ISSN: 3067-8358
Research Article
Journal of Global Perspectives on Society, Culture, and Development

Intelligent Fault Detection in GPON Networks Using K-NN Algorithm

Zegdou Khaoula[1*] & Ahmed Garadi[2]

[1]*Department of Electronics, Faculty of Technology, University of Saida Dr Moulay Tahar, Algeria*
[2]*Laboratory Technology of Communications, University of Saida Dr Moulay Tahar, Algeria*

[*]**Corresponding author:** Zegdou Khaoula, Department of Electronics, Faculty of Technology, University of Saida Dr Moulay Tahar, Algeria

Submitted: 31 May 2025 **Accepted:** 06 June 2025 **Published:** 13 June 2025

https://doi.org/10.63620/MKJGPSCD.2025.1015

Citation: *Zegdou, K., Garadi, A. (2025). Intelligent Fault Detection in GPON Networks Using K-NN Algorithm. J Glob Perspect Soc Cult Dev, 1(2), 01-08.*

Abstract
This research aims to improve fault detection and classification in Gigabit Passive Optical Networks (GPON) by utilizing machine learning, focusing on the K-Nearest Neighbors (K-NN) algorithm. The GPON network is extensively simulated using OptiSystem, where essential performance metrics—Optical Power (dBm), Bit Error Rate (BER), and Signal-to-Noise Ratio (SNR)—are analyzed under various fault and interference scenarios. The collected data undergoes preprocessing and normalization before classification with the K-NN algorithm implemented in MATLAB, using the Euclidean distance metric to measure similarity.

Classification results evaluated via confusion matrices show accuracy rates between 63.16% and 75.00% across different Optical Network Units (ONUs). ONU 2 and ONU 8 achieved the highest accuracies of 75.00% and 72.73%, respectively, while ONU 1 and ONU 7 recorded lower accuracies of 63.64% and 63.16%. Additionally, a detailed analysis of fiber attenuation effects on BER reveals significant signal degradation with increased attenuation. This effect is notably more severe in the segment between the splitter and ONUs compared to the path from the Optical Line Terminal (OLT) to the splitter.

These findings highlight the effectiveness of K-NN-based fault diagnosis systems in automating detection and enhancing GPON reliability, thus reducing downtime and operational costs. Future work may explore more advanced machine learning classifiers, improved feature selection, and real-time monitoring techniques to boost detection accuracy and network resilience.

Keywords: G-PONm, K Nearest Neighbor algorithm (K-NN), BER, SNR, ML

Introduction

The ever-increasing demand for high-speed, reliable, and scalable internet access has positioned fiber-optic communication as a fundamental pillar of modern digital infrastructure. Optical systems, leveraging the properties of light for data transmission, have enabled transformative progress in fields such as telecommunications, medical imaging, industrial automation, and scientific instrumentation. Among the leading fiber-based access technologies, Gigabit Passive Optical Networks (GPON) have emerged as a key solution to meet broadband needs for both residential and enterprise environments.

GPON networks utilize a point-to-multipoint topology enabled by passive optical components, allowing a single optical fiber to be shared among multiple users without the need for active elements in the distribution segment. This design reduces operational costs while ensuring high throughput and service quality. Standard GPON implementations operate at downstream rates of 2.5 Gbps and upstream rates of 1.25 Gbps, employing time-sharing protocols such as TDM and TDMA for efficient bandwidth management. However, despite their robustness, GPON systems are not immune to performance degradation resulting from fiber impairments, component aging, misconfiguration, and environmental factors.

Traditional diagnostic techniques, such as Optical Time-Domain Reflectometry (OTDR), provide vital tools for fault localization. Yet, they often require manual interpretation, lack predictive capability, and may not scale effectively in large, dynamic networks. With the growing complexity of optical infrastructure, there is a critical need for intelligent monitoring and autonomous fault management.

Recent advances in Machine Learning (ML) offer powerful frameworks for enhancing the resilience of optical access networks. By mining data generated from performance monitoring systems, ML algorithms can detect patterns, classify abnormal behaviors, and forecast potential failures. This enables a shift from reactive maintenance toward proactive and even preventive network operation. Techniques such as supervised classification, anomaly detection, clustering, and deep learning have been successfully applied to identify signal impairments, optimize resource allocation, and support self-healing mechanisms in optical networks.

In this context, the K-Nearest Neighbors (K-NN) algorithm is investigated for its effectiveness in GPON fault analysis. K-NN is a non-parametric, instance-based learning method known for its simplicity and versatility. It can be employed to classify various types of optical faults, assess signal degradation, and assist in predictive maintenance workflows by comparing real-time network metrics against historical data patterns. Its ease of implementation and adaptability to diverse data distributions make it an appealing choice for real-world deployment in optical systems.

This work proposes a methodology that integrates K-NN into GPON monitoring systems for fault detection and performance assessment. Section 2 introduces the architecture of GPON, emphasizing upstream and downstream transmission dynamics, along with a mathematical formulation of key metrics such as Signal-to-Noise Ratio (SNR) and Bit Error Rate (BER). Section 3 presents the theoretical basis and application of the K-NN algorithm in the context of network fault classification. Section 4 details the experimental framework, including simulations conducted using MATLAB and OptiSystem platforms, and discusses the performance results. Section 5 concludes with an evaluation of the proposed method's effectiveness and outlines future directions for research in intelligent optical network management.

System Model and Analysis

An optical network typically comprises three core components: the transmitter, the transmission medium, and the receiver. The transmitter section includes a light source—such as a laser diode or LED—along with a driving circuit that modulates electrical signals into optical ones. The transmission medium mainly consists of optical fiber and may also include additional components like regenerators, splitters, couplers, multiplexers, and connectors, which support signal propagation and enable flexible distribution. At the receiving end, a photodetector—commonly a PIN or avalanche photodiode—converts the incoming optical signal back into an electrical form. This signal is then amplified and processed by a receiver circuit to recover the transmitted data accuratel

Optical Transmission in GPON-Based Access Networks

In fiber-optic access networks—particularly within the "last mile" segment—the infrastructure extends from the service provider's central office to end-user locations such as residences, commercial buildings, and multi-dwelling units. This segment typically adopts a Passive Optical Network (PON) architecture, which utilizes passive optical splitters to distribute signals through a point-to-multipoint tree topology. Such a design enables efficient and cost-effective sharing of optical fiber among multiple subscribers.

Gigabit Passive Optical Network (GPON), standardized under ITU-T G.984, is among the most widely deployed PON technologies. It offers high-capacity broadband access, delivering downstream speeds of up to 2.5 Gbps and upstream speeds of up to 1.25 Gbps. These capabilities support bandwidth-demanding services such as high-definition video streaming, Voice over IP (VoIP), and high-speed internet access.

GPON uses Time Division Multiplexing (TDM) for downstream communication from the Optical Line Terminal (OLT) to Optical Network Units (ONUs), while upstream transmission relies on Time Division Multiple Access (TDMA). This arrangement ensures efficient and fair bandwidth allocation among users. Furthermore, Forward Error Correction (FEC) techniques are incorporated to enhance signal integrity and counteract common transmission impairments such as attenuation and chromatic dispersion.

Thanks to its scalability, reliability, and cost-effectiveness, GPON has become the preferred technology for modern Fiber-to-the-Home (FTTH) and Fiber-to-the-Business (FTTB) deployments. An overview of the general architecture of a GPON system is illustrated in Figure 1.

Signal attenuation in GPON systems can be described using the general optical fiber loss equation:

$$P_{out} = P_{in}\, 10^{-\alpha L/10} \qquad (1)$$

Here, Pout denotes the output power, Pin the input power, α the fiber attenuation coefficient measured in decibels per kilometer (dB/km), and L the fiber length in kilometers (km)

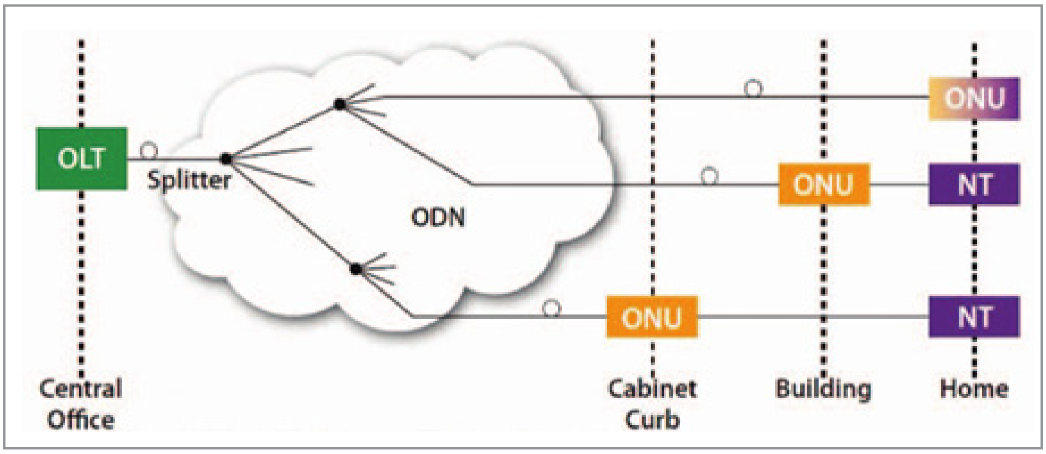

Figure 1: The general structure of a GPON network

Fault detection in Gigabit Passive Optical Networks (GPON)-based Fiber-To-The-Home (FTTH) systems presents a significant challenge due to the absence of physical-layer monitoring at the individual user level. GPON utilizes a point-to-multipoint topology, where numerous Optical Network Units (ONUs) share a single Optical Line Terminal (OLT) through passive splitters. This shared medium limits the operator's ability to isolate faults affecting specific users, unlike active optical networks where each user benefits from a dedicated fiber link and full signal visibility. Conventional diagnostic tools such as Optical Time-Domain Reflectometers (OTDRs) are commonly employed to locate fiber faults; however, their effectiveness diminishes in the shared distribution segment beyond the splitter. Additionally, faults caused by fiber degradation, contaminated connectors, or rogue ONUs inducing signal interference often require labor-intensive field inspections, which escalate operational costs and prolong service outages.

To mitigate these limitations, machine learning offers a promising approach for intelligent, automated fault detection in GPON-based FTTH environments. By monitoring and analyzing key network performance indicators—such as received optical power, bit error rate, latency, and signal-to-noise ratio—data-driven models can detect and classify anomalies with high precision. In this study, we adopt the K-Nearest Neighbors (K-NN) algorithm, leveraging supervised learning techniques trained on historical fault datasets (e.g., fiber cuts, signal degradation, ONU misbehavior). This model enables real-time pattern recognition by comparing live measurements against known fault signatures, thereby facilitating proactive fault localization. The proposed method significantly improves diagnostic accuracy, reduces the reliance on manual intervention, supports timely alerts, and contributes to lower operational expenditures and enhanced network reliability [10]

B.K-Nearest Neighbors (K-NN) Algorithm

The K-Nearest Neighbors (K-NN) algorithm is a widely used supervised machine learning technique applicable to both classification and regression tasks. As a lazy learning method, K-NN does not involve a separate training phase; instead, it retains the entire training dataset. When a new input instance is introduced, the algorithm identifies the K most similar samples—referred to as the nearest neighbors—based on a predefined distance metric, commonly the Euclidean distance.

For classification problems, the algorithm assigns the input to the most frequent class label among its K nearest neighbors. In regression contexts, it returns the average of the output values corresponding to those neighbors. K-NN's simplicity, interpretability, and effectiveness in handling non-linear decision boundaries make it a popular choice in various practical applications [4-5]

Formally, given a training dataset {(x_n,y_n)}_(n=1)^N, the algorithm computes the distance between the test sample and each training point, selects the K nearest samples, and makes a prediction accordingly. Since K-NN is non-parametric and instance-based, its performance depends heavily on the choice of K and the structure of the feature space. The Euclidean distance, often used to measure similarity, is calculated as shown in Equation (2) :

$$\text{distance}\ (x_0, x_n) = \sqrt{(x_0 - x_n)^T (x_0 - x_n)} \tag{2}$$

The query point x is classified based on the majority vote among its k nearest neighbors, as shown in Equation (3) :

$$y_0 = \arg_y \ \max \sum_{(x_n,\, y_n)} - \delta(y = y_n) \tag{3}$$

Where: y_0 represents the class label, and y_i corresponds to the class label of the i-th nearest neighbour. The indicator function . δ(y=y_n), returns a value of one if the class y_i of the neighbour x_i matches the class y_0 , and zero otherwise.

G-PON systems are highly recommended because of their straightforward design and strong performance in managing nonlinear data distributions

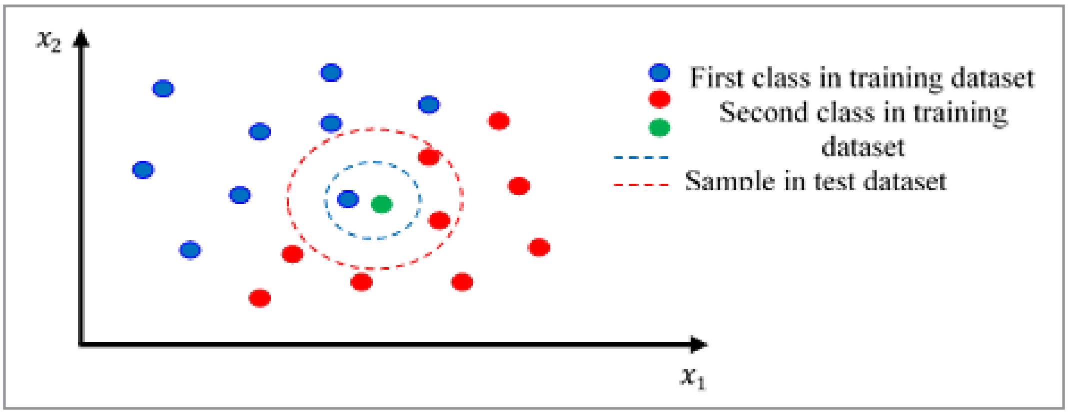

Figure 2: The K-NN algorithm when K=1 and K=3

G-PON System Model Enhanced with K-NN for Fault Detection

Figure 3 illustrates a practical model of a Gigabit Passive Optical Network (G-PON) architecture enhanced with machine learning for proactive fault detection. By integrating the K-Nearest Neighbors (K-NN) classification algorithm, the system effectively identifies signal degradation and predicts potential optical fiber failures. This integration significantly enhances network reliability and optimizes maintenance operations.

The model consists of the following key components

1. **Optical Line Terminal (OLT):** Located at the leftmost part of the architecture, the OLT serves as the central transmission unit, distributing optical signals to multiple Optical Network Units (ONUs).
2. **Power Splitter:** A cascaded 1×2 splitter divides the optical signal into multiple branches, each directed to a separate ONU. While essential for signal distribution,the splitter inherently introduces attenuation, closely simulating real-world operational conditions.
3. **Optical Fiber Links:** Each ONU connects through a dedicated fiber segment ranging from 0.028 km to 0.05 km. These variations reflect realistic deployment scenarios where distance-dependent losses impact signal integrity. Additionally, the total length between the OLT and the power splitter is fixed at 25 km, aligning with practical G-PON implementations

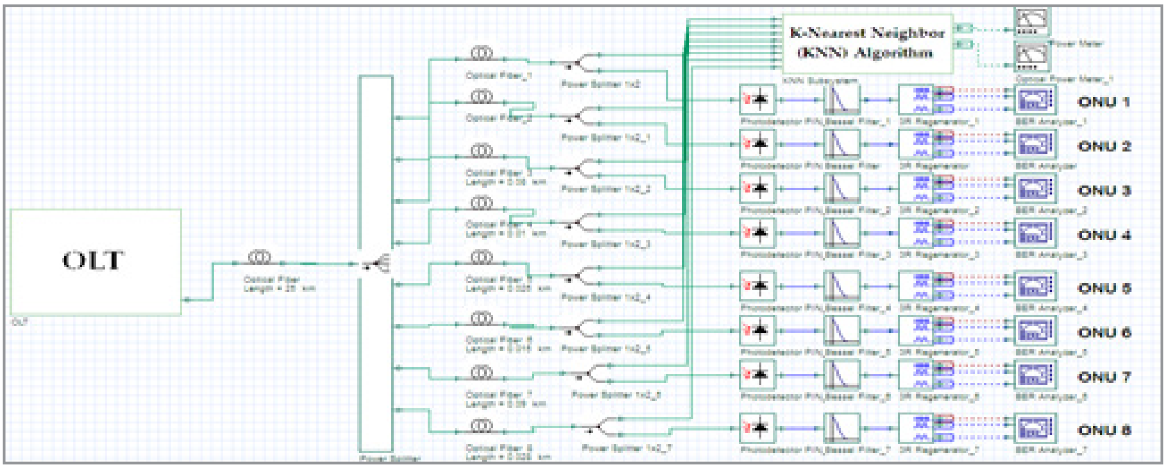

Figure 3: Illustration of the G-PON with K-NN model

The G-PON system integrated with the K-NN algorithm is examined through its classification scheme and performance outcomes. Initially, the input parameters for K-NN, which are the received power values, undergo data preprocessing before applying the machine learning algorithm. The labeled output data (targets) consist of eight classes: ONU1, ONU2, ..., ONU8, each representing different attenuation conditions. Afterward, the K-NN algorithm is applied and evaluated using confusion matrices to determine the accuracy of optical fiber distortion prediction for each user condition.

Prior to applying the K-NN algorithm, data extraction and normalization are essential steps. Data normalization addresses issues related to high variance and outlier detection [3]. In this context, two parameters are extracted from Figure 3: the received power for each individual ONU and the aggregate received power across all ONUs. These parameters are normalized using the Z-score method.

Normalization resolves these issues by producing transformed values that maintain the original data distribution while scaling all values consistently across the dataset. The Z-score is defined as follows:

$$Z_n = \frac{x_n - \overline{Z}}{\sigma_Z} \quad (4)$$

Where Z is the mean deviation and σ_Z is the standard deviation of each feature n.

In this study, since the dataset includes predefined classes corresponding to eight known users, supervised learning techniques are applied. Among various machine learning classifiers, both the K-Nearest Neighbors (K-NN) algorithm and Support Vector Machine (SVM) are selected due to their proven effectiveness in optical communication classification tasks [5]. These classifiers were chosen for their simplicity, computational efficiency, and robustness. Furthermore, their implementation is straightforward and accessible, particularly using software platforms such as MATLAB

Results and Discussion

An effective method for simulating fiber damage involves increasing the attenuation coefficient at specific locations or segments of the optical fiber. This technique allows the modeling of both localized and distributed types of degradation. Localized defects are simulated by applying a sharp attenuation spike at a defined point along the fiber, emulating faults such as microbends, cracks, or splice losses. In contrast, distributed damage is modeled by gradually increasing the attenuation coefficient over a certain length of the fiber, reflecting broader physical stress, bending, or environmental aging, which collectively impact a larger region of the fiber. These impairments lead to signal degradation, reduced power levels, and a decline in the signal-to-noise ratio (SNR). In this study, we concentrate on the analysis of distributed damage, aiming to assess its impact on system performance. The simulation results, obtained through OptiSystem—a comprehensive optical system design software—are presented in the first section, based on the configuration shown in Figure 3.

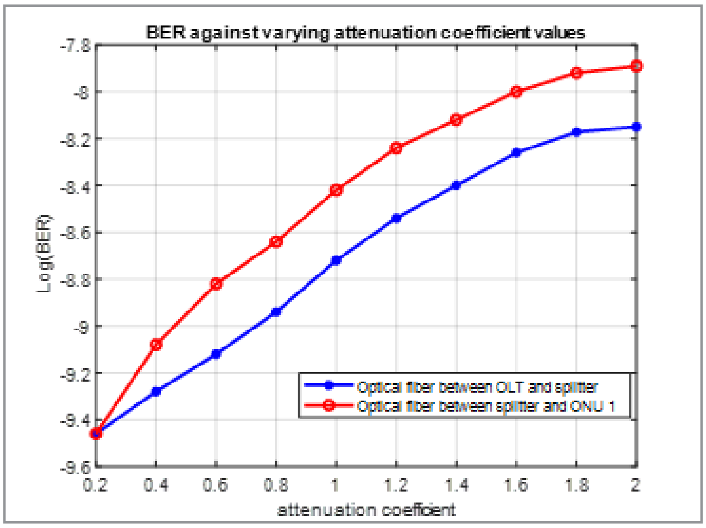

Figure 4. Effect of different optical fiber damage positions under varying attenuation coefficient values

Figure 4 illustrates the effect of fiber damage at two critical segments of the G-PON infrastructure—between the Optical Line Terminal (OLT) and the splitter, and between the splitter and the first Optical Network Unit (ONU)—on system performance when employing the K-Nearest Neighbors (K-NN) classification algorithm. The system's Bit Error Rate (BER) is used as the primary performance metric.

The results demonstrate that acceptable BER levels can be maintained when the attenuation coefficient remains below 0.5 dB/km for the fiber segment between the OLT and splitter, and below 0.8 dB/km for the segment between the splitter and the first ONU.

Moreover, the graph reveals a clear trend: BER deteriorates with increasing attenuation in both segments. Notably, the fiber segment connecting the splitter to ONU 1 (represented by the red curve) exhibits a higher BER compared to the segment between the OLT and splitter (blue curve). This discrepancy is attributed to additional power losses introduced by the optical splitter as well as longer fiber lengths, which collectively contribute to greater signal degradation on the ONU side

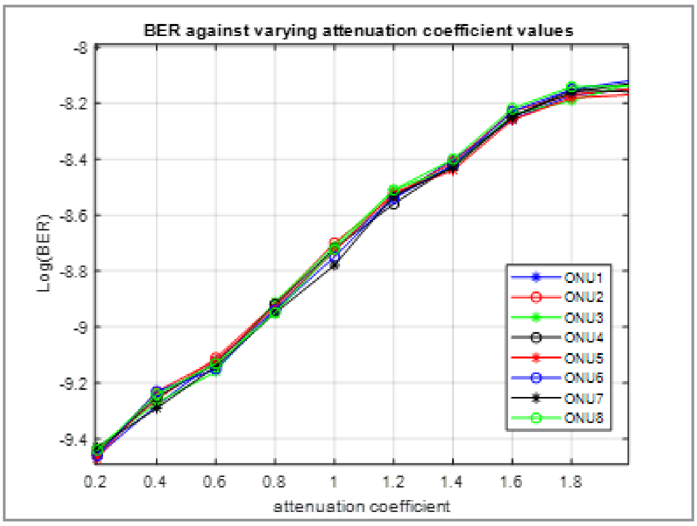

Figure 5: Effect of different optical fiber damage users under varying attenuation values

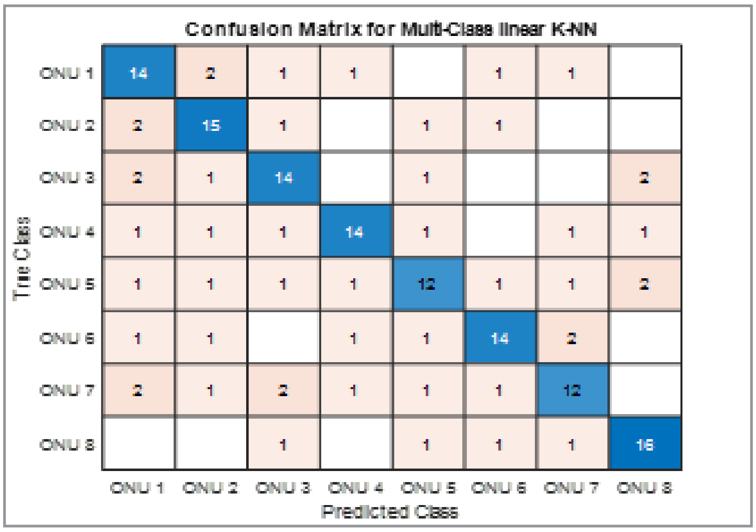

Figure 6: Confusion matrices of K-NN classifier for optical fiber damage prediction

Figure 5 presents comparable BER values for different ONUs under varying attenuation coefficients, with slight variations resulting from the differing distances between the splitter and each ONU. This observation highlights the significant effect of attenuation on the Bit Error Rate (BER) within a multi-ONU optical communication system. As the attenuation increases, the BER correspondingly degrades across all ONUs, following a nearly identical pattern. This consistency indicates that the primary cause of signal degradation is attenuation rather than individual differences among the ONUs.

The confusion matrix further confirms the effectiveness of the classification model, showing accurate identification for the majority of ONUs. Misclassification errors mainly occur between neighboring ONUs, which can be attributed to the similarity in their signal features. Notably, ONU 8 achieved perfect classification with 16 correct identifications out of 20 samples, while other ONUs exhibited minor classification inaccuracies. Overall, the model demonstrates robust performance, although there is potential for improvement through enhanced feature selection or the adoption of more sophisticated machine learning algorithms to minimize misclassification.

$$\text{Accuracy}_{\text{ONU}_i} = \frac{\text{True Positives (TP)}}{\text{Total Samples for ONU}_i} \tag{5}$$

Where:

- True Positives (TP): Represented by the diagonal elements of the confusion matrix, indicating correct classifications.
- Total Samples for ONU: The sum of all entries in the respective row, representing the total number of classification attempts for that ONU.

The classification performance across the eight ONUs demonstrates generally good accuracy, ranging from 63.16% to 75.00%. ONU 2 achieved the highest accuracy (75.00%), followed closely by ONU 8 (72.73%), suggesting that these ONUs exhibit more distinct signal features, making them easier for the model to identify. In contrast, ONU 1 (63.64%) and ONU 7 (63.16%) showed the lowest classification accuracy, indicating a higher rate of confusion with other ONUs. The remaining ONUs recorded intermediate accuracies between 66.67% and 70.00%, reflecting moderate classification reliability.

The confusion matrix reveals a noticeable degree of misclassification, likely due to overlapping feature distributions or signal similarities among neighboring ONUs. Despite this, the model demonstrates reasonable overall performance. However, further enhancements—such as more discriminative feature extraction or the implementation of advanced classification algorithms—could improve accuracy, particularly for ONUs with lower classification rates

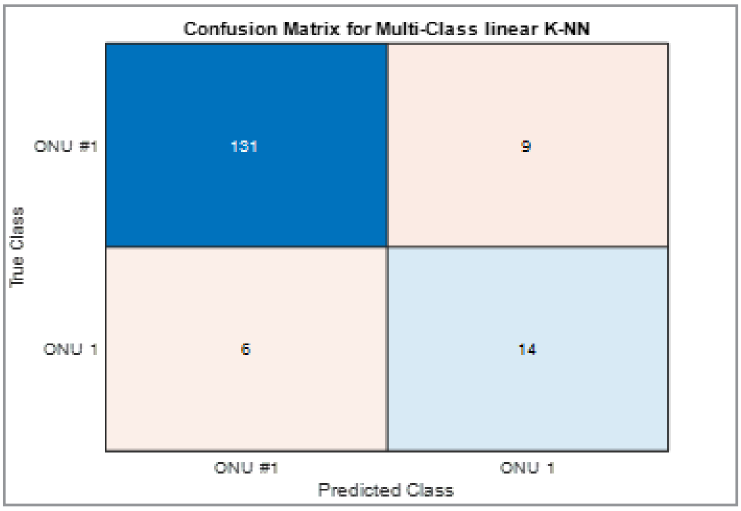

Figure 7: Confusion matrices of K-NN classifier for ONU 1

To enhance the accuracy of the system, the classification was refined to distinguish only between the desired ONU and the other undesired ONUs. The confusion matrix demonstrates the classification performance between ONU #1 and ONU 1, achieving an overall accuracy of 90.63%. The model correctly identified 131 instances of ONU #1 and 14 instances of ONU 1, while 9 ONU #1 samples were misclassified as ONU 1, and 6 ONU 1 samples were misclassified as ONU #1. Although the classifier performs well, some misclassifications remain, particularly affecting ONU 1. Improving feature extraction or optimizing the model could further enhance classification accuracy.

Conclusion

This study highlights the effectiveness of using machine learning-based fault detection in Gigabit Passive Optical Networks (GPON) through the application of the K-Nearest Neighbors (K-NN) algorithm. By conducting simulations in OptiSystem and performing data classification in MATLAB, the system efficiently detects and classifies signal degradations across multiple Optical Network Units (ONUs). Analysis of the confusion matrices indicates classification accuracy ranging from 63.16% to 75.00%, with ONU 2 and ONU 8 achieving the highest accuracy levels, while ONU 1 and ONU 7 are more susceptible to misclassification. Furthermore, the study underscores the impact of attenuation on the Bit Error Rate (BER), demonstrating that increased attenuation causes significant signal degradation, especially in the fiber segment between the splitter and the ONUs.

Overall, integrating the K-NN algorithm for fault detection in GPON significantly improves network reliability by automating fault diagnosis and minimizing downtime. Nevertheless, there remains room for enhancement in classification performance through more sophisticated feature extraction methods, hyperparameter optimization, and the incorporation of advanced deep learning techniques. Additionally, future research could investigate the benefits of real-time network monitoring and adaptive machine learning models to dynamically optimize GPON performance under varying operational conditions. These advancements have the potential to further strengthen fault resilience and service quality in optical access network

References

1. Kalera, R., Kaler, R. S. (2011). Simulation of fiber to the home at 10 Gbit/s 10-GPON architecture. Optik - International Journal for Light and Electron Optics, 122(17), 1362-1366. https://doi.org/10.1016/j.ijleo.2010.06.019
2. Wang, J., Wang, G., Zhang, L., Li, H. (2017). Ground simulation method for arbitrary distance optical transmission of a free-space laser communication system based on an optical fiber nanoprobe. IEEE/OSA Journal of Optical Communications and Networking, 9(12), 1131-1137. https://doi.org/10.1364/JOCN.9.001131
3. Shumate, W. P. (2008). Fiber-to-the-home. Journal of Lightwave Technology, 26(9), 1093-1103. https://doi.org/10.1109/JLT.2008.920769
4. Sun, J. W., Du, W., Shi, N. (2018). A survey of kNN algorithm. In Information engineering and applied computing 416-420. Springer. https://doi.org/10.1007/978-3-319-93846-2_57
5. Cunningham, P., Delany, S. J. (2007). k-nearest neighbour classifiers. ACM Computing Surveys, 39(1), 3. https://doi.org/10.1145/1216370.1216375

6. Suyal, M., Goyal, P. (2022). A review on analysis of K-Nearest Neighbor classification machine learning algorithms based on supervised learning. International Journal of Engineering Trends and Technology, 70(7), 43-48. https://doi.org/10.14445/22315381/IJETT-V70I7P206
7. Wang, H. (2002). Nearest neighbours without k: A classification formalism based on probability (Technical report). University of Ulster, Faculty of Informatics. https://link.springer.com/chapter/10.1007/3-540-32370-8_12
8. Karthikeya, K., Sudarshan, K. H. (2019). Prediction of agriculture crops using KNN algorithm. International Journal of Innovative Science and Research Technology, 5(5), 1422-1424.
9. Krishnamoorthy, N., Umarani, N. (2021). Diabetes prediction in healthcare using KNN algorithm. International Journal of Multidisciplinary Educational Research, 10(5), 36-39.
10. Chen, Y., Zhang, Y., Liu, H. (2021). Machine learning-based fault detection in GPON fiber access networks. IEEE Access, 9, 118034-118045. https://doi.org/10.1109/ACCESS.2021.3108530