Machine Learning Applications for Fault Tracing and Localization in Optical Fiber Communication Networks: A Review

Shiny Pearl D. Abdula[1], Mary Joy D. Llagas[2], Alessandra May D. Fernandez[3], Edwin R. Arboleda[4]

[1]Department of Computer, Electronics, and Electrical Engineering, Cavite State University, Indang, Cavite, Philippines.
Email: shinypearl.abdula@cvsu.edu.ph

[2]Department of Computer, Electronics, and Electrical Engineering, Cavite State University, Indang, Cavite, Philippines.
Email: maryjoy.llagas@cvsu.edu.ph

[3]Department of Computer, Electronics, and Electrical Engineering, Cavite State University, Indang, Cavite, Philippines.
Email: alessandramay.fernandez@cvsu.edu.ph

[4]Department of Computer, Electronics, and Electrical Engineering, Cavite State University, Indang, Cavite, Philippines. Email:
edwin.r.arboleda@cvsu.edu.ph

A B S T R A C T

The review aims to assess fifteen (15) academic literature sources, highlighting the application of machine learning algorithms in the maintenance operations of optical fiber networks. It exhibits the collection of data using PRISMA methodology—Preferred Reporting Item for Systems Review and Meta-Analyses. The application, results, and performance metrics are discussed based on the collected observations, computations, and statistics in the studies, which revealed records of high accuracy degrees ranging from 86% to 98% on average and quality ML models including Neural Networks (NNs), Support Vector Machines (SVMs), and LSTM, as well as deep learning models that disclosed effective results of determining challenges and problems within the optical fiber lines. The review mainly centralized on superior machine learning technologies that surpass traditional techniques in fault detection and localization for improved optical fiber networks' operations while providing insights into the limitations and challenges encountered in real-world applications of these models, offering a comprehensive perspective on the optical fiber network's domain.

Keywords: Machine Learning, Optical Fiber Networks, Anomaly Identification, and Localization

1. INTRODUCTION

1.1. Overview of Fiber Optic Communication Networks

In modern communication, many industries are interconnected creating real-time communication that exchanges information, enabling to delivery of rich media content, and handling diverse types of data without lagging or buffering. Fast data transfer across long distances provided by fiber optic networks has become one of the main pillars of power communication with its high capacity, anti-interference, low cost, and fast speed in a long transmission distance. As these networks failure occurs, they significantly impact the effective operation and secure production of the power system as the primary transmission medium of the power communication highway [9].

Optical fiber is a technology that involves transmitting data using light pulses through a long glass or plastic fiber [27]. It serves as a light guide, typically in the infrared range, transmitting optical signals over long distances with minimal signal loss. Due to the non-conducting nature of this fiber, no electrical voltages or currents can be associated. These special properties allow the optical fiber to be used in a sensor design that helps manipulate signals effectively. This leads to the development of highly sensitive sensors that produce faster and clearer signals [25]. However, making fiber optic-based sensors can be challenging because of the need to integrate interferometers and carefully manage amplitude modulation in the output spectrum. Despite significant progress, challenges persist in the manufacturing and operation of these sensors. These challenges include complexity, associated costs, and technical limitations in both sensor design and manufacturing processes [7]. Any disruption of different types of anomalies, including a fiber cut or unauthorized access through eavesdropping can be enormous and must be responded to immediately. The manual discovery of these incidents occurring in the fiber requires considerable knowledge and probing time until a fault is identified [16].

1.2. Machine Learning Role in Addressing Challenges

Machine learning (ML) has a significant role in the management of failures in optical fiber networks, enhancing the reliability and efficiency of these networks. It introduces the automated methods that are transforming traditional, mostly manual approaches to failure management. The different key aspects where ML contributes:

1. **Failure Detection:** Monitoring network performance, ML can detect anomalies that may indicate a failure. It can analyse the amounts of network data to identify patterns and anomalies that may not be apparent to human operators. The algorithms used can quickly process data in real-time, enabling rapid detection of anomalies in network performance that may indicate potential failures, such as changes in signal strength or transmission speed. ML helps operators identify and allows for faster response times and reduces the impact of network disruptions.

2. **Failure Localization:** Refers to the process of identifying the specific location or component within an optical fiber network that is experiencing a failure or anomaly. This is a critical step in network maintenance and troubleshooting, as it allows operators to quickly pinpoint the source of the problem and take appropriate corrective actions. By analysing network data and patterns, ML algorithms can identify the specific segment or device that is experiencing issues, speeding up the repair process and minimizes downtime and ensure the reliability of the optical fiber network.

3. *Failure Identification:* Involves identifying the specific type of failure, which can be a complex task given the variety of potential issues in optical networks. ML aids in accurately diagnosing the problem by identifying patterns associated with different types of failures and analyzing historical data. By leveraging real-time monitoring, ML algorithms can detect anomalies from normal network behavior. It helps to differentiate between different types of failures including a fiber cuts, and signal degradation, based on patterns in network data.

1.3. Significance of Fault tracing and Localization

This review critically examines machine learning applications for anomaly detection and localization, providing significant insight into their applicability for specific applications. This section highlights several reasons of the importance of anomaly detection and localization:

- Data Loss Prevention: Anomaly detection and localization help prevent major data loss by quickly finding and fixing issues like fiber cuts or eavesdropping attempts that can disrupt data transmission in fiber optic networks [16].

- Service Continuity: By detecting and pinpointing anomalies, network operators can ensure that thousands of customers continue to receive uninterrupted service. Ensuring the reliability of communication networks is crucial, especially during critical situations [16].

- Enhancement of Security: Anomaly detection and localization enhance network security by identifying and addressing potential security breaches like invasions or attacks, protecting sensitive data transmitted across the network [16].

- Cost Reduction: Timely identification and localization of anomalies can lead to significant cost savings by preventing downtime, reducing the need for lengthy troubleshooting, and facilitating quick repair work [16].

- Efficient Network Management: Machine learning techniques, such as autoencoders and attention-based algorithms, improve the accuracy and speed of anomaly detection and localization. This enables network operators to proactively manage optical fiber communication networks, ensuring their security, reliability, and uninterrupted service to users [16].

1.4. Objectives

The primary objective of the literature review is to provide a systematic exposition of data based on the integrated potentials given on ML-based systems in examining, tracing, and localizing the irregularities/anomalies in optical fiber communication networks. It intends to fulfill the following:

- Analyse the existing related literature to observe trends in machine learning applications for abnormalities in optical fiber communication networks.

- Determine the methodologies employed in previous studies to assess their efficacies in detecting and localizing faults along optical fiber networks.

- Classify leading machine learning algorithms on anomaly identification and localization in optical fibers networks.

- Observe the performance metrics results of ML-based defect detection models in optical fiber networks.

- Identify the manifested challenges in implementing machine learning algorithms and architectures for anomaly detection and localization in optical fiber communication networks.

2. METHOD

The accumulated published papers have undergone thorough screening, enabling a centralized focus on the review objectives via the PRISMA methodology—Preferred Reporting Item for Systems Review and Meta-Analyses [17]. The filtration process involved discerning sets of literature integrated with machine learning for fault identification and localization within an optical fiber network.

2.1. Literature Search

The initial review conforms to standardized protocols, entailing formal criteria.

2.2. Principles for ML-Based Fault Identification in Optical Fiber Networks

The gathered studies were submitted for extensive evaluation, adhering to a framework that emphasizes ML-based systems with objectives to perform fault diagnosis and localization for optical fiber networks sourced from legitimate scholarly databases. The established systems and experimentations between 2019 to 2024 were regarded in the review.

Reputable online research repositories—ResearchGate, IEEE Xplore, and Google Scholar—were utilized to obtain reliable academic literature and verified ML-based configurations. Direct input of the terms in the search, such as 'Machine Learning,' 'Fault Detection,' and 'Optical Fiber Network,' in the titles, keywords, and abstracts, was used at the outset of the collection. Alternative words such as 'Anomaly Diagnosis,' 'Defect Detection,' and 'Optical Fiber Cable' were also considered and entered within each database to delve further into the topic of concern—particularly in the status of optical fiber transmission lines—and maximize the information yielded on the platforms.

Studies incorporated persistent repetition of terms during scrutiny, such as 'Optical Fiber', 'Machine Learning', and 'Fault Detection,' were subjugated to discreet scanning of the materials to avoid inconsistency with the central topic. Literature containing such terms was found; however, any identified discrepancies were automatically rejected from the review.

The preliminary procedure in checking the integrity of the retrieved literature progressed via the following phases of evaluation:

The researchers conducted the following steps as preliminaries to assess the integrity of the searched studies:

1. **Identification:** The data collection involved a straightforward registration of the terms— 'Machine Learning,' 'Optical Fiber,' and 'Fault Detection'—in each database to ensure a direct filtration of the topics. Parallel terms such as 'Anomaly,' 'Neural Network,' 'Defect Detection,' 'Optical Fiber Network,' 'Optical Fiber Cable,' and 'Optical Fiber Communication' were used universally to maintain coherence and evenness in the processed studies on the platforms. These words were also regarded as references to find related titles, abstracts, and keywords.

2. **Screening:** In this stage, measurements are taken cautiously when correlating keywords/terms in the accumulated scholarly papers to refrain from acquiring false arguments, especially from optical fiber networks using ML-based fault detection models. Disparate topics unrelated to the scope of the review were excluded; additionally, methodologies applied tangentially, contradicting the deployment of machine learning in the localization of faults within optical fiber networks, were not exempted from the procedure.

3. **Eligibility:** Appropriate applications of Machine Learning-based systems in optical fiber networks from found literature are thoroughly explored and subjected to in-depth analysis. This process serves to determine the integrity of the relationship between the engineered configuration and the transmission line, aligning with the primary focus of the review. Furthermore, it centralizes the assessment around the advantages of various Machine Learning techniques in anomaly diagnosis and monitoring schemes in optical fiber networks.

By completing the listed phases—identification, screening, and eligibility—adequate literature is obtained, and searched studies are guaranteed to meet the set requirements in the review. These processes ensure a comprehensive and reliable foundation, guiding the evaluation to draw robust findings and a systematic presentation of data in the field of optical fiber networks.

2.3. Data Extraction

The indexed related literature selected from each database underwent substantial channels—specifically, a two-stage review—aimed at classifying the applied methodologies, paradigms, and the alignment between the objectives of the chosen studies and the scope of the review on optical fiber networks.

The initial assessment of the literature, according to the review criteria, involves the identification, screening procedure, and eligibility check. Then, a random sample of seven ML-based fault detection systems on optical fiber cables was exposed to an extensive examination, with technical documentation being executed to develop an extraction grid, founding the review of the fifteen (15) related systems.

The data extraction recorded the significant output patterns, techniques (Machine Learning), and descriptive statistics (accuracy rates, parameters, and results) demonstrated by the reviewed literature.

1. **Extraction Procedure:** Literature with diverging objectives from the coverage of the review, albeit associated with the terms 'Optical Fiber Networks,' 'Machine Learning,' and 'Fault detection' in the titles, abstracts, and keywords, is rejected; otherwise, assessed with caution.

Afterward, the literature that passes is shifted into the second screening phase, which is determined by the preliminary technical documentation—involvement in the matrix system sorted by relevant withdrawn outputs and figures.

2. **Performance Metrics:** Accuracy rates and valuable numerical records are regarded in the grid.

3. **Limitations:** Registered hindrances and set thresholds within the selected literatures are addressed and interpreted to emphasize the integrity and reliability of the literature to gain insights.

2.4. Analysis of Key Variables in ML-based Fault Detection Systems

Expounding the fundamental variables discovered in the searched literature is a vital part of the review process. This includes the maximized descriptive reports generated through the experimentations and devised ML-based fault detection models, which provide insights into their accuracy rates and characteristics on the optical fiber networks. A systematic and transparent exposition of data is pivotal to meeting the set standards of the review; hence, discussing the analysis must be precise.

The extraction grid for reviewed published articles focuses on the following key factors:

1. **Machine Learning Scheme:** The techniques, configuration models, and administration modes of specific Machine Learning algorithms used on optical fiber transmission lines are analysed based on their results and intended purpose.

2. **Anomaly Localization Technique:** The implementation and localization of defects, damages, and faults using machine learning are described in detail, highlighting the developed system's capabilities.

3. **Performance Evaluation Metrics:** The accuracy rates and obtained numerical figures pertinent to the documentation of the Machine Learning are also analysed to provide a comprehensive understanding of the systems' performance.

3. RESULTS

3.1. Overview of Anomaly Detection and Localization in Optical Fiber

Anomaly detection and localization in optical fiber communication maintains the efficiency, privacy and reliability of communication networks which enable users to recognize, localize, and resolve issues guaranteeing uninterrupted communication services. When physical damage, environmental changes, or intrusion attempts occur in optical fiber systems, anomaly detection detects them. Localization of network anomalies is made possible by wireless sensors and methods like TDR, which enhance security, reliability, and maintenance. This leads to early detection, cost savings, and reliability of the network.

3.2. Challenges Associated with Traditional Techniques

Data accessibility and reliability concerns make machine learning model training difficult, necessitating advancements in physics-informed machine learning and digital twin technology for better failure management. [6] Due to issues with soil, cable installation, and environmental factors, traditional fault prediction algorithms have trouble locating faults and are not scalable since they cannot be made to work in new locations or with different network configurations. [20] Due to their weak algorithms and large amounts of data, traditional perimeter security systems have trouble handling complex sensing events. As a result, there are delays, missed threats, reduced accuracy, and problems scaling as security systems get bigger. [29] Due to problems with manual feature extraction, neural network constraints, excess fitting, delayed processing, and classification, traditional DOVS approaches have difficulty identifying intrusions. [13] Challenges including concentrating on pattern recognition, maximizing sensor performance, and evaluating new sensors arise when multiplexers and machine learning are integrated. They will reach their full potential in a variety of disciplines if they can overcome these barriers [14].

3.3. Machine Learning as a Detection Instrument of Optical Fiber Networks

Machine learning has been progressively adapted for fault tracing in optical fiber networks, predicting its performance for complex network management [20]. The continuous modification of artificial intelligence (AI) in suppressing traditional algorithms from emerging, enabling improved and accurate identification of damages along the transmission line. This operation is associated with training high-dimensional ML models with prior data, recognizing patterns and impairments that indicate serious faults.

Machine learning has several methods to perform fault diagnosis and offers a wide range of algorithms to improve accuracy rate. However, when identifying compound anomalies in optical fiber cables, there are two learning methods provide robust performances, independent from each other: supervised learning and unsupervised learning [23].

1. **Supervised Learning:** A machine learning model can be trained using a cluster of categorized or classified data [12], where the input data is correlated with the output. The machine learns to identify the output for new or unseen inputs.

2. **Unsupervised Learning:** A machine learning model can be trained on unlabelled network data [12], where the input is not linked to the output data. The machine learns to find patterns and relationships out of an abstract data.

Different machine-learning techniques are employed to develop defect detection models for optical fiber transmission lines. Supervised learning algorithms: Support Vector Machine (SVM) [10], decision trees [10], and neural networks are among the prominent techniques valued to categorize anomalies based on labelled data. Unsupervised methods, including Principal Component Analysis (PCA) [2], K-means clustering [2] and autoencoders [27] as well as isolation forest [3], are also a few of the methods utilized to recognize patterns and anomalies in unlabelled data for fault detection without initial support to generate specific outputs.

Supervised learning algorithms with inputs having identified features to be processed:

- Support Vector Machine (SVM): Applied mainly for classification. Its algorithms work on learning to find boundaries [10] between two data points to yield accurate predictions.

- Decision Trees: A classification-type algorithm whose architecture depicts a tree-like structure [10], consisting of decisions and consequences, presents nodes and pathways to process outcome.

- Neural Networks: Integrated with algorithms based on the functions of the human brain. Excels at performing complex executions and nonlinear relationships, particularly notable for classification processes [24].

Unsupervised learning techniques designed for monitoring procedure and localizing discrepancies:

- K-means clustering: A classification-type algorithm that partition a dataset into K distinct–non-overlapping clusters based on the diagnosed characteristic of cluster's centroid [2].

- Principal Component Analysis (PCA): Utilize for dimensionality reduction, involving principal components which uses orthogonal projections. It is functional to diagnose optical fiber networks for anomalies by projecting the data onto the principal components to discern patterns from the origin [1].

- Isolation Forest: An unsupervised machine learning, specifically to localize fault manifestations via generation of decision trees group, attempting to isolate anomalous data points from the rest the data [3].

- Autoencoder: Utilize a neural network architecture to develop a compressed input for reconstruction, which machine learning will harness to learn that can be applied for anomaly detection and localization [27].

3.4. Flow of ML-based Anomaly Localization in Optical Fiber Networks Assessment

Shown in Figure 1 is the PRISMA-based assessment system [17] of the review. The evaluation of selected ML-based systems progressed with 350 studies, with 300 duplicate copies rejected from academic repositories—IEEE Xplore, ResearchGate, and Google Scholar during the preliminary search. From the filtration process, 46 remain due to invalid retrieval of the full text of a few related studies and dissimilar contexts. Then, by excluding 31 papers for having contrasting coverage from the review objectives, 15 legitimate scholarly papers discussing ML-based fault tracing in optical fiber networks are engaged for the final review process.
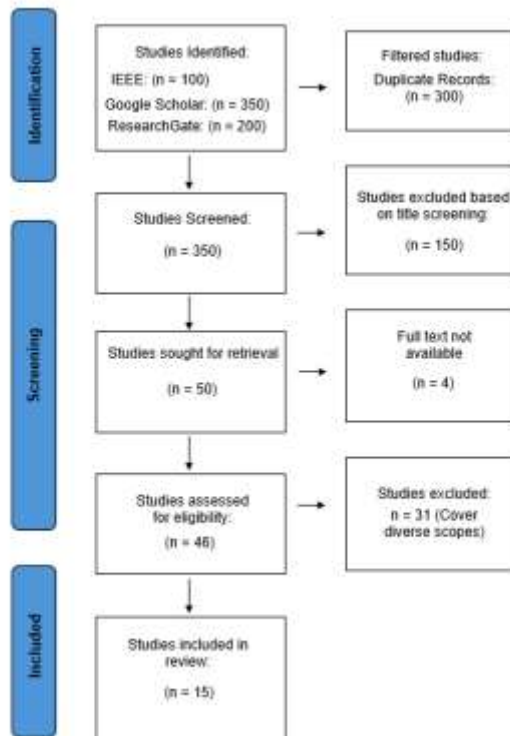
Figure 1. PRISMA Flow Chart of Study Selection

3.5. Review of Integrated Machine Learning-Based Fault Tracing and Localization Models in Optical Fiber Communication Networks.

The following table shows the tabulated extracted data of the sets of literature sourced from academic databases, employing ML-based algorithms for optical fiber networks:

.

Table 1 - **Review of Integrated Machine Learning-Based Fault Tracing and Localization Models in Optical Fiber Communication Networks**

Literature Titles, Leading Authors, and Years	Machine Learning Technique(s)		Model(s)	Anomaly Localization Technique(s)	Application and Finding(s)	Accuracy Rate(s)
	Supervised	Unsupervised				
Application of Neural Network in Fault Location of Optical Transport Network - Liu et al. (2019)	• Back Propagation Neural Network (BPN) • Long Short-Term Memory (LSTM) neural network	N/A	LSTM model	Machine Learning-Based Algorithm	The proposed models used in the article are to apply neural networks in solving problems of fault location in optical communication networks. However, the LTSM model is innovated by using techniques like gradient clipping and weight regularization. LSTM model outperforms the standard BPNN in terms of faster localization time and higher F1-score, meeting the accuracy and real-time requirements for OTN fault location. The developed model shows advantages over traditional methods.	The LSTM model achieved a score of approx. 0.96. While BP neural network has approx. 0.93. Since the literature did not provide specific accurate ratings, the results were based on F1-scores. The LSTM model had a more stable and higher F1-score curve compared to the BP neural network.
A review of machine learning-based failure management in optical networks - Wang et al. (2022)	• Support Vector Machine (SVM) • Decision Tree • Naïve Bayes • Extreme Gradient Boosting (XGBoost) • Artificial Neural Network (ANN)	• Autoencoder (AE) • Gaussian Process • Generative Adversarial Networks (GAN) • Graph Neural Network (GNN)	No Particular Model	• Machine Learning-Based Algorithm • TDR-base Localization	The literature findings related to optical network failure analysis are managed and recorded accordingly. It mentioned investigations on different varieties of machine learning-based algorithms for optical network failure prediction, localization, etc. Included in these are: ANN, SVM, Decision Tree, etc. Experimental procedures	The Accuracy rates were evaluated and registered. Binary-SVM, random forest, multiclass SVM, and single-layer neural networks showed a consistency of 98%. The LSTM-based model's fault mechanism flexed with 93% accuracy,

	• Convolutional Neural Network (CNN) • Long Short-Term Memory (LSTM) • Bayesian Neural Network (BNN) • Recurrent Neural Network (RNN)				were also demonstrated and listed to show highly accurate predictions and classification in optical fiber networks. It showcased the advantages of ML-based algorithms in improving the reliability and efficiency of optical network systems.	outperforming the conventional OTDR analysis techniques. Overall, the cognitive fault management models, which employ ML for autonomous failure detection, achieved superior performances based on the analysis.
Predicting the actual location of faults in underground optical networks using linear regression **- Nyarko-Boateng et al. (2020)**	• Binary support vector • machine (SVM) • random forest • multi-class SVM • Neural Networks (NNs) • Linear Regression	N/A	• Simple Linear Regression model • Single-layer Perceptron Neural Network	Machine Learning-Based Algorithm	The paper proposed an actual fault identifier in underground fiber networks using mainly linear regression and neural network. By utilizing 334 fiber network failures, the study generated models that contribute to reducing failures along the lines and contrasted these ML-based models to discuss the highest efficient model in the repair operations in underground fiber optics networks.	The SLR model showed a high-R-squared value of 97% indicating a good index for the data. However, compared to the SLP neural network model, the results achieved a high accuracy rate better than SLR with 98%, accompanying complex computational resources.

An Optical Communication's Perspective on Machine Learning and Its Applications **- Khan et al. (2019)**	• Artificial Neural Networks (ANN) • Support Vector Machines (SVMs) • Decision Trees (DTs) • Random Forests (RFs) • Gaussian Mixture Models (GMMs) • Convolutional Neural Networks (CNNs) • Recurrent Neural Networks (RNNs)	• K-Means Clustering • Expectation-Maximization (EM) Algorithm • Spectral Clustering • Principal Component Analysis (PCA) • Independent Component Analysis (ICA) • Non-negative Matrix Factorization (NMF)	No Particular Model	Machine Learning-Based Algorithm	The literature discusses the exploration of machine learning (ML) algorithms and their beneficial advantages in the field of optical communications and networking. It found observations that ML techniques can enhance nonlinear transmission systems, optical performance monitoring, etc. Proactive fault detection using ML can significantly improve the performance of optical fiber networks.	The paper provides accuracy ratings of 94.48%, 93.05%, and 95.53% respectively; it displayed the efficiencies and high-profile ratings of ML techniques in monitoring OSNR, CD, DGD, AND MFI in optical networks.
Experimental Study of Machine-Learning-Based Detection and Identification of Physical-Layer Attacks in Optical Networks **- Natalino et al. (2019)**	• Artificial Neural Network (ANN) • Support Vector Machine (SVM) • Gaussian Process (GP) • Decision Tree (DT) • Random Forests (RF) • Naive Bayes (NB) • Quadratic Discriminant Analysis (QDA)	N/A	ALL mentioned Supervised Learning	Machine Learning-Based Algorithm	The primary objectives of the literature are to detect and identify physical-layer attacks in optical networks. The paper generated an Attack Detection Identification (ADI) framework, optimizing ML techniques where the ANN classifier secured the highest classification accuracy rate among the other ML-based classifiers.	ANN achieved 99.9% accuracy on average and had the lowest standard deviation. GP and RF performed well, garnering a high-test accuracy, however, ANN outperformed them. Regardless, the QDA classifier had the lowest classification accuracy.

	• Nearest Neighbors (NN)				
Neural network-based fiber optic cable fault prediction study for power distribution communication network **- Zhang, Yan, et al. (2023)**	• Memory Feature Generating Convolutional Neural Network (MFG-CNN) • Weighted Sequential Pattern Mining Algorithm (DWSPM)	Generative Adversarial Networks (GANs)	Memory Feature Generating Convolutional Neural Network (MFG-CNN)	Machine Learning-Based Algorithm	The literature has developed an effective fault prediction model for fiber optic cables, utilizing enhanced data mining and deep learning techniques to improve the accuracy and efficiency of fault prediction, and demonstrates a practical approach to reducing repair time and improving network reliability.	The average accuracy that MFG-CNN obtained for fault diagnosis method is 98.68%.
Machine Learning Applications in Optical Fiber Sensing: A Research Agenda **- Reyes-Vera et al. (2024)**	• Support Vector Machine (SVM) • Artificial Neural Network (ANN) • Long Short-Term Memory (LSTM) • Convolutional Neural Network (CNN) • Support Vector Regression by Least Squares	• Principal Component Analysis (PCA) • Clustering Algorithms • Self-Organizing Maps (SOM) • Nonlinear Principal Component Analysis	No Particular Model	Machine Learning-Based Algorithm	The main point of the literature is to discuss the variations of machine learning techniques, including Neural Networks (NNs), random forests, Support Vector Machines (SVM), and semi-supervised learning to upgrade the performance, accuracy, and security of fiber optic systems across various applications–structural health monitoring, leak detection, telecommunications, etc.	It highlights the general analysis and high potential of covered machine learning techniques, involving their quality performance in different system domains.
Optical Fiber Distributed Vibration Sensing Using Grayscale Image and Multi-	• Convolutional Neural Network (CNN)	N/A		Machine Learning-Based Algorithm	The developed deep learning model is designed for multi-event recognition in optical fiber. The 2DCNN-LSTM model enables the effective	2DCNN-LSTM hybrid deep learning model demonstrated an accuracy rate of 97.0% on the vibration

Class Deep Learning Framework for Multi-Event Recognition **- Sun et al. (2021)**	• Long Short-Term Memory (LSTM) neural network • SoftMax classifier		2DCNN-LSTM model		recognition and classification of different sensing events in an optical fiber-distributed vibrating sensing system for security applications. The model can extract automatic features without relying on predefined parameters.	pattern recognition task.
Fault Monitoring in Passive Optical Networks using Machine Learning Techniques **- Abdelli et al. (2023)**	Long Short-Term Memory (LSTM)	N/A	LSTM-based Model	• Machine Learning-Based Algorithm • TDR-based Algorithm	The literature suggests two machine learning approaches for fault detection and localization in passive optical networks (PONs). The first approach employs an LSTM architecture to classify and localize reflection and event types in PON through supervised learning. The second method involves an LSTM-based autoencoder for localizing various types of anomalies. The paper provides a detailed analysis of these two techniques, which have shown high levels of accuracy in fault localization.	LSTM-based autoencoder extracted a diagnostic accuracy of 97% while maintaining low prediction errors. However, the LSTM network model classifies different types of reflection with an accuracy test of only 95%, which provides relatively small errors but is not superior to the second method ML-based model.

Machine learning methods for optical communications -Usman, H. M. (2020).	• Support Vector Machine (SVM) • K-Nearest Neighbors (KNN) • Artificial Neural Networks (ANNs) • Deep Neural Networks (DNNs)	N/A	No Particular Model	Machine Learning-Based Algorithm	The literature highlights the categories of applications where machine learning methods have been successfully employed, such as non-linearity mitigation, performance monitoring, network planning, and performance prediction.	The article does not provide specific accuracy data points. However, it presents a comparative evaluation of machine learning techniques such as RL and SVM. These techniques aim to mitigate nonlinear effects in fiber-optic systems and offer a higher degree of accuracy compared to traditional methods.
Deep learning-based fault diagnosis and localization method for fiber optic cables in communication networks - Zhang, Gao, et al. (2023)	Convolutional neural network (CNN)	Generative adversarial network (GAN)	DCGAN-CNN fault diagnosis model	Machine Learning-Based Algorithm	The study intends to test deep learning models to diagnose and localize faults in fiber optic cables in communication networks. The DCGAN-CNN technology can achieve better fault diagnosis with an accuracy rate of 98.5% by utilizing the characteristics of GAN to generate simulation data and the classification ability of CNN.	The DCGAN-CNN achieved 98.5% compared to other methods. The SDGAN-FM utilized a large amount of unlabeled data to complete the diagnosis with an accuracy rate of 91.1%, making the DCGAN-CNN model better as a fault detector overall.
Machine learning framework for timely soft-failure detection and localization in elastic optical networks	Encoder-Decoder Long Short-Term Memory	N/A	Encoder-Decoder Long Short-Term Memory (ED-LSTM) model	Machine Learning-Based Algorithm	The ED-LSTM model can predict hard-failures up to 4 days in advance when modeling soft-failure evolution over 1-2 lightpaths.	The accuracy of the ED-LSTM model varied depending on the number of lightpath sequences. The soft-failure evolution model of 2

- Behera et al. (2023)					The overall framework reduces operational expenses by triggering repair actions only, when necessary, based on the predicted soft-failure evolution, rather than relying on fixed QoT thresholds	lightpaths achieves an accuracy of 4.5x10^7. It was identified as the most effective approach.
Pattern Recognition for Distributed Optical Fiber Vibration Sensing: A Review **- Li et al. (2021)**	- Support Vector Machine (SVM) - Artificial Neural Network (ANN) - Deep Learning (including Convolutional Neural Network, Convolutional Long Short-Term Memory Network) - Relevant Vector Machine (RVM) - Linear Discriminant Analysis (LDA) - Gaussian Mixture Model (GMM) - Random Forest (RF) - Echo State Network (ESN)	Sparse auto-encoders algorithm (deep learning)	Models and Algorithms used in DOVS systems: - Support Vector Machine (SVM) - Relevance Vector Machine (RVM) - Linear Discriminant Analysis (LDA) - Gaussian Mixture Model (GMM)	Machine Learning-Based Algorithm	The article provides a performance comparison of different pattern recognition methods applied to DOVS applications. It shows that techniques like SVM, RVM, and deep learning can manage to score over 90% in defining types of intrusions/threats, leaks/etc.	The CNN model: 90%. GMM: 97.67%. ESN: 98.75%. Random Forest Classifier: 96.58%. CLDNN: 97%. Hierarchical Convolutional LSTM: 90%. The overall accuracy rate report ranges from around 85% to 97%, demonstrating high performance.
Machine Learning-Aided Optical Performance Monitoring Techniques: A Review	- Support Vector Machine (SVM) - K-Nearest Neighbors (KNN)	K-Means Clustering	No Particular Model	Machine Learning-Based Algorithm	The literature explored the works of the diverse range of ML models in indexing cost-effective, real-time, and multi-impairment monitoring tools in optical communication	It recorded correlation coefficients ranging from 0.91 – 0.99. For other studies, the literature noted accuracy rates, scoring

- Tizikara et al. (2022)	• Decision Tree • Artificial Neural Network (ANN)	• Principal Component Analysis (PCA)			networks. It assessed the previous observations of ML algorithms in fault management in optical fiber networks and established generalizations on their high-performing aspects.	95% in simulation and 60% in experimental procedures. The results demonstrate that ML techniques for simultaneous monitoring of multiple physical layer impairment in optical networks are incomparable to traditional techniques.
Machine Learning-Based Anomaly Detection in Optical Fiber Monitoring - Abdelli et al. (2022b)	• Attention-based bidirectional gated recurrent unit (A-BiGRU) • BiLSTM, • BiLSTM-CNN, • BiGRU,	• Autoencoder • GRU-AE-BiGRU	A-BiGRU model	Machine Learning-Based Algorithm	Autoencoder is applied to quickly detect any anomalies or faults in the optical fiber, such as fiber cuts and optical eavesdropping attacks, while Attention-based BiGRU is utilized to diagnose the type of detected fiber fault (e.g. fiber cut, eavesdropping) and localize the fault position once an anomaly is detected by the autoencoder. The integrated approach combining the autoencoder and BiGRU models outperformed standalone BiGRU models, demonstrating the benefits of the two-stage framework.	Anomaly Detection Model (GRU-AE) for the optimal threshold of 0.008, the precision, recall, and F1 scores are around 96.9%, indicating excellent separability between normal and faulty classes. However, A-BiGRU achieves over 97% accuracy in diagnosing fault types. The accuracy increases with higher SNR, reaching close to 100%.

4. DISCUSSION

4.1 Table Analysis

The literature reviewed discusses the utilization of machine learning-based algorithms in detecting and localizing faults in optical fiber networks for security and maintenance applications. As shown in Table 1, ML models for diagnostic testing, including their performance metrics, applications, and results, are covered to provide systematic descriptions of how effective these frameworks are in the field of optical fiber networks. The machine learning techniques are separated into two categories, specifying the methods used in each study to address challenges within the networks: unsupervised learning and supervised learning. The included models are evaluated based on the assessment procedures of each article, highlighting those with the best performance in detection and localization. Applications and outcomes of these machine learning models are also evaluated based on predictions, accuracy rates, and overall tracing abilities regarding leaks, anomalies, or irregularities in optical networks. Lastly, only notable accuracy levels are included, and these are verified with caution to ensure the authenticity of the inputs recorded in the literature papers.

4.2 Outstanding Machine Learning in Optical Fiber Network for fault Diagnosis and Localization

Several evaluated machine learning techniques for optical fiber sensing revealed promising skills and excellent progression from fault analysis, establishing solutions in the management of the network's operation system. However, from the provided studies, the most capable machine learning models include Support Vector Machines (SVMs), Neural Networks (NNs), specifically Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM), Random Forest (RF), Decision Tree, Gaussian Mixture Model (GMM), and Autoencoders. Others, like K-Nearest Neighbors (KNN), Principal Component Analysis (PCA), and Linear Discriminant Analysis (LDA), have also been discovered to outperform traditional methodologies in detecting and localizing faults in the line. The performance metrics of these machine learning algorithms are mentioned as possessing high-level scores ranging from 86% to 98% on average, indicating positive remarks for their advantages and beneficial applications in the preservation of optical fiber networks.

4.3 Accuracy Metrics of Machine Learning Algorithms

Machine learning technologies in the review, surpassing the capabilities embedded within traditional applications have been assessed and found to produce sharp recognition programs. These techniques have achieved highly documented accuracy values and low error predictions, demonstrating robust algorithms and frameworks suitable for the field. Accuracy percentages ranging from 90% to 98% have been observed and verified.

Some of the acknowledged ML-based models from the review with high accuracy include Convolutional Neural Networks (CNNs), which have proven highly effective in localizing and tracing faults in optical fiber networks, achieving diagnostic accuracy rates of up to 96-97% compared to Time Domain Reflectometry (TDR) methodologies. Support Vector Machines (SVMs) achieved a score of 98%, autoencoders exhibited an accuracy rate that can increase up to 97.61%, LSTM-based models garnered an accuracy rating of 95.53%, and GMM-based models demonstrated an overall accuracy of 97.67%.

However, despite the mentioned accuracy rates, the potential and effectiveness of machine learning techniques in achieving 90% ratings are specific to the studies conducted. These may vary based on different variables. Hence, considering the type of sensor, data, and error being analysed is valuable in obtaining an average performance rating above 90%.

4.4 Challenges and Limitations

While studies mentioned observable high accuracy rates and scores for machine learning models in fault management and monitoring in optical fiber networks, the existing challenges and issues associated with the creation of these applications still prevail. Some of these challenges significantly affect the performance and the extent to which these machine learning models can be applied. Nonetheless, it is worth noting the following issues that are inevitable to consider:

1. **Limited Data Availability**- Machine learning algorithms naturally require large amounts of high-quality data to achieve the highest degree of accuracy. However, data accumulation in optical networks is complex, and preparations to create a complete visual of the system for this type of line consume time. Hence, the training of data is limited.

2. **Model Complexity** – Some machine learning frameworks are superior in design and require advanced computations, which take time to train. Implementation in real-time or resource-constrained environments is difficult and limited.

3. **Heterogenous and Dynamic Data**- Optical networks mostly produce large volumes of heterogeneous and dynamic data, which deeply affects the structural composition of machine learning models. The data can be influenced by various factors such as signal noise, fiber attenuation, and environmental factors which vary frequently, making the prediction and operation hard due to these diverse behavioral activities within the networks.

5. CONCLUSION

The literature review extracted different machine learning-based algorithms for fault tracing and localizing in optical fiber networks. These machine learning techniques are categorized into two main types for coherent analysis: unsupervised and supervised learning. Each literature paper discusses the different machine learning models, employing a diverse scope of algorithms and schemes in figuring out solutions to the challenges and problems such as faults, leakages, and security issues within the network's domain and operation. Promising ML models are highlighted and briefly evaluated. Machine learning such as SVM, CNN, LSTM, RF, Decision Tree, GMM, and AE are a few of the models that outperformed the traditional technologies in optical networks for anomaly diagnosis and localization, obtaining accuracy rates ranging from 86%-98%. Also, the review provides analysis and systematic exposition display of the results of these models. Performance metrics of the models are addressed and verified, as well as their limitations, applications, and challenges encountered in real-world applications. Overall, the review provides a comprehensive synthesis of 15 distinct academic papers about ML-based strategies applicable to the operation of optical fiber communication networks.

References

[1] A. Biswal, "What is Principal Component Analysis?," *Simplilearn.com*, Nov. 07, 2023. https://www.simplilearn.com/tutorials/machine-learning-tutorial/principal-component-analysis

[2] A. Khalfe, "Unsupervised machine learning: Clustering, dimensionality reduction, and anomaly detection techniques. - The Talent500 Blog," *The Talent500 Blog*, Aug. 04, 2023. https://talent500.co/blog/unsupervised-machine-learning-clustering-dimensionality-reduction-and-anomaly-detection-techniques/

[3] C. Maklin, "Isolation Forest - Cory Maklin - Medium," *Medium*, Jul. 15, 2022. [Online]. Available: https://medium.com/@corymaklin/isolation-forest-799fceacdda4

[4] C. Natalino, M. Schiano, A. Di Giglio, L. Wosinska, and M. Furdek, "Experimental study of Machine-Learning-Based detection and identification of Physical-Layer Attacks in optical networks," *Journal of Lightwave Technology*, vol. 37, no. 16, pp. 4173–4182, Aug. 2019, doi: 10.1109/jlt.2019.2923558.

[5] D. K. Tizikara, J. Serugunda, and A. Katumba, "Machine Learning-Aided Optical Performance Monitoring Techniques: A review," *Froxtiers in Communications and Networks*, vol. 2, Jan. 2022, doi: 10.3389/frcmn.2021.756513.

[6] D. Wang, C. Zhang, W. Chen, H. Yang, M. Zhang, and A. P. T. Lau, "A review of machine learning-based failure management in optical networks," *Science China. Information Sciences/Science China. Information Sciences*, vol. 65, no. 11, Oct. 2022, doi: 10.1007/s11432-022-3557-9.

[7] E. Reyes-Vera, A. Valencia-Arías, V. G. Pineda, E. F. A. Vigo, H. Á. Vásquez, and G. Sánchez, "Machine Learning Applications in Optical Fiber Sensing: a research agenda," *Sensors*, vol. 24, no. 7, p. 2200, Mar. 2024, doi: 10.3390/s24072200.

[8] F. N. Khan, Q. Fan, C. Lu, and A. P. T. Lau, "An Optical Communication's perspective on machine learning and its applications," *Journal of Lightwave Technology*, vol. 37, no. 2, pp. 493–516, Jan. 2019, doi: 10.1109/jlt.2019.2897313.

[9] H. G. Çitil, "Important notes for a fuzzy boundary value problem," *Applied Mathematics and Nonlinear Sciences*, vol. 4, no. 2, pp. 305–314, Jul. 2019, doi: 10.2478/amns.2019.2.00027.

[10] H. H. Nguyen, "A complete view of decision trees and SVM in machine learning," *Medium*, Dec. 07, 2021. [Online]. Available: https://towardsdatascience.com/a-complete-view-of-decision-trees-and-svm-in-machine-learning-f9f3d19a337b

[11] H. M. Usman, "Machine learning methods for optical communications," *Trends in Computer Science and Information Technology*, pp. 055–057, Sep. 2020, doi: 10.17352/tcsit.000023.

[12] J. Delua, "Supervised vs. Unsupervised Learning: What's the Difference? - IBM Blog," *IBM Blog*, Mar. 12, 2021. https://www.ibm.com/blog/supervised-vs-unsupervised-learning/

[13] J. Li *et al.*, "Pattern Recognition for Distributed Optical Fiber Vibration Sensing: A review," *IEEE Sensors Journal*, vol. 21, no. 10, pp. 11983–11998, May 2021, doi: 10.1109/jsen.2021.3066037.

[14] J. Yang, S. Li, Z. Wang, and G. Yang, "Real-Time tiny part defect detection system in manufacturing using deep learning," *IEEE Access*, vol. 7, pp. 89278–89291, Jan. 2019, doi: 10.1109/access.2019.2925561.

[15] K. Abdelli, C. Tropschug, H. Grießer, and S. Pachnicke, *Fault Monitoring in Passive Optical Networks using Machine Learning Techniques*. 2023. doi: 10.1109/icton59386.2023.10207489.

[16] K. Abdelli, J. Y. Cho, F. Azendorf, H. Grießer, C. Tropschug, and S. Pachnicke, "Machine-learning-based anomaly detection in optical fiber monitoring," *Journal of Optical Communications and Networking*, vol. 14, no. 5, p. 365, Apr. 2022, doi: 10.1364/jocn.451289.

[17] K. Kolasa, B. Admassu, M. Hołownia-Voloskova, K. Kędzior, J. Poirrier, and S. Perni, "Systematic reviews of machine learning in healthcare: a literature review," *Expert Review of Pharmacoeconomics & Outcomes Research*, vol. 24, no. 1, pp. 63–115, Nov. 2023, doi: 10.1080/14737167.2023.2279107.

[18] L. Zhang, W. Gao, and L. Yan, "Deep learning-based fault diagnosis and localization method for fiber optic cables in communication networks," *Applied Mathematics and Nonlinear Sciences*, vol. 9, no. 1, Aug. 2023, doi: 10.2478/amns.2023.2.00241.

[19] L. Zhang, L. Yan, W. Shen, F. Li, J. Wu, and W. Liang, "Neural network-based fiber optic cable fault prediction study for power distribution communication network," *Applied Mathematics and Nonlinear Sciences*, vol. 9, no. 1, Nov. 2023, doi: 10.2478/amns.2023.2.01278.

[20] O. Nyarko-Boateng, A. F. Adekoya, and B. A. Weyori, "Predicting the actual location of faults in underground optical networks using linear regression," *Engineering Reports*, vol. 3, no. 3, Nov. 2020, doi: 10.1002/eng2.12304.

[21] R. Y, "Optical fiber," *Circuit Globe*, Jan. 25, 2023. https://circuitglobe.com/optical-fiber.html

[22] S. Behera, T. Panayiotou, and G. Ellinas, "Machine learning framework for timely soft-failure detection and localization in elastic optical networks," *Journal of Optical Communications and Networking*, vol. 15, no. 10, p. E74, Sep. 2023, doi: 10.1364/jocn.490008.

[23] Seldon, "Supervised vs Unsupervised Learning Explained," *Seldon*, Jan. 23, 2024. https://www.seldon.io/supervised-vs-unsupervised-learning-explained

[24] "SVM vs Neural network," *Baeldung*, Mar. 18, 2024. https://www.baeldung.com/cs/svm-vs-neural-network

[25] T. Agarwal, "Optical Fiber : working principle, types, advantages and disadvantages," *ElProCus - Electronic Projects for Engineering Students*, Aug. 26, 2019. https://www.elprocus.com/optical-fiber-working-and-its-applications/

[26] T. Liu, H. Mei, Q. Sun, and H. Zhou, "Application of neural network in fault location of optical transport network," *China Communications*, vol. 16, no. 10, pp. 214–225, Oct. 2019, doi: 10.23919/jcc.2019.10.014.

[27] T. L. Singal, *Optical Fiber Communications: Principles and applications*. 2017. [Online]. Available: https://www.amazon.com/Optical-Fiber-Communications-Principles-Applications/dp/1316610047

[28] "What is an autoencoder? | IBM." https://www.ibm.com/topics/autoencoder

[29] Z. Sun *et al.*, "Optical fiber distributed vibration sensing using Grayscale image and Multi-Class Deep Learning framework for Multi-Event recognition," *IEEE Sensors Journal*, vol. 21, no. 17, pp. 19112–19120, Sep. 2021, doi: 10.1109/jsen.2021.3089004.