Domain Adaptation and Transfer Learning for Failure Detection and Failure-Cause Identification in Optical Networks Across Different Lightpaths

FRANCESCO MUSUMECI[1,*], VIRAJIT G. VENKATA[1], YUSUKE HIROTA[2], YOSHINARI AWAJI[2], SUGANG XU[2], MASAKI SHIRAIWA[2], BISWANATH MUKHERJEE[3], AND MASSIMO TORNATORE[1]

[1] *Politecnico di Milano, Italy, Piazza Leonardo da Vinci, 32, 20133, Milano, Italy*
[2] *National Institute of Information and Communications Technology, 4-2-1, Nukui-Kitamachi, Koganei, Tokyo 184-8795, Japan.*
[3] *University of California, Davis, 3037 Kemper Hall, CA 95616, USA.*
[*] *Corresponding author: francesco.musumeci@polimi.it*

Optical Network Failure Management (ONFM) is a promising application of Machine Learning (ML) to optical networking. Typical ML-based ONFM approaches exploit historical monitored data, retrieved in a specific domain (e.g., a link or a network), to train supervised ML models and learn failures characteristics (a signature) that will be helpful upon future failures occurrence in that domain. Unfortunately, in operational networks, data availability often constitutes a practical limitation to the deployment of ML-based ONFM solutions, due to scarce availability of labeled data comprehensively modeling all possible failure types. One could purposely inject failures to collect training data, but this is time consuming and not desirable by operators. A possible solution is Transfer Learning (TL), i.e., training ML models on a Source Domain (SD), e.g., a lab testbed, and then deploy trained models on a Target Domain (TD), e.g., an operator network, possibly fine-tuning the learned models by re-training with few TD data. Moreover, in those cases when TL re-training is not successful (e.g., due to intrinsic difference of SD and TD), another solution is domain adaptation, which consists of combining unlabeled SD and TD data before model training. We investigate domain adaptation and TL for failure detection and failure-cause identification across different lightpaths leveraging real OSNR data. We find that, for the considered scenarios, up to 20 percentage points of accuracy increase can be obtained with domain adaptation for failure detection, while for failure-cause identification only combining domain adaptation with model re-training provides significant benefit, reaching 4-5 percentage points of accuracy increase in the considered cases.

1. INTRODUCTION

Machine Learning (ML) is currently under investigation as a promising enabler for automated design, control and management in Optical Networks (ONs). It has been shown [1] that ML can be used to extract useful information from the high volume of data retrievable in ON monitors. In particular, ML applied for automated Optical Network Failure Management (ONFM) is of paramount importance, as it goes beyond limitations (e.g., in terms of cost and required time) of traditional troubleshooting procedures, which are heavily based on manual observation of network alarms by human domain experts.

As discussed in [2], several studies have demonstrated the potential of ML and data analytics in leveraging field data to automate ONFM and effectively perform Quality of Transmission (QoT) monitoring [3], failure detection [4, 5], failure-cause identification [6, 7], failure localization [8–11] and failure prediction [12]. Most ML-based ONFM approaches rely on supervised learning techniques and on monitoring of signal-quality data, e.g., Optical Signal-to-Noise Ratio (OSNR) and/or Bit Error Rate (BER), made available by modern coherent receivers or by Optical Spectrum Analysers (OSAs). The idea behind such approaches is that ML models learn the "signature" of past failures observing historical data. For such historical data, a correspondence between system characteristics (i.e., the features) and an output label, indicating known failure scenarios (e.g., the presence/absence of a failure, or the failure cause) is available. Then, this signature is recognized in future observations of similar failures. In general, ML has been demonstrated to accurately

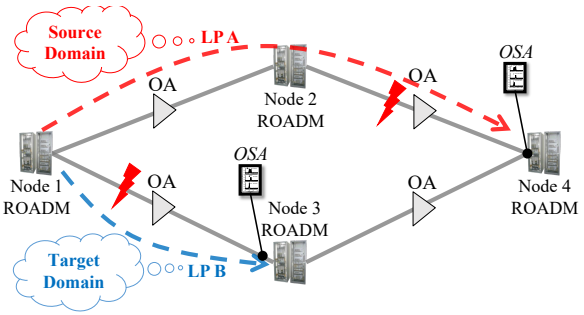

Fig. 1. Example of TL for lightpath failure detection.

recognize such failure signatures under specific assumptions (e.g., a single lightpath, a certain number of spans in an optical transmission system, etc.) [4, 6]. However, a main limitation to practical ML deployments for ONFM comes from the fact that, in real network scenarios, availability of historical failure data can be scarce for several reasons (e.g., lack of monitoring equipment at every network node, high cost of acquisition of large datasets, high resilience of optical transmission systems, which makes failures occurrence rare phenomena, etc.).

Let us consider the example in Fig. 1, which shows a simple ON constituted by four Reconfigurable Optical Add/Drop Multiplexers (ROADMs) interconnected through four optical fiber links, equipped with inline Optical Amplifiers (OAs). Two lightpaths (LPs) are routed in the network, namely, *LP A* through nodes 1-2-4, and *LP B* between nodes 1 and 3.

In this example, we assume that the OSNR at the receivers of LPs A and B can be degraded due to, e.g., an unexpected OA gain decrease, and therefore a failure may occur for any of the two lightpaths. Our objective is to train a ML model for failure detection on *LP B*, i.e., we aim at understanding, e.g., by observing lightpath's OSNR at the receiver (e.g., through an Optical Spectrum Analyzer (OSA)), if an anomaly is affecting *LP B*, which would lead to a failure. In case no historical data is available for failures on *LP B*, a supervised ML model cannot be trained to achieve our goal. So, one alternative is to train the ML model on a different domain (e.g., another lightpath, as *LP A* in Fig. 1, where failures are purposely injected to collect data) and then to deploy the trained model and apply it to the domain of interest, i.e., *LP B*. However, often the Source Domain (SD), i.e., *LP A* in Fig. 1, can be significantly different from the Target Domain (TD), i.e., *LP B* in in Fig. 1, e.g., in terms of path length, types/number of ROADMs, Wavelength Selective Switches (WSSes) and OAs along the route, used wavelength, baud rate, modulation format, etc. Therefore, the failure signatures learned by training on SD are only partially useful when applied to TD.

As a consequence, further data from the TD must be collected to re-train the ML model originally trained on SD. This is a costly solution, as, if complete re-training of the ML model is necessary each time the model is applied to a different lightpath, a very large re-training overhead will be necessary on a real operational network. For these situations, Transfer Learning (TL) is considered as a possible solution. TL allows to exploit the existing ML models trained on the SD, and then, by collecting only a small amount of data from the TD, TL allows to fine-tune the original models, enabling faster and more efficient adaptation of ML models to various lightpaths.

However, in some cases, TL might not be sufficient to gain satisfactory performance, as the SD and TD might have very different characteristics. Another solution is domain adaptation, which consists of combining unlabeled SD and TD data before model training with SD data, in order to achieve a more effective utilization of the available data from SD and TD.

In view of this, TL [13, 14] and domain adaptation techniques [15] constitute two promising tools to reduce the amount of additional data from the TD to fine-tune the ONFM models that were originally trained only with SD data.

In this paper, we investigate the application of TL to failure detection and failure-cause identification in ONs, focusing on two distinct types of failures, i.e., 1) *extra attenuation* and 2) *excessive filtering*. We consider real OSNR data observed at a lightpath receiver as inputs for the ML-based failure detection and failure-cause identification, which are modeled as ML classification problems. This paper extends our previous work [16] by including the following additional contributions:

- we formally describe the adopted domain adaptation and TL methods and discuss how we applied them to our ONFM use cases;
- we consider two ONFM use cases: failure detection and failure-cause identification ([16] covered only failure detection);
- we describe in detail the proposed ML-based framework for failure detection and failure-cause identification exploiting domain adaptation and TL across different lightpaths; and
- we provide extensive numerical results for both use cases.

The rest of the paper is organized as follows. Section 2 discusses related work and clarifies the novelty of our work. Sec. 3 formally defines the ML-based failure detection and failure-cause identification problems. Sec. 4 discusses the domain-adaptation algorithm considered in this paper. The various steps constituting the ONFM framework are described in Sec. 5. We provide numerical evaluation in Sec. 6. Conclusion and possible future extensions are discussed in Sec. 7.

2. RELATED WORK

Several studies have adopted ML for ONFM (see, e.g., ref. [17] for a review on the topic) and many other ON use cases [1, 18, 19]. In particular, the interest on TL has increased in the past few years for applications at both physical layer and control/management of an ON, especially concerning QoT estimation and prediction. For example, authors in [20] exploit evolutionary TL to optimize the weights and the architecture of a neural network for QoT estimation in multi-domain Elastic Optical Networks (EONs). TL for QoT estimation has been also investigated in [21, 22], where authors use TL to transfer knowledge across different network topologies, and in [23], where the performance of TL-based domain adaptation and active learning approaches, working with limited-size datasets, are evaluated and compared with standard ML-based QoT estimators. Ref. [24] adopts TL to reduce uncertainty in the Generalized Signal-to-Noise Ratio (GSNR) computation across different lightpaths sharing the same types of equipment. A similar evaluation is performed in [25] to predict QoT in terms of Q-factor, considering systems with different modulation formats, transmission distances, and fiber types. TL for QoT estimation is also discussed in [26], where authors present a life-cycle management of an ML algorithm deployed in an optical network. Interestingly,

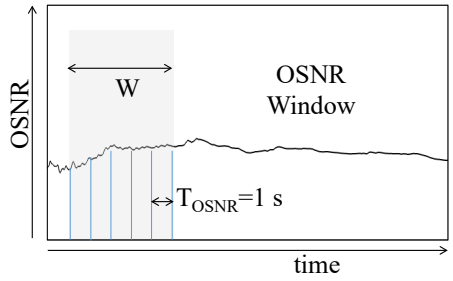

Fig. 2. OSNR window of duration W including OSNR observations with T_{OSNR} sampling period.

in [27] knowledge transfer is applied between ML algorithms of different natures, i.e., Recurrent and Feedforward Neural Networks, and TL is used to address nonlinear equalization in short-reach optical links.

Focusing on networking problems, authors of [28] adopt TL to perform Routing, Modulation and Spectrum Allocation in EONs to reduce learning time and improve network blocking performance. In [29], TL is used to predict spectrum defragmentation and perform spectrum optimization in multi-core EONs.

To the best of our knowledge, no previous study has considered TL in the context of ONFM, in particular concerning failure detection and failure-cause identification use cases. In this paper, using real data coming from fault injection on a lab testbed, we investigate for the first time failure detection and failure-cause identification with domain adaptation and TL across different lightpaths, considering as source and target domains lightpaths with different number of traversed ROADMs (i.e., number of hops) and optical fiber links.

3. ML-BASED FAILURE DETECTION AND FAILURE-CAUSE IDENTIFICATION PROBLEM STATEMENT

We model failure detection and failure-cause identification as ML-based binary classification problems. In both problems, we concentrate on a given lightpath for which we are given the OSNR at the receiver, monitored with sampling period of T_{OSNR} seconds (e.g., $T_{OSNR} = 1$ s). Classification is performed considering OSNR windows of duration W seconds, each one including a sequence of $1 + W/T_{OSNR}$ consecutive OSNR observations (also including the OSNR samples at the start and end times of the window), as qualitatively shown in Fig. 2.

According to the problem at hand (i.e., either failure detection or failure-cause identification), given an OSNR window as input, we classify it in the following way:

- for the failure detection problem, we distinguish between classes *1) normal* and *2) failure*;
- for the failure-cause identification problem, we discriminate between *1) extra attenuation* and *2) excessive filtering* failure causes.

Note that, for the failure-cause identification problem, we assume as inputs (in both ML training and test phases) only OSNR windows corresponding to some failure, i.e., we neglect normal windows. Development of multi-class classifiers, able to distinguish between normal (i.e., non-failed), extra attenuation, and excessive filtering at once, is an open problem for future work.

4. DOMAIN ADAPTATION METHOD

Pure application of TL, i.e., training ML models on a SD and testing/deploying the learned model on the TD, possibly with partial re-training with limited amounts of TD data, may be insufficient in some cases, and lead to ML models' performance deterioration. This is mainly due to the intrinsic differences between SD and TD data distributions, i.e., in our case, the different lightpaths constituting the two domains can have different path lengths, types/number of ROADMs, WSSes and OAs along the route, used wavelength, baud rate, modulation format, etc. Therefore, in this paper we also adopt a domain-adaptation algorithm, called CORrelation ALignment (CORAL), originally proposed in [30]. CORAL is an unsupervised domain-adaptation approach which aims at aligning the features distribution of data points in SD and TD. Although several other domain-adaptation approaches can be used [13], we adopt CORAL, because, besides SD data, it requires only unlabeled TD data, which is a more realistic scenario in the context of ONFM.

In short, CORAL calculates the covariance of TD data and then uses it to transform the statistical distribution of the features in SD data such that they are distributed as in the TD. Finally, the transformed SD data are used to train the ML model in the TD in a supervised manner. More formally, let X_S and X_T be the matrices including the values of the features for data points in SD and TD, respectively; the basic idea of CORAL is to transform SD features distribution X_S and obtain a new distribution X_S^*, which is aligned to the TD features distribution by following the steps below:

1. compute SD and TD covariance matrices, i.e.,
$$C_S = X_S X_S^t \quad \textbf{(1)}$$
$$C_T = X_T X_T^t \quad \textbf{(2)}$$
where X_S^t (respectively, X_T^t) is the transpose of matrix X_S (resp., X_T);
2. decorrelate SD features distribution using SD covariance matrix (i.e., perform so-called features whitening) as:
$$\hat{X}_S = C_S^{-1/2} X_S \quad \textbf{(3)}$$
3. align the decorrelated (whitened) SD features distribution using TD covariance matrix (i.e., perform so-called features re-coloring) as:
$$X_S^* = C_T^{1/2} \hat{X}_S \quad \textbf{(4)}$$

After CORAL is applied, TL can be adopted by training ML models with SD data considering new features distribution X_S^* associated with the original labels in the SD, and then perform partial retraining with few TD data, where original unchanged features distribution X_T is considered and associated with TD labels.

This process guarantees that information on features distribution in the TD is leveraged during first model training performed with labeled SD data. This scenario is in line with realistic situations in which ONFM should be applied to TD where large amounts of unlabeled data is available (e.g., because OSNR is continuously monitored in a deployed TD lightpath), whereas only a limited amount of labeled TD data is available, e.g., because the operational TD lightpath has been rarely affected by failures. Therefore, the large amounts of unlabeled SD/TD data can be used before the initial ML model training, when CORAL is used to perform SD/TD domain adaptation; then, the available labeled TD data will be leveraged at a second step to fine-tune the ML model.

5. ONFM FRAMEWORK WITH DOMAIN ADAPTATION AND TL ACROSS DIFFERENT LIGHTPATHS

In this section, we describe the proposed ONFM framework used to perform failure detection and failure-cause identification with domain adaptation and TL across different lightpaths. The main building blocks of the ONFM framework are shown in Fig. 3 as a sequence of steps of our framework. In the figure we detail the main decisions taken at each step, and also show how SD and TD data are involved in the various steps. In the following subsections the various building blocks are described in more detail.

A. OSNR Data Collection and Normalization

As first step, OSNR traces are collected from lightpaths with different characteristics, in terms of number of hops, presence/absence of a failure, type of failure (i.e., either extra attenuation or excessive filtering), and failure location along the route. For each observed lightpath l, we assume a total of D OSNR samples are collected, and we identify the sequence of samples as $OSNR_l(t_k)$, where $k = 0, 1, ..., (D-1)$ indicates the k-th observation instant, i.e., $t_k = k \cdot T_{OSNR}$.

As we will discuss in Sec. 6, different SD/TD pairs are identified in order to perform different analysis on the effectiveness of our TL method. As an example, in case of failure-cause identification, data in SD can be extra attenuation and excessive filtering OSNR data for lightpath l_1, and data in TD can be extra attenuation and excessive filtering OSNR data for lightpath l_2. Recall that, as discussed in Sec. 1, in realistic scenarios TD datasets are typically much smaller than SD datasets. Although we collected the same amount of OSNR data for all the lightpaths, we emulate the typical difference of size between SD and TD datasets by ignoring part of the TD data during numerical evaluation.

In this phase, raw OSNR data are normalized as follows:

$$OSNR_l(t_k) \leftarrow \frac{OSNR_l(t_k) - min_k[OSNR_l(t_k)]}{max_k[OSNR_l(t_k)] - min_k[OSNR_l(t_k)]} \quad \textbf{(5)}$$

Such OSNR normalization provides modified OSNR values ranging between 0 and 1, and allows building ML classifiers which are independent of specific OSNR absolute (i.e., raw) values collected at the receivers. In fact, for different lightpaths, even if they may have similar failure characteristics (e.g., two lightpaths in normal conditions, or two lightpaths both suffering from excessive filtering), OSNR values may fluctuate around very different mean values, according to system settings in the two lightpaths (e.g., spans length, number of hops, center wavelength, types/number/gain of the traversed optical amplifiers, etc.). Therefore, in such situations, simple observation of OSNR raw data would not be sufficient to perform failure detection and failure-cause identification independently from the specific lightpath.

B. OSNR Window Formation and Features Extraction and Scaling

For each lightpath l in both SD and TD datasets, the sequence of D OSNR samples is converted into a set of windows of duration W (which is varied in the different experiments carried out in this paper), including OSNR samples collected every T_{OSNR} seconds, as discussed in Sec. 3. Note that different windows may overlap, e.g., if window "a" contains samples from t_1 to t_{10}, the successive window "b" contains samples from t_2 to t_{11}. Once windows are formed, they are treated independently one from another at train/test phases, as we concentrate on the detection and failure-cause identification of failures observing a "snapshot"of the OSNR windows. In other words, we do not include any temporal relation between any two distinct windows in our ML classifiers, and it may even happen that two overlapping windows are included in separate training and test sets, respectively[1].

Each window is characterized by a certain label $\mathbf{y}$, which represents the lightpath failure status, i.e., *1) normal vs. 2) failure* for failure detection, or *1) extra attenuation vs. 2) excessive filtering* for failure-cause identification, respectively.

For both failure detection and failure-cause identification problems, we consider the same features vector $\mathbf{x}$ characterizing each OSNR window and constituted by the following 16 features:

- $x_1 = min$: minimum OSNR value in the window;
- $x_2 = max$: maximum OSNR value in the window;
- $x_3 = mean$: mean OSNR value in the window;
- $x_4 = std$: OSNR standard deviation in the window;
- $x_5 = p2p$: "peak-to-peak" OSNR, i.e., $p2p = max - min$;
- $x_6 = RMS$: OSNR root mean square in the window;
- $x_7 - x_{16}$: the ten strongest spectral components in the window, extracted by applying Fast Fourier Transform (FFT) on the OSNR window.

These features have been selected as they are similar to those used in our previous works [17] and [31], where BER windows have been considered to perform ML-based failure detection, failure-cause identification and failure-magnitude estimation. In these previous works, sensitivity analysis has been performed to manually select features and in particular to select the amount of relevant FFT components.

Note that, to avoid that specific features provide different impact on ML algorithm training due to their different scale, before training we perform feature normalization so as to obtain features ranging between -1 and 1. More formally, assuming training set S_{train} consists of N_{train} data points, for a given feature x_i we calculate its mean and range considering data points in S_{train} as follows:

$$mean[x_i] = \frac{1}{N_{train}} \sum_{j \in S_{train}} x_i^j, \quad \forall i = 1, 2, ..., 16 \quad \textbf{(6)}$$

$$range[x_i] = \max_{j \in S_{train}} x_i^j - \min_{j \in S_{train}} x_i^j, \quad \forall i = 1, 2, ..., 16 \quad \textbf{(7)}$$

where x_i^j indicates the value of the i-th ($i = 1, 2, ..., 16$) feature and for the j-th ($j \in S_{train}$) training sample. Then, we apply features normalization by modifying feature values for all data points (i.e., in sets S_{train}, S_{val} and S_{test} sets, i.e., training, validation and test sets, respectively) as follows:

$$x_i^j \leftarrow \frac{x_i^j - mean[x_i]}{range[x_i]}, \quad \forall i = 1, 2, ..., 16; \ \forall j \in S_{train} \cup S_{val} \cup S_{test} \quad \textbf{(8)}$$

Note that the values of $mean[x_i]$ and $range[x_i]$ are calculated considering training points only, and are used to perform feature normalization also for data points in validation and test sets.

[1] Note that temporal dependencies between consecutive windows can be leveraged, e.g., adopting time-series ANNs like recurrent neural networks, gated recurrent units or long-short-term-memory ANNs. Considering this aspect is left for future work.

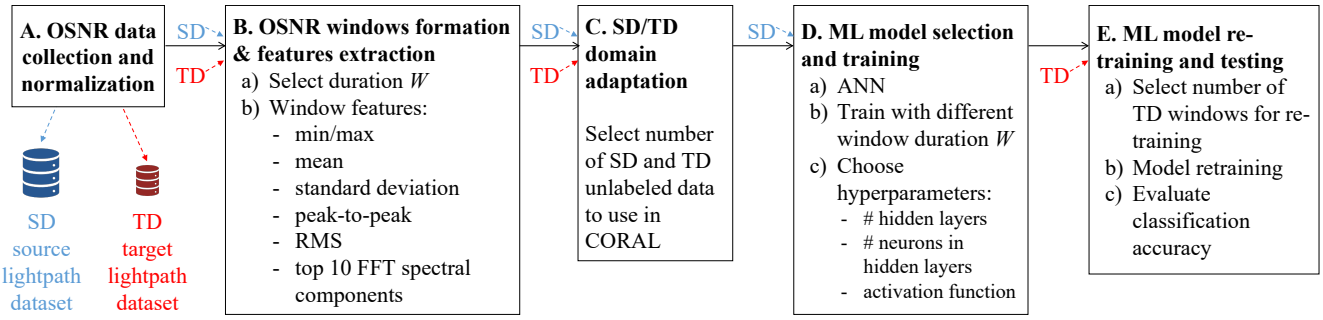

Fig. 3. Block diagram of the TL-based ONFM framework.

C. SD/TD Domain Adaptation

Domain adaptation is then applied to align features distribution of SD and TD data using CORAL algorithm as discussed in Sec. 4. In this phase, we consider several scenarios in which we vary the amount of SD and TD unlabeled data to perform domain adaptation, in order to assess the impact of this parameter on the accuracy provided after applying CORAL and TL. Note that, to apply domain adaptation, every time the TD lightpath is changed, a certain amount of unlabeled data is necessary from this domain, which can be collected in a commissioning phase before the TD lightpath is operational and actually transporting user traffic. After this phase, new training needs to be performed once the features of the SD lightpath have been re-colored. Although this solution may appear ineffective, we observe that the new training on SD data can be performed in the very early stages of TD lightpath lifetime, during which the available ML models can be still used to perform failure management on TD data.

D. ML Model Selection and Training

For both failure detection and failure-cause identification, we consider ML classifiers based on Artificial Neural Networks (ANN), as in [17].

For each value of W considered to generate our SD and TD datasets (see Sec. 5.B), we perform ML model selection by applying 5-fold cross-validation on SD data and testing different ANNs with different number of hidden layers, hidden neurons per layer and activation functions. Note that, in this phase, only SD data are used to design the initial ML models. More specifically, we consider a number of hidden layers between 1 and 7, 3 to 15 hidden neurons per layer and ReLu, Tanh and Sigmoid activation functions [17]. After testing the different combinations of hyperparameters, the final ANN structures are selected as the ones with highest accuracy and limited training duration (i.e., below 30 minutes, considering the used hardware/software).

In our experiments, the best performing ANNs obtained for failure detection have 2 hidden layers, each with 9 hidden neurons and ReLu activation function, whereas sigmoid activation function has been used in the output layer. The same ANN hyperparameters have been selected after cross-validation also for the failure-cause identification problem, with the only difference being that in this latter case we consider 10 hidden neurons per layer, instead of 9.

After model selection, all SD data points are then used to perform model training before applying TL.

E. ML Model Re-training and Testing

As final step, we select a limited amount of TD data to fine-tune the trained model. Our main objective is to assess how many additional data points from the TD are necessary to approximate the classification performance of analogous ML models trained only with the TD dataset. Therefore, we consider different amounts of TD data to perform model fine-tuning.

Then, considering an independent test set, constituted by TD data not used for ML model fine-tuning, we evaluate the performance of knowledge transfer from SD to TD in terms of classification accuracy.

Note that fine-tuning the ML model by leveraging a limited amount of TD data consists of performing additional steps of the ANN training to update its weights in order to fit the TD data distribution, starting from the knowledge (i.e., the ML model) obtained after the initial training using SD data.

6. NUMERICAL RESULTS

A. Testbed Setup

We perform TL-based failure detection and failure-cause identification using real data obtained on a testbed of the National Institute of Information and Communications Technology (NICT) located in Sendai (Japan). The testbed is shown in Fig. 4 and consists of 4 ROADMs, identified as Node 2, Node 3, Node 4 and Node 1, interconnected through optical fibers, and equipped with one pre-amplifier and one booster (OA in the figure) at their input and output, respectively. Each fiber link can emulate fiber spans of up to 80 km using a Variable Optical Attenuator (VOA) with maximum 20 dB attenuation. Two failure scenarios, i.e., 1) *extra attenuation* and 2) *excessive filtering*, are emulated by including in a fiber span either an attenuator with extra 11 dB attenuation (emulating the *extra attenuation* scenario) or a WSS with passing bandwidth of 12.5 GHz (emulating the *excessive filtering* scenario). In the latter case, due to the WSS insertion loss of 14 dB, the VOA attenuation is reduced to 6 dB to compensate the effect of the WSS on the last span overall attenuation.

Figure 4 also shows the routing of one of the lightpaths considered in our numerical analysis, routed through nodes 2, 3, 4 and 1, and where failure (either extra attenuation or excessive filtering) is introduced in the last optical span (i.e., between nodes 4 and 1).

The set of lightpaths considered in our analysis is shown in Tab. 1, together with the location of failures emulated in each case. To perform data collection, each lightpath is set-up in the testbed separately, i.e., no other lightpath is simultaneously set-up in the testbed. For each lightpath, the same 100 GHz bandwidth with central frequency 194.8 THz is used to transmit a 10

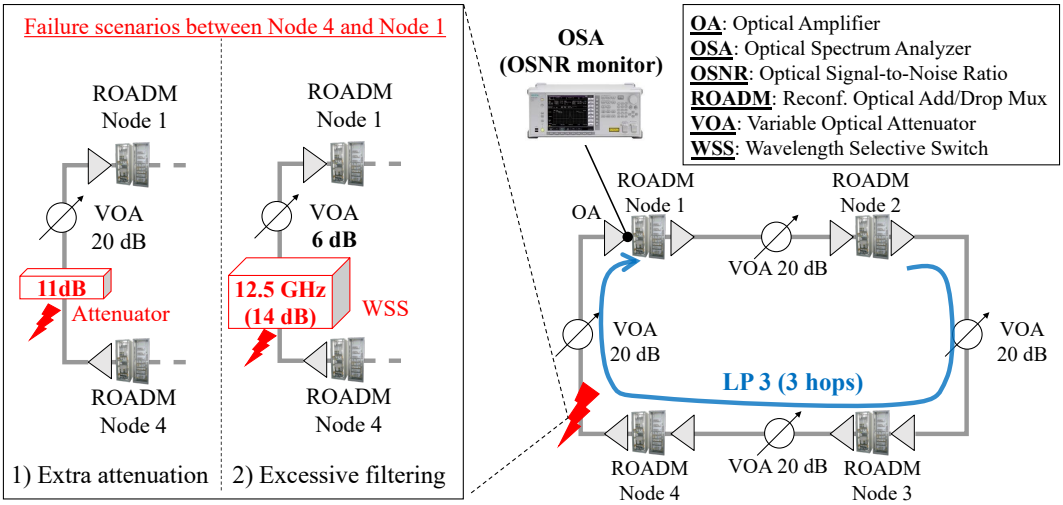

Fig. 4. Schematic of NICT's Sendai Testbed setup and example of the emulated failure scenarios.

Gbit/s signal using OOK modulation format. For each lightpath, we consider 3 cases (i.e., no failure, extra attenuation and excessive filtering) and, in each case, OSNR samples are collected for 6 hours at a sampling period of $T_{OSNR} = 1$ second, so the entire dataset consists of around 72 hours of OSNR monitoring at lightpaths receivers and therefore for each lightpath, around 20000 OSNR windows are available. To generate datasets used for the TL-based failure detection and failure-cause identification, we use distinct values of window duration, i.e., $W = 10$, $W = 30$ and $W = 50$ seconds.

B. Benchmark Scenarios

We numerically evaluate our TL-based failure detection and failure-cause identification framework using CORAL algorithm (simply referred to as *CORAL* in the following) and compare it to the following benchmark scenarios:

- *SD only*: this is a lower-bound scenario where we train the ML model using 5000 windows from SD data and then test on TD data without any re-training or SD/TD domain adaptation;
- *Pure TL*: in this case, we train the ML model with 5000 windows from SD data and then perform partial re-training with a small amount of labeled TD data, yet without applying CORAL for domain adaptation; similarly to *Pure TL* case, in the *CORAL* scenario we still train the ML model using 5000 windows from SD data, but before this step we perform domain adaptation using a certain amount of SD/TD unlabeled data;
- *TD only*: this scenario acts as an upper bound, as we perform the entire training of the ML model with 5000 windows taken from TD data.

In all the following numerical results, the test set consists of 5000 "independent" windows from TD data, (i.e., windows which were not used either for TL re-training or for CORAL-based domain adaptation), constituted by different lightpaths, according to the specific analysis to be performed.

Table 1. Set of lightpaths considered.

ID	No. of hops	Source node	Dest. node	Location of failure
LP1a	1	Node 4	Node 1	Link 4-1
LP1b	1	Node 2	Node 3	Link 2-3
LP2	2	Node 3	Node 1	Link 4-1
LP3	3	Node 2	Node 1	Link 4-1

C. Discussion: Failure Detection

We start our analysis by considering failure detection and concentrate on *extra attenuation* failure type, where 11dB of additional attenuation has been introduced in one of the optical links of the lightpaths, as explained in Sec. 6.A. Similar results (not reported here for sake of conciseness) have been obtained for the detection of *excessive filtering* failure type.

We provide numerical results in terms of classification accuracy (i.e., percentage of OSNR windows correctly classified as *normal* or *failure* for failure detection, and *extra attenuation* or *excessive filtering* for failure-cause identification) for the TD lightpath, which varies according to the specific scenario under analysis.

C.1. Impact of CORAL algorithm and TL re-training

Figure 5 shows TD failure detection accuracy for TL across lightpaths LP1a and LP3, obtained when performing CORAL-based domain adaptation and partial re-training using increasing amounts of TD labeled data (*CORAL* scenarios in the figure). We compare accuracy values with the benchmark scenarios described in Sec. 6.B, i.e., *SD only*, *Pure TL* and *TD only*. More specifically, in *SD only* and *TD only*, we do not perform domain adaptation with CORAL algorithm, and consider a fixed ML model trained only with the SD or TD data, respectively, while in *Pure TL*, the model obtained in *SD only* is the starting point for partial re-training with increasing amount of TD data.

On the other hand, we consider three *CORAL* cases where different amounts of unlabeled SD and TD data are used to perform domain adaptation before ML algorithm training. More specifically, from each of the two domains (SD and TD), 1000, 3000 or 5000 OSNR windows are taken, and these scenarios are referred to as "CORAL (1000)", "CORAL (3000)", and "CORAL (5000)" in the following, respectively. In the three cases, the

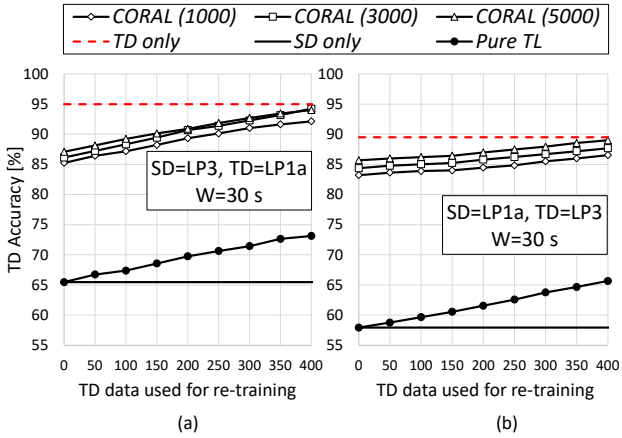

Fig. 5. Failure detection (*normal vs. extra attenuation*): TD accuracy for increasing amount of TD data used for re-training and different SD/TD scenarios (window duration $W = 30$ s): (a) SD=LP3, TD=LP1a; (b) SD=LP1a, TD=LP3.

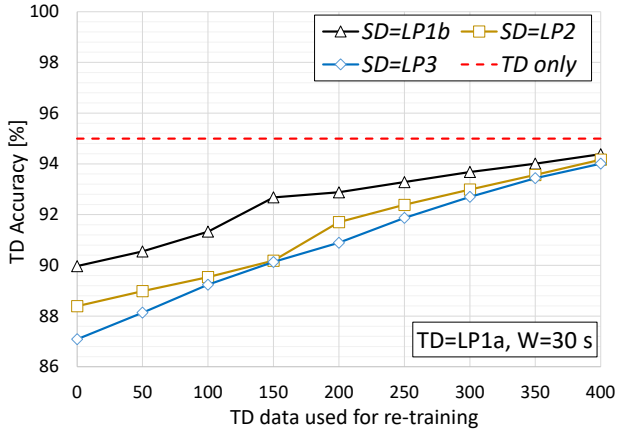

Fig. 6. Failure detection (*normal vs. extra attenuation*): LP1a TD accuracy for increasing amount of TD data used for re-training and different SD lightpaths (window duration $W = 30$ s).

same SD data points, together with their labels, are used to train the ML models which are then re-trained using few TD data points as shown the x-axis[2]. In this analysis, window duration is assumed as $W = 30$ seconds in all cases.

Comparing Figs. 5(a) and 5(b), we can observe that accuracy for LP1a is higher compared to LP3. This is due to the higher discrepancy in OSNR variations that can be observed comparing normal and failure cases for shorter lightpaths (such as LP1a). In both TL cases (i.e., LP3-to-LP1a and LP1a-to-LP3), partial re-training with TD data performed with *Pure TL* brings limited accuracy improvements compared to the poor performance of *SD only*, i.e., up to around 9 percentage points for TD=LP1a and 10 percentage points for TD=LP3, respectively, when 400 TD data points are used for re-training. A more significant amount of TD data would be necessary to perform re-training, resulting in time-consuming re-training process which reduces the benefits of the TL approach.

Noticeably, a high benefit compared to *SD only* is provided by domain adaptation with CORAL, even without re-training using TD data. As observed for the values $x = 0$ in the abscissa of Fig. 5, remarkable accuracy improvement of at least 20 and 16 percentage points (i.e., from 65% to 85% and from 57% to 83%, respectively) is obtained for TD=LP1a and TD=LP3, respectively. Then, according to the amount of SD data points used, re-training with TD data leads to an additional accuracy improvement (linearly increasing with the amount of used TD data) of around 7-9 percentage points for TD=LP1a, and 3-4 percentage points for TD=LP3, when 400 TD data points are used, approaching accuracy of 95% and 89.5% in the two *TD-only* scenarios.

For a given amount of TD data used for re-training (i.e., for a given value in the x-axes of Figs. 5(a)-(b)), accuracy variation is only slightly affected by an increase in the amount of SD/TD data used for domain adaptation (see the three *CORAL* cases in the figures), and is always below 2.1 and 2.6 percentage points in the two TL scenarios of Figs. 5(a)-(b), respectively.

This analysis strongly supports the use of domain-adaptation algorithms like CORAL when performing TL for failure detection, since, for the considered lightpaths and optical transmission characteristics, leveraging limited amounts of TD data to perform *Pure TL* might be not sufficient to a satisfactory increase in detection accuracy. For this reason, in the following, we do not consider further cases with *Pure TL*.

C.2. Impact of SD selection

Now we concentrate on the impact of the SD lightpaths when applying CORAL and TL to a fixed TD lightpath. To this end, we focus on the case with $W = 30$ seconds and TD=LP1a, and show in Fig. 6 how TD accuracy of LP1a changes when considering different SD, i.e., LP1b, LP2 and LP3 and for increasing TD data points used for re-training. As expected, regardless of the specific SD, detection accuracy increases with increasing amount of TD data used for re-training. Moreover, we observe that the case when SD=LP1b is the one with the highest accuracy, due to the fact that SD and TD have the most similar characteristics, i.e., both LP1a (=TD) and LP1b (=SD) are single-hop lightpaths. This occurs even if SD and TD lightpaths are entirely disjoint, i.e., despite the fact that, when training on SD data, the failure signature is learned on a link (i.e., Link 2-3), which is different from the failed link in the TD (i.e., Link 4-1).

D. Discussion: Failure-Cause Identification

In this section, we discuss failure-cause identification, i.e., we consider the performance of classifiers designed to distinguish between 11dB *extra attenuation* and 12.5 GHz *excessive filtering* failure types.

D.1. Impact of OSNR window duration W

As a first analysis, we evaluate the performance of the proposed failure-cause identification framework by comparing it with *SD-only* and *TD-only* baseline scenarios, varying the amount of SD/TD data used for domain adaptation, the number of TD data used for retraining, and for different values of window duration, i.e., $W = 10$, 30, and 50 seconds, as shown in Fig. 7. For this case, we consider a fixed SD/TD scenario, i.e., we assume SD=LP1a and TD=LP3, but similar considerations have been obtained for other SD/TD pairs. Comparing the different values of window duration W (i.e., comparing Figs. 7(a)-(b)-(c)), we first

[2]Note that we assume that unlabeled SD/TD data, used to perform domain adaptation with CORAL, are collected and exploited before ML model selection and training. Then, since a history of failures (i.e., labels from the TD) is not needed to perform CORAL, TD data used to perform domain adaptation has a different role compared to TD labeled data used for ML model fine-tuning, and therefore the amounts of unlabeled and labeled data exploited in the domain adaptation and TL phases are assumed as independent.

observe that, as expected, increasing W leads to an improvement in TD accuracy, as more information is included in windows with longer duration. Moreover, comparing *CORAL* cases with *SD only* and *TD only*, we observe that, independently from the value of W, the accuracy improvement provided by domain adaptation (i.e., CORAL algorithm) using 5000 SD data points is in the order of 2 percentage points, which is comparable with the improvement provided by re-training with a few hundreds of TD data.

This shows that, unlike the failure-detection scenario, where domain adaptation provides a much more significant improvement compared to re-training with TD data (see Sec. 6.C), in some cases, it is worth to combine domain adaptation and model re-training to have more significant increase in classification accuracy, mainly due the fact that, in such cases, the baseline accuracy provided by *SD only* is already satisfactory.

D.2. Impact of unlabeled data used for CORAL

We now evaluate in more detail how failure-cause identification accuracy is affected by the amount of SD/TD data points used for domain adaptation, shown in the x axes of graphs in Fig. 8. Here we consider fixed TD, i.e., LP1a, and different SDs, i.e., LP1b, LP2 and LP3, in subfigures (a), (b) and (c), respectively, and a fixed value of $W = 10$ seconds.

We observe that the most significant improvement compared to *SD only* is obtained for the case when SD=LP3 (see Fig. 8(c), where curves increase in a steeper manner for increasing amount of SD/TD data points used for domain adaptation. In this case, the accuracy in *SD only* has the lowest value among the different SDs, confirming that CORAL-based domain adaptation is more beneficial when failure-cause-identification knowledge acquired from the SD is poor. Observing the different curves laying between *SD only* and *TD only*, differing from the amount of TD data used for re-training, we see that a small accuracy, within 2 percentage points in most cases, is obtained when passing from 0 to 400 labeled TD data points used for re-training.

D.3. Impact of TD selection

We conclude our discussion analysing failure-cause-identification performance when transferring knowledge from a fixed SD lightpath (LP1a in our case) to different TD lightpaths, i.e., LP1b, LP2 and LP3, as shown in Fig. 9, considering different amounts of SD/TD data used in CORAL, increasing labeled TD data used for re-training, and assuming $W = 10$ seconds.

As usual, the highest improvement of CORAL over baseline scenario is obtained in the case of lower *SD-only* accuracy, i.e., when TD=LP1b (see Fig. 9(a)). Moreover, since starting accuracy of the *SD only* is already high for all TD cases (always above 91%), we again find that the global improvement, which allows approaching performance of *TD only*, is almost equally distributed between domain adaptation based on CORAL and model retraining with TD data. As an example, let us consider the case of Fig. 9(c), and let us observe the case when no re-training is performed (i.e., line "No Labeled TD data" in the figure), although similar considerations can be drawn for other scenarios: comparing this curve with *SD only* (straight solid line), we observe that, applying CORAL with 5000 SD/TD data, accuracy increases from 92.88% to 94.94% (i.e., around 2 percentage points); then, performing re-training with 400 labeled TD data points (see line "400 Labeled TD data" in the figure) provides an additional increase of ∼2.5 percentage points, which is comparable to the improvement provided by domain adaptation with CORAL.

7. CONCLUSION AND FUTURE WORK

We provided ML-based algorithms performing domain adaptation and transfer learning across different lightpaths for failure detection and failure-cause identification in optical networks. The proposed framework uses unlabeled data from source and target domains/lightpaths and implements CORAL algorithm for domain adaptation, and it also leverages small amounts of labeled TD data to perform model fine-tuning (i.e., knowledge transfer).

Considering real OSNR traces emulating normal and failed lightpath states (due to excess filtering or attenuation) in a 10 Gbit/s testbed, we found that significant improvements are obtained in both failure detection and failure-cause identification by combining the effect of 1) domain adaptation (based on CORAL algorithm) and 2) model re-training. This is true, in particular, when accuracy provided with simpler knowledge-transfer approaches, i.e., train on a SD and directly apply the model on a different TD (*SD only* in this paper) is already high, which happens for failure-cause identification in our case. Here, we found that the two approaches (domain adaptation and model re-training) provide a comparable accuracy improvement, in the order of 2-2.5 percentage points each.

Conversely, when performance of *SD only* is poor, as in our failure-detection scenarios, the most significant advantage is brought by domain adaptation, which leads up to 20 percentage points of accuracy improvement over *SD only*, whereas model re-training still provides limited increase of accuracy.

We finally observe that several further investigations could be developed, which are left for future work. For example, TL and domain adaptation can be applied to different ONFM tasks, e.g., for failure localization and/or for failure magnitude estimation. Also the problems of failure detection and failure-cause identification studied here could be investigated by including other meaningful dimensions, such as, e.g., performing TL across different network topologies, transmission technologies (e.g., direct vs. coherent detection, different baud rates, modulation formats, coding rates, etc.) or equipment vendors. Furthermore, another possible extension can target an extension of the binary classification developed in this paper into a multi-class scenario, where ML models are able to distinguish between normal, extra attenuation and excessive filtering cases at once.

ACKNOWLEDGEMENTS

The work leading to these results has been supported by USA National Science Foundation Award no. 1818972 (JUNO2 program).

REFERENCES

1. D. Rafique and L. Velasco, "Machine Learning for Network Automation: Overview, Architecture, and Applications," J. Opt. Commun. Netw. **10**, D126–D143 (2018).
2. F. Musumeci, "Machine Learning for Failure Management in Optical Networks (Invited Tutorial)," in *Proceedings of Optical Fiber Communication Conference (OFC)*, (2021).
3. T. Tanimura, S. Yoshida, S. Oda, K. Tajima, and T. Hoshida, "Advanced data-analytics-based fiber-longitudinal monitoring for optical transport networks," in *2020 European Conference on Optical Communications (ECOC)*, (2020), pp. 1–4.
4. L. Shu, Z. Yu, Z. Wan, J. Zhang, S. Hu, and K. Xu, "Dual-Stage Soft Failure Detection and Identification for Low-Margin Elastic Optical

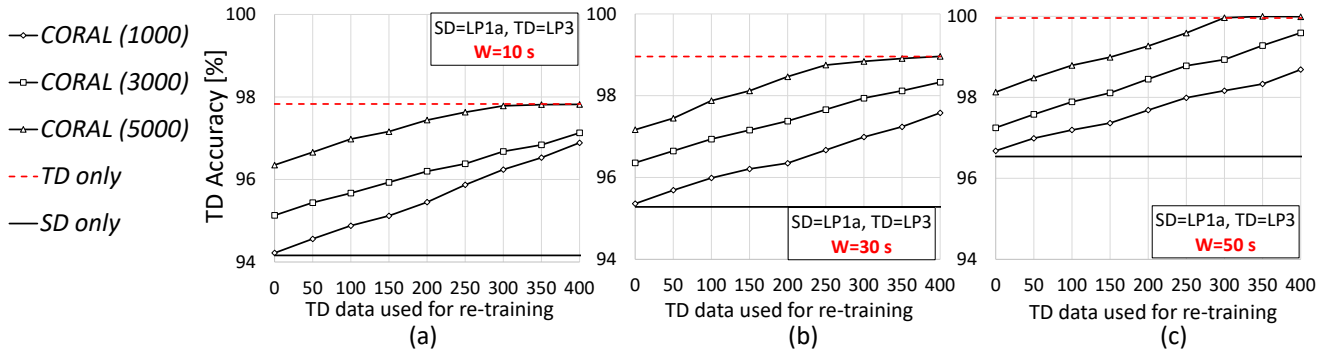

Fig. 7. Failure-cause identification (*extra attenuation vs. excessive filtering*): LP3 TD accuracy for increasing amount of labeled TD data used for re-training and for different OSNR window duration (SD=LP1a): (a) $W = 10$ s; (b) $W = 30$ s; (c) $W = 50$ s.

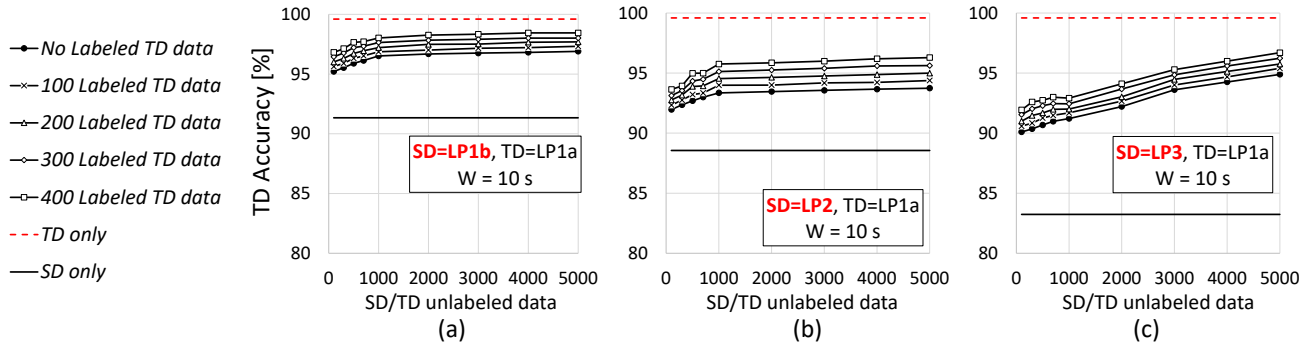

Fig. 8. Failure-cause identification (*extra attenuation vs. excessive filtering*): LP1a TD accuracy for increasing amount of unlabeled SD/TD data used in CORAL and labeled TD data used for re-training and for different SD lightpaths (window duration $W = 10$ s): (a) SD=LP1b; (b) SD=LP2; (c) SD=LP3.

Network by Exploiting Digital Spectrum Information," IEEE/OSA J. Light. Technol. **38**, 2669–2679 (2020).

5. C. Delezoide, K. Christodoulopoulos, A. Kretsis, N. Argyris, G. Kanakis, A. Sgambelluri, N. Sambo, P. Giardina, G. Bernini, D. Roccato, A. Percelsi, R. Morro, H. Avramopoulos, E. Varvarigos, P. Castoldi, P. Layec, and S. Bigo, "Pre-Emptive Detection and Localization of Failures Towards Marginless Operations of Optical Networks," in *2018 20th International Conference on Transparent Optical Networks (ICTON),* (2018), pp. 1–4.
6. 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 J. Light. Technol. **38**, 2992–2999 (2020).
7. A. P. Vela, M. Ruiz, F. Fresi, N. Sambo, F. Cugini, G. Meloni, L. Potì, L. Velasco, and P. Castoldi, "BER Degradation Detection and Failure Identification in Elastic Optical Networks," IEEE/OSA J. Light. Technol. **35**, 4595–4604 (2017).
8. H. Lun, Y. Wu, M. Cai, X. Liu, R. Gao, M. Fu, L. Yi, W. Hu, and Q. Zhuge, "ROADM-Induced Anomaly Localization and Evaluation for Optical Links Based on Receiver DSP and ML," IEEE/OSA J. Light. Technol. **39**, 2696–2703 (2021).
9. H. Lun, X. Liu, M. Cai, M. Fu, Y. Wu, L. Yi, W. Hu, and Q. Zhuge, "Anomaly Localization in Optical Transmissions Based on Receiver DSP and Artificial Neural Network," in *Optical Fiber Communications Conference and Exhibition (OFC),* (2020), pp. 1–3.
10. S. Barzegar, M. Ruiz, and L. Velasco, "Soft-Failure Localization and Time-Dependent Degradation Detection for Network Diagnosis," in *22nd International Conference on Transparent Optical Networks (ICTON),* (2020), pp. 1–4.
11. Z. Li, Y. Zhao, Y. Li, S. Rahman, F. Wang, X. Xin, and J. Zhang, "Fault Localization based on Knowledge Graph in Software-Defined Optical Networks," IEEE/OSA J. Light. Technol. **39**, 4236–4246 (2021).
12. C. Zhang, M. Wang, M. Zhang, D. Wang, C. Song, L. Guan, and Z. Liu, "Adaptive Failure Prediction Using Long Short-term Memory in Optical Network," in *24th OptoElectronics and Communications Conference (OECC) and 2019 International Conference on Photonics in Switching and Computing (PSC),* (2019), pp. 1–3.
13. S. J. Pan and Q. Yang, "A Survey on Transfer Learning," IEEE Transactions on Knowl. Data Eng. **22**, 1345–1359 (2010).
14. C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, and C. Liu, "A survey on deep transfer learning," in *Artificial Neural Networks and Machine Learning – ICANN 2018,* V. Kůrková, Y. Manolopoulos, B. Hammer, L. Iliadis, and I. Maglogiannis, eds. (Springer International Publishing, Cham, 2018), pp. 270–279.
15. G. Wilson and D. J. Cook, "A Survey of Unsupervised Deep Domain Adaptation," ACM Transactions on Intell. Syst. Technol. **11** (2020).
16. F. Musumeci, V. G. Venkata, Y. Hirota, Y. Awaji, S. Xu, M. Shiraiwa, B. Mukherjee, and M. Tornatore, "Transfer Learning across Different Lightpaths for Failure-Cause Identification in Optical Networks," in *2020 European Conference on Optical Communications (ECOC),* (2020), pp. 1–4.
17. F. Musumeci, C. Rottondi, G. Corani, S. Shahkarami, F. Cugini, and M. Tornatore, "A Tutorial on Machine Learning for Failure Management in Optical Networks," J. Light. Technol. **37**, 4125–4139 (2019).
18. R. Gu, Z. Yang, and Y. Ji, "Machine learning for intelligent optical networks: A comprehensive survey," J. Netw. Comput. Appl. **157**, 102576 (2020).
19. J. Mata, I. de Miguel, R. J. Durán, N. Merayo, S. K. Singh, A. Jukan, and M. Chamania, "Artificial intelligence (AI) methods in optical networks: A comprehensive survey," Opt. Switch. Netw. **28**, 43–57 (2018).
20. C.-Y. Liu, X. Chen, R. Proietti, and S. J. B. Yoo, "Performance studies of evolutionary transfer learning for end-to-end QoT estimation in multi-domain optical networks [Invited]," IEEE/OSA J. Opt. Commun. Netw. **13**, B1–B11 (2021).

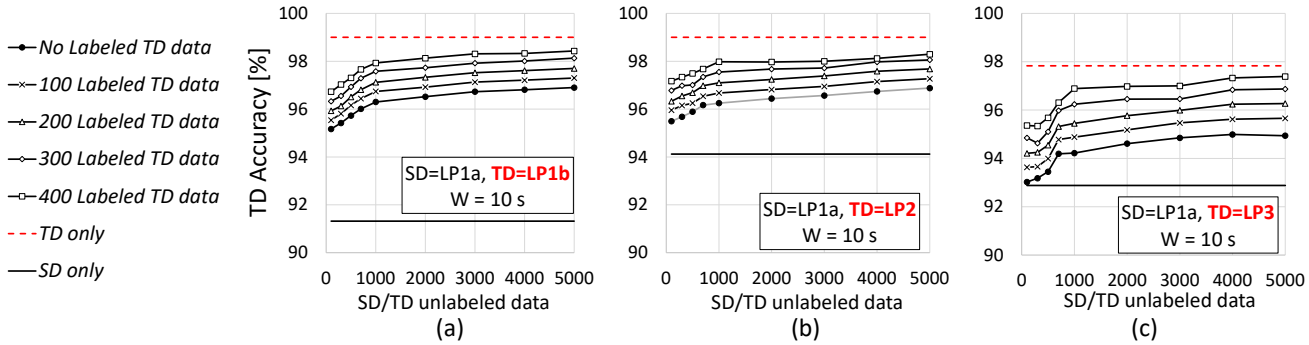

Fig. 9. Failure-cause identification (*extra attenuation vs. excessive filtering*): accuracy for different TD lightpaths, increasing amount of unlabeled SD/TD data used in CORAL and labeled TD data used for re-training (window duration $W = 10$ s): (a) TD=LP1b; (b) TD=LP2; (c) TD=LP3.

21. J. Pesic, M. Lonardi, T. Zami, N. Rossi, and E. Seve, "Transfer Learning Using ANN for G-OSNR Estimation in WDM Network Topologies," in *OSA Advanced Photonics Congress (AP) 2020 (IPR, NP, NOMA, Networks, PVLED, PSC, SPPCom, SOF),* (Optical Society of America, 2020), p. NeM3B.3.
22. J. Pesic, M. Lonardi, E. Seve, N. Rossi, and T. Zami, "Transfer Learning from Unbiased Training Data Sets for QoT Estimation in WDM Networks," in *2020 European Conference on Optical Communications (ECOC),* (2020), pp. 1–4.
23. D. Azzimonti, C. Rottondi, A. Giusti, M. Tornatore, and A. Bianco, "Comparison of domain adaptation and active learning techniques for quality of transmission estimation with small-sized training datasets [Invited]," IEEE/OSA J. Opt. Commun. Netw. **13**, A56–A66 (2021).
24. I. Khan, M. Bilal, M. Umar Masood, A. D'Amico, and V. Curri, "Lightpath QoT computation in optical networks assisted by transfer learning," IEEE/OSA J. Opt. Commun. Netw. **13**, B72–B82 (2021).
25. J. Yu, W. Mo, Y.-K. Huang, E. Ip, and D. C. Kilper, "Model transfer of QoT prediction in optical networks based on artificial neural networks," IEEE/OSA J. Opt. Commun. Netw. **11**, C48–C57 (2019).
26. P. Paudyal, S. Shen, S. Yan, and D. Simeonidou, "Toward Deployments of ML Applications in Optical Networks," IEEE Photonics Technol. Lett. **33**, 537–540 (2021).
27. Z. Xu, C. Sun, T. Ji, J. H. Manton, and W. Shieh, "Feedforward and Recurrent Neural Network-Based Transfer Learning for Nonlinear Equalization in Short-Reach Optical Links," J. Light. Technol. **39**, 475–480 (2021).
28. X. Chen, R. Proietti, C.-Y. Liu, Z. Zhu, and S. J. Ben Yoo, "Exploiting Multi-Task Learning to Achieve Effective Transfer Deep Reinforcement Learning in Elastic Optical Networks," in *2020 Optical Fiber Communications Conference and Exhibition (OFC),* (2020), pp. 1–3.
29. Q. Yao, H. Yang, A. Yu, and J. Zhang, "Transductive Transfer Learning-Based Spectrum Optimization for Resource Reservation in Seven-Core Elastic Optical Networks," J. Light. Technol. **37**, 4164–4172 (2019).
30. B. Sun, J. Feng, and K. Saenko, "Return of frustratingly easy domain adaptation," in *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence,* (AAAI Press, 2016), AAAI'16, p. 2058–2065.
31. S. Shahkarami, F. Musumeci, F. Cugini, and M. Tornatore, "Machine-Learning-Based Soft-Failure Detection and Identification in Optical Networks," in *Optical Fiber Communications Conference and Exposition (OFC) 2018,* (2018), pp. 1–3.