Machine Learning enabled Fault-Detection Algorithms for Optical Spectrum-as-a-Service Users

Sai Kireet Patri*†, Isabella Dick*◇, Kaida Kaeval*◇, Jasper Müller*†, Jose-Juan Pedreno-Manresa*, Achim Autenrieth*, Jörg-Peter Elbers*, Marko Tikas‡ and Carmen Mas-Machuca†

* ADVA, Martinsried, Germany
Email: {spatri, jmueller, jpedrenomanresa, aautenrieth, jpelbers}@adva.com
† Chair of Communication Networks, School of Computation, Information and Technology, Technical University of Munich (TUM), Germany
Email: cmas@tum.de
‡ Transmission Networks Tele2 Estonia AS Tallinn, Estonia
Email: marko.tikas@tele2.com

***Abstract*—The growing usage of high-bandwidth, low-latency applications has led to a significant increase in data traffic in recent years. To meet this demand, optical network operators have begun upgrading to Elastic Optical Networks (EONs), powered by Flexible Bandwidth Variable Transceivers (Flex-BVTs). Encouraged by the disaggregation trend, where Flex-BVTs and Open Line Systems (OLS) are owned and controlled by different parties, the operators are introducing new service models like Optical Spectrum-as-a-Service (OSaaS) in their networks. The OSaaS user in this service model perceives OLS as a transparent light tunnel with no monitoring points other than the Flex-BVTs. As multiple OSaaS users share the same OLS, these networks are more susceptible to failures caused by power degradation or channel interference. To reduce system disruptions and repair costs, it is therefore crucial to detect, identify, and counter such failures timely.**

In this work, we investigate the methods for OSaaS users to detect and identify failures as early as possible using only the telemetry data available from the end Flex-BVTs. Deploying and monitoring five Flex-BVTs within a 400-GHz dedicated OSaaS channel on a pan-European live network for 45 days, we evaluate the applicability of two Machine Learning (ML) based algorithms for EON failure detection, namely, an Artificial Neural Network (ANN) model with dynamic threshold calculation, and a One-Class Support Vector Machine (OCSVM) model. Our results show that the ANN-based approach can detect all artificially introduced failures, with a misclassification rate of 0.01% as compared to the OCSVM-based approach which was unable to detect up to one-third of artificially introduced failures, along with a misclassification rate of 0.6%.

***Index Terms*—Optical Spectrum-as-a-Service, Optical Networks, Network Monitoring, Fault Detection, Machine Learning**

This work has been partially funded in the framework of the CELTIC-NEXT project AI-NET-PROTECT (Project ID C2019/3-4) by the German Federal Ministry of Education and Research (**#16KIS1279K**). Carmen Mas-Machuca acknowledges the support of the Federal Ministry of Education and Research of Germany (BMBF) in the programme "Souverän. Digital. Vernetzt." joint project 6G-life (**#16KISK002**).

◇ Isabella Dick and Kaida Kaeval contributed to this work during their affiliation with ADVA Germany, and are currently affiliated with TWAICE Technologies GmBH, Munich, Germany, and Tallinn Technical University, Tallinn, Estonia respectively.

I. Introduction

In recent years, there has been a significant rise in data traffic growth due to the increasing usage of high-bandwidth, low-latency applications such as ultra-high-definition video streaming, online gaming, and mission-critical applications. To address this demand, optical network operators have begun upgrading from fixed-grid optical networks to Elastic Optical Networks (EONs) [1]. EONs allow network operators to efficiently adjust the optical bandwidth of each optical transponder in the network, based on the demand's bit rate and transmission distance [2]. However, despite their advantages, EONs are more susceptible to soft failures than fixed-grid optical networks due to the increased complexity of the individual network components and various link impairments [3]. Soft failures are defined as those failures that affect the signal quality through slowly varying phenomena that manifest themselves as anomalies in the Optical Performance Monitoring (OPM) data [4]. These anomalies can gradually raise the Bit Error Rate (BER), leading to potential service downtime. Therefore, detecting, identifying, and counteracting these soft failures can significantly reduce system disruptions and repair costs.

To leverage the full potential of the infrastructure of EONs through flexible resource utilization, the concept of Optical Spectrum as a Service (OSaaS) has emerged [5]. In OSaaS, optical transceivers are owned and controlled by the service end user, while the Open Line System (OLS), which provides signal equalization, transportation, and amplification, is controlled by the optical network operator. This method enables flexible resource utilization inside the dedicated customer spectrum of the OSaaS and eliminates unnecessary optical-electrical-optical conversions [6].

In light of the anticipated high demand for OSaaS, this study addresses the question of how OSaaS users can detect and identify failures at an early stage while only using information that is available at the end-user flexible bandwidth variable transceiver modules (Flex-BVTs). The approach is completely

978-3-903176-54-6

independent of OLS element telemetry data, as OLS-related parameters and configurations are not known to the OSaaS service user and will thus not be considered in the scope of this study.

In this study, we measure OPM data from five commercial Flex-BVTs deployed on a live long-haul pan- European network. We act as OSaaS users with five Flex- BVTs within our control but possess no information of the underlying network. To enable single-ended measurements from the test site, a physical optical loop-back is featured in the far-end ROADM in the network, efficiently doubling the optical distance. The total transmission distance of the channels under test (CUTs) is 1792 km and they are continuously monitored for 45 days. The input data points are considered to be the OPM data of all the CUTs.

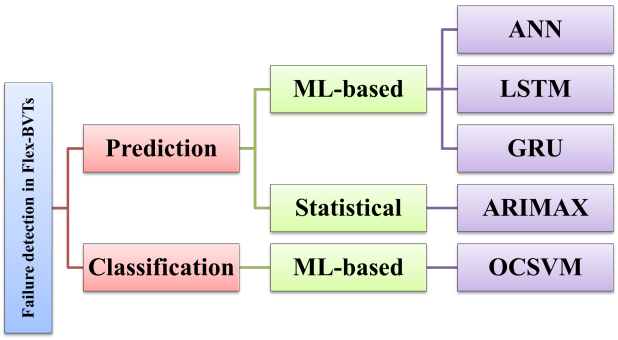


Fig. 1: Failure detection algorithms investigated in this work.

In long-haul networks, soft failures resulting in the gradual degradation of the signals occur infrequently compared to hard failures, introduced by fiber cuts, power outages, fatal equipment failure, or human errors in handling fibers. Thus, soft failure detection can be conceptualized as anomaly detection. A common method in anomaly detection is to predict the expected value of a parameter of interest and compare it to the monitored value. If the difference between the prediction and the monitored value exceeds a certain threshold, it is defined as a failure.

As shown in Fig. 1, failure detection can be broadly divided into two categories: detecting failures based on the prediction of certain input parameters, such as ORP or BER; and detecting failures based on classification algorithms. Due to the large amount of time-series data generated by continuous monitoring, ML-based algorithms are well-suited for predicting or classifying failures. To compare the benefits of using ML-based algorithms for ORP prediction, we also implement a statistical-based algorithm, namely, Autoregressive Integrated Moving Average with Exogenous Variables (ARIMAX). Additionally, the OCSVM is implemented as a classification-based approach.

In our work, we first compare the prediction-based ML algorithms (ANN, LSTM, GRU) among themselves and the statistical-based ARIMAX algorithm. We then compare the best-performing prediction-based ML algorithm with the classification-based OCSVM in terms of accuracy of prediction/classification, misclassification rate, and time taken for predicting or classifying a failure.

The remainder of the paper is organized as follows: Chapter II identifies related work in this field. Chapter III deals with the measurement setup and data collection for the CUTs. Chapter IV explains the various steps taken to implement the algorithms for failure detection. Chapter V discusses the results of the evaluation and finally, Chapter VI summarizes and concludes our findings.

II. State-of-the-Art

In recent years, OSaaS margins and Quality of Transmission (QoT) have been experimentally characterized using several manual network testing techniques [6]. The results of such studies enable network operators to operate the transponders in a low-margin regime. However, the amount of data that can be collected is limited to a few devices, and finding patterns manually becomes increasingly difficult.

In optical network failure detection, several methods have been analyzed, including the use of SVMs for detecting and identifying filter shift and tight filtering [7]. The residual-based SVM was found to perform best in terms of accuracy and robustness. Additionally, a new method of analyzing constellation diagram images using a Convolutional Neural Network (CNN) and the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm was presented [8]. However, these methods are reported to be computationally intensive, causing failure detection delays.

A method for detecting, localizing, and identifying potential faults using Software Defined Network (SDN) integrated knowledge was also proposed [9]. The failure detection uses optical power level abnormalities and the localization works by network topology mapping. However, questions remain regarding whether proposed solutions also work on real field data, and which ML techniques are best suited for this purpose. Furthermore, there are currently no studies evaluating failure detection for OSaaS users, where the only knowledge of the network is available from OPM data at the end-user Flex-BVTs.

III. Measurement Setup and Data Collection

To collect data for training and evaluation of the ML models, a commercial pan-European live network owned and operated by Tele2 Estonia spanning 2869 km is used. Besides the live channels carrying production traffic, five ADVA TeraFlex™ Flex-BVTs [10] were installed to generate data for testing. To create training data for the underlying study, five test channels were inserted into a dedicated add/drop port of the ROADM in the test site using an 8-port splitter/combiner module. Then, a 400-GHz wide OSaaS with a central frequency of 193.95 THz within the OLS C-band was configured in all traversed ROADMs. A physical loopback was

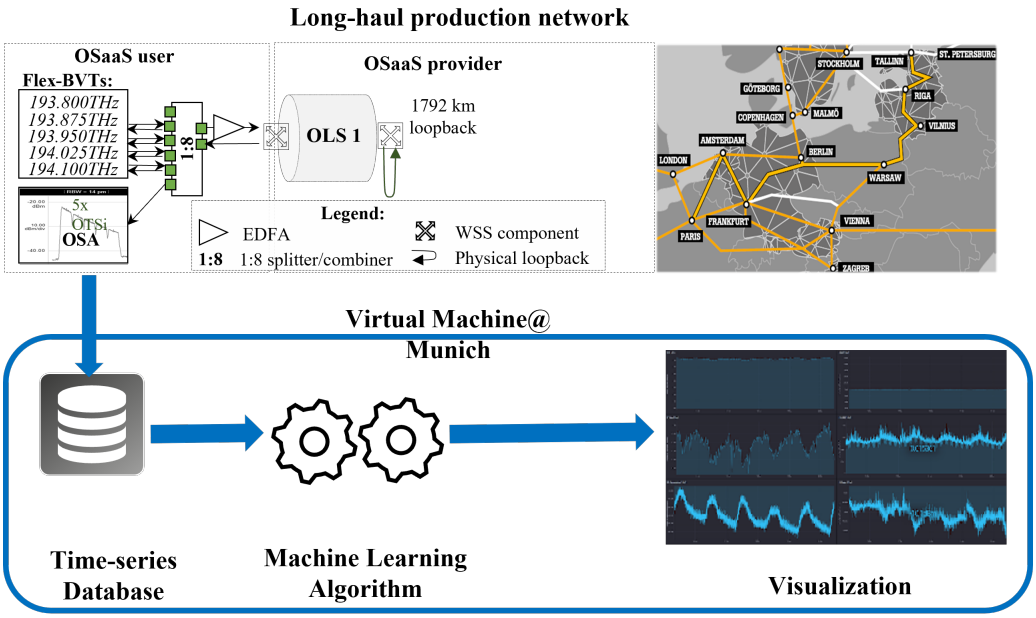


Fig. 2: Overview of the production network and the automated data collection framework.

featured in the far end ROADM location, to enable single-ended measurements from the test site and create a link length of 1792 km. Flex-BVT configurations 200 Gbps DP-QPSK 69 GBd and 200 Gbps DP-16QAM 34 GBd were used in data collection, maintaining the nominal power spectral density of the network, when switched between configurations.

A. Collecting OPM Data

Fig. 2 depicts an overview of the data collection framework, using a Python-based script to extract monitoring telemetry via NETCONF requests, polling the devices every 30 seconds, and then storing the returned information in a time-series database. From each CUT, a total of 7 OPM parameters were extracted, namely, Carrier Frequency Offset (CFO), Chromatic Dispersion Compensation (CDC), Differential Group Delay (DGD), Pre-Forward Error Correction Bit Error Rate (pre-FEC BER), Optical Received Power (ORP), Optical Signal to Noise Ratio (OSNR), Q-factor, and the electrical Signal to Noise Ratio (SNR).

B. Generating Failure Data

Two types of soft failures are artifically introduced on the CUTs in our analysis, namely, power degradation and inter-channel interference.

To simulate gradual power degradation, the transmit power of each CUT was reduced by 0.2, 0.5, and 1 dBm every minute until the signal cannot be decoded. While the ORP of each CUT was gradually decreasing, the pre-FEC BER also increased. Henceforth, the power degradation failure is referred to as **Failure 1**.

Inter-channel interference occurs when two neighbouring channels are too close to each other. This failure was artifically introduced by shifting the tuned frequency of the CUT first to the left and then to the right by 6.25 GHz frequency steps, reducing the channel spacing to the left and the right neighbouring channel, respectively. It can be seen that the BER of the overlapping CuTs is increasing. This failure is henceforth referred to as **Failure 2**.

IV. Methodology

As discussed earlier, soft failures occur rarely and it is difficult to build a large enough failure data set to train a supervised ML model on data including failures. Hence, the failure detection part of this work is considered as an anomaly detection exercise using supervised and semi-supervised learning methods based on data without failures.

A. Data Analysis

The OPM data used for training and hyperparameter optimization covers the period from 18th November 2020 until 2nd January 2021 and thus spans 45 days. Further, the OPM data within this time period was stable with minimal human interventions and hard failures. An additional 4.5 days of data was collected and used as the testing data for the trained models.

For analyzing the OPM data from all CUTs, we first used correlation methods to measure the relationship between each OPM parameter and provide non-correlated features as input to the ML models. We use the Pearson Correlation, Spearman Index, and Maximum Information Co-efficient to remove correlated parameters. From Pearson correlation coefficients, it is observed that the strongest positive linear correlations are between the Q-factor, SNR, and OSNR. These three variables behave very similarly in terms of linearity. The ORP follows similar patterns as the SNR but is weakly correlated.

Looking at negative correlations, it is noticeable that pre-FEC BER is strongly negative linear dependent on the four previously mentioned variables. From the Spearman ranks, we infer the existence of a non-linear correlation between ORP and pre-FEC BER. Finally, the results of Maximum Information Correlation suggest a weak correlation of CDC with the strongly correlated parameters of SNR, OSNR, pre-FEC BER, and ORP.

Although the seasonality and effect of ambient temperature on various OPM parameters were also studied, they are not included in the scope of this work since the equipment is always operating in temperature-controlled spaces.

B. Feature Scaling and Selection

As all parameters monitored have different ranges, the data is normalized to ensure unitless input for the ML algorithms. This is necessary since most algorithms expect scaled input data. For failure detection, the OPM data is transformed such that each feature individually ranges from zero to one instead of from its minimum to maximum. Every entry x_i of each feature vector $\mathbf{x}$ is converted to [11]:

$$x_{i,MinMaxScaled} = \frac{x_i - min(\mathbf{x})}{max(\mathbf{x}) - min(\mathbf{x})} \tag{1}$$

Choosing the right features as input plays an important role in terms of accuracy. Fewer parameters not only lead to decreased computational time and complexity but also better interpretability. Of all 7 extracted OPM parameters (ref. Sec. III-A), only pre-FEC BER and ORP are selected as input features. This is justified based on a combination of data analysis and existing literature [4]. Other parameters such as pre-FEC BER, OSNR, Q-factor, and SNR provide information to augment ORP. However, since the correlation analysis shows that they all carry similar information, we select pre-FEC BER which is reliable and up-to-date since the number of error bits is always counted.

C. Hyperparamater Optimization

As shown in Fig. 1, failure detection can be formulated either as an ORP prediction problem or as a classification problem. For ORP prediction, ARIMAX algorithm is used as a statistical baseline to compare ML algorithms. For ARIMAX, hyperparameters p, q, and d have to be chosen [12]. A grid search is performed to iteratively find optimal values of the hyperparameters. The optimal set of parameters is defined by the set with the lowest Akaike Information Criterion (AIC) value [13]. It gives a relative measure of whether the model is a good fit for the data, also taking into account the model's complexity. The results of the grid search indicate that a combination of low values of p and d as well as a high value for q result in the lowest AIC value. Table I shows the values chosen for ARIMAX, given the collected data.

Since training an ANN takes longer compared to other algorithms and large hyperparameter space, a combination of knowledge about the general behavior of ANNs and randomized search is used to find the optimal hyperparameters.

Model	Hyperparameter	Value
ARIMAX	p	0
	d	1
	q	4
ANN, LSTM, GRU	Hidden Layers	3
	Activation function (ANN only)	ReLU
	Optimizer	Adam
	Batch size	50
	Epochs	100
	Test split	0.25
	Validation split	0.2
OCSVM	Kernel	4th degree polynomial
	ν	0.0005
	k	5

TABLE I: Hyperparameters for ARIMAX, ANN, and OCSVM models.

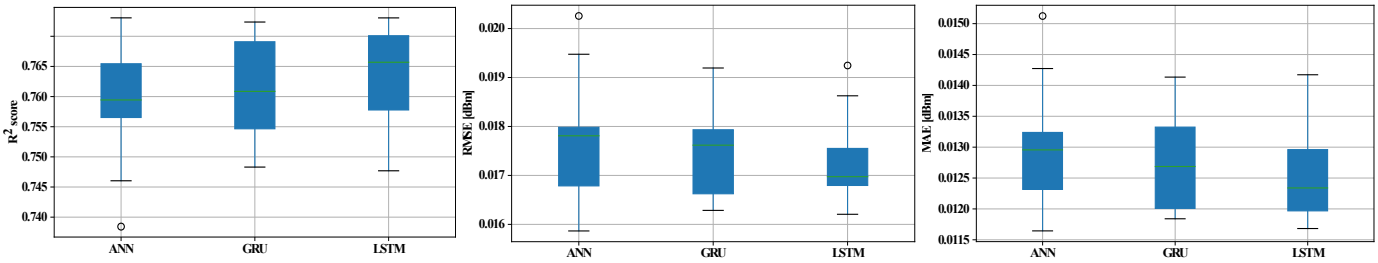
Randomized search evaluates a given number of random combinations by selecting a random value for every hyperparameter at each iteration. Therefore, this search method can find better models within a limited evaluation time by effectively searching a larger configuration space [14]. The architecture of the ANN needs to be pre-defined and cannot be set by a randomized search. Deeper networks containing more hidden layers have a much higher parameter efficiency, resulting in fewer neurons and thus less computation time for training. Therefore, we choose three hidden layers in addition to a flattening and a dropout layer. The flattening layer undertakes dimensional reduction of the time series and the dropout layer avoids model over-fitting. After a 33-run randomized search, an R^2 score of 0.9955 is achieved. It is assumed that the same hyperparameters (excluding the activation function) are ideal for the LSTM and the GRU models. The train-test split for the data is selected as 80%-20% and data are shuffled before splitting to ensure homogeneity.

To find the best values for the hyperparameters of the OCSVM, a grid search with the validation split parameter k (i.e. number of datasets of equal size), the ratio of outliers ν, and kernel function is performed. Validation split is used to assess how a model can predict or classify new data of an independent data set not used for estimating it, in order to reduce over-fitting and bias during the training. From the results it becomes clear, that kernel functions with the best performance are polynomial functions with a degree of two and four. They both perform with 100% accuracy (no misclassified data points) for small ν. Therefore, the hyperparameters for OCSVM are selected as: kernel = polynomial with degree = 4, ν = 0.05%, and k = 5.

Once the ML models were trained, we collected data from the production network for an additional 4.5 days. This "unseen" data, containing 13000 datapoints were injected with 6 artifically introduced failures (four of type Failure 1 and two of type Failure 2). Each trained ML model is run independently on this test data while predicting the ORP at each time step.

D. Dynamic Threshold Calculation

The ANN model is further augmented by a dynamic threshold calculation to reduce the prediction error [12]. From the prediction error of historical values, a single-dimensional



(a) R^2 values (b) RMSE values (c) MAE values

Fig. 3: Analysis of different ML algorithms for ORP prediction method.

vector $\mathbf{e}$ is created and each value is further smoothed using an exponentially-weighted moving average in order to obtain $\mathbf{e}_s$. A set of candidate thresholds is computed as $\epsilon = \mu(\mathbf{e}_s) + \mathbf{z}\sigma(\mathbf{e}_s)$, where $\mathbf{z}$ is an ordered vector of the k highest deviations of $\mathbf{e}_s$ above the mean $\mu(\mathbf{e}_s)$. k is a tunable hyperparameter between 2-10. Therefore, the actual threshold can be selected as shown:

$$t = \frac{\frac{\Delta\mu(\mathbf{e}_s)}{\mu(\mathbf{e}_s)} + \frac{\Delta\sigma(\mathbf{e}_s)}{\sigma(\mathbf{e}_s)}^2}{|\mathbf{e_a}| + |\mathbf{E}_{seq}|} \tag{2}$$

$$\text{where} \ : \Delta\mu(\mathbf{e}_s) = \mu(\mathbf{e}_s) - \mu(\{e_s \in \mathbf{e}_s | e_s < \epsilon\}), \tag{3}$$

$$\Delta\sigma(\mathbf{e}_s) = \sigma(\mathbf{e}_s) - \sigma(\{e_s \in \mathbf{e}_s | e_s < \epsilon\}), \tag{4}$$

$$\mathbf{e}_a = \{e_s \in \mathbf{e}_s | e_s < \epsilon\} \tag{5}$$

$$\mathbf{E}_{seq} = \text{continous sequences of } e_a \in \mathbf{e}_a \tag{6}$$

The error threshold for ORP prediction with ANN is calculated one timestep ahead, thereby allowing reducing the overall prediction errors and improving the accuracy.

V. Results

Once the hyperparameters of the ML models have been chosen, we compare the ML algorithms in two methods. First, we compare the R^2, Root Mean Square Error (RMSE), and Mean Absolute Error (MAE) values of the ORP prediction-based ML models. Further, the three models are compared with ARIMAX in terms of computation time. Finally, the best-performing prediction-based ML method is compared with the classification-based OCSVM method in order to detect artifically introduced failures in the network.

To evaluate the accuracy of the ORP prediction approaches, R^2 score, RMSE and MAE are computed with 10-fold Cross Validation (CV). The results shown in Fig. 3 are based on the mean of ten different sets of unseen test data. x As seen in Fig. 3a, the R^2 score is highest for the LSTM model, with a 0.005 increase in the median as compared to ANN and GRU. From Fig. 3b it becomes clear, that again LSTM performs better compared to the other two algorithms. Fig. 3c shows, that the LSTM model has the lowest MAE whereas The GRU outperforms the fully connected ANN by about 0.005 dBm. In summary, LSTM outperforms the other two neural networks slightly for all three accuracy metrics. The GRU and the fully connected ANN are very similar in the context of R^2 score and RMSE, but GRU has a slightly lower MAE.

Due to the fact that the ARIMAX model has very a high computation time for large datasets, which is shown in the next section, it can not be applied directly to the entire unseen data. Nevertheless, a piecewise comparison using five different datasets each of 1000 points, has revealed 17%, 15%, and 14% higher RMSE values than ANN, LSTM, and GRU, respectively. Furthermore, the MAE of ARIMAX is 16%, 14%, and 13% higher compared to ANN, LSTM, and GRU, respectively.

Fig. 4 shows, how the calculation time for each model increases with increasing amounts of points to predict. While ARIMAX quickly escalates its calculation time in order of minutes (right y-axis), GRU and LSTM stay in the range of 1-12 seconds for 5000 data points (left y-axis).

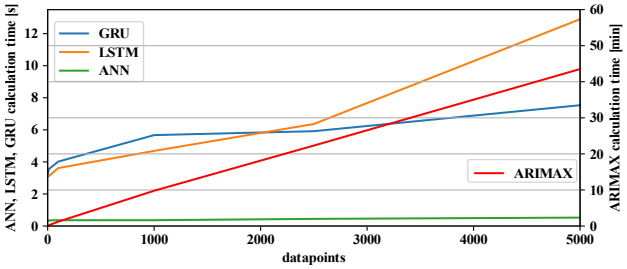


Fig. 4: Computation time for ORP prediction algorithms.

In terms of failure detection, the ANN model in combination with the dynamic threshold is compared to OCSVM. To calculate the detection accuracies, six artifically introduced CUT failures (ref. Sec. III-B) as well as 13000 unseen faultless data points are used. Fig. 5 presents the detection rates for each failure. For failure 1, only three out of six errors are detected by OCSVM. This is because a correctly classified failure is defined as a binary operation between *failure* and *no failure*. Further, for three measurements of Failure 1, the ORP and pre-FEC BER values before and after the failures occurred are misclassified as failures by OCSVM and hence no failure is detected, when an actual failure occurs. All other failures are correctly detected by both algorithms. For ANN with a dynamic threshold, 2 out of 13000 (¡ 0.01%) datapoints are misclassified as failures indicated by an ORP prediction error lying above the dynamic threshold. OCSVM, misclassified 112 out of 13000 datapoints (approx. 0.6%), thereby leading to a higher number of false positives. The low performances of

OCSVM can be explained by the fact that the model is adapted to the training data. Therefore, the OCSVM algorithm is not capable of adapting to new ORP ranges and trends, which is the case for the real field ORP data.

Fig. 5 shows the total duration of each failure in comparison with the time until it is detected by each algorithm. For all failures that are detected by OCSVM, the detection happens earlier or at the same time as the detection with ANN. The mean of the difference between failure start to detection divided by the total failure duration is 24.29% for OCSVM and 47.54% for the ANN with a dynamic threshold approach. Hence, the ANN approach performs sufficiently well as well, detecting the failure almost always within the first half of the total failure duration. The computation time is 0.0057 seconds for OCSVM and 0.7262 seconds for the detection with ANN including the dynamic threshold calculation.

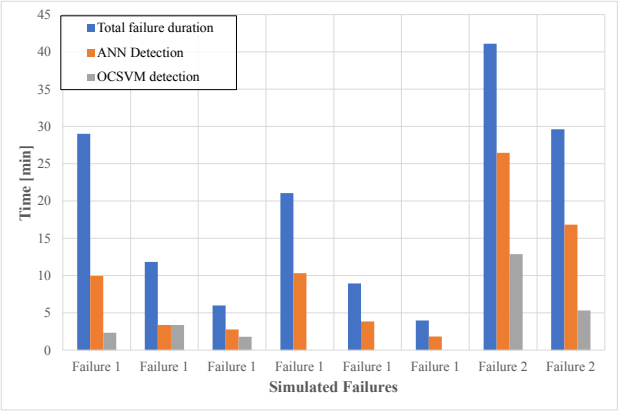


Fig. 5: Computation for times for various failure detections.

VI. Conclusions

In this work, several EON fault detection and identification algorithms were compared and evaluated based only on the OPM data available at the end user's transceiver. Overall, five algorithms were evaluated on OPM data from a 1792 km loopback link from a live production network, running for 45 days. Furthermore, two different types of failures were recreated to evaluate the models. For the failure detection, ORP values were predicted with three different neural network architectures as well as with the statistical approach ARIMAX. Based on the resulting prediction error, a dynamic threshold was used to augment the ANN's decision.

Furthermore, as a second approach for failure detection, an OCSVM was built based on faultless data. The evaluation showed, that although LSTM has the highest accuracy amongst the ORP prediction algorithm, the fully connected ANN has the shortest calculation time. Comparing the ANN model augmented by a dynamic threshold calculation with OCSVM, failures are detected earlier with OCSVM, but with a higher accuracy using the ANN approach. While the ANN approach has a 100% detection rate and 99.989% accuracy rate for failures and faultless data, respectively, the OCSVM achieved a 62.5% failure detection rate and a 99.45% accuracy rate.

Nevertheless, as external conditions might change over time, it is recommended to retrain the prediction ANN as soon as the standard deviation of one day exceeds 1 dBm. To overcome problems with false failure detection, a failure should be recognized only if two consecutive failures have been detected by the algorithm. This would however delay the time of detection by 30 seconds.

Our work shows that certain degradations in the network can be detected quickly without having complete knowledge of the underlying network. This can help in quick service restoration and root cause analyses. Machine learning algorithms can benefit from the augmentation of non-ML-based algorithms to improve the accuracy of prediction. Future work includes failure identification for OSaaS users using additional input parameters.

References

[1] S. Hardy, "*SNCF Reseau subsidiary Terralpha deploys Adva TeraFlex*." https://t.ly/tC_9, [Online] Accessed: April 2022.

[2] J. M. Simmons and G. N. Rouskas, "Routing and Wavelength (Spectrum) Assignment," in *Springer Handbook of Optical Networks*, pp. 447–484, Springer, 2020.

[3] H. Lun, M. Fu, X. Liu, Y. Wu, L. Yi, W. Hu, and Q. Zhuge, "Soft failure identification for long-haul optical communication systems based on one-dimensional convolutional neural network," *IEEE/OSA Journal of Lightwave Technology*, vol. 38, no. 11, p. 2992–2999, 2020. LFL+20.

[4] S. Shahkarami, F. Musumeci, F. Cugini, and M. Tornatore, "Machinelearning- based soft-failure detection and identification in optical networks," in *2018 Optical Fiber Communications Conference and Exposition (OFC)*, 2018.

[5] K. Grobe, C. Furst, A. Autenrieth, and T. Szyrkowiec, "Flexible spectrum-as-a-service," in *TERENA Networking Conference (TNC)*, 2014-05.

[6] K. Kaeval, T. Fehenberger, J. Zou, S. L. Jansen, K. Grobe, H. Griesser, J.-P. Elbers, M. Tikas, and G. Jervan, "Qot assessment of the optical spectrum as a service in disaggregated network scenarios," *J. Opt. Commun. Netw.*, vol. 13, pp. E1–E12, Oct 2021.

[7] Z. Wang, M. Zhang, D. Wang, C. Song, M. Liu, J. Li, L. Lou, and Z. Liu, "Failure prediction using machine learning and time series in optical network," *Optics Express*, vol. 25, no. 16, p. 18553–18565, 2017-08.

[8] C. Natalino, A. Udalcovs, L. Wosinska, O. Ozolins, and M. Furdek, "Spectrum anomaly detection for optical network monitoring using deep unsupervised learning," *IEEE Communications Letters*, pp. 1–1,, 2021. NUW+21.

[9] D. Rafique, T. Szyrkowiec, A. Autenrieth, and J. Elbers, "Analytics-driven fault discovery and diagnosis for cognitive root cause analysis," in *2018 Optical Fiber Communications Conference and Exposition (OFC)*, 2018. RSAE18.

[10] ADVA, "*TeraFlex*™." https://www.adva.com/en/products/open-optical-transport/fsp-3000-open-terminals/teraflex. [Last accessed: January 2023].

[11] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in python," *Journal of Machine Learning Research*, vol. 12, no. 85, p. 2825–2830, 2011.

[12] R. Hyndman and G. Athanasopoulos, "Forecasting: principles and practice," *OTexts*, 2018.

[13] J. deLeeuw, "Introduction to Akaike (1973) information theory and an extension of the maximum likelihood principle," in *Breakthroughs in statistics*, pp. 599–609, Springer, 1992.

[14] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," *The Journal of Machine Learning Research*, vol. 13, no. 281-305, p. 03, 2012.