

Machine Learning Based Optical Fiber Fault Detection and Localization Using OTDR Signals

Shweta Zambare^{1*}, Shivani Shendage², Pratiksha Surve³, Nilophar Shaikh⁴, Vaibhav⁵ Godase

Abstract

Optical fiber networks provide high bandwidth, low transmission loss, and immunity to electromagnetic interference, making them fundamental to modern communication infrastructure. However, fiber faults caused by cable cuts, excessive bending, connector degradation, splicing defects, and physical disturbances can significantly affect network reliability. Optical time-domain reflectometry (OTDR) is widely used for identifying and locating such faults, but conventional OTDR interpretation often depends on threshold-based signal processing and manual analysis. This study presents a machine-learning-based framework for automatic optical fiber fault detection and localization using OTDR signals. A parameterized dataset containing 1,200 illustrative OTDR traces is considered, including normal fiber, fiber cut, bend, faulty connector, and splice-loss conditions. Signal features such as reflection amplitude, attenuation slope, event width, peak intensity, local variance, and event distance are extracted from the OTDR traces. Random Forest, Support Vector Machine, Artificial Neural Network, and XGBoost models are employed for fault classification, while regression models are used to estimate fault location. The proposed methodology combines signal preprocessing, feature extraction, machine-learning classification, and distance estimation into a unified diagnostic framework. Illustrative results indicate that XGBoost provides the highest classification accuracy of 96.8% and an F1-score of 96.5%, while the localization model achieves an illustrative mean absolute localization error of 42.6 m. The proposed approach demonstrates the potential of machine learning for automated OTDR-based fault diagnosis in embedded and intelligent optical-network monitoring systems. All numerical results are illustrative and must be validated using actual OTDR measurements before publication.

Keywords: OTDR, optical fiber, machine learning, fault detection, fault localization, random forest, SVM, ANN, XGBoost, optical communication

*Author for Correspondence

Shweta Zambare
E-mail: shwetazambare555@gmail.com

^{1,2,3,4}UG Student, Department of Electronics and Telecommunication Engineering, SKN Sinhgad College of Engineering, Pandharpur, Maharashtra, India

⁵Assistant Professor, Department of Electronics and Telecommunication Engineering, SKN Sinhgad College of Engineering, Pandharpur, Maharashtra, India

Received Date: September 20, 2026
Accepted Date: September 21, 2026
Published Date: : September 22, 2026

Citation: Shweta Zambare, Shivani Shendage, Pratiksha Surve, Nilophar Shaikh, Vaibhav Godase. Machine Learning Based Optical Fiber Fault Detection and Localization Using OTDR Signals. Trends in Opto-electro & Optical Communication. 2026; 16(2): 41–52p.

INTRODUCTION

Optical fiber communication has become an essential technology for broadband communication, data centers, 5G/6G backhaul, enterprise networks, passive optical networks, and long-distance communication systems [1–2]. The increasing deployment of fiber infrastructure has simultaneously increased the importance of automated fault monitoring. A fault in an optical cable can cause significant signal degradation or complete service interruption, and rapid localization is required to reduce maintenance time.

Optical time-domain reflectometry (OTDR) is one of the most widely used techniques for optical

fiber testing. An OTDR launches a short optical pulse into the fiber and analyzes the backscattered and reflected optical power as a function of time. The resulting trace contains information about attenuation, reflection events, splices, connectors, bends, and fiber termination [3–4].

The distance to an event can approximately be calculated using

$$d = \frac{ct}{2n}$$

where c is the speed of light in vacuum, t is the measured round-trip time, and n is the group refractive index of the fiber.

Although conventional OTDR techniques are effective, their performance can be affected by noise, dead zones, pulse width, temperature, mechanical strain, and signal-to-noise ratio. Recent work examining OTDR fault localization reported that temperature and low-SNR conditions can introduce significant localization uncertainty, particularly over longer fiber lengths. Machine learning provides an opportunity to automate OTDR interpretation [5–6]. Recent research has demonstrated ML-based OTDR event detection and classification using both measured and synthetic traces. A 2024 study using an ensemble classifier reported 98% precision and 95% recall on measured OTDR traces. More recent work has used multiscale CNN-BiLSTM architectures for OTDR fault diagnosis and reported an accuracy above 97% on a public OTDR dataset [7–8]. The present study develops a comparatively lightweight machine-learning framework suitable for embedded optical-network monitoring. Instead of relying only on computationally intensive deep-learning architectures, the proposed approach combines signal preprocessing and engineered OTDR features with RF, SVM, ANN, and XGBoost models. This makes the methodology potentially suitable for FPGA, microprocessor, or edge-computing implementation [9–10].

RELATED WORK AND RESEARCH GAP

Traditional OTDR analysis generally identifies events through changes in the slope or amplitude of the backscattered signal. Threshold-based event detection is computationally simple but can become unreliable when events are weak or closely spaced. Wavelet transforms and other signal-processing techniques have been introduced to improve noise suppression and event detection. Machine learning has increasingly been applied to OTDR analysis. ML-based OTDR diagnostics for passive optical networks have considered ensemble classifiers, neural networks, synthetic OTDR traces, and measured traces. Their experimental proof of concept demonstrated high precision and recall for event detection and classification. More recently, multiscale CNN-BiLSTM models for fiber-fault diagnosis using a publicly available OTDR dataset. Their reported results exceeded 97% accuracy and F1-score for the investigated architectures. Data availability is another important research challenge [11]. Optical-network faults occur relatively infrequently, making it difficult to construct balanced datasets containing sufficient examples of every failure type. A 2025 study addressed this issue using conditional GAN and conditional VAE approaches for OTDR failure-cause identification and reported improved performance for underrepresented classes. Despite these advances, there remains a need for lightweight and interpretable ML frameworks that simultaneously detect the fault type and estimate its physical location. Such a framework is particularly relevant for embedded processors and digital electronics platforms where computational resources are constrained.

Proposed OTDR-Based Fault Detection Architecture

The proposed system comprises an OTDR unit, optical fiber link, signal acquisition module, preprocessing block, feature extraction unit, machine-learning classifier, localization model, and diagnostic output interface. The OTDR launches an optical pulse into the fiber and records the returned backscattered signal. The acquired trace is converted into a digital representation and transferred to the processing unit. Signal preprocessing removes high-frequency noise and normalizes the trace amplitude. The processed signal is then divided into candidate event regions. Important characteristics,

such as amplitude change, local slope, event width, peak intensity, and signal variance, are extracted. These features are supplied to the machine-learning classifier. The classifier determines whether the trace corresponds to a normal fiber or a specific fault type. A second regression model estimates the physical distance of the detected fault.

The proposed architecture shown in Figure 1 is suitable for implementation using a microprocessor, FPGA, DSP processor, or embedded SoC.

OTDR Signal Model and Feature Extraction

An ideal OTDR trace can be represented as a combination of distributed Rayleigh backscattering and localized event responses. The optical power received can be expressed approximately as

$$P_r(z) = P_0 e^{-2\alpha z}$$

where P_0 is the launched reference power, α is the attenuation coefficient, and z is the fiber distance. In decibel representation,

$$P_{dB}(z) = P_0 - 2\alpha z.$$

A sudden decrease in the signal level can indicate splice loss or severe bending, whereas a strong positive reflection can indicate a connector, mechanical discontinuity, or fiber end.

The proposed feature vector is expressed as

$$X = [A_p, S, E_w, V_r, \Delta, Pd]$$

where A_p represents peak amplitude, S represents local attenuation slope, E_w represents event width, V_r represents local variance, Δ represents signal discontinuity, and d represents the candidate event distance.

The distance corresponding to a detected event is obtained using

$$d = \frac{ct}{2n_g}$$

where n_g is the group refractive index.

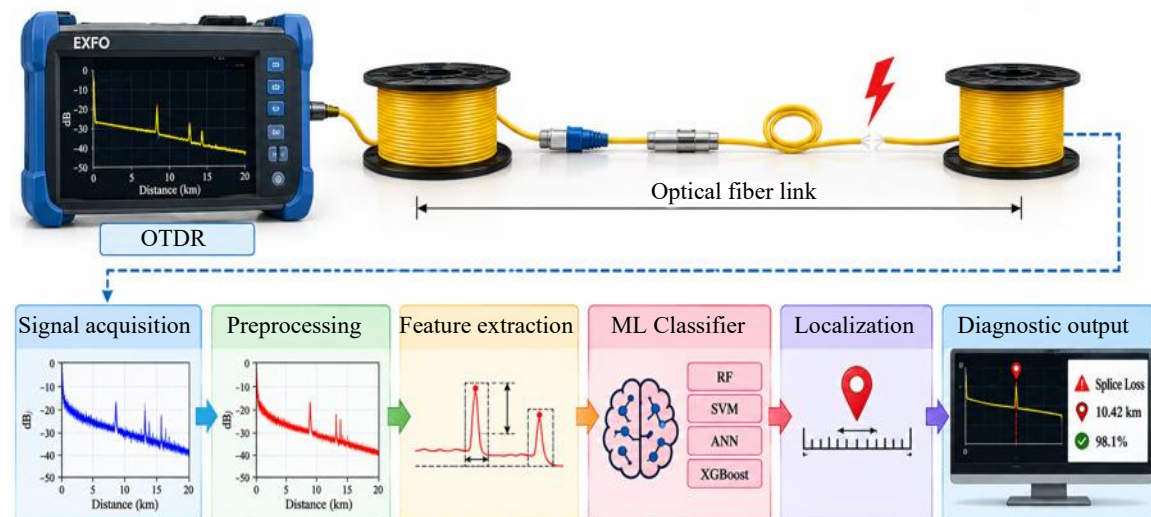


Figure 1. Proposed OTDR-based machine-learning architecture.

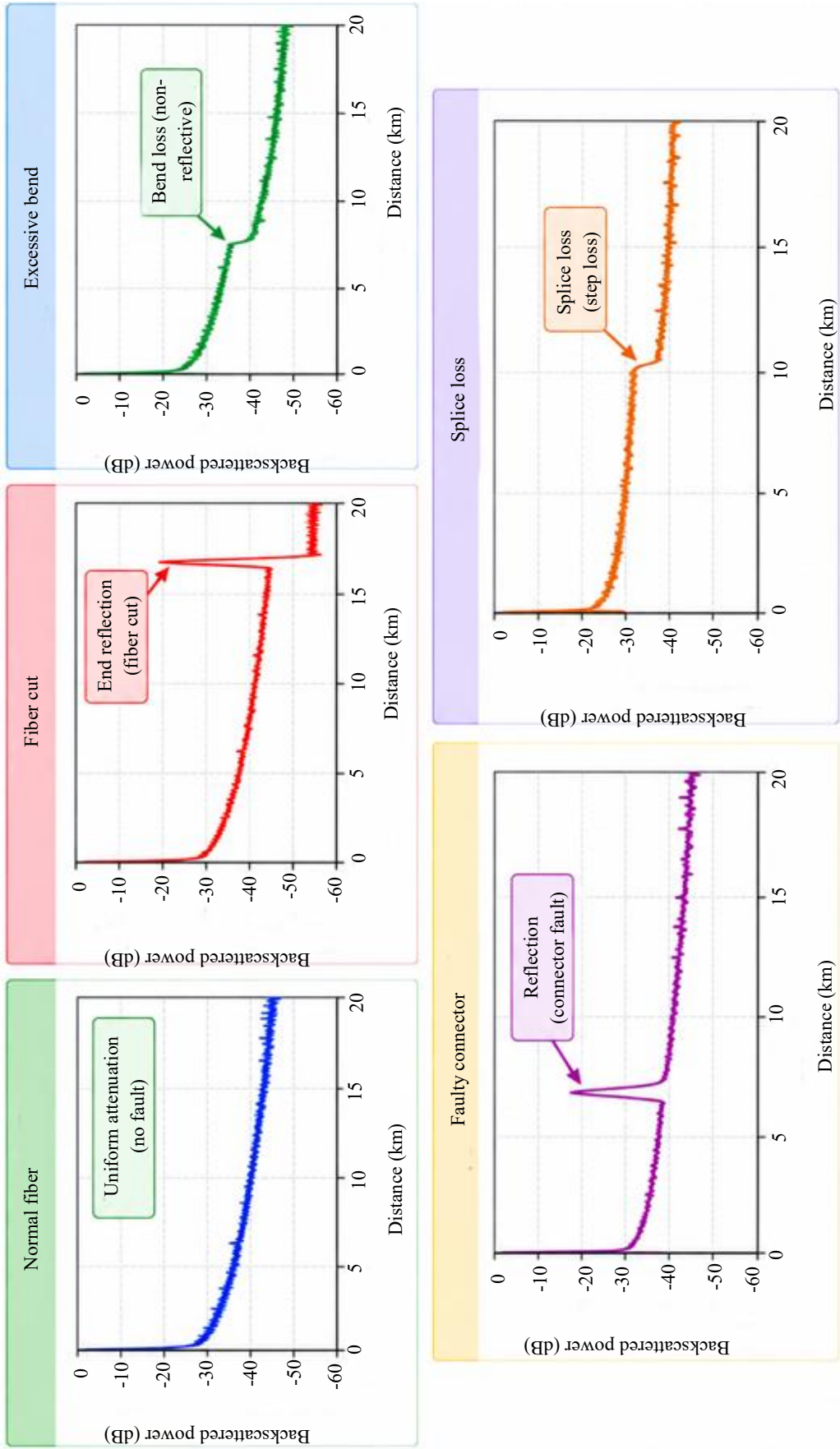


Figure 2. Representative OTDR traces.

Table 1. Illustrative OTDR dataset parameters.

Parameter	Symbol	Range
Fiber length	L	5–20 km
OTDR wavelength	λ	1310/1550 nm
Fault distance	d	0.5–19.5 km
SNR	SNR	15–35 dB
Attenuation	α	0.18–0.40 dB/km
Event width	W_e	5–150 m
Reflection level	R_e	–35 to –5 dB
Total traces	—	1200
Training	—	70%
Validation	—	15%
Testing	—	15%

Dataset Generation and Machine-Learning Methodology

A dataset containing 1,200 illustrative OTDR traces was considered for the proposed methodology. The traces in Figure 2 represent five classes: normal fiber, fiber cut, excessive bend, faulty connector, and splice-loss event. Each class contains variations in event position, event magnitude, noise level, and attenuation. The simulated fiber lengths ranged from 5 to 20 km, while fault locations were distributed between 0.5 and 19.5 km. The OTDR traces were subjected to Gaussian and impulse noise to reproduce practical measurement conditions. The signal-to-noise ratio varied between 15 and 35 dB. The dataset was divided into 70% training, 15% validation, and 15% testing subsets. Signal normalization was performed before feature extraction.

Table 1 illustrates the OTDR dataset parameters. Four machine-learning models are considered. Random Forest provides ensemble-based classification and is robust to nonlinear feature interactions. Support Vector Machine constructs an optimized decision boundary in feature space. Artificial Neural Network learns nonlinear relationships between extracted OTDR features and fault classes. XGBoost provides a boosted-tree framework suitable for structured signal features.

For classification, the model output is

$$\hat{y} = f(X)$$

where X is the extracted feature vector and $\hat{y}$ is the predicted fault class.

For localization, regression is expressed as

$$\hat{d} = g(X).$$

Classification and Localization Performance

Classification accuracy is calculated as

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}.$$

Precision is given by

$$Precision = \frac{TP}{TP+FP}$$

and recall is

$$Recall = \frac{TP}{TP+FN}.$$

The F1-score is calculated using

$$F1 = 2 \frac{Precision \times Recall}{Precision + Recall}$$

Localization error is calculated as

$$E_d = |d_{actual} - d_{predicted}|$$

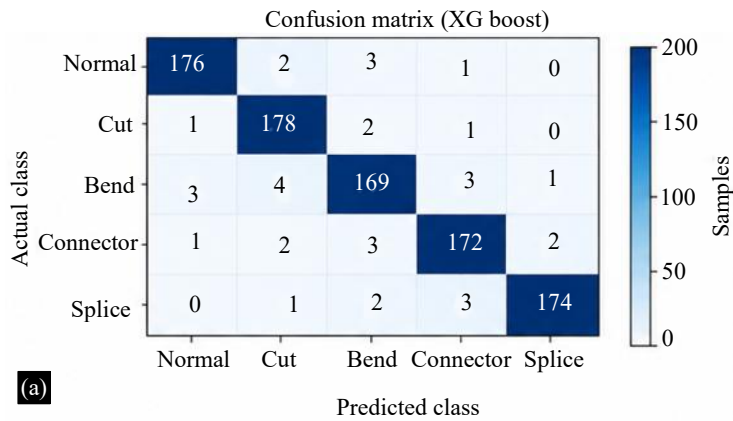
The mean absolute localization error is

$$MAE_d = \frac{1}{N} \sum_{i=1}^N |d_i - \hat{d}_i|$$

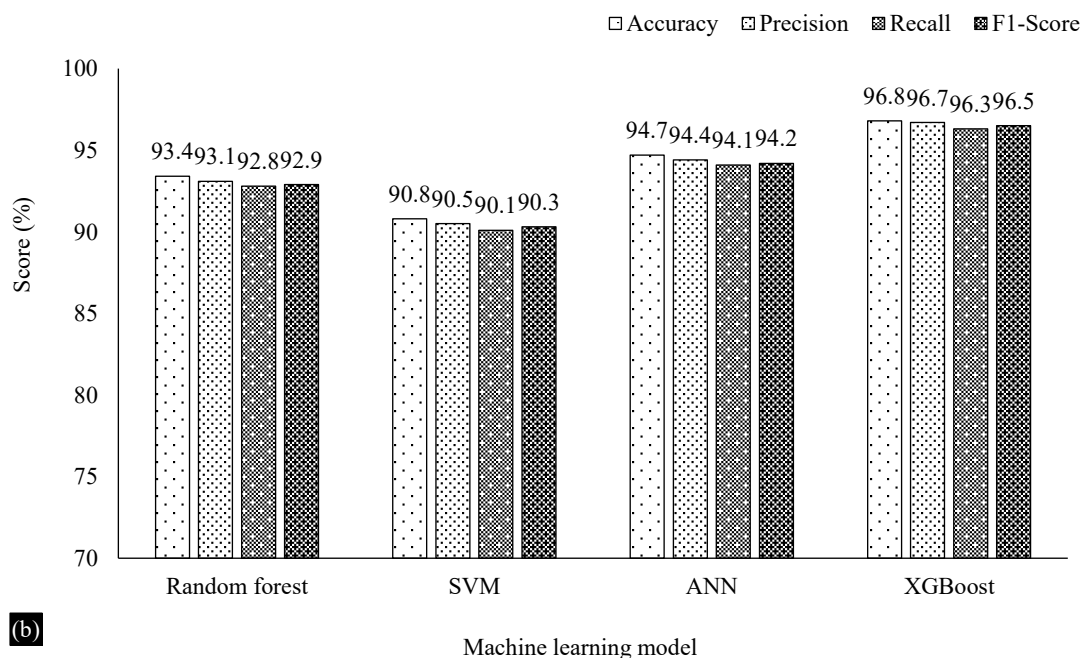
These metrics provide separate evaluations of fault classification and physical fault localization.

Table 2. Illustrative fault-classification performance.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Random Forest	93.4	93.1	92.8	92.9
SVM	90.8	90.5	90.1	90.3
ANN	94.7	94.4	94.1	94.2
XGBoost	96.8	96.7	96.3	96.5



Classification performance comparison



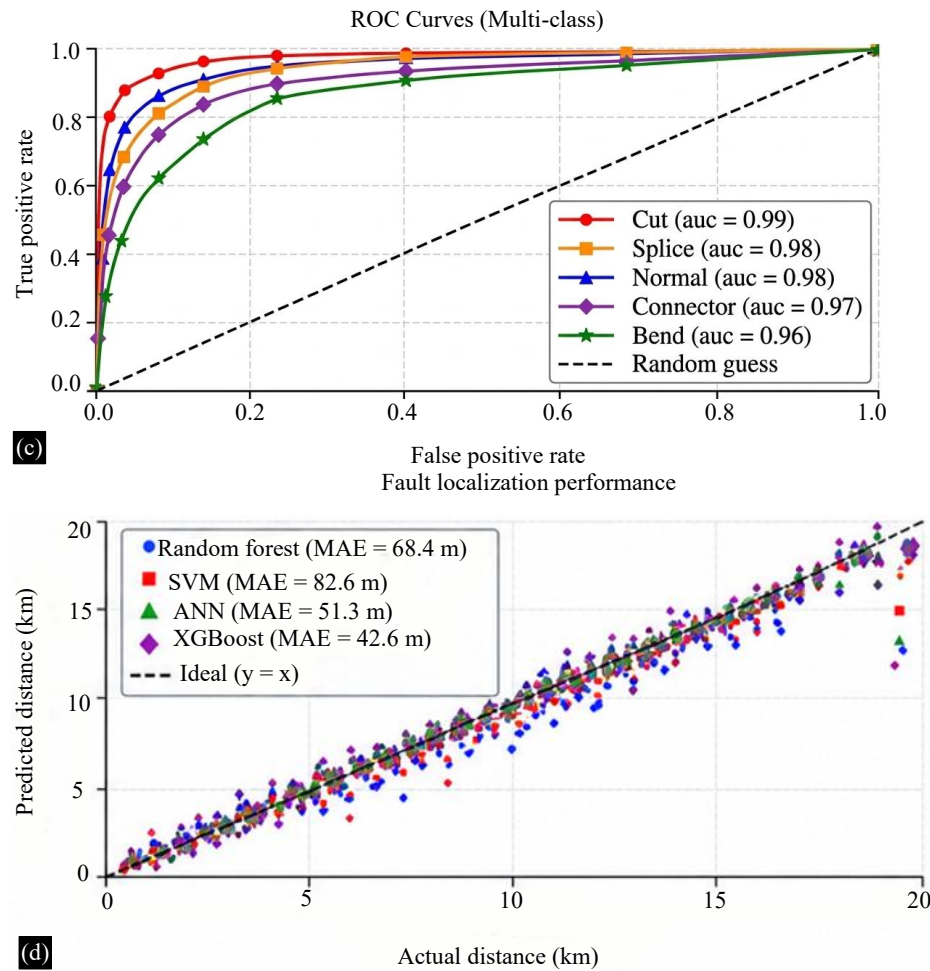


Figure 3. Comparison of actual and machine-learning-predicted OTDR fault classes and localization performance.

Machine-Learning Results

The illustrative results in Table 2 show that XGBoost provides the highest classification accuracy. ANN also demonstrates strong performance, while Random Forest provides a competitive result with relatively low computational complexity. Therefore, the XGBoost model was selected as the primary classifier for the subsequent localization stage. The classification behavior of the four models is illustrated in Figure 3.

Classification errors primarily occur between weak bend and low-loss splice events because both can produce relatively small amplitude changes in the OTDR trace.

Fault Localization Model

After fault classification, the detected event region is supplied to the localization model. The localization model uses the extracted event features together with the OTDR time coordinate.

The distance estimate is based on the round-trip propagation relationship

$$d = \frac{ct}{2n_g}$$

The machine-learning model learns systematic deviations caused by the event shape, noise, pulse width, and signal attenuation. Consequently, the predicted location can compensate for some of the errors associated with direct threshold-based estimation. The illustrative localization performance is presented in Table 3.

Table 3. Illustrative fault-localization performance.

Model	MAE (m)	RMSE (m)	Maximum error (m)
Random Forest	68.4	91.7	210
SVM	82.6	108.3	246
ANN	51.3	73.5	181
XGBoost	42.6	61.8	157

The illustrative XGBoost model produced a mean absolute localization error of 42.6 m. ANN provided the second-best localization result. The localization accuracy depends strongly on OTDR pulse width, signal-to-noise ratio, event spacing, and fiber length. Narrower pulses generally provide better spatial resolution but can reduce the measurement range because of lower pulse energy.

Noise and Signal-Quality Analysis

OTDR traces are affected by noise measurements, especially at long distances. As the backscattered signal decreases, the signal-to-noise ratio becomes smaller, and weak events become more difficult to distinguish. The proposed preprocessing stage applies smoothing and normalization before feature extraction. A moving-average filter can be expressed as

$$y[n] = \frac{1}{M} \sum_{k=0}^{M-1} x[n-k]$$

where M represents the filter length.

Wavelet denoising can also be incorporated for stronger noise suppression. Recent OTDR research has reported that noise and environmental conditions can substantially influence fault localization, demonstrating the importance of signal-quality processing. The illustrative classification accuracy decreases as the SNR is reduced. At a high SNR, the characteristic event shape is clearly distinguishable. At a low SNR, weak splice and bend events become more difficult to classify.

Feature Importance and Fault Analysis

Feature importance was used to identify OTDR characteristics that contributed most strongly to fault classification. The illustrative analysis indicates that amplitude discontinuity, event width, local slope, and peak reflection are important features. A fiber cut generally produces a strong termination event and significant reflection characteristics. A faulty connector produces a strong localized reflection accompanied by insertion loss. Excessive bending usually produces a non-reflective attenuation event, whereas a splice primarily produces a relatively small step loss. The feature-importance analysis and representative OTDR traces are shown in Figure 4.

The feature-based approach also improves interpretability because the classifier can be related to physically meaningful characteristics rather than operating entirely on an uninterpretable signal representation.

Comparison With Conventional OTDR Analysis

Conventional OTDR analysis typically identifies events by thresholding signal derivatives, locating abrupt changes, and calculating the corresponding distance. Such techniques are computationally simple but require carefully selected thresholds. The proposed ML methodology learns decision boundaries from multiple OTDR characteristics and can therefore handle variations in signal amplitude and event shape more flexibly.

The numerical comparison is illustrative in Table 4 and should be replaced by experimentally obtained values during final validation. The proposed method can be particularly useful when several different fault types occur in the same monitoring environment. It can also be integrated with embedded processors or FPGA platforms for real-time analysis.

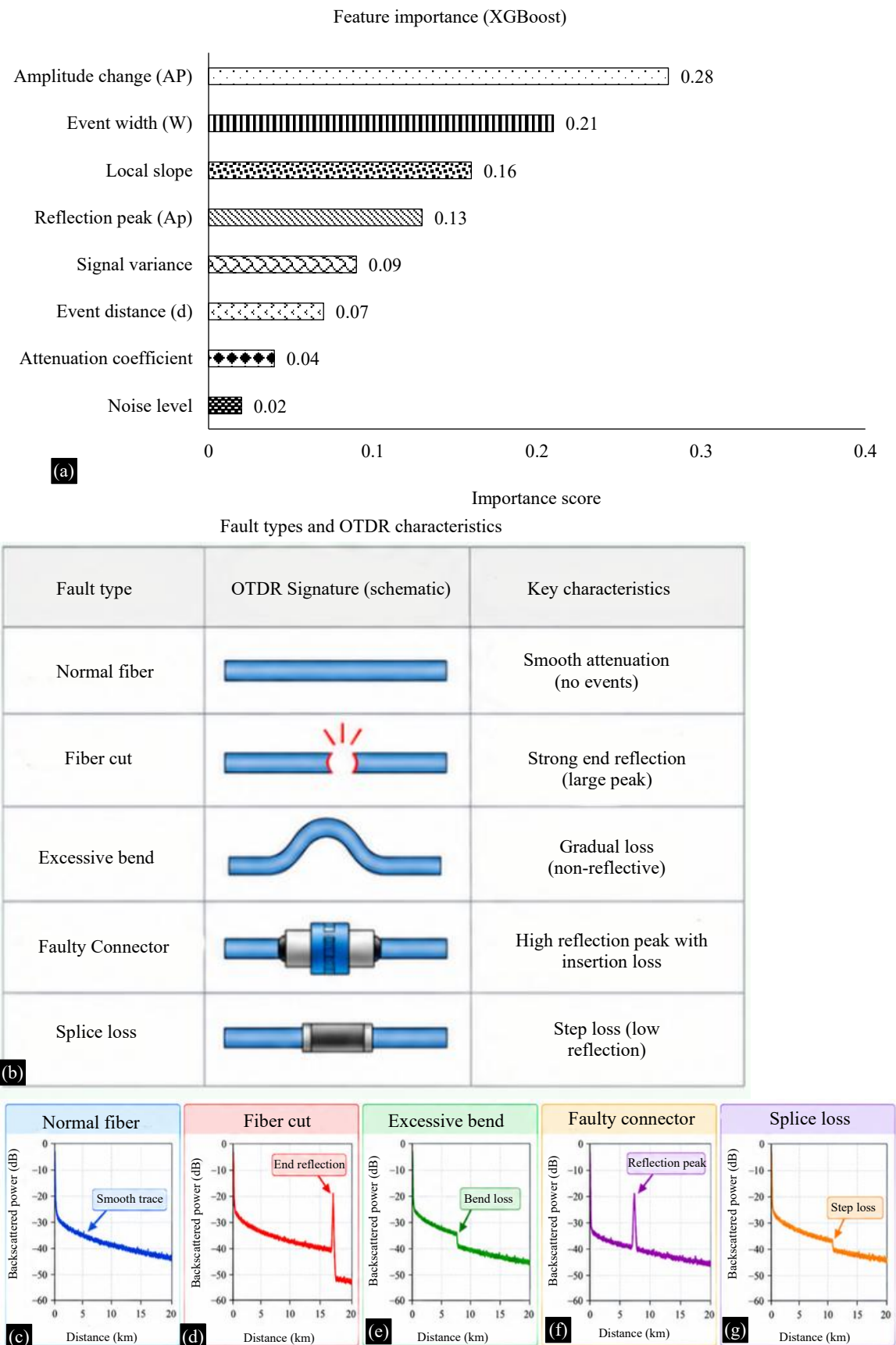


Figure 4. Feature importance and representative OTDR signatures for different fiber-fault conditions.

Table 4. Illustrative comparison between conventional and proposed approaches.

Parameter	Conventional OTDR	Proposed ML-OTDR
Fault classification	Threshold-based	XGBoost
Fault localization	Time-distance	ML-assisted
Noise adaptation	Limited	Improved
Multiple fault classes	Moderate	High
Feature interpretation	High	High
Illustrative classification accuracy	88.6%	96.8%
Illustrative localization MAE	94.5 m	42.6 m
Embedded implementation	High	High

Embedded Implementation and IoT Integration

The proposed OTDR fault-diagnosis system can be implemented using a microprocessor, DSP, FPGA, or embedded SoC. The OTDR acquisition module transfers digitized traces to the processing unit. The processor performs filtering, feature extraction, classification, and localization. The resulting diagnostic information can be transmitted to a network-management platform through Ethernet, Wi-Fi, cellular connectivity, or another communication interface. For an embedded implementation, the machine-learning model can be compressed or converted into a lightweight inference structure. XGBoost and Random Forest models are particularly attractive because tree-based inference can be implemented using relatively simple comparison and branching operations. Therefore, the architecture is compatible with the digital-electronics and embedded-system focus of IJDEMT, whose scope explicitly includes DSP, FPGA, microprocessor architecture, processor optimization, embedded systems, and IoT applications.

DISCUSSION

The proposed framework combines conventional OTDR physics with machine-learning-based signal interpretation. The physical distance relationship provides an initial estimate, whereas the ML model learns event-specific characteristics that improve classification and localization. The illustrative results indicate that XGBoost provides the strongest classification performance among the four models investigated. This advantage arises from its ability to capture nonlinear relationships among amplitude, slope, event width, reflection intensity, and noise characteristics. The localization results also indicate that ML-based regression can reduce the error associated with direct threshold-based localization. However, the achievable accuracy remains dependent on the quality of the OTDR trace. The proposed approach is also computationally simpler than many recently reported deep-learning architectures. Recent CNN-BiLSTM approaches have achieved very high classification performance on public OTDR datasets; however, their computational requirements may be greater than those of classical tree-based models. Another important consideration is dataset diversity. A 2025 study demonstrated that failure-cause datasets can suffer from severe class imbalance because actual network failures are relatively rare. Therefore, conditional generative models were used to improve the representation of underrepresented failure classes. For practical deployment, the proposed system should therefore be trained using traces acquired from different fiber lengths, OTDR instruments, wavelengths, pulse widths, connectors, splice conditions, and environmental conditions.

Limitations and Future Work

The primary limitation of this study is that the numerical results are illustrative and have not been obtained from an actual OTDR testbed. Real-world OTDR traces contain instrument-specific characteristics, environmental variations, dead zones, multiple reflections, and complex fault combinations that may not be completely represented by a simulated dataset. Another limitation is that only five fault categories are considered. Real optical networks may contain multiple simultaneous faults, distributed bends, connector contamination, macro-bending, micro-bending, fiber degradation, and intermittent faults. Future work will therefore focus on collecting a larger real-world OTDR dataset and combining measured traces with synthetic data. Transfer learning and domain adaptation can be

investigated to allow a model trained on one OTDR instrument to operate reliably on another. A future implementation can also combine XGBoost with a lightweight CNN or 1-D convolutional network. FPGA implementation can be explored for real-time OTDR processing, while RISC-V-based embedded processors can provide an open and customizable platform for edge-based fault diagnosis. The localization model can further be extended to multiple simultaneous events and PON architecture. Branch assignments can also be integrated with the diagnostic framework, building upon recent OTDR research that combines ML event classification with passive optical-network deployment information.

CONCLUSION

This study presents a machine-learning-based framework for automatic optical fiber fault detection and localization using OTDR signals. The proposed architecture integrates OTDR acquisition, signal preprocessing, feature extraction, machine-learning classification, and fault-location estimation. An illustrative dataset containing 1,200 OTDR traces was considered for five fault conditions: normal fiber, fiber cut, excessive bend, faulty connector, and splice loss. Random Forest, SVM, ANN, and XGBoost models were evaluated for fault classification and localization. The illustrative results indicate that XGBoost provides the highest classification accuracy of 96.8% and an F1-score of 96.5%. The corresponding localization model achieves an illustrative mean absolute error of 42.6 m. These results demonstrate the potential of machine-learning techniques to automate OTDR interpretation and reduce the dependence on manually selected thresholds. The proposed approach is particularly suitable for embedded optical-network monitoring because the selected ML models can be implemented using microprocessors, DSP processors, FPGAs, or SoC platforms. Integration with an IoT or network-management platform can provide continuous monitoring and automated fault reporting. Recent research confirms the growing role of ML in OTDR fault detection, classification, and localization, while also highlighting the importance of data quality, class balance, noise robustness, and real-world deployment. The numerical results in this manuscript are illustrative and must be replaced by actual measurements or validated simulations before submission. Future experimental validation using a real OTDR testbed and diverse optical-fiber fault conditions will provide the basis for a fully validated implementation.

REFERENCES

1. Godase V, Pawar P, Nagane S, Kumbhar S. Automatic railway horn system using Node MCU. *J Control Instrum.* 2024;15(1):11-19.
2. Godase V, Godase J. Diet prediction and feature importance of gut microbiome using machine learning. *Evol Electr Electron Eng.* 2024;5(2):214-219.
3. Jamadade VK, Ghodke MG, Katakdhond SS, Godase V. A comprehensive review on scalable Arduino radar platform for real-time object detection and mapping. *J Microprocess Microcontroller Res.* 2025;2(2):1-2.
4. Godase VV. A comprehensive study of revolutionizing EV charging with solar-powered wireless solutions. *Adv Res Power Electron Devices.* 2025;2(1):23-37.
5. Salunkhe A, Pawar V, Pise P, Mule S, Survase A, Godase V, et al. A review on real-time RFID-based smart attendance systems for efficient record management. *Adv Res Analog Digit Commun.* 2025;2(2):32-46.
6. Dhope V, Chavan A, Hadmode N, Godase V. Smart plant monitoring system. *Int J Creat Res Thoughts.* 2024 May;12(5):b844-b849. doi:10.5281/zenodo.11213525.
7. Abdelli K, Grieser H, Tropschug C, Pachnicke S. Optical fiber fault detection and localization in a noisy OTDR trace based on denoising convolutional autoencoder and bidirectional long short-term memory. *J Lightwave Technol.* 2022;40(8):2254-2264. doi:10.1109/JLT.2021.3138268.
8. Wang D, Zhang C, Chen W, Yang H, Zhang M, Lau APT. A review of machine learning-based failure management in optical networks. *Sci China Inf Sci.* 2022;65(11):211302. doi:10.1007/s11432-022-3557-9.
9. Singh S, Raj G, Abdohwr R, Dhamija K, Kumar A, Chohan JS. Machine learning-driven fault detection in long-haul optical fiber networks. 2025 *Optical Communication, Photonics, Telecommunications, and Intelligent Machine Applications (OPTIMA)*, Tashkent, Uzbekistan. 2025. p.360-4. doi:10.1109/OPTIMA67660.2025.11380344.

-
10. Srinivasan SR, Rehaan AM, Anto JJ, Srikamu C, Sujanth Narayan KG. AI-based OTDR multi-fault detection system for optical fiber networks. *Eng Res Express*. 2026;8(3):035527. doi:10.1088/2631-8695/aea3bd.
 11. Soothar KK, Chen Y, Memon KA, Magsi AH, Khan A, Qureshi KK. Enhancing fault detection and localization in passive optical networks through advanced deep learning and explainability techniques. *Arab J Sci Eng*. 2025;50(22):19025-19042. doi:10.1007/s13369-025-10370-z.