning in optical networks," *IEEE Communications Surveys & Tutorials*, vol. 25, no. 2, 2023.

[2] F. Musumeci *et al.*, "An overview on application of machine learning techniques in optical networks," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 2, 2019.

[3] O. Gerstel *et al.*, "Elastic optical networking: A new dawn for the optical layer?" *IEEE Commun. Mag.*, vol. 50, no. 2, 2012.

[4] V. Lopez *et al.*, "On the benefits of multilayer optimization and application awareness," *J. Light. Technol.*, vol. 35, no. 6, 2017.

[5] M. Klinkowski *et al.*, "Solving large instances of the RSA problem in flexgrid elastic optical networks," *J. Opt. Commun. Netw.*, vol. 8, no. 5, 2016.

[6] Ć. Rožić *et al.*, "A survey of multi-layer network optimization," ONDM 2016.

[7] P. Almasan *et al.*, "Network digital twin: Context, enabling technologies, and opportunities," *IEEE Commun. Mag.*, vol. 60, no. 11, 2022.

[8] F. Nourmohammadi *et al.*, "Using convolutional neural networks for blocking prediction in elastic optical networks," *Applied Sciences*, vol. 14, no. 5, 2024.

[9] P. Ksieniewicz *et al.*, "Pattern recognition model to aid the optimization of dynamic spectrally-spatially flexible optical networks," ICCS 2020.

[10] M. Klinkowski *et al.*, "Application of linear regression in latency estimation in packet-switched 5G xHaul networks," ICTON 2023.

[11] A. Knapińska *et al.*, "Application of ensemble regression methods in elastic optical network optimization," PP-RAI 2024.

[12] A. Adadi *et al.*, "Peeking inside the black-box: a survey on explainable artificial intelligence (XAI)," *IEEE Access*, vol. 6, 2018.

[13] G. O. Ferreira *et al.*, "Forecasting network traffic: A survey and tutorial with open-source comparative evaluation," *IEEE Access*, vol. 11, 2023.

[14] M. Ibrahimi *et al.*, "Machine learning methods for quality-of-transmission estimation," in *Machine Learning for Future Fiber-Optic Communication Systems*. Elsevier, 2022.

[15] P. Lechowicz, "Regression-based fragmentation metric and fragmentation-aware algorithm in spectrally-spatially flexible optical networks," *Comput. Commun.*, vol. 175, 2021.

[16] M. Malekzadeh, "Performance prediction and enhancement of 5G networks based on linear regression machine learning," *EURASIP Journal on Wireless Communications and Networking*, vol. 74, no. 1, 2023.

[17] Z. Zhong *et al.*, "Routing without routing algorithms: an AI-based routing paradigm for multi-domain optical networks," OFC 2019.

[18] L. M. M. Zorello *et al.*, "Black-box optimization for anticipated baseband-function placement in 5G networks," *Comput. Netw.*, vol. 245, 2024.

[19] O. Karandin *et al.*, "If not here, there. explaining machine learning models for fault localization in optical networks," ONDM 2022.

[20] O. Ayoub *et al.*, "On using explainable artificial intelligence for failure identification in microwave networks," ICIN 2022.

[21] A. Morichetta *et al.*, "Explain-it: Towards explainable AI for unsupervised network traffic analysis," Big-DAMA 2019.

[22] A. Knapińska *et al.*, "Explainable artificial intelligence-guided optimization of ML-based traffic prediction," ONDM 2024.

[23] O. Ayoub *et al.*, "Towards explainable reinforcement learning in optical networks: The RMSA use case," OFC 2024.

[24] O. Ayoub *et al.*, "Towards explainable artificial intelligence in optical networks: the use case of lightpath QoT estimation," *J. Opt. Commun. Netw.*, vol. 15, no. 1, 2022.

[25] A. Knapińska *et al.*, "On advantages of traffic prediction and grooming for provisioning of time-varying traffic in multilayer networks," ONDM 2023.

[26] M. Jinno *et al.*, "Distance-adaptive spectrum resource allocation in spectrum-sliced elastic optical path network," *IEEE Commun. Mag.*, vol. 48, no. 8, 2010.

[27] Ciena, https://www.ciena.com/insights/data-sheets/800g-wavelogic-5-extreme-motr-module.html.

[28] J. Perelló *et al.*, "Reducing the number of transceivers with probabilistic constellation shaping in flex-grid over MCF optical backbone networks," ICC 2023.

[29] W. Borek *et al.*, "Inductive parallel learning for multiple classification problems," IJCNN 2022.

[30] Y. Fan *et al.*, "Linear regression vs. deep learning for signal quality monitoring in coherent optical systems," *IEEE Photonics Journal*, vol. 14, no. 4, 2022.

[31] S. M. Lundberg *et al.*, "A unified approach to interpreting model predictions," NeurIPS 2017.

[32] W. E. Marcílio *et al.*, "From explanations to feature selection: assessing SHAP values as feature selection mechanism," SIBGRAPI 2020.

[33] D. Fryer *et al.*, "Shapley values for feature selection: The good, the bad, and the axioms," *IEEE Access*, vol. 9, 2021.