Comparative Analysis on Cluster-based Algorithms for Failure Management in Optical Networks

Rafael Sales, Andrei Ribeiro, Fabrício Lobato, João CWA Costa, and Moisés Silva

***Abstract*— The occurrence of failures in transparent optical networks (TONs) potentially compromises their reliability, leading to loss of information or even link disruption. In that regard, as the conventional methods for failure management become difficult with the increasing complexity of TONs, the use of machine learning algorithms has emerged as an alternative strategy given their sole capabilities concerning network scalability and self-awareness. Therefore, in this paper, the failure management performance of three cluster-based algorithms (k-means, fuzzy c-means, and Gaussian mixture models) are compared on a real-world testbed in terms of failure classification errors.**

***Keywords*— Clustering, Failure Management, Unsupervised Learning, Optical Networks.**

I. Introduction

Failure management in transparent optical networks (TONs) is critical for network reliability maintenance. Optical transmission is affected by the occurrence of failures, leading to degradation in the quality of transmission and service level agreements (SLA) violations [1], [2]. In that sense, the fulfillment of SLA can ensure an effective monitoring of the physical layer regarding the occurrence of failures [3].

Powered by the increasing advances in several fields (e.g. cloud computing, internet of things, 5G applications), the parameter complexity of TONs has grown, as the number of optical devices (e.g. EDFAs, repeaters, transponders, etc.) increases. Consequently, the conventional methods for failure management become inefficient, as they are commonly based on simplified threshold and probabilistic statistical models. Correspondingly, machine learning (ML) algorithms have been attracting attention for failure management in optical networks. These techniques can automate complex tasks, ranging from pattern recognition to forecasting, which often require human labor, and are also time-consuming. Hence, the use of ML algorithms can enable automated self-awareness networks capable of performing effective failure management [4], [5].

However, most of the ML-driven studies for failure management in TONs are focused on the use of supervised learning approaches, such as artificial neural networks [6], [7], tree-based algorithms and support vector machines [8], [9]. In these cases, the models need prior knowledge of data labels, requiring information about all possible failure conditions. However, for optical networks, the datasets collected from the practical operating system often present an extremely imbalanced nature, where the volume of data under normal conditions (i.e. without failures) is much larger than the volume of data under failure conditions (if any). This limitation makes supervised learning approaches unfeasible as they need a considerable number of failure samples to be trained on.

Therefore, unsupervised learning algorithms [10], [11] have become a promising alternative due to their ability to be trained with only data under normal conditions, disregarding the acquisition of data from equipment that is undergoing failures or to simulate additional data from statistical models. Specifically, the main goal of unsupervised learning models is to learn hidden patterns under normal operation conditions by recognizing underlying information. Leveraging that fact, when anomalies (failures) are presented, they can be identified from the residual errors, as the models were trained with data only from the system's normal behavior. Correspondingly, cluster-based techniques are presented as one of the most known unsupervised learning algorithms. The rationale of these models is to group data points by learning their inner distributions. Upon grouping, these data groups (clusters) may be used to identify samples arising from a different distribution unseen by the models.

In that regard, in this paper, three cluster-based algorithms, namely as k-means, fuzzy c-means (FCM), and Gaussian mixture model (GMM), are used to perform failure detection and localization in a transparent optical network. False-positive (type I) and false-negative (type II) indications of failures are used for algorithm performance evaluation.

The rest of the paper is structured as follows. In Section 2 we present the background of the three cluster-based algorithms under study. Section 3 carries out the proposed approach to failure detection and localization. Section 4 describes the testbed and the dataset used in our experiments, and the failure management results of the three cluster-based algorithms. Finally, in Section 5 we draw our conclusions.

II. Clustering Algorithms

A. *K-means*

The k-means algorithm is one of the most common unsupervised clustering techniques, widely used to cluster large data sets. The algorithm assigns each of the data samples to one of the K clusters generated by the method. Considered a hard-clustering algorithm, each sample belongs to only one cluster, the one that has the smallest Euclidean distance from the sample. The idea of the k-means clustering is to update the centroids by calculating the mean of the samples belonging to the cluster and repeat the relocating-and-updating process until convergence criteria are satisfied. The number of clusters is defined by the user and the initial positions of the clusters are generally random [12].

Rafael Sales, Andrei Ribeiro, Fabrício Lobato, and João CWA Costa, Insitute of Technology, Federal University of Pará, Belém-Pará, Brazil, e-mail: [rafael.sales, andrei.ribeiro]@itec.ufpa.br, [frl, jweyl]@ufpa.br; Moisés F. Silva, Los Alamos National Laboratory, Los Alamos-NM, USA, e-mail: mfelipe@lanl.gov

K-means aims to minimize the Equation 1, where x is the sample belonging to the k-th cluster C and μ_k is the centroid of the cluster C_k. Some iterations are necessary until the algorithm reaches a local minimum.

$$J = \sum_{k=1}^{K} \sum_{x \in C_k} \|x - \mu_k\|^2, \tag{1}$$

In this work, a failure indicator (FI) is used to analyze whether the sample was drawn from a failure condition. Different clustering approaches may follow different approaches to generate their FIs. Considering a dataset $X_{n \times m}$, with n samples of m dimensions, the k-means FIs are given using the following Equation:

$$FI(x_i, \mu_j) = \sqrt{\sum_{d=1}^{m} (x_{id} - \mu_{jd})^2}, \tag{2}$$

where x_i is a sample from the dataset and μ_j is the closest centroid to the sample.

B. Fuzzy C-means

Fuzzy c-means is an unsupervised clustering technique based on centroids similar to k-means. However, in contrast to k-means, FCM is a soft-clustering algorithm [13]. This means that instead of each sample belonging to just one cluster, the sample belongs to all clusters with a certain degree of membership. This makes the equation for the centroids of each cluster a weighted average, given by:

$$c_k = \frac{\sum_{i=1}^{n} w_{ik}^p x_i}{\sum_{i=1}^{n} w_{ik}^p}, \tag{3}$$

where x_i is the i-th sample, w_{ki} is the degree of membership of the i-th sample to the k-th cluster and p is the fuzziness parameter. Membership values are initially set randomly within the range allowed by the algorithm (0 to 1) and updated after the cluster center is calculated. The updating is given by Equation 4 as follows:

$$w_{ik} = \frac{1}{\sum_{j=1}^{n} \left(\frac{\|x_i - c_k\|}{\|x_i - c_j\|} \right)^{\frac{2}{p-1}}}. \tag{4}$$

The FCM algorithm updates the membership values until Equation 5 reaches a local minimum or a pre-defined number of iterations. Note that k-means aims to minimize the same equation, just restricting the membership values.

$$J = \sum_{i=1}^{N} \sum_{k=1}^{K} w_{ik}^p \|x_i - \mu_k\|^2. \tag{5}$$

The FIs for the FCM are the same as that of k-means since both algorithms are based on the same clustering approach. For the FCM, FIs are given using the Equation 2.

C. Gaussian Mixture Model

The Gaussian mixture model algorithm performs model-based data clustering. For this purpose, multivariate finite mixture models are used, which aim to capture the main clusters. On this wise, the GMM can learn non-linear relationships, assuming that the data can be modeled by a set of finite multivariate Gaussian distributions. For a GMM, each component $g(x|\theta_k)$ is represented as a Gaussian distribution,

$$g(x|\theta_k) = \frac{exp\{-\frac{1}{2}(x - \mu_k)^T \Sigma_k^{-1} (x - \mu_k)\}}{(2\pi)^{m/2} \sqrt{det(\Sigma_k)}}, \tag{6}$$

being each component denoted by the parameters, $\theta_k = \{\mu_k, \Sigma_k\}$, composed by the mean vector, μ_k and the covariance matrix, Σ_k. Thus, a GMM is completely specified by a set of parameters $\Theta = \{\alpha_1, \alpha_2, ..., \alpha_K, \theta_1, \theta_2, ..., \theta_K\}$.

Hence, a finite mixture model, $g(x|\Theta)$, is the weighted sum of $K > 1$ components $g(x|\theta_k)$ in $\mathbb{R}^m$,

$$g(x|\Theta) = \sum_{k=1}^{K} \alpha_k g(x|\theta_k), \tag{7}$$

where α_k corresponds to the weight of each component. These weights are positive $\alpha_k > 0$ with $\sum_{k=1}^{K} \alpha_k = 1$.

To estimate the GMM parameters, the expectation–maximization (EM) local search method is one of the most used [14], [15]. This method consists of two steps: i) expectation and ii) maximization. In order for the log-likelihood (LogL), $log(g(X|\Theta)) = log(\prod_{i=1}^{n} g(x_i|\Theta))$ to converge to a local optimum, the two steps are applied alternately. The performance of the EM algorithm directly depends on the choice of initial parameters Θ [16], as a poor choice of the initial parameter can result in many replications of this method during an execution.

The FIs for GMM differs from other algorithms presented. In this case, the squared Mahalanobis distance is used to calculate the FIs. Unlike the Euclidean distance, it takes into account the distribution of the data, for which the FIs are given using the following equation:

$$FI(x_i|\theta_k) = (x_i - \mu_k) \Sigma_k^{-1} (x_i - \mu_k)^T, \tag{8}$$

where x_i is a sample from the dataset and θ_k is the closest component to the sample.

III. Cluster-based Failure Management

To detect and locate failures in the optical network, this paper uses the approach shown in Fig. 1. The first phase consists on model training, where network data under normal conditions is used for this task. The reason for using only normal data is that cluster models can group normal behavior to later distinguish anomalous behavior.

Training is based on generating thresholds from the normal condition. For this, the model groups the normal data into clusters and generates FIs (different algorithms can have different ways of generating FIs). Thresholds are generated using the 99th percentile of the FIs since clustering algorithms based on centroids (k-means, FCM) and distributions (GMM) are sensitive to outliers. For localization purposes, FIs and

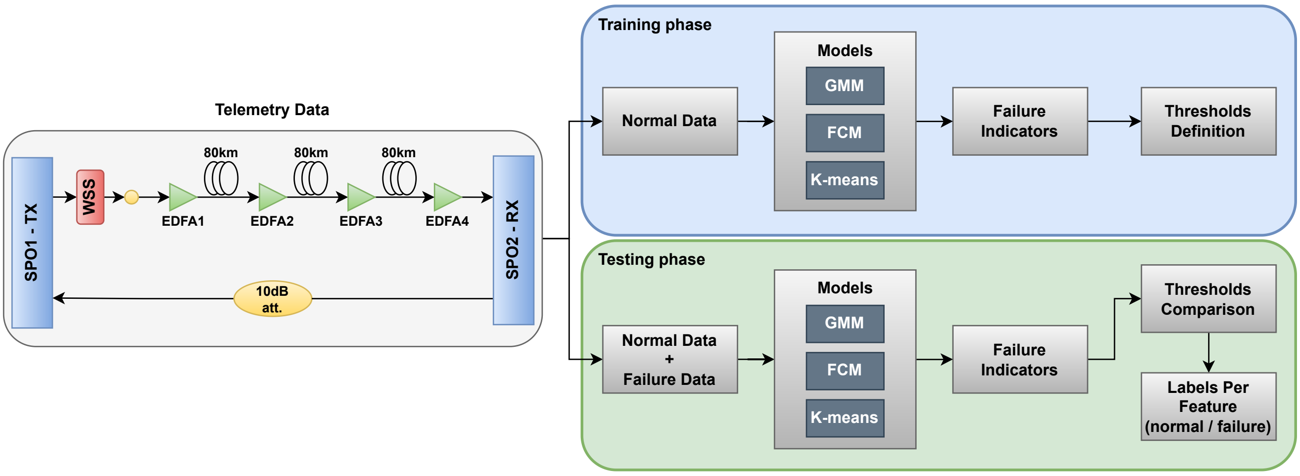


Fig. 1: Proposed approach.

thresholds are generated for each feature. Since the detection is done by feature, the failure localization for each span is made by analyzing the errors in the input power of the EDFA.

In the testing phase, unlabeled data from the network is presented to the trained model. The model generates FIs for each feature in the sample. After that, each feature has its FI compared to its specific thresholds previously defined in the training phase. If the FI exceeds the thresholds, then it is considered under failure condition, otherwise, it is considered normal. Since the model was trained only on the features of the EDFAs, this approach allows us to detect the failure and localize it through the EDFAs.

Properly mapping the normal condition of the network is important to ensure adequate failure management. For this reason, it is extremely important to define an appropriate number of clusters so that the models can better capture the normal behavior of the network. In this work, we use the Akaike information criteria (AIC) to define the number of clusters [17]. AIC uses the negative log-likelihood and adds a penalizing term associated with the number of variables, penalizing more complex models. The AIC is given by:

$$AIC = -2L(M) + 2v(M), \tag{9}$$

where $L(M)$ is the log likelihood function and $v(M)$ is the number of free parameters in the model M.

IV. Results

A. Testbed and Dataset

The experimental data is collected from the TON testbed (Fig. 1) available in [18]. It includes two Ericcson SPO 1400 devices, one wavelength selective switch (WSS), and four EDFA amplifiers. At the end of the WSS, an attenuator is installed to simulate failures. The optical link between the SPO-TX and SPO-RX consists of 3 spans of 80 km each. The data is collected with a sampling frequency of 3.5 seconds and 4 features are used in this work, corresponding to the 4 EDFAs input powers. The dataset consists of 8252 data points (10 hours), where the first 7067 samples under normal conditions (8 hours) are used for training and the remaining 1185 samples under failure and normal conditions (2 hours) are used for testing. It is important to note that failure samples are presented only in the percentage of data corresponding to the test phase. Fig. 2 presents the four features composing the dataset and splitting it into training and testing sets. As the attenuator is placed between WSS and EDFA1, only the input power of EDFA1 presents significant variations in the testing set.

Moreover, one can note in Fig. 2 two distinct distributions along the training set. The first 6 hours are composed of data under real normal network condition, with no changes applied with the WSS. While the remaining 2 hours are still composed of data under normal conditions, but with variations in attenuation. The WSS is used to randomly change the attenuation every 10 seconds in a range from 0 to 18dB. These variations are used for model generalization concerns as they provide different information from the normal condition of network traffic. However, in the two hours of the testing set, WSS is used to add 25dB attenuation every 40 seconds, thus placing the network under a failure condition for 10 seconds (hard-failure). After that, the WSS is reconfigured so that the network works correctly again.

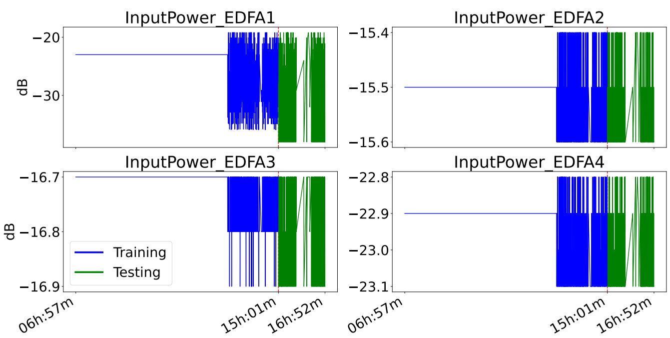


Fig. 2: Optical dataset along with the training and test data.

B. Comparison Analysis

In this subsection, the analysis is divided into two parts. Firstly, as a previous step of the optical failure management, the AIC (Equation 9) is calculated for each clustering model to choose its respective number of components. A set of number of components ranging from 2 to 10 components is evaluated. The AIC is executed 20 times for each number of components and the average value is considered at the end, since the algorithms are seed sensitive.

As shown in Fig. 3, the k-means, FCM, and GMM models showed the same optimal number of components (2), as the minimum AIC values were drawn for the first model. Moreover, k-means and FCM presented almost the same AIC values for each number of components. In this sense, to provide specific results with respect to a failure detection strategy, the best models are selected based on the AIC. Therefore, k-means, FCM, and GMM with 2 components are evaluated in the second part of this subsection. The outlier detection performance of these compared techniques is evaluated in terms of type I (false-positive) and type II (false-negative) indications of failures based on a linear threshold defined for the 99th percentile of the training data.

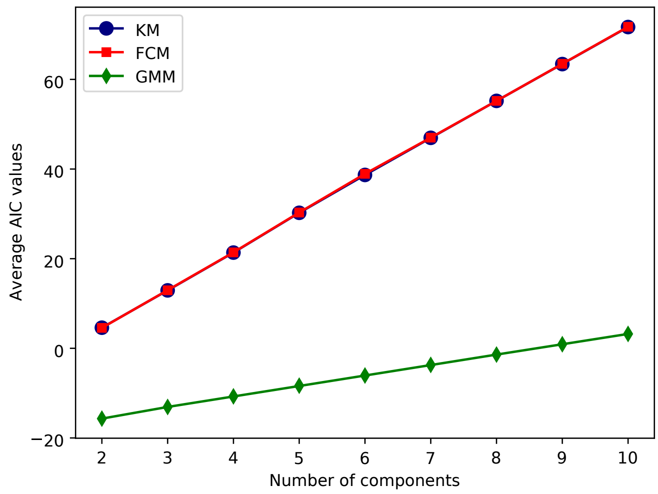


Fig. 3: Average AIC values for 20 trials of each cluster-based algorithm for varying number of components.

The failure detection performance of the three approaches is summarized in Table I. Note that type II indications are not presented for EDFA2, EDFA3, and EDFA4, as no failures were simulated at these equipments. The table shows that all models present marginal differences in results over the two comparison metrics. However, one may note in Table I that the models misclassified some failure samples presented in EDFA1, corresponding to the respective type II errors. Part of this is due to the simulated process of failures in the dataset. When the failure condition was simulated, although the label of the respective sample was set to failure, the system gradually changed from one state to the other, creating a few samples labeled as "failure". Thus, those not detected failure samples have characteristics related to normal conditions.

Another important detail is the relative high rate of type I errors in EDFA3. However, as shown in Fig. 2, unlike other EDFAs, the EDFA3 input power values do not present a significant amount of data with variation from the normal condition during the period used for training. Therefore, this fact affects the failure detection performance of the clustering models by reducing model generalization when submitted for failure scenarios in the test phase.

Model	Amplifier	Type I (%)	Type II (%)
K-means	EDFA1	0.8812	3.3616
	EDFA2	0	-
	EDFA3	27.4151	-
	EDFA4	0	-
FCM	EDFA1	0.6854	3.4595
	EDFA2	0	-
	EDFA3	27.4151	-
	EDFA4	0	-
GMM	EDFA1	0	3.4922
	EDFA2	0	-
	EDFA3	26.5339	-
	EDFA4	0	-

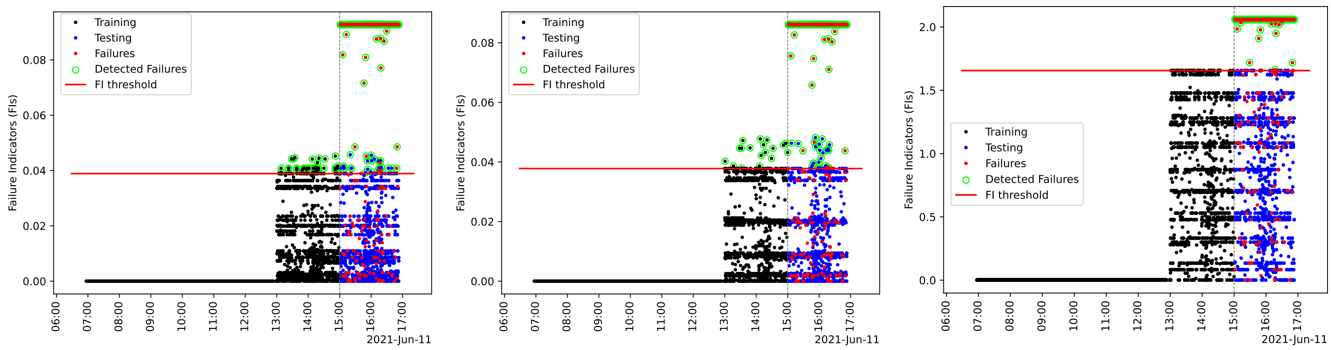
TABLE I: Failure detection performance of the k-means, the FCM and the GMM.

In that regard, for further analysis, the FIs for the EDFA1 derived from the three clustering approaches are shown along the threshold and over time in Fig. 4. From a general perspective, all the models successfully detect/locate the failure presented at the EDFA1 equipment. Red points in the Fig. 4 (EDFA1 for the three clustering models) are majority arranged above the threshold line, corresponding to around 85% of the 711 failure samples.

In fact, all the previous analysis demonstrate all the approaches achieved marginal differences in results. To find the most appropriate model, we closely look at type I and II errors. For type I errors, the model accuses the network of being in a fault condition, even though the network is in a normal state. From an economic point of view, the maintenance costs associated with a recurring type I error can be a problem, increasing the OPEX. In this scenario, GMM may be the most appropriate option, since it presented a marginally better result. However, the situation changes when we look at the type II error. In this scenario, the model indicates normality while the network is in a fault condition. This error affects the availability of network services. For a scenario where network availability is absolutely critical, k-means may be the most suitable as it presents marginally better results than other algorithms for type II errors.

V. Conclusions

In this paper, we proposed a comparison study of three cluster-based algorithms for failure detection and localization in TONs. The application of these unsupervised models allows the use of this technique for cases when data under fault conditions is a problem. Moreover, their performance comparison in terms of type I/II errors demonstrates marginal differences in tests for the real-world testbed evaluated. Hence, a more detailed analysis of type I and II errors was provided to choose the most appropriate model. For cases where maintenance is expensive, the GMM algorithm outperforms the others, as it



(a) K-means (b) FCM (c) GMM

Fig. 4: Failure detection performance of the proposed approaches for EDFA1.

achieved the best results in type I errors. For cases where network availability is a absolutely critical, k-means has an advantage over other algorithms in terms of type II errors.

In future works, new studies to evaluate the performance of other existing clustering algorithms, in terms of failure management in optical networks, should be carried out. Furthermore, failure forecasting will be further evaluated by analysis related to the temporal characteristics of the data, providing crucial resources for failure prediction.

ACKNOWLEDGEMENTS

This work was supported by the Federal University of Pará (UFPA), National Council for Scientific and Technological Development (CNPq) and Institute for Amazon Sustainability with Science and Innovation (ISACI). Thanks to Andrea Sgambelluri and Luca Valcarenghi from Scuola Superiore Sant'Anna for making the dataset available.

REFERENCES

[1] X. Chen et al., "Flexible Availability-Aware Differentiated Protection in Software-Defined Elastic Optical Networks," in Journal of Lightwave Technology, vol. 33, no. 18, pp. 3872-3882, 15 Sept.15, 2015, doi: 10.1109/JLT.2015.2456152.

[2] Z. Cheng, X. Zhang, S. Shen, S. Yu, J. Ren and R. Lin, "T-trail: link failure monitoring in software-defined optical networks," in Journal of Optical Communications and Networking, vol. 10, no. 4, pp. 344-352, April 2018, doi: 10.1364/JOCN.10.000344.

[3] A. P. Vela et al., "Soft failure localization during commissioning testing and lightpath operation," in Journal of Optical Communications and Networking, vol. 10, no. 1, pp. A27-A36, Jan. 2018, doi: 10.1364/JOCN.10.000A27.

[4] Musumeci, Francesco et al. "A Survey on Application of Machine Learning Techniques in Optical Networks." ArXiv abs/1803.07976 (2018).

[5] X. Chen, R. Proietti, H. Lu, A. Castro and S. J. B. Yoo, "Knowledge-Based Autonomous Service Provisioning in Multi-Domain Elastic Optical Networks," in IEEE Communications Magazine, vol. 56, no. 8, pp. 152-158, August 2018, doi: 10.1109/MCOM.2018.1701191.

[6] D. Rafique, T. Szyrkowiec, H. Grießer, A. Autenrieth and J. -P. Elbers, "Cognitive Assurance Architecture for Optical Network Fault Management," in Journal of Lightwave Technology, vol. 36, no. 7, pp. 1443-1450, 1 April1, 2018, doi: 10.1109/JLT.2017.2781540.

[7] A. P. Vela et al., "BER Degradation Detection and Failure Identification in Elastic Optical Networks," in Journal of Lightwave Technology, vol. 35, no. 21, pp. 4595-4604, 1 Nov.1, 2017, doi: 10.1109/JLT.2017.2747223.

[8] S. Shahkarami, F. Musumeci, F. Cugini and M. Tornatore, "Machine-Learning-Based Soft-Failure Detection and Identification in Optical Networks," 2018 Optical Fiber Communications Conference and Exposition (OFC), San Diego, CA, USA, 2018, pp. 1-3.

[9] B. Shariati, M. Ruiz, J. Comellas and L. Velasco, "Learning From the Optical Spectrum: Failure Detection and Identification," in Journal of Lightwave Technology, vol. 37, no. 2, pp. 433-440, 15 Jan.15, 2019, doi: 10.1109/JLT.2018.2859199.

[10] X. Chen, B. Li, R. Proietti, Z. Zhu and S. J. B. Yoo, "Self-Taught Anomaly Detection With Hybrid Unsupervised/Supervised Machine Learning in Optical Networks," in Journal of Lightwave Technology, vol. 37, no. 7, pp. 1742-1749, 1 April1, 2019, doi: 10.1109/JLT.2019.2902487.

[11] H. Lun et al., "A GAN Based Soft Failure Detection and Identification Framework for Long-Haul Coherent Optical Communication Systems," in Journal of Lightwave Technology, vol. 41, no. 8, pp. 2312-2322, 15 April15, 2023, doi: 10.1109/JLT.2022.3227719.

[12] Jin, X., Han, J. (2017). K-Means Clustering. In: Sammut, C., Webb, G.I. (eds) Encyclopedia of Machine Learning and Data Mining. Springer, Boston, MA. https://doi.org/10.1007/978-1-4899-7687-1_431

[13] R.Suganya, R.Shanthi (2018); "Fuzzy C- Means Algorithm- A Review"; Int J Sci Res Publ 2(11) (ISSN: 2250-3153)

[14] G. McLachlan and D. Peel, Finite Mixture Models, Wiley Series in Probability and Statistics (Wiley, 2004).

[15] Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977), "Maximum Likelihood from Incomplete Data Via the EM Algorithm." Journal of the Royal Statistical Society: Series B (Methodological), 39: 1-22. https://doi.org/10.1111/j.2517-6161.1977.tb01600.x

[16] M. A. T. Figueiredo and A. K. Jain, "Unsupervised learning of finite mixture models," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 3, pp. 381-396, March 2002, doi: 10.1109/34.990138.

[17] Lord, D., Qin, X., and Geedipally, S. R. (2021). Chapter 2 - fundamentals and data collection. In Lord, D., Qin, X., and Geedipally, S. R., editors, Highway Safety Analytics and Modeling, pages 17–57. Elsevier.

[18] InRete Lab, "Optical failure dataset," https://github.com/Network-And-Services/optical-failure-dataset (2021). Accessed: 2024-01-13