CHALMERS
UNIVERSITY OF TECHNOLOGY

Modulated vs. Unmodulated Polarization Signatures for ML-Based Fiber Sensing

Citation for the original published paper (version of record):
Sadighi, L., Natalino Da Silva, C., Karlsson, S. et al (2026). Modulated vs. Unmodulated Polarization Signatures for ML-Based Fiber Sensing. Journal of Lightwave Technology, In Press.
http://dx.doi.org/10.1109/JLT.2026.3660791

(article starts on next page)

DP-16QAM Modulated vs. Unmodulated Polarization Signatures for ML-Based Fiber Sensing

Leyla Sadighi, *Member, IEEE,* Carlos Natalino, *Senior Member, IEEE,* Stefan Karlsson, *Member, IEEE,* Lena Wosinska, *Senior Member, IEEE,* Eoin Kenny, Venkata Virajit Garbhapu, *Member, IEEE,* Marco Ruffini, *Senior Member, IEEE,* Marija Furdek, *Senior Member, IEEE*

***Abstract*—Fast and accurate detection of various physical layer threats that target optical networks is key to secure and reliable global communications. Conventional monitoring methods often fail to detect subtle anomalies, which requires advanced sensing techniques. Machine Learning (ML) analysis of the State of Polarization (SOP) of unmodulated signals was shown to successfully detect such disturbances. However, real-world networks typically operate with high-speed modulated signals, which may alter SOP behavior and challenge the applicability of ML techniques developed for unmodulated signals. This paper investigates the implications of signal modulation supporting high-data rates on the interpretability of SOP signatures. We perform the first experimental comparison of anomaly detection approaches based on SOP for Dual-Polarization 16-Quadrature Amplitude Modulation (DP-16QAM) modulated and unmodulated optical signals subjected to identical physical perturbations caused by fiber tapping and vibrations. We analyze four representative events under both signal modalities and assess the impact of modulation on SOP dynamics using a 63.4 km fiber link in a real-world metro network. We design four datasets that isolate, merge, and jointly classify the different signal modalities, and compare the performance of ten best-performing supervised ML techniques in each case. Our findings indicate that modulated signals tend to exhibit smoother SOP trajectories, likely due to the temporal averaging effects introduced by high symbol rates, wherein rapid symbol transitions suppress high-frequency polarization noise. Importantly, this smoothening does not obscure the slower, event-induced polarization drifts observed during physical disturbances, allowing ML models to reliably differentiate between different physical events (e.g., bending, vibrations) and signal modalities (modulated vs. unmodulated), achieving accuracy values between 97.12% and 98.47%.**

***Index Terms*—Stokes parameters, State of Polarization (SOP), Modulation, Physical layer security, Machine Learning (ML), Vibrations, Fiber bending, Fiber sensing.**

I. Introduction

OPTICAL fiber networks are the cornerstone of global connectivity, enabling high-capacity and low-latency communication for global internet services, cloud computing, financial systems, and critical infrastructure. As these networks scale in speed and complexity, their role in supporting critical services grows accordingly. Vulnerability of optical fibers to various threats, including mechanical stress, accidental damage, and deliberate intrusions such as fiber tapping [1], poses a risk of security violations and service disruptions. Therefore, ensuring the integrity and security of optical transmission networks is of paramount importance, particularly as cyber-physical security becomes a growing concern in next-generation network deployments.

One of the key components of optical network security management is the timely detection of physical-layer anomalies. Techniques such as Optical Time Domain Reflectometry (OTDR) are commonly used to identify severe faults, like fiber cuts or sharp bends, by analyzing Rayleigh backscattering [2]–[4]. They offer relatively precise fault localization and are well established in field deployments. However, their practical use is often constrained by deployment cost and limited scalability [5], and they typically lack the sensitivity needed to detect more subtle disturbances, such as low vibrations or small mechanical deformations. Other approaches, including Distributed Fiber Optic Sensing (DFOS) for intrusion monitoring [6], offer higher sensitivity but require specialized hardware and complex processing, which significantly increases the cost and complicates

L. Sadighi, V. Garbhapu, and M. Ruffini are with the Department of Computer Science and Statistics, CONNECT Center, Trinity College Dublin, Ireland (e-mail: {sadighil, garbhapv, marco.ruffini}@tcd.ie).

C. Natalino, L. Wosinska, and M. Furdek are with the Department of Electrical Engineering, Chalmers University of Technology, Gothenburg, Sweden (e-mail: {carlos.natalino, wosinska, furdek}@chalmers.se).

S. Karlsson is with Micropol Fiberoptics AB, Stockholm, Sweden (e-mail: stefan.karlsson@vikinginvent.se).

Eoin Kenny is with HEAnet CLG, Dublin, Ireland. (e-mail: eoin.kenny@heanet.ie).

This work was supported by Taighde Éireann – Research Ireland under Grant No. 18/RI/5721: OpenIreland Research Infrastructure, the Swedish Research Council (2023–05249), the European Commission's Digital Europe Programme (101127973) through the 5G-TACTIC project, and the European Union's Horizon Europe research and innovation program (10113933) through the ECO-eNET project.

their adoption. State of Polarization (SOP) analysis has emerged as a powerful tool for monitoring and securing optical fiber networks. The statistical and dynamic properties of the SOP carry valuable information about environmental disturbances and physical-layer threats such as fiber tapping or mechanical perturbations. Unlike traditional methods that rely on backscatter or reflection signatures, SOP-based techniques exploit the intrinsic polarization sensitivity of optical signals [7]. This enables the detection of subtle anomalies with minimal modifications of the existing infrastructure. As a result, SOP-based monitoring provides a cost-effective and streamlined alternative, avoiding the hardware complexity and deployment overhead associated with DFOS and OTDR, albeit without the ability of event localization along a fiber. SOP-based sensing in coherent transmission systems can be implemented using either an external polarization analyzer, as adopted in our previous work [8], or by leveraging the internal Digital Signal Processing (DSP) of a coherent receiver. While external analyzers directly measure polarization fluctuations without accessing signal data, coherent receivers inherently estimate and compensate fiber birefringence as part of their DSP pipeline. This internal tracking can be exploited for polarization sensing, as demonstrated in [9], where the authors compare DSP-based phase and polarization sensing in a deployed metro network. Notably, this approach naturally resolves the low-Degree of Polarization (DoP) limitation encountered by standalone analyzers when observing rapidly modulated signals.

Accurate monitoring and detection of polarization variations induced by external events is crucial for identifying fiber disturbances and preserving network reliability. Traditional SOP-based monitoring approaches that depend on static thresholds or heuristic rules, such as the one in [10] often fall short when facing complex or evolving physical-layer threats [11]. To address these limitations, recent research has turned increasingly to data-driven methodologies. By leveraging Machine Learning (ML), such approaches enable automated, scalable, and adaptive analysis of polarization behavior, empowering the system to recognize intricate disturbance patterns directly from SOP data.

Despite the proliferation of ML-based techniques for accurate SOP monitoring, most existing studies use simplified experimental setups with limited conditions, i.e., the absence of signal modulation. This exclusive reliance on unmodulated signals raises an important question about the applicability of these methods in practical, real-world systems. Modern optical networks predominantly operate with modulated signals, particularly in coherent systems using modulation formats like Quadrature Phase-Shift Keying (QPSK) or Quadrature Amplitude Modulation (QAM). It should be noted that the modulated signals considered in this paper are coherent polarization-multiplexed 16-ary Dual-Polarization 16-Quadrature Amplitude Modulation (DP-16QAM) channels generated and detected using coherent transceivers. This is fundamentally different from Intensity-Modulated Direct-Detection (IM-DD) systems, which typically employ Non-Return-to-Zero (NRZ) modulation and exhibit nearly fully polarized optical carriers (DoP $\approx$100% at 1 ms integration), resulting in polarization behavior opposite to that of DP-16QAM signals. Furthermore, existing studies lack a direct experimental comparison between SOP signatures of modulated and unmodulated signals under similar physical and environmental conditions. Without such comparison, it remains unclear whether the presence of modulation fundamentally alters the SOP variations in a way that impacts the effectiveness of monitoring techniques like ML-based SOP fiber sensing.

Unmodulated signals maintain a nearly fixed SOP apart from slow environmental drifts [12], which offer the advantage of relatively clean polarization trajectories, making them particularly amenable to DSP and ML-based classification for anomaly detection. Unlike unmodulated signals, a modulated data signal's SOP can fluctuate on sub-nanosecond timescales due to the rapid symbol transitions at multi-Gbps rates. In effect, SOP is no longer represented by a single, stable point on the Poincaré sphere, but hops among a continuum of states dictated by the bit sequence [13].

The fast polarization fluctuations in modulated signals present new challenges for sensing and monitoring. They can act as a high-frequency noise floor that masks the more gradual SOP rotations caused by physical disturbances. Fiber bends, vibrations, or taps typically induce SOP changes on the Hz-kHz scale, whereas symbol-rate polarization changes occur at the GHz scale. When a modulated channel is observed at a slower timescale than its symbol rate, as is often the case with photodiode-based or low-cost polarization analyzers, the rapidly fluctuating SOP appears scrambled. The analyzer effectively averages over many symbols, and this temporal averaging reduces the measured DoP, resembling the effect of a polarization scrambler [14]. This means that, paradoxically, a high bit rate signal might exhibit a smoother SOP trajectory when viewed in aggregate.

Namely, a slow polarimeter does not provide instantaneous SOP values, but outputs a *time-averaged* Stokes estimate $\bar{\mathbf{S}}(t) = \frac{1}{T}\int_{t-T/2}^{t+T/2} \mathbf{S}(\tau)\, d\tau$, where T is the effective acquisition time. For high-baud-rate polarization-multiplexed signals (e.g., $\sim$30 Gbaud DP-16QAM), the SOP varies at (or above) the symbol rate, i.e., much faster than $1/T$. Therefore, the polarization-dependent components largely cancel each other within the averaging window and the reported $\mathrm{DoP} = |\bar{\mathbf{S}}(t)|/\bar{S}_0$

can approach zero. However, DoP ≈ 0 only indicates a small mean Stokes vector magnitude on the polarimeter time scale, without implying zero values of $\bar{S}_1, \bar{S}_2, \bar{S}_3$. Instead, the polarimeter outputs noisy, time-averaged Stokes samples with non-zero variance. When an external disturbance such as bending changes the fiber birefringence, it introduces slow polarization rotations that modulate the statistics of these time-averaged Stokes estimates, yielding observable low-frequency excursions even in the presence of strong modulation-induced fluctuations.

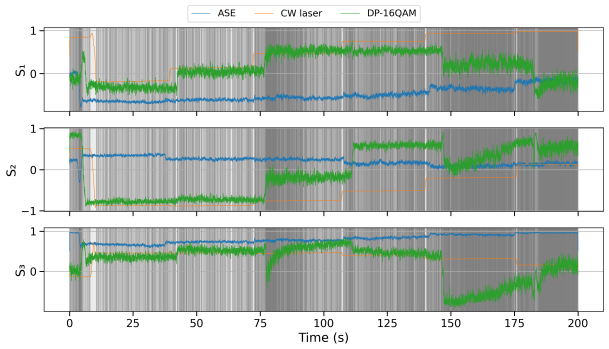


Fig. 1. Temporal evolution of normalized Stokes parameters (S_1, S_2, S_3) under automated bending for three optical excitations: ASE, CW, and DP-16QAM. The time axis corresponds to a controlled disturbance sequence: fiber relaxed from 0–60 s, bent from 60–120 s, relaxed from 120–180 s, and bent again from 180–200 s. The ASE signal exhibits randomized Stokes fluctuations with no clear correlation to the applied disturbance. The CW laser shows smooth and well-defined Stokes trajectories that track the bending-induced polarization evolution. The DP-16QAM traces are substantially noisier, yet their low-frequency evolution remains synchronized with the bending-induced drift observed for the CW reference.

This effect is shown in Fig. 1, where the horizontal axis represents the measurement time during a controlled automated bending sequence: the fiber is relaxed from 0–60 s, bent from 60–120 s, relaxed again from 120–180 s, and bent once more from 180–200 s. Compared to the smooth CW laser traces, the DP-16QAM Stokes components are substantially noisier (consistent with DoP $\approx$ 0), yet their low-frequency evolution remains synchronized with the bending-induced drift observed for the CW reference, indicating that disturbance-related polarization dynamics remain observable for polarization-multiplexed coherent signals. The Amplified Spontaneous Emission (ASE) signal shown in the figure provides a contrasting baseline: its Stokes components exhibit randomized fluctuations with no clear correspondence to the applied disturbance, consistent with true depolarization. The disparity between ASE and DP-16QAM indicates that the observed disturbance-locked drift is not a generic consequence of stochastic Stokes fluctuations, but is instead linked to the presence of a polarization-multiplexed coherent optical field with a well-defined phase relationship between orthogonal polarization components, which is absent in the ASE case.

Any anomaly detection scheme must ensure that the residual fast SOP variations in modulated signals do not trigger false alarms or confound true event signatures. To address the gap in the literature, we perform a comprehensive experimental comparison between modulated and unmodulated optical signals under similar physical and environmental conditions. We systematically evaluate the impact of modulation on SOP dynamics and assess its implications for ML-based anomaly detection. The main contributions include detailed statistical analyses comparing the behavior of polarization signatures in modulated and unmodulated signals, offering quantitative evidence of how modulation affects polarization dynamics. We collect a real-world experimental dataset from a 63.4 km fiber link in the HEAnet [15] metro network, capturing eight representative polarization event signatures: relaxed fiber, soft bending, eavesdropping attempt, and 80 Hz vibration, each recorded for modulated and unmodulated signal conditions. These events were selected to represent a diverse mix of normal (e.g., relaxed, soft bending) and potentially harmful or malicious (e.g., eavesdropping, vibration) fiber conditions commonly encountered in operational environments and security-sensitive scenarios. We then evaluate the performance of a range of supervised ML classifiers for these events and for different signal modalities. We analyze four scenarios, each corresponding to a distinct dataset (i.e., separated, mixed, and joint modality classification), assessing the impact of modulation on the ability of ML models to learn and separate polarization signatures corresponding to modulated and unmodulated signals.

The remainder of the paper is organized as follows. Section II reviews prior work on SOP analysis for modulated and unmodulated signals. Section III presents the experimental testbed and data collection process. Section IV details the events used for SOP signature analysis and a statistical comparison of modulated vs. unmodulated signals. Section V introduces the dataset configurations and preprocessing. Section VI reports experimental results, and Section VII concludes the paper.

II. Related Work

A wide spectrum of ML approaches has been developed to enhance the interpretability and adaptability of SOP-based monitoring. Supervised Learning (SL) techniques are often applied to detect and classify known physical disturbances and malicious activities such as fiber tapping and mechanical intrusions with high accuracy [8], [16]–[20]. Unsupervised Learning (USL) methods, such as clustering and outlier detection,

allow the identification of previously unseen or emerging anomalies in SOP trajectories without requiring labeled data [21]–[24]. Semi-supervised Learning (SSL) approaches strike a balance by leveraging a small set of labeled normal data along with abundant unlabeled samples, including unknown disturbances, to improve generalization under sparse annotation scenarios [24], [25]. The majority of ML-driven SOP analyses rely on CW light sources, particularly unmodulated Continuous Wave Distributed Feedback (CW-DFB) lasers to simplify the polarization behavior and avoid the complexities introduced by high-speed modulation [7], [8], [18], [19], [26], [27]. The review of polarization-based fiber sensing methods and anomaly detection algorithms [7] demonstrated that polarization measurements from a CW source, obtained via a polarimeter, reflect only the disturbance-induced SOP changes, whereas modulated signals introduce additional SOP estimation noise, confirming that unmodulated light yields clean polarization signatures that can be readily interpreted for anomaly detection. Our prior experimental studies [8], [19], [28] demonstrated successful use of supervised ML techniques to detect and classify disturbances like harmful vibrations and eavesdropping in controlled setups with unmodulated signals, with the impact of noise in real-world environments considered in [18], [26].

Modulation-induced polarization effects have received comparatively little attention, primarily approached from the communications perspective. Recent work by Karlsson *et al.* demonstrated SOP-based event detection using coherent transceivers in live networks. A real-time monitoring system detected polarization precursors before a cable break [29]. Follow-up analysis showed that mechanical disturbances produce distinct SOP patterns separable from noise [30]. In coherent communications, polarization tracking algorithms follow the time-varying SOP of a high-speed modulated signal so that the receiver can correctly demultiplex the polarization-multiplexed channels. For instance, a stochastic polarization drift model that treats the SOP evolution as a random walk on the Poincaré sphere can be found in [31]. Empirical studies have shown that environmental factors can indeed induce rapid SOP changes in deployed systems. SOP fluctuations on ms time scale in aerial fibers under varying climatic conditions were reported in [32], while cyclic SOP oscillations in overhead fiber cables caused by wind gusts and power-line electromagnetic interference were observed in [33]. Lightning-induced abrupt, fast polarization rotations, were shown to necessitate sub-ms reaction times in coherent receivers [34]. Real-time polarization tracking in modern coherent transceivers' DSP, e.g., the adaptive filters that continuously correct for polarization rotation and mode dispersion, allowing the receiver to lock on the signal's SOP [35], [36], treat polarization fluctuations purely as impairments to be corrected, rather than indicators of physical network disturbances. The effect of modulation on polarization sensing has not been explicitly addressed, and existing research lacks a systematic investigation of how modulated signals compare to unmodulated lights in terms of polarization disturbance detectability. In this work, our objective is to identify statistically meaningful patterns in SOP dynamics that correlate with external disturbances, which is key to understanding the role of ML-based polarization monitoring as a sensing tool.

III. Experimental Setup

A. Network topology and setup

To evaluate the effects of optical signal modulation on the behavior of polarization signatures under a realistic noisy environment, we use the experimental setup in Fig. 2, illustrating the 63.4 km test channel route of the HEAnet Dublin metro ring, connected to the OpenIreland [37] testbed infrastructure. The route traverses six Reconfigurable Optical Add-Drop Multiplexers (ROADMs) in sequence: TCD-1 (Lloyd Institute, Trinity College Dublin (TCD)), UCD-2 and UCD-1 (two nodes located at University College Dublin (UCD), Belfield Campus), CWT-2 (Citywest), PW-2 (Park West), and TCD-2 (Pearse Street, TCD), with the corresponding fiber segment lengths indicated in Fig. 2. The optical signal is added at TCD-1 and dropped at TCD-2 before returning to the OpenIreland Lab via a cross-campus fiber patch, where controlled disturbances are applied. The HEAnet ring is lightly loaded, and the experiment runs over a 400 GHz Optical Spectrum as a Service (OSaaS) window in the C-band (192.8–193.2 THz). The SOP sensing is carried out over two wavelength channels injected from a Lumentum ROADM in OpenIreland into the HEAnet entry ROADM. The first channel (λ_m, centered at 193 THz) is a coherent DP-16QAM 200 Gbps channel generated by an Adtran Teraflex transceiver, while the other (λ_u, centered at 193.1 THz) is an unmodulated signal generated by an External Cavity Laser (ECL) source. The two channels are within the same 400 GHz OSaaS window, allowing for a fair comparison focusing on the impact of modulation. Only one of the two signals is active at a time.

The signals traverse the same path, which ensures that both unmodulated and modulated channels experience similar propagation conditions. Here, the signals are exposed to controlled physical disturbances while traveling on a fiber patch that is part of the equipment shown in Fig. 2, including soft bending, eavesdropping by bending the fiber, and 80 Hz vibrations, which are introduced to evaluate the impact on polarization signatures. The signals are then demultiplexed by a

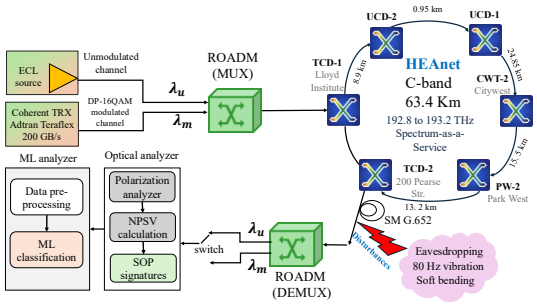


Fig. 2. Schematic of the experimental and analytical setup used to investigate the impact of signal modalities on SOP dynamics under controlled perturbations.

second Lumentum ROADM (DEMUX) and sent to the polarization analyzer. Our polarization analyzer instrument is a commercial polarization sensing module [8] (a "black box") capable of measuring all three Stokes polarization components (S_1, S_2, S_3) variations. This device uses a carefully designed arrangement of passive optical components to project the signal onto different polarization bases. The complete processing pipeline, including additional optical analyzer components and analysis stages, is described in the following subsection.

B. Data Collection Process

Upon transmission over the same path and exposure to similar physical disturbances, each optical signal is directed to the optical analyzer. This analyzer implements a structured processing pipeline for extracting and analyzing the Normalized numerical Polarization State Variation (NPSV) data and generating SOP signatures for each disturbance event and signal modality. In the first stage of processing, the optical polarization analyzer captures the temporal evolution of the SOP variations by projecting the received signal onto the Poincaré sphere. For each experimental run, the signal is continuously sampled for 15 minutes at 0.5 ms intervals, resulting in approximately 1.8 million NPSV samples per experiment. At each time slot t, the value NPSV_t serves as a scalar indicator of the magnitude of SOP variation between two adjacent sampling points, i.e., during the interval $[t-1, t]$. To quantify this, we define the polarization intensity (the norm of the Stokes vector) at time instance τ i.e $S_{0,\tau}$ as $S_{0,\tau} = \sqrt{S_1^2(\tau) + S_2^2(\tau) + S_3^2(\tau)}$, where S_1, S_2, and S_3 are the normalized Stokes parameters corresponding to the horizontal/vertical, diagonal/anti-diagonal, and right-/left-handed circular polarization components, respectively. The normalized polarization magnitude at time t is then given by

$$A_t = \frac{S_1^2(t) + S_2^2(t) + S_3^2(t)}{S_{0,t}} \tag{1}$$

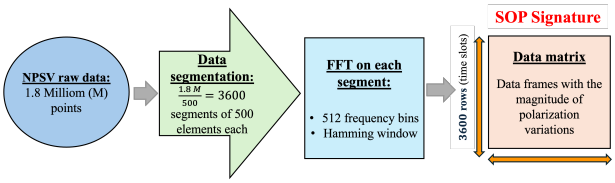


Fig. 3. Data collection and DSP processing pipeline, from raw NPSV data to the SOP signature used for ML classification.

and likewise for the previous sample:

$$A_{t-1} = \frac{S_1^2(t-1) + S_2^2(t-1) + S_3^2(t-1)}{S_{0,t-1}} \tag{2}$$

The resulting normalized polarization state variation is computed as

$$\text{NPSV}t = A_t - A_{t-1}. \tag{3}$$

Although the 200 Gbps Dual-Polarization (DP)-16-Quadrature Amplitude Modulation (16QAM) modulated signal exhibits rapid polarization fluctuations at the symbol rate (tens of GHz), these are effectively averaged out by our optical analyzer, which operates at a sampling rate of 2 ksamples/s and bandwidth of 1–2 kHz. As a result, our measurements isolate only the slow SOP drifts caused by environmental or mechanical disturbances below 5 kHz. These low-frequency effects are preserved and comparable in both modulated and unmodulated channels. In both cases, these vibrations cause a rotation of the polarization state along the fiber, causing again a movement of the NPSV SOP state, which we can detect using the variation of S1, S2 and S3 parameters, which are measured by our device.

To analyze the spectral features of the SOP variations and transfer the signal from the time to the frequency domain, each 0.5-second segment (comprising 500 NPSV values) undergoes Fast Fourier Transform (FFT) processing with a Hamming window [38]. Each FFT generates 512 frequency components. This results in a time-frequency representation of size $3{,}600 \times 512$, where each row corresponds to a 0.5-second time slot and each column to a specific frequency bin. These spectral profiles form what we refer to as *SOP signatures*. Fig. 3 summarizes this data-processing pipeline, illustrating the transformation from raw NPSV measurements (1.8 million samples) into 3,600 time segments, followed by FFT-based spectral decomposition with 512 frequency bins. The resulting $3{,}600 \times 512$ dataset captures the spectral magnitude of polarization variations over time and constitutes the final SOP signature used in the analysis.

The generated SOP signatures are forwarded to the ML analyzer for event and signal modality classification. The ML analysis is conducted in two stages. First, a data pre-processing pipeline aggregates and prepares the input

SOP signatures for ML investigation, including partitioning of the dataset into training and testing subsets. In the second stage, a suite of supervised ML classification algorithms, ranging from ensemble methods to kernel-based and linear models, is evaluated to identify the most suitable classifier for each dataset. The selection is based on the highest classification accuracy and overall performance across precision, recall, and F1-score, as detailed in Section VI. The top-performing model is then used to infer the class of unseen samples and compute the corresponding performance metrics.

IV. Experimental Scenarios

A. Types of Disturbances

Using the described experimental setup, we collect SOP signatures for eight types of events encompassing normal operating conditions and abnormal, harmful events. We consider two classes of normal fiber activity: relaxed fiber (*rlx*) and soft bending (*sbd*). The relaxed condition serves as a baseline reference, capturing only routine background noise in the absence of any deliberate physical disturbances. In contrast, soft bending reflects benign mechanical interactions that are typically encountered during routine handling and maintenance of fiber installations. To simulate such conditions typically found in patch panels, a fiber segment was gently bent by hand to a curvature radius of approximately 2 cm. This action was repeated at 10-second intervals, mimicking handling behavior commonly exhibited by data center technicians.

The abnormal events we consider encompass eavesdropping attempts and potentially harmful vibrations. To simulate an eavesdropping attempt (*eav*), we adopt the mechanical interaction described in [39], where an attacker breaches the outer layer of a standard G.652 fiber and introduces a controlled bend in the internal core with a curvature radius of 4 mm and a bending angle of 25 degree. This controlled deformation enables signal leakage suitable for covert interception. This configuration allows for successful signal tapping while remaining nearly undetectable under standard power monitoring.

To simulate harmful vibrations, we introduce a mechanical disturbance at 80 Hz (*80vb*), which is equivalent to a frequency commonly associated with heavy machinery such as excavators. These machines present a tangible threat to optical fiber infrastructure, as their activity may unintentionally damage or sever fiber cables. The dominant vibration frequency stems from the rotational speed of the engine, which typically operates at around 4,800 Revolutions Per Minute (RPM), equivalent to 80 Hz. To replicate this real-world scenario in a controlled environment, we positioned a loudspeaker 2 to 4 cm away from the fiber under test and generated an 80 Hz acoustic signal. The signal's intensity corresponds to normal human conversation and is a conservative approximation, considering that actual excavators generate significantly higher vibration amplitudes. The goal of this setup is to evaluate the system's ability to detect early-stage mechanical interference before it escalates into service disruption.

TABLE I
Statistical properties of NPSV obtained for modulated and unmodulated signals in different scenarios.

Event	Mean (μ)	Standard Deviation (σ)	Skewness (γ)	Kurtosis (κ)
rlx_u	3.77	18.38	-9.64	276.24
rlx_m	4.01	3.60	-0.29	-0.59
eav_u	3.93	11.00	-4.97	232.43
eav_m	4.03	2.53	-0.12	0.13
sbd_u	-58.12	144.54	-2.52	5.59
sbd_m	4.00	3.01	0.10	-0.14
$80vb_u$	3.90	19.20	-0.08	0.98
$80vb_m$	4.00	3.23	0.01	-0.50

B. Statistical Analysis of the SOP Signatures of Unmodulated vs. Modulated Signals

To evaluate the effect of signal modulation on the statistical behavior of polarization signatures, we perform a comparative analysis of the four collected signatures for unmodulated (denoted by rlx_u, eav_u, sbd_u, $80vb_u$) and modulated (denoted by rlx_m, eav_m, sbd_m, $80vb_m$) signal configurations. We analyze the NPSV distribution for each event and extract four key statistical parameters from these distributions: the mean, standard deviation, skewness, and kurtosis, which offer a comprehensive description of the shape and dynamics of the distribution. The mean μ quantifies the average magnitude of variation and serves as a measure of the central tendency of the distribution. The standard deviation σ captures the extent of dispersion around the mean, reflecting the variability in polarization dynamics. Skewness γ describes the asymmetry of the distribution; negative values indicate a longer tail on the low NPSV values (i.e., more frequent low-magnitude deviations), while positive values suggest a longer tail on high NPSV values. Finally, kurtosis κ describes how sharply peaked a distribution is and how often extreme values occur. It measures the tendency of a distribution to produce outliers by describing the heaviness of its tails relative to a normal distribution ($\kappa = 0$). High kurtosis value means that the distribution has a sharp peak and more values far from the average (i.e., more outliers). Low or negative kurtosis value means that the distribution is flatter, with fewer extreme values.

Table I summarizes the statistical characteristics of the NPSV distributions for the modulated and unmodulated signal modalities under the four collected event signatures. In all conditions, the modulated signals con-

sistently exhibit markedly lower standard deviations, reduced kurtosis, and skewness values closer to zero than those of the unmodulated counterpart. These patterns suggest that modulation plays a significant role in suppressing stochastic polarization fluctuations, yielding more symmetric and approximately Gaussian-like NPSV distributions. This suppression effect is not a consequence of intrinsic polarization behavior, but rather stems from limitations in the resolution of the polarization analyzer. The mechanical disturbances investigated in this paper, such as fiber bending and vibrations, induce SOP variations at much lower frequencies at rates in the kHz, which is a very slow variation compared to the 200 Gbps high-speed modulation scheme of DP-16QAM. As a result, modulation indirectly contributes to the reduction of apparent stochasticity by decorrelating slowly varying polarization states and leading to smoother NPSV signatures in this case.

In the relaxed condition, the unmodulated signal displays considerable variability ($\sigma = 18.38$) and an extremely peaked and heavy-tailed ($\kappa = 276.24$), indicating heavy tails and the presence of rare, large deviations from the mean. In contrast, the modulated counterpart is significantly more stable ($\sigma = 3.60$) and near-Gaussian ($\kappa = -0.59$), with a slightly elevated mean ($\mu = 4.01$ vs. 3.77). Similar stabilizing effects of modulation are observed in the eavesdropping and 80 Hz vibration scenarios, where modulation reduces standard deviation by approximately 77% and 83%, respectively.

The soft bending event reveals the most striking contrast. The unmodulated signal exhibits a dramatically negative mean ($\mu = -58.12$) and very high dispersion ($\sigma = 144.54$), reflecting intense and irregular polarization disturbances. Conversely, the modulated case maintains a stable distribution with a baseline mean ($\mu = 4.00$) and low variance ($\sigma = 3.01$), closely resembling the relaxed condition. However, this superficial similarity in statistics can be misleading: while the modulated signal appears stable in terms of mean and variance, the full polarization trajectory may still contain subtle but relevant temporal patterns.

Fig. 4 illustrates the statistical trends discussed in Table I by showing the NPSV histograms and their Gaussian fits for modulated and unmodulated signals under the four events. The plot has been scaled uniformly in the two axis to allow for a comparative analysis. In each subplot, the unmodulated distributions exhibit broader, often asymmetric profiles with heavier tails. By contrast, the modulated signal distributions are consistently narrower, more symmetric, and closely aligned with their Gaussian approximations, indicating reduced variability and improved stability in capturing polarization behavior. These visual observations support the numerical findings in Table I and highlight the benefits of modulation in achieving consistent polarization signature profiles, indicating that it might be more suitable for anomaly detection.

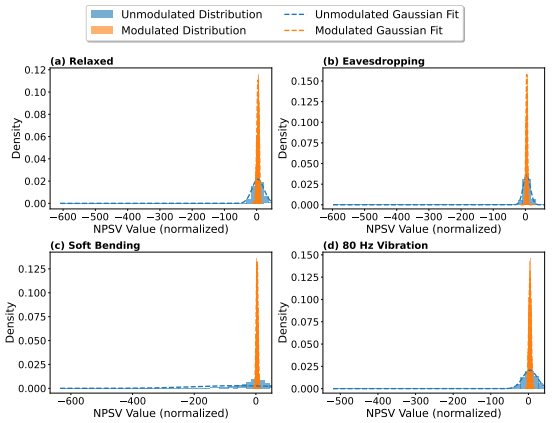


Fig. 4. NPSV histograms and Gaussian fits for unmodulated and modulated conditions in four event types: (a) relaxed, (b) eavesdropping, (c) soft bending, and (d) 80 Hz vibration.

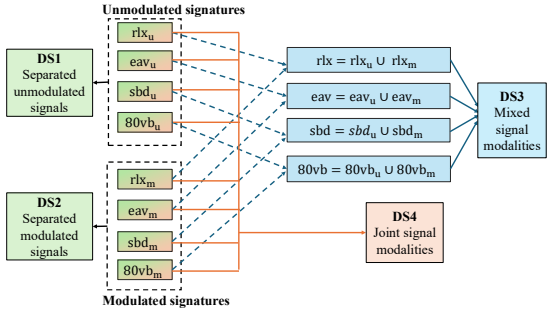


Fig. 5. The four datasets for evaluating ML-based classification of SOP signatures in modulated and unmodulated signals.

V. Dataset Definition and Pre-Processing

This section introduces four systematically designed datasets, each developed to investigate a distinct aspect of ML-driven classification for SOP-based optical fiber monitoring. The goal is to assess how signal modulation influences the effectiveness of ML techniques in distinguishing physical events, learning polarization-based features, and generalizing for different signal modalities. The datasets are constructed to evaluate the performance of the ML classification task: from separated analysis of signal types, to mixed-modality configuration, and finally to a fully discriminative and signal-aware framework. For each dataset, a corresponding data pre-processing pipeline is implemented to support its specific structure and objective. The four dataset configurations are illustrated in Fig. 5 and are described as follows.

A. Datasets 1 and 2: Separate Signal Modalities

Datasets 1 (DS1) and 2 (DS2) represent a scenario where the modulated and unmodulated optical signals are analyzed separately. Each dataset contains four SOP event signatures. This separation enables a direct comparison of how signal modulation influences the discriminability of physical disturbances in the SOP signatures and the learnability of their spectral patterns by ML models. The research question we address with these datasets is: *Do the inherent differences in polarization dynamics between modulated and unmodulated signals lead to measurable variations in the performance of ML-based classification of SOP signatures?*

An identical pre-processing pipeline was applied separately to the modulated and unmodulated signal sets. As detailed in Section III-B, each signature comprises 3,600 samples per event, with each sample corresponding to 0.5 ms of SOP variations and represented by 512 frequency-domain features derived from spectral analysis. This results in 14,400 samples for each signal type. The two datasets were independently shuffled and partitioned using an 80/20 split, yielding 11,520 training samples and 2,880 testing samples per set.

B. Dataset 3: Mixed Signal Modalities

Dataset 3 (DS3) extends the scope of the classification pipeline by combining modulated and unmodulated polarization signatures into a unified dataset. In this setup, the model is trained and assessed on a combined scenario comprising all eight collected SOP signatures, grouped into four distinct event classes, each defined by the union of its modulated and unmodulated instances.

Unlike DS1 and DS2, DS3 introduces mixed signal modalities during training and inference, i.e., equivalent classes from the modulated and unmodulated scenarios are merged into a single class. The goal is to evaluate whether a unified classifier can effectively learn class boundaries from both signal types, regardless of their underlying characteristics. This mixed modality setup reflects practical deployment scenarios where the signal format may vary or be unknown. The research question we aim to address by models trained with DS3 is: *Can a single ML model accurately classify SOP signatures when modulated and unmodulated signals are present during training and inference?* The results of a model trained with this dataset will provide insights into the performance of ML-driven SOP-based fiber sensing in heterogeneous signal environments.

To prepare DS3, each of the eight collected SOP signatures was first independently partitioned into training and testing subsets using an 80/20 split. This yielded 2,880 training samples and 720 testing samples per signature. After splitting, samples from modulated and unmodulated signals were merged within each event type to form four final event classes, each comprising 5,760 training samples and 1,440 testing samples. This results in a unified dataset containing 23,040 training samples and 5,760 testing samples.

C. Dataset 4: Joint Signal Modalities

Dataset 4 (DS4) contains eight classes, where each collected SOP signature is treated as a distinct class, explicitly distinguishing both the event type and its associated signal modality in the class. Unlike the DS3, which merges modulated and unmodulated variants of the same event into a single class, DS4 should capture fine-grained distinctions between signatures influenced by signal modality. This structure reflects a deployment scenario where both the nature of the event and the signal type are relevant and unknown thereby requiring classification. The central research question to be answered by training a model with DS4 is: *Can an ML model simultaneously distinguish both event type and signal modality when polarization signatures from both domains are presented as separate classes?* This approach will allow for a detailed investigation into the model's ability to distinguish event variations introduced by modulation effects.

To prepare DS4, the eight collected SOP signatures were partitioned individually using the same pre-processing pipeline of DS1 and DS2where each signature was partitioned into training and testing subsets using an 80/20 split, yielding 2,880 training samples and 720 testing samples per class. In total, the dataset contains 23,040 training and 5,760 testing samples.

VI. Results

This section presents the evaluation results for four scenarios trained with the datasets introduced in Section V. To identify the most suitable classification algorithm for each scenario, we conduct a comprehensive benchmarking of ten supervised ML classifiers available in the Scikit-learn library. The evaluated methods span a diverse range of learning paradigms, including ensemble learners: Random Forest (RF), Extra Trees Classifier (ETC), Histogram Gradient Boosting (HGB), eXtreme Gradient Boosting (XGBoost), Gradient Boosting (GB); kernel-based models: Support Vector Machine (SVM); linear classifiers: Logistic Regression (LR), Linear Discriminant Analysis (LDA); distance-based techniques: k-Nearest Neighbors (KNN); and tree-based learners: Decision Tree (DT). These classifiers are selected based on their previously documented performance in time-frequency domain tasks and their compatibility with the spectral representations extracted from the SOP data. For each dataset, the classifiers were trained and tested independently using 5-fold cross-validation to ensure robust

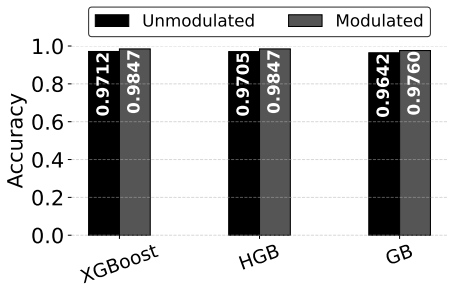
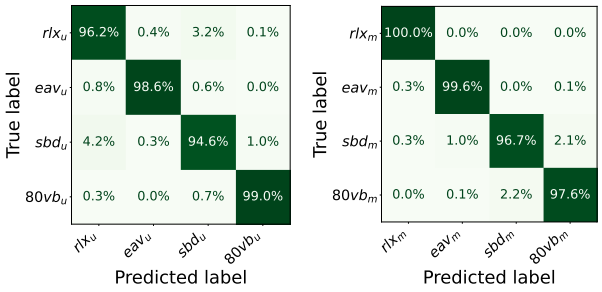


Fig. 6. Classification accuracy on separated signal modalities scenario for the top three performing models: eXtreme Gradient Boosting (XGBoost), Histogram Gradient Boosting (HGB), and Gradient Boosting (GB).



(a) Unmodulated (b) Modulated

Fig. 7. Confusion matrices of the eXtreme Gradient Boosting (XGBoost) classifier for (a) unmodulated and (b) modulated signals.

performance assessment. The ML classifier yielding the highest test accuracy was selected as the representative classifier for that dataset, and its confusion matrix is reported in the corresponding subsection.

A. Scenarios 1 and 2: Classification Performance for Separated Signal Modalities (DS1 and DS2)

Scenarios 1 and 2 evaluate the performance of ML-based SOP classification when polarization signatures are analyzed separately for modulated and unmodulated signals. Independent classifiers were trained using their default hyperparameters for DS1 and DS2. Fig. 6 shows the classification accuracy in the test set for the three ML classifiers that achieve the highest accuracy across both DS1 and DS2: XGBoost, HGB, and GB. Among them, XGBoost demonstrates superior performance, attaining an accuracy of 97.12% for unmodulated and 98.47% for modulated signals. A deeper insight into the performance of XGBoost is offered by the confusion matrices in Fig. 7. In both cases, the classifiers exhibit strong class separability with minimal confusion. Few misclassifications were observed primarily between the *sbd* and *80vb* events. Nonetheless, the impact of these errors on overall detection reliability is negligible.

These results indicate that signal modulation, such as the DP-16QAM 200 Gbps signal used in the experiment, does not hinder the ability of ML classifiers to distinguish between SOP signatures associated with different physical events. In fact, modulated signals in this study exhibited more stable SOP behavior than their unmodulated counterparts, contributing to consistently high classification performance. While modulated signals are often associated with higher structural complexity at the symbol level, their temporal averaging effect appears to suppress high-frequency polarization noise, leading to smoother SOP trajectories. Consequently, low-frequency SOP variations caused by external disturbances remain clearly distinguishable. These findings reinforce the feasibility of ML-based SOP analysis for anomaly detection in coherent optical networks, where modulation is an inherent aspect of system operation.

B. Scenario 3: Classification Performance for Event Classification in Mixed Signal Modalities (DS3)

Scenario 3 investigates the performance of ML classifiers in a unified classification task where both modulated and unmodulated signals are included for each physical event class. Table II summarizes the performance of the tested classifiers in terms of accuracy, precision, recall, F1-score, and computational cost (training and inference times). The classifiers are sorted in descending order of accuracy. Compared to the other models, HGB and XGBoost achieved the two highest accuracy values of 98.09% and 97.93%, respectively. The confusion matrix of the best-performing HGB is shown in Fig. 8. The model achieves excellent class separability for all four event types. The relaxed (*rlx*) and eavesdropping (*eav*) classes are recognized with near-perfect accuracy (98.8% and 99.0%, respectively). Minor confusion is observed between soft bending (*sbd*) and 80 Hz vibration (*80vb*), with at most 2.2% of misclassified samples.

These results demonstrate that ML models are able to learn discriminative features of both unmodulated and modulated signals. This supports the feasibility of deploying signal-agnostic SOP-based ML monitoring systems in network environments with unmodulated and modulated signals.

C. Scenario 4: Classification Performance for Joint Signal Modalities (DS4)

Scenario 4 evaluates the classification performance in a fine-grained, eight-class scenario, where each unique event-modality combination (e.g., rlx_u, rlx_m) is treated as a distinct class. Table III presents the performance of all evaluated classifiers for this setup. Among the tested models, XGBoost achieved the highest overall accuracy, followed closely by HGB, and GB. In addition to

TABLE II
PERFORMANCE BENCHMARKING OF SUPERVISED ML CLASSIFIERS FOR SCENARIO 3 (MIXED SIGNAL MODALITIES CLASSIFICATION).

Classifier	Accuracy	Precision	Recall	F1-score	Training Time (s)	Inference Time (s)
Histogram Gradient Boosting	0.9809	0.9809	0.9809	0.9809	18.38	0.057
XGBoost	0.9793	0.9794	0.9793	0.9793	7.87	0.024
Gradient Boosting	0.9641	0.9640	0.9641	0.9640	1555.49	0.042
Random Forest	0.9616	0.9617	0.9616	0.9615	40.46	0.066
Extra Trees Classifier	0.9498	0.9509	0.9498	0.9490	7.29	0.109
SVM Classifier	0.9479	0.9477	0.9479	0.9476	55.99	17.99
Decision Tree	0.9028	0.9030	0.9028	0.9028	18.89	0.005
K-Nearest Neighbors	0.8365	0.8417	0.8365	0.8298	0.023	0.655
Logistic Regression	0.8193	0.8194	0.8193	0.8193	40.51	0.021
Linear Discriminant Analysis	0.7814	0.7813	0.7814	0.7810	1.70	0.012

TABLE III
PERFORMANCE BENCHMARKING OF SUPERVISED ML CLASSIFIERS FOR SCENARIO 4 (JOINT SIGNAL MODALITIES CLASSIFICATION).

Classifier Name	Accuracy	Precision	Recall	F1-score	Training Time (s)	Inference Time (s)
XGBoost	0.9804	0.9804	0.9804	0.9804	13.0668	0.0401
Histogram Gradient Boosting	0.9793	0.9794	0.9793	0.9793	21.8151	0.0631
Gradient Boosting	0.9688	0.9687	0.9688	0.9687	2394.0043	0.0598
Random Forest	0.9658	0.9661	0.9658	0.9658	39.8580	0.0678
SVM Classifier	0.9483	0.9483	0.9483	0.9481	24.5029	17.2565
Extra Trees Classifier	0.9467	0.9488	0.9467	0.9462	7.4505	0.1216
Decision Tree	0.9057	0.9057	0.9057	0.9056	17.8165	0.0053
Linear Discriminant Analysis	0.8807	0.8817	0.8807	0.8794	1.0988	0.0115
Logistic Regression	0.8806	0.8807	0.8806	0.8805	56.5408	0.0043
K-Nearest Neighbors	0.8365	0.8458	0.8365	0.8300	0.0272	0.8149

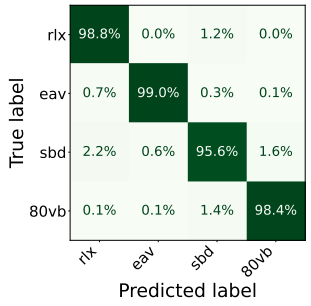


Fig. 8. Confusion matrix of the Histogram Gradient Boosting (HGB) classifier for mixed signal modalities classification scenario.

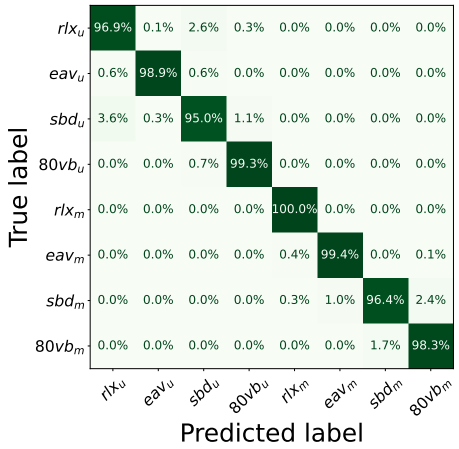


Fig. 9. Confusion matrix for scenario 4 showing classification performance using the eXtreme Gradient Boosting (XGBoost) classifier.

delivering the best classification performance, XGBoost maintained a relatively modest computational footprint, requiring 13.07 s for training and only 0.04 s for inference. The confusion matrix of XGBoost is shown in Fig. 9. XGBoost achieves high discrimination for nearly all classes. The modulated signature classes have near-perfect accuracy above 98.9%, with 100% accuracy for the rlx_m class. The worst-case misclassification in the unmodulated signals is observed for the sbd_u class, where 3.6% of samples are incorrectly labeled as rlx_u. In the modulated signal, the highest confusion occurs between sbd_m and $80vb_m$, with a misclassification rate of 2.4%. These errors indicate strong class separability even when both signal modality and event type are jointly classified. Interestingly, there is no confusion between unmodulated and modulated signals (100% accuracy), i.e., misclassification is only observed within the same modality. These results confirm that it is possible for a classifier to accurately classify both modality and physical events simultaneously. The findings reinforce the conclusion that modulation does not significantly affect the performance of ML-based classifiers.

VII. CONCLUSIONS

This paper investigates the impact of optical signal modulation on ML classification of harmful events by us-

ing SOP signatures. While prior studies have largely focused on unmodulated light sources, our work addresses the critical gap of evaluating SOP-based monitoring under practical, high-speed modulated conditions. Using a real-world experimental setup on a metro network, we collected polarization signatures from modulated and unmodulated signals subjected to identical physical disturbances. Statistical analysis revealed that modulation stabilizes the SOP distribution by suppressing fluctuations. A suite of classifiers was assessed on four datasets with separated, mixed, and joint modalities. The assessed ML algorithms consistently achieved high classification accuracy in benign and malicious events, confirming that the spectral features extracted from SOP dynamics remain distinguishable by ML algorithms regardless of modulation effects. While real-world deployment challenges like large-scale polarization data acquisition, modulation diversity, and ML control-plane integration remain, our results show that accurate anomaly classification is feasible under realistic traffic conditions.

REFERENCES

[1] D. Dahan and U. Mahlab, "Security threats and protection procedures for optical networks," *IET Optoelectronics*, vol. 11, no. 5, pp. 186–200, 2017.

[2] W. Lee, S. I. Myong, J. C. Lee, and S. Lee, "Identification method of non-reflective faults based on index distribution of optical fibers," *Optics express*, vol. 22, no. 1, pp. 325–337, 2014.

[3] K. Abdelli, H. Grießer, C. Tropschug, and S. Pachnicke, "Optical fiber fault detection and localization in a noisy OTDR trace based on denoising convolutional autoencoder and bidirectional long short-term memory," *IEEE Journal of Lightwave Technology*, vol. 40, no. 8, pp. 2254–2264, 2021.

[4] 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, pp. 365–375, 2022.

[5] B. Steinar, "Locating disturbances in optical fibres," U.S. Patent WO2 022 185 075A1, Sep 9, 2022.

[6] Y. Aono, E. Ip, and P. Ji, "More than communications: environment monitoring using existing optical fiber network infrastructure," in *Optical Fiber Communications Conference and Exhibition (OFC)*, 2020, p. W3G.1.

[7] S. Pellegrini, L. Minelli, L. Andrenacci, G. Rizzelli, D. Pilori, G. Bosco, L. D. Chiesa, C. Crognale, S. Piciaccia, and R. Gaudino, "Overview on the state of polarization sensing: application scenarios and anomaly detection algorithms," *J. Opt. Commun. Netw.*, vol. 17, no. 2, pp. A196–A209, Feb 2025.

[8] L. Sadighi, S. Karlsson, C. Natalino, and M. Furdek, "Machine learning-based polarization signature analysis for detection and categorization of eavesdropping and harmful events," in *Optical Fiber Communications Conference and Exhibition (OFC)*, 2024, p. M1H.1.

[9] L. Andrenacci, D. Pilori, S. Pellegrini, L. Minelli, G. Bosco, C. Crognale, S. Piciaccia, and R. Gaudino, "Comparison between phase and polarization sensing using coherent transceivers over deployed metro fibers," in *Optical Fiber Communication Conference*. Optica Publishing Group, 2024, pp. M2K–2.

[10] S. Karlsson, M. Andersson, R. Lin, L. Wosinska, and P. Monti, "Detection of abnormal activities on a SM or MM fiber," in *Optical Fiber Communication Conference (OFC)*, 2023, p. M3Z.6.

[11] D. Rafique, T. Szyrkowiec, H. Grießer, A. Autenrieth, and J.-P. Elbers, "Cognitive assurance architecture for optical network fault management," *Journal of Lightwave Technology*, vol. 36, no. 7, pp. 1443–1450, 2018.

[12] EXFO Inc, "Tunable DFB Laser Sources – Application Note 012," available here.

[13] B. Szafraniec, T. S. Marshall, and B. Nebendahl, "Performance monitoring and measurement techniques for coherent optical systems," *Journal of Lightwave Technology*, vol. 31, no. 4, pp. 648–663, 2013.

[14] L. Moeller and B. Bakhshi, "Polarization modulation of supervisory signals for reducing interference with data signals," Mar. 12 2019, uS Patent 10,230,472.

[15] HEAnet, "Ireland's National Education and Research Network." [Online]. Available: https://www.heanet.ie/about

[16] K. Abdelli, M. Lonardi, J. Gripp, S. Olsson, F. Boitier, and P. Layec, "Breaking boundaries: harnessing unrelated image data for robust risky event classification with scarce state of polarization data," in *European Conference on Optical Communications (ECOC)*, vol. 2023, 2023, pp. 924–927.

[17] K. Abdelli, M. Lonardi, F. Boitier, D. Correa, J. Gripp, S. Olsson, and P. Layec, "Vision transformers for anomaly classification and localization in optical networks using SOP spectrograms," *Journal of Lightwave Technology*, vol. 43, no. 4, pp. 1902–1914, 2024.

[18] L. Sadighi, S. Karlsson, C. Natalino, L. Wosinska, M. Ruffini, and M. Furdek, "Detection and classification of eavesdropping and mechanical vibrations in fiber optical networks by analyzing polarization signatures over a noisy environment," in *European Conference on Optical Communication (ECOC)*, 2024, pp. 527–530.

[19] L. Sadighi, S. Karlsson, L. Wosinska, and M. Furdek, "Machine learning analysis of polarization signatures for distinguishing harmful from non-harmful fiber events," in *International Conference on Transparent Optical Networks (ICTON)*, 2024.

[20] W. Qin, Q. Zhang, W. Hou, X. Zhang, and X. Gong, "Convolutional neural networks for fiber-bending eavesdropping attacks detection in coherent optical communication systems," in *International Conference on Ubiquitous Communication (Ucom)*, 2024, pp. 342–345.

[21] K. Abdelli, M. Lonardi, J. Gripp, S. Olsson, F. Boitier, and P. Layec, "Unsupervised anomaly detection and localization with generative adversarial networks," *arXiv preprint arXiv:2409.03657*, 2024.

[22] X. Chen, B. Li, R. Proietti, Z. Zhu, and S. J. B. Yoo, "Self-taught anomaly detection with hybrid unsupervised/supervised machine learning in optical networks," *Journal of Lightwave Technology*, vol. 37, no. 7, pp. 1742–1749, 2019.

[23] H. Song, R. Lin, L. Wosinska, P. Monti, M. Zhang, Y. Liang, Y. Li, and J. Zhang, "Cluster-based unsupervised method for eavesdropping detection and localization in WDM systems," *Journal of Optical Communications and Networking*, vol. 16, no. 10, pp. F52–F61, 2024.

[24] L. Sadighi, S. Karlsson, C. Natalino, and M. Furdek, "ML-based state of polarization analysis to detect emerging threats to optical fiber security," *TechRxiv*, 2025, DOI: 10.36227/techrxiv.175099843.35967234.v1.

[25] A. Tomasov, P. Dejdar, P. Munster, T. Horvath, P. Barcik, and F. Da Ros, "Enhancing fiber security using a simple state of polarization analyzer and machine learning," *Optics & Laser Technology*, vol. 167, p. 109668, 2023.

[26] L. Sadighi, S. Karlsson, C. Natalino, L. Wosinska, M. Ruffini, and M. Furdek, "Deep learning for detection of harmful events in real-world, noisy optical fiber deployments," *Journal of Lightwave Technology*, pp. 1–9, 2025.

[27] C. J. Carver and X. Zhou, "Polarization sensing of network health and seismic activity over a live terrestrial fiber-optic cable," *Communications Engineering*, vol. 3, no. 1, p. 91, 2024.

[28] L. Sadighi, C. Natalino, S. Karlsson, L. Wosinska, M. Ruffini, and M. Furdek, "AI/ML-based state-of-polarization monitoring in optical networks: Concepts and challenges," in *Optical Fiber Communication Conference (OFC) 2025*, 2025, p. M3F.6.

[29] M. Mazur, D. Wallberg, L. Dallachiesa, E. Börjeson, R. Ryf, M. Bergroth, B. Josefsson, N. K. Fontaine, H. Chen, D. T. Neilson, J. Schröder, P. Larsson-Edefors, and M. Karlsson, "Real-

time monitoring of cable break in a live network using a coherent transceiver prototype," in *2024 Optical Fiber Communications Conference and Exhibition (OFC)*, 2024, pp. 1–3.

[30] A. Rode, M. Farsi, V. Lauinger, M. Karlsson, E. Agrell, L. Schmalen, and C. Häger, "Machine learning opportunities for integrated polarization sensing and communication in optical fibers," *Optical Fiber Technology*, vol. 90, p. 104047, 2025.

[31] C. B. Czegledi, M. Karlsson, E. Agrell, and P. Johannisson, "Polarization drift channel model for coherent fibre-optic systems," *Scientific Reports*, vol. 6, no. 1, p. 21217, 2016, DOI: 10.1038/srep21217.

[32] D. Waddy, P. Lu, L. Chen, and X. Bao, "Fast state of polarization changes in aerial fiber under different climatic conditions," *IEEE Photonics Technology Letters*, vol. 13, no. 9, pp. 1035–1037, 2001.

[33] J. Wuttke, P. Krummrich, and J. Rosch, "Polarization oscillations in aerial fiber caused by wind and power-line current," *IEEE Photonics Technology Letters*, vol. 15, no. 6, pp. 882–884, 2003.

[34] P. M. Krummrich, D. Ronnenberg, W. Schairer, D. Wienold, F. Jenau, and M. Herrmann, "Demanding response time requirements on coherent receivers due to fast polarization rotations caused by lightning events," *Opt. Express*, vol. 24, no. 11, pp. 12 442–12 457, May 2016.

[35] S. J. Savory, "Digital filters for coherent optical receivers," *Opt. Express*, vol. 16, no. 2, pp. 804–817, Jan 2008.

[36] K. Kikuchi, "Performance analyses of polarization demultiplexing based on constant-modulus algorithm in digital coherent optical receivers," *Opt. Express*, vol. 19, no. 10, pp. 9868–9880, May 2011.

[37] CONNECT Centre for Future Networks and Communications, "Open Ireland Testbed," available here.

[38] P. Podder, T. Z. Khan, M. H. Khan, and M. M. Rahman, "Comparative performance analysis of hamming, hanning and blackman window," *International Journal of Computer Applications*, vol. 96, pp. 1–7, 2014.

[39] S. Karlsson, R. Lin, L. Wosinska, and P. Monti, "Eavesdropping G.652 vs. G.657 fibres: a performance comparison," in *International Conference on Optical Network Design and Modeling (ONDM)*, 2022, p. Tu1.4.