Explaining Aggregated Network Traffic Predictors

Aleksandra Knapińska[1,2] and 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, krzysztof.walkowiak}@pwr.edu.pl

***Abstract*—Network traffic prediction is essential for the intelligent management of modern backbone networks. In application-aware settings, it becomes crucial to generate detailed forecasts for each traffic class to ensure they are handled with appropriate care. To address scalability and survivability challenges, models built using data aggregation techniques offer an effective solution. In this paper, we examine how such models operate to make successful forecasts for diverse traffic classes in real and semi-synthetic data, incorporating Explainable Artificial Intelligence (XAI) tools. The analysis reveals interesting trends in how various regressors capture cross-class intricacies and correlations, highlighting the potential of aggregated models.**

***Index Terms*—traffic prediction, data aggregation, machine learning, explainable AI**

I. Introduction

Network traffic prediction is essential for the intelligent management of modern backbone networks [1]. In such architectures, prior knowledge of upcoming traffic changes enables early resource preparation and more effective administration of connected clients. Identifying traffic patterns also contributes to energy efficiency, as resources can be better managed and activated only when needed [2]. Furthermore, in application-aware settings, detailed forecasts for each traffic class are necessary to ensure they are handled with appropriate care [3]. However, as the number of distinct traffic classes increases, scalability becomes a concern. Models built using data aggregation techniques address both scalability and survivability challenges by improving generalization and reducing the number of models required.

Two main approaches to data aggregation in network traffic prediction have been proposed in the literature. The first is *multioutput regression* [4], where a single model simultaneously forecasts traffic across all data classes. The second is *agnostic prediction* [5], where the same model is applied to each traffic class individually, one after another. Previous studies have highlighted the strong capabilities of both approaches and demonstrated how their prediction quality compares favorably with that of dedicated models for each traffic class. However, since Machine Learning (ML) models operate as black-box systems, the details of their internal reasoning remain unknown.

Explainable Artificial Intelligence (XAI) methods provide researchers with a window into these black boxes, helping them understand how the models operate [6]. By doing so, XAI methods enable the extraction of valuable information, particularly regarding the importance of various model inputs, which can enhance both the understanding and performance of the models. This can be achieved in two main ways: by using interpretable models (e.g., decision trees), or by explaining black-box models in a *post-hoc* manner, i.e., after they have been trained. The former offers a clean solution, allowing users to understand the model's reasoning, but such models are often less powerful than state-of-the-art ones. The latter, on the other hand, typically provides an approximate insight into the model's operation, but with much better prediction quality. Moreover, *post-hoc* explanations can be applied to a wide range of ML models.

In the field of backbone network optimization, the explainability of applied ML models is crucial. In practical settings, simple heuristic methods are often favored, as more sophisticated ML-based methods are difficult to fully understand and are therefore considered too risky. Consequently, the research community continues to focus on developing more transparent models that support network operations to enhance their adoption. Efforts toward explainable ML-based Quality of Transmission (QoT) estimation [7] serve as a prime example in this area. However, in the context of backbone network traffic prediction, only introductory research has been conducted so far [8].

Therefore, in this paper, we address the issue of explaining aggregated network traffic prediction models. We examine the operation of *multioutput* and *agnostic* approaches with various base regressors on both real and semi-synthetic datasets. Our analysis reveals interesting dependencies that enhance the understanding of their operation and may serve as a basis for further optimizations. Furthermore, it demonstrates how data aggregation techniques improve the forecasting capabilities of models by leveraging cross-class correlations and dependencies.

The remainder of the paper is organized as follows. Section II discusses the current literature related to explainability of ML models in networking problems. Section III presents the aggregated traffic prediction models we consider. Section IV gives an overview of the datasets used in our study. Following that, Section V discusses the prediction performance of the aggregated models on these datasets. Section VI dives into the internal reasoning of the models. Finally, Section VII concludes this work.

II. Related Work

In this section, we review recent literature related to XAI in the context of communication networks. Arguably, the most extensively studied area is QoT estimation [7]. Several studies

ISBN 978-3-903176-72-0

have focused on quantifying the contribution of input features [9], simplifying models [10], [11], and validating uncertain decisions [12]. These analyses have not only improved the understanding of black-box models but also supported their refinement, primarily through effective feature engineering.

Another significant area of research is anomaly and failure management, which encompasses several core tasks, including the detection of anomalies [13]–[15], the localization of faults [16], [17], and, finally, the identification and diagnosis of issues [18]–[20].

A more recent research direction is ML-based and ML-assisted Routing and Spectrum Allocation (RSA). Initial efforts to understand autonomous network management using reinforcement learning have recently emerged [21]. Similarly, attempts to enhance heuristic-based approaches through the application of XAI are gaining traction [22], [23]. Notably, the first attempt to explain a network regression model estimating an RSA heuristic has also been recently proposed [24].

In contrast, the field of explainable network traffic prediction remains virtually unexplored. A recent study made initial efforts to understand and optimize traffic forecasters using XAI-based feature selection [8]. Additionally, explainable traffic identification was addressed in another study [25]. However, to the best of our knowledge, no prior work has examined explainability in the context of aggregated network traffic prediction models. To fill this gap, we conduct such an analysis to identify common trends and enable future optimization of these models.

III. Aggregated Traffic Prediction Models

In this Section, we introduce the details of two aggregated traffic prediction models considered in this work.

A. Multioutput Regression

When the number of expected outputs (e.g., traffic types) is known and constant, *multioutput regression* is one way to reduce the number of required models. To achieve this, we create a joint training set with all the traffic types labeled. This approach enables the simultaneous forecasting of all traffic types, reducing the number of models needed to just one. Additionally, because the relationships between the targets are available to the model, multioutput regression methods generally offer better predictive performance compared to using multiple single-output models [26]. One potential downside of this approach is its limited scalability when new traffic types are introduced.

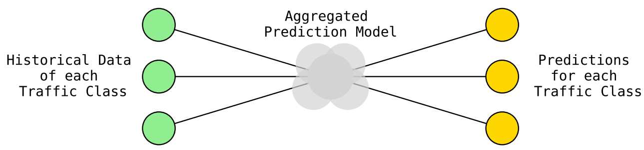

Fig. 1. Illustration of the *multioutput* aggregated traffic prediction model.

The idea behind the model is illustrated in Fig. 1, which shows an example with three traffic classes. The model takes as inputs historical traffic measurements across all considered traffic types and generates forecasts for each of these types. The aggregated training set contains traffic samples, with corresponding features added as subsequent columns. The multi-target model outputs predictions for all traffic classes simultaneously, accounting for their history and the correlations between them.

B. Agnostic Prediction

When the number of traffic classes to be forecasted is unknown, non-constant, or very large, the second model, *agnostic prediction*, can serve as a potential solution. It can be tailored, for example, to forecast individual connection requests for various traffic classes in application-aware networks. The idea behind this approach is to create a general aggregated prediction model that is agnostic to traffic type. Instead of training separate dedicated models or using a multioutput regressor, all available historical traffic measurements from multiple classes are combined into a single, general training set. In other words, all historical traffic data are merged into one single-output model. To avoid increasing complexity, traffic class labels are not retained. Instead, all traffic samples are treated equally, as though they belong to the same traffic class.

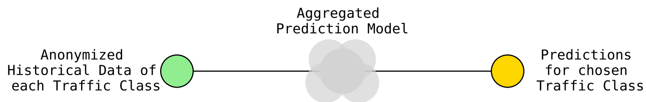

Fig. 2. Illustration of the *agnostic* aggregated traffic prediction model.

The concept of agnostic prediction is illustrated in Fig. 2. For instance, when considering three traffic classes, the proposed model is trained with three times more data than traditional dedicated models. Additionally, only one model is created, acting as a versatile predictor. As a result, there is no need for traffic identification (i.e., an additional classification module) to select the appropriate model. Furthermore, the training set is significantly augmented compared to traditional dedicated models, providing more data for model development.

Given the high diversity among traffic classes in the network, the samples in such a general training set may differ substantially. On one hand, this could make it challenging for the model to identify clear patterns and generalize effectively. On the other hand, the increased size and diversity of the training set could help the predictor adapt to changes in each traffic type over time and potentially enable the forecasting of previously unseen traffic classes.

C. Features and Regressors

Each of the considered models is trained using input features extracted from the raw traffic data. Following the guidance in [8], we create two features for each traffic class: a traffic sample taken a day before the forecasted one (*yesterday*), and a traffic sample taken a week before the forecasted one (*last_week*).

Since the aggregated models considered in this work are generic, they can be used with any regression algorithm. In this work, we explore their performance with three distinct base regressors: a simple model – Linear Regression (LR); a tree-based model – Random Forest (RF); and a neural-network-based model – Multilayer Perceptron (MLP).

IV. Datasets

In this Section, we introduce the datasets considered in this work.

A. Semi-Synthetic Traffic–Sandvine Report and Traffic Weaver

The first source of data is based on the Sandvine Mobile Internet Phenomena Report [27]. The report contains information from multiple client networks across various continents and presents the data in an aggregated form. It provides extensive statistics on the daily usage patterns of various network-based applications (e.g., YouTube, Zoom, TikTok), presented as bar plots with averaged information about the time-variability of service usage within 24 hours, using 1-hour intervals.

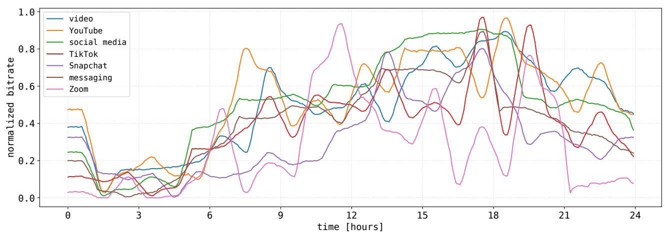

Fig. 3. Sandvine dataset illustration – daily traffic patterns of various network-based services and applications generated with `Traffic Weaver` [28].

Since the data provided by Sandvine is primarily demonstrative, it serves as an excellent base for creating *semi-synthetic* datasets, where real patterns are used as the basis for a generator. To this end, we recreate the shapes visible in the provided plots as numerical sequences and use the *Traffic Weaver* Python package [28] to generate a continuous signal with finer granularity. For experimental purposes, we also introduce Gaussian noise. The daily patterns of the seven traffic classes considered in this work are illustrated in Fig. 3. To conduct the experiments, we generate a two-month-long semi-synthetic dataset containing signals for each of these patterns.

B. Real Traffic – Seattle Internet Exchange Point

The second source comprises real data from the Seattle Internet Exchange Point (SIX), published in [29]. SIX interconnects hundreds of networks and data centers, carrying more than 2 Tbps of peak bitrate. The SIX website has provided traffic statistics for over twenty-five years, showing how traffic varies throughout the day and over extended periods. In this work, we consider a nearly five-month-long traffic dataset with 5-minute sampling intervals. The data includes information about traffic across 13 frame sizes, representing traffic classes denoted by the letters *a*–*m*, as in [4].

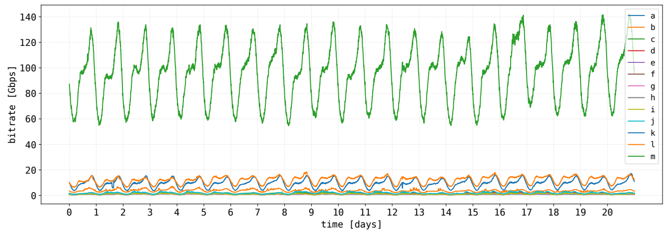

Fig. 4. SIX dataset illustration, representative three-week fragment.

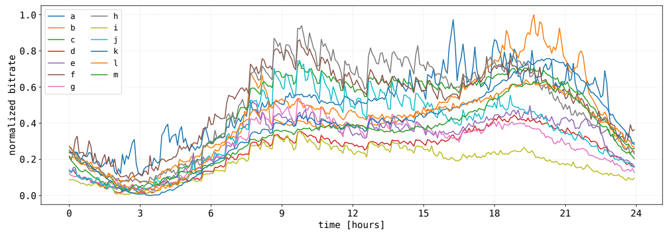

Fig. 5. SIX dataset illustration, representative day with normalized bitrate.

All traffic classes exhibit clear daily patterns, though they differ in terms of bitrate, as illustrated in Fig. 4. Furthermore, Fig. 5 shows a representative day with normalized bitrate, highlighting numerous daily micro-fluctuations.

C. Experimental Setup

For the semi-synthetic Sandvine dataset, we use the first 30 days as the training set. For the SIX dataset with real data, we train the models on the first two months' worth of traffic measurements (60 days). To recall, we use a 5-minute sampling interval. In both datasets, the models are tested on multiple subsets of day-long demands constructed from the remaining data. Specifically, after training the models, the remaining traffic data with corresponding features are divided into 288-sample (24-hour) sections. For each section, the input features are provided to the trained models, and the 24-hour forecasts are compared with actual traffic values. The motivation behind this experimental design is to evaluate the ability of the proposed models to forecast the entire traffic pattern. Furthermore, existing studies (e.g., [30]) argue that lightpaths might be set up for connection requests several hours in advance, based on forecasts. Thus, it is essential to predict the bitrate of each demand several samples ahead. We discuss the averaged errors across all demands within traffic classes and datasets in the following Section.

V. Prediction Quality

In this Section, we discuss the prediction quality of the considered aggregated models in comparison to the baseline dedicated models for each traffic class. We use the Mean Absolute Percentage Error (MAPE) as a metric to directly compare performance across traffic classes and datasets.

The MAPE for each regressor, averaged across traffic classes, is presented in Tab. I–II for the Sandvine and SIX datasets, respectively. First, we observe excellent prediction quality, with very low errors of around 5% for each regressor across both datasets. It is also evident that, in almost all

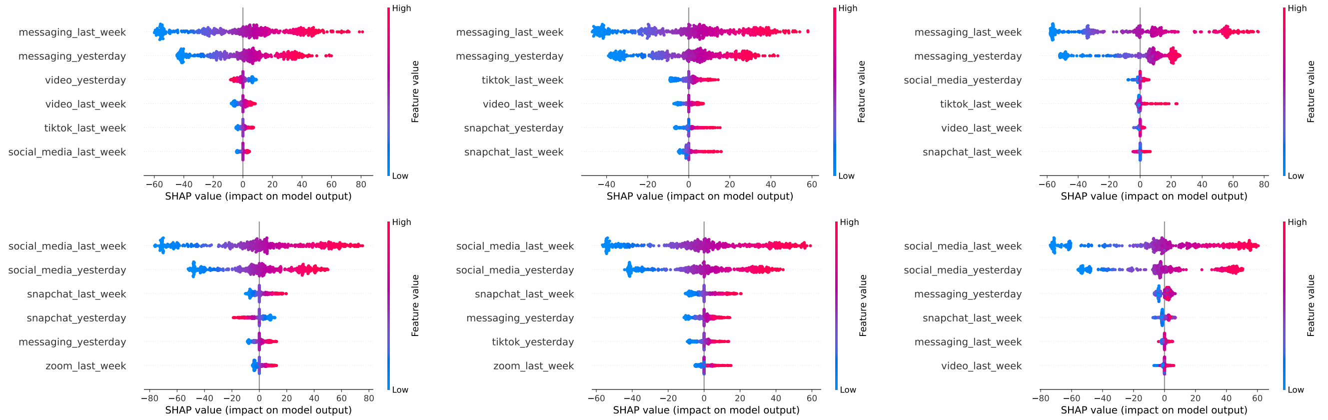

Fig. 6. SHAP summary plots for the Sandvine dataset, multioutput model predicting the *messaging* (first row) and *social media* (second row) traffic. LR (left), MLP (middle), and RF (right) regressors.

cases, either of the data aggregation approaches helps reduce prediction errors across traffic classes. With the exception of the MLP for the SIX dataset, aggregating all traffic classes into one model not only reduces the number of models but also improves prediction quality. Additionally, we observe interesting trends between regressors: RF generalizes better with the multioutput aggregation, while LR performs better with the agnostic approach.

TABLE I
MAPE FOR THE SANDVINE DATASET AVERAGED OVER TRAFFIC CLASSES.

ML algorithm	dedicated models	multioutput model	agnostic model
LR	0.0529	0.0538	**0.0527**
MLP	0.0527	0.0562	**0.0526**
RF	0.0572	**0.0491**	0.0565

TABLE II
MAPE FOR THE SIX DATASET AVERAGED OVER TRAFFIC CLASSES.

ML algorithm	dedicated models	multioutput model	agnostic model
LR	0.0434	0.0581	**0.0413**
MLP	**0.0415**	0.0641	0.0452
RF	0.0517	**0.0495**	0.0510

These results demonstrate the validity of data aggregation models for backbone network traffic prediction. The flexibility gained from using a single predictor is coupled with improved prediction consistency and quality. However, given the lack of insight into the internal reasoning of the models, operators may remain skeptical about implementing forecasters in their networks. Thus, in the next Section, we attempt to explain how the regressors arrive at their predictions.

VI. Model Internal Reasoning

To gain insight into the internal reasoning of the models used with various base regressors, we employ SHapley Additive exPlanations (SHAP) [31] – a game-theory-based framework that explains the output of any ML model by estimating the contribution of each feature. SHAP calculates the values for a specific model using the trained ML model and the corresponding training data. In regression tasks, the SHAP value of a feature represents its contribution to the model's prediction. A positive (or negative) SHAP value indicates that the feature has a positive (or negative) impact on the prediction, i.e., it increases (or decreases) the predicted value.

Below, we analyze SHAP summary plots, which rank features for a particular model from most (top) to least (bottom) influential. For space constraints, we limit the plots to the top-6 features. The x-axis represents the impact of each feature on the model's output, i.e., how much the value of a given feature shifts the prediction in either the positive or negative direction. The color of each point corresponds to the actual value of the feature. Finally, for each feature, each point on the plot represents the SHAP value assigned to that feature for a particular prediction. The visually wider vertical sections of the plot indicate that more features in the dataset have been assigned a corresponding SHAP value along the x-axis.

A. *Multioutput*

In the case of multioutput models, we can analyze the contribution of features to each output. Starting with the semi-synthetic Sandvine dataset, we obtained results that were largely as expected (see representative cases plotted in Fig. 6). The most influential features were overwhelmingly those describing the past measurements of the forecasted traffic class. This dependence was observed across all regressors and traffic classes in this dataset. However, the impact of other traffic types was non-zero (as seen in the example plots), suggesting that including additional traffic classes helps the model reason and influences its predictions.

However, a similar analysis conducted on the real SIX dataset revealed much more nuanced trends. Fig. 7 presents example SHAP summary plots for various traffic classes. In this case, for some outputs, features related to *other* traffic types appear more influential. Notably, for the MLP regressor, the most important feature is consistently associated with

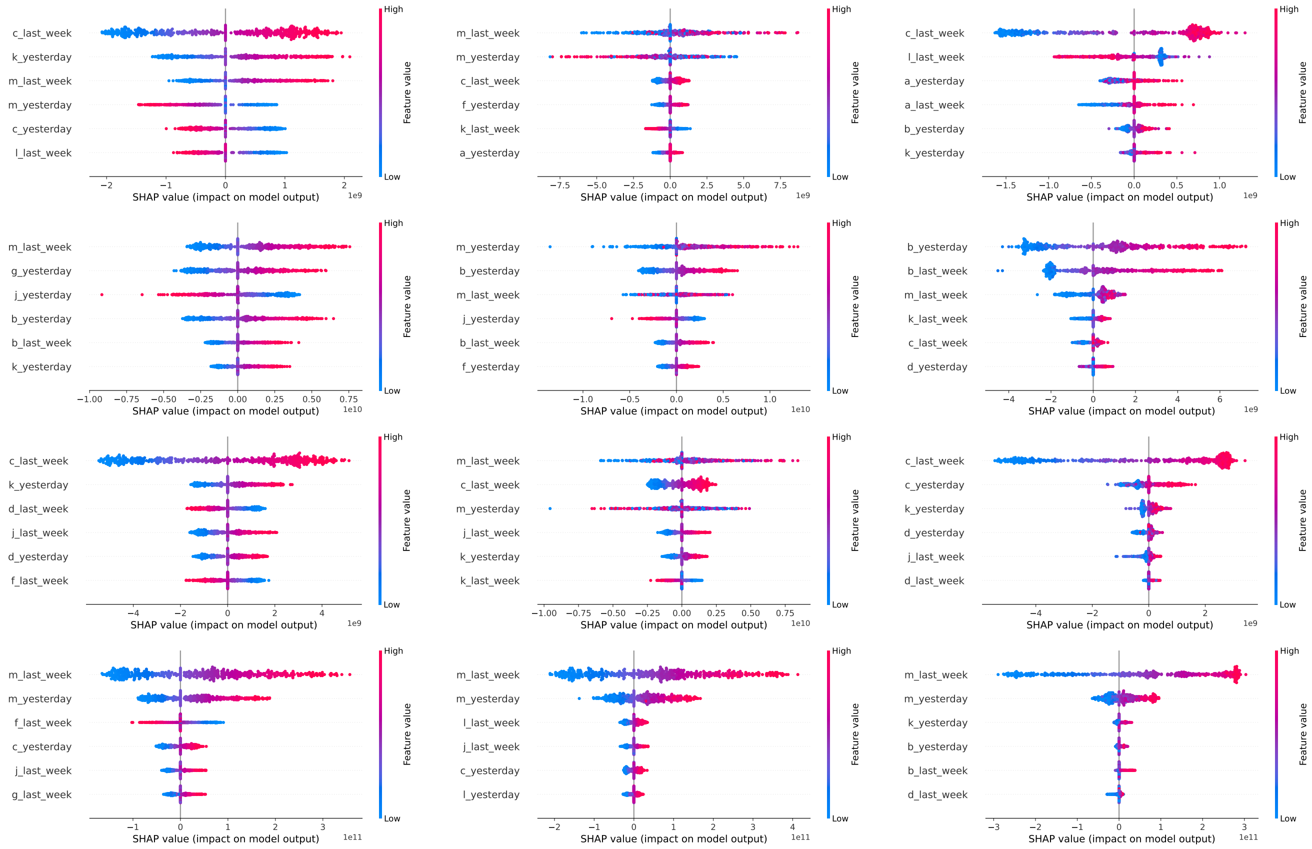

Fig. 7. SHAP summary plots for the SIX dataset, multioutput model predicting the *a* (first row), *b* (second row), *c* (third row), and *m* (fourth row) traffic. LR (left), MLP (middle), and RF (right) regressors.

traffic class *m*, which has by far the highest bitrate (recall Fig. 4). This suggests that the neural-network-based model heavily relies on patterns from this high-volume traffic class when forecasting others. Interestingly, for low-bitrate traffic classes, an increase in *m*-class traffic does not necessarily correspond to an increase in the predicted bitrate for those classes. This can be inferred from the distribution of point colors in the plots, indicating that the MLP regressor captures subtle interdependencies rather than relying on direct proportional relationships between traffic types.

This behavior, however, differs for the other regressors. The RF model generally focuses on the historical data of the forecasted traffic class, while the LR model often assigns importance to features from traffic classes *c* and *m*.

When forecasting traffic *m* itself, the most influential features across all regressors are, unsurprisingly, those directly related to class *m*. In particular, the MLP and RF models show very limited influence from other traffic types, reinforcing the dominant role of *m* in the dataset.

This analysis illustrates that aggregated multioutput models can effectively leverage information from the historical traffic of other classes to improve their predictions. By combining all traffic classes into a single joint model, the learning algorithm can identify cross-class relationships and exploit shared temporal patterns within the network to achieve better generalization.

B. *Agnostic*

In contrast to the previous model, agnostic prediction does not retain traffic class labels. As a result, it is not possible to analyze the influence of specific traffic classes on the model's final prediction. Nevertheless, the consistently low prediction errors suggest that the models generalize well and are capable of identifying patterns within such diversified training datasets.

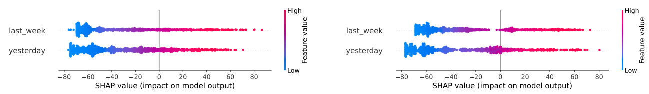

Fig. 8. SHAP summary plots for the Sandvine dataset, agnostic. MLP (left) and RF (right) regressors.

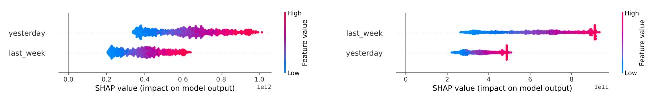

Fig. 9. SHAP summary plots for the SIX dataset, agnostic. MLP (left) and RF (right) regressors.

Fig. 8–9 present SHAP summary plots for the agnostic models applied to the Sandvine and SIX datasets, respectively. For the Sandvine dataset, both features – traffic measured one day prior and one week prior – exert a similar influence on the model's predictions. In contrast, for the SIX dataset, one of these features typically dominates (as indicated by a more pronounced spread along a given row in the plot), although the dominant feature varies depending on the regressor.

Although the agnostic model offers greater flexibility in handling a larger number of traffic classes and forecasting previously unseen ones, it is clearly less transparent than the multioutput model. In terms of explainability, the reasoning behind the multioutput approach is thus easier to interpret.

VII. Conclusions

In this paper, we addressed the problem of traffic prediction in backbone communication networks by exploring two data aggregation approaches for forecasting multiple traffic classes. We conducted an XAI analysis of the internal workings of *multioutput* and *agnostic* models, each combined with diverse base regressors, to predict traffic types in both real and semi-synthetic datasets. The study revealed several noteworthy trends, showing how these models effectively leverage historical data from different traffic types to enhance the quality of future predictions across classes. As a result, both aggregation approaches not only reduce the number of required models but also maintain or improve the prediction quality for each traffic class in a stable and consistent manner.

Acknowledgment

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

References

[1] T. Panayiotou, M. Michalopoulou, and G. Ellinas, "Survey on machine learning for traffic-driven service provisioning in optical networks," *IEEE Communications Surveys & Tutorials*, vol. 25, no. 2, pp. 1412–1443, 2023.

[2] H. Khodakarami, B. S. G. Pillai, and W. Shieh, "Quality of service provisioning and energy minimized scheduling in software defined flexible optical networks," *Journal of Optical Communications and Networking*, vol. 8, no. 2, pp. 118–128, 2016.

[3] L. Velasco, S. Barzegar, F. Tabatabaeimehr, and M. Ruiz, "Intent-based networking and its application to optical networks," *Journal of Optical Communications and Networking*, vol. 14, no. 1, pp. A11–A22, 2022.

[4] A. Knapińska, P. Lechowicz, and K. Walkowiak, "Machine-learning based prediction of multiple types of network traffic," in *21st International Conference on Computational Science (ICCS)*, 2021.

[5] A. Knapińska, P. Lechowicz, S. Spadaro, and K. Walkowiak, "Agnostic prediction of multiple types of time-varying traffic in optical networks," in *IEEE Global Communications Conference (GLOBECOM)*, 2023.

[6] A. Adadi and M. Berrada, "Peeking inside the black-box: A survey on explainable artificial intelligence (XAI)," *IEEE Access*, vol. 6, pp. 52 138–52 160, 2018.

[7] O. Ayoub, S. Troia, D. Andreoletti, A. Bianco, M. Tornatore, S. Giordano, and C. Rottondi, "Towards explainable artificial intelligence in optical networks: the use case of lightpath QoT estimation," *Journal of Optical Communications and Networking*, vol. 15, no. 1, pp. A26–A38, 2022.

[8] A. Knapińska, O. Ayoub, C. Rottondi, P. Lechowicz, and K. Walkowiak, "Explainable artificial intelligence-guided optimization of ML-based traffic prediction," in *28th International Conference on Optical Network Design and Modelling (ONDM)*, 2024.

[9] O. Ayoub, D. Andreoletti, S. Troia, S. Giordano, A. Bianco, and C. Rottondi, "Quantifying features' contribution for ML-based quality-of-transmission estimation using explainable AI," in *European Conference on Optical Communication (ECOC)*, 2022.

[10] O. Ayoub, A. Bianco, D. Andreoletti, S. Troia, S. Giordano, and C. Rottondi, "On the application of explainable artificial intelligence to lightpath QoT estimation," in *Optical Fiber Communication Conference (OFC)*, 2022.

[11] H. Fawaz, F. Arpanaei, D. Andreoletti, I. Sbeity, J. A. Hernández, D. Larrabeiti, and O. Ayoub, "Reducing complexity and enhancing predictive power of ML-based lightpath QoT estimation via SHAP-assisted feature selection," in *28th International Conference on Optical Network Design and Modeling (ONDM)*, 2024.

[12] H. Houssiany, O. Ayoub, C. Rottondi, and A. Bianco, "Using SHAP values to validate model's uncertain decision for ML-based lightpath quality-of-transmission estimation," in *23rd International Conference on Transparent Optical Networks (ICTON)*, 2023.

[13] P. Barnard, N. Marchetti, and L. A. DaSilva, "Robust network intrusion detection through explainable artificial intelligence (XAI)," *IEEE Networking Letters*, vol. 4, no. 3, pp. 167–171, 2022.

[14] T. Senevirathna, B. Siniarski, M. Liyanage, and S. Wang, "Deceiving post-hoc explainable AI (XAI) methods in network intrusion detection," in *21st Consumer Communications & Networking Conference (CCNC)*, 2024, pp. 107–112.

[15] N. Moustafa, N. Koroniotis, M. Keshk, A. Y. Zomaya, and Z. Tari, "Explainable intrusion detection for cyber defences in the internet of things: Opportunities and solutions," *IEEE Communications Surveys & Tutorials*, vol. 25, no. 3, pp. 1775–1807, 2023.

[16] F. Ezzeddine, O. Ayoub, D. Andreoletti, M. Tornatore, and S. Giordano, "Vertical split learning-based identification and explainable deep learning-based localization of failures in multi-domain nfv systems," in *IEEE Conference on Network Function Virtualization and Software Defined Networks (NFV-SDN)*, 2023.

[17] O. Karandin, O. Ayoub, F. Musumeci, Y. Hirota, Y. Awaji, and M. Tornatore, "If not here, there. Explaining machine learning models for fault localization in optical networks," in *26th International Conference on Optical Network Design and Modeling (ONDM)*, 2022.

[18] M. Moulay, R. G. Leiva, P. J. R. Maroni, F. Diez, V. Mancuso, and A. F. Anta, "Automated identification of network anomalies and their causes with interpretable machine learning: The CIAN methodology and TTrees implementation," *Computer Communications*, vol. 191, pp. 327–348, 2022.

[19] O. Ayoub, N. Di Cicco, F. Ezzeddine, F. Bruschetta, R. Rubino, M. Nardecchia, M. Milano, F. Musumeci, C. Passera, and M. Tornatore, "Explainable artificial intelligence in communication networks: A use case for failure identification in microwave networks," *Computer Networks*, vol. 219, p. 109466, 2022.

[20] O. Ayoub, F. Musumeci, F. Ezzeddine, C. Passera, and M. Tornatore, "On using explainable artificial intelligence for failure identification in microwave networks," in *25th Conference on Innovation in Clouds, Internet and Networks (ICIN)*, 2022.

[21] O. Ayoub, C. Natalino, and P. Monti, "Towards explainable reinforcement learning in optical networks: The RMSA use case," in *Optical Fiber Communications Conference and Exhibition (OFC)*, 2024.

[22] R. Goścień, "Traffic prediction-and explainable artificial intelligence-based dynamic routing in software-defined elastic optical networks," in *IFIP Networking Conference*, 2024.

[23] R. Goścień, "Explainable artificial intelligence-based framework for efficient content placement in elastic optical networks," *Expert Systems with Applications*, vol. 262, p. 125541, 2025.

[24] K. Duszyńska, P. Polski, M. Włosek, A. Knapińska, P. Lechowicz, and K. Walkowiak, "XAI-guided optimization of a multilayer network regression model," in *IFIP Networking Conference*, 2024.

[25] A. Morichetta, P. Casas, and M. Mellia, "EXPLAIN-IT: Towards explainable AI for unsupervised network traffic analysis," in *3rd ACM CoNEXT Workshop on Big DAta, Machine Learning and Artificial Intelligence for Data Communication Networks (Big-DAMA)*, 2019.

[26] H. Borchani, G. Varando, C. Bielza, and P. Larrañaga, "A survey on multi-output regression," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 5, no. 5, pp. 216–233, 2015.

[27] Sandvine, "The Mobile Internet Phenomena Report," May 2021.

[28] P. Lechowicz, A. Knapińska, A. Włodarczyk, and K. Walkowiak, "Traffic Weaver: Semi-synthetic time-varying traffic generator based on averaged time series," *SoftwareX*, vol. 28, p. 101946, 2024.

[29] Seattle Internet Exchange Point. https://www.seattleix.net/statistics/.

[30] A. Włodarczyk, A. Knapińska, P. Lechowicz, and K. Walkowiak, "Machine learning assisted provisioning of time-varying traffic in translucent optical networks," *IEEE Access*, vol. 12, pp. 110 193–110 212, 2024.

[31] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," *Advances in neural information processing systems (NeurIPS)*, vol. 30, 2017.