Optical Performance Monitoring in Mode Division Multiplexed Optical Networks

Waddah S. Saif, Amr M. Ragheb, Tariq A. Alshawi, and Saleh A. Alshebeili

***Abstract*—This article considers, for the first time, optical performance monitoring (OPM) in few mode fiber (FMF)-based optical networks. 1-D features vector, extracted by projecting a 2-D asynchronous in-phase quadrature histogram (IQH), and the 2D IQH are proposed to achieve OPM in FMF-based network. Three machine learning algorithms are employed for OPM and their performances are compared. These include support vector machine, random forest algorithm, and convolutional neural network. Extensive simulations are conducted to monitor optical to signal ratio (OSNR), chromatic dispersion (CD), and mode coupling (MC) for dual polarization-quadrature phase shift keying (DP-QPSK) at 10, 12, 16, 20, and 28 Gbaud transmission speeds. Besides, M-ary quadrature amplitude modulation (M = 8 and 16) is considered. Also, the OPM accuracy is investigated under different FMF channel conditions including phase noise and polarization mode dispersion. Simulation results show that the proposed 1D projection features vector provides better OPM results than those of the widely used asynchronous amplitude histogram (AAH) features. Furthermore, it has been found that the 2D IQH features outperform the 1D projection features but require larger number of features samples. Additionally, the effect of fiber nonlinearity on the OPM accuracy is investigated. Finally, OPM using the 2D IQH features has been verified experimentally for 10 Gbaud DP-QPSK signal. The obtained results show a good agreement between both simulation and experimental findings.**

***Index Terms*—Coherent optical communication, few mode fiber, optical performance monitoring.**

I. Introduction

THE rapid advances in information technology and data usage intensify new challenges and limitations on the single-mode fiber(SMF)-based optical networks [1]. This motivated research and development (R&D) agencies to propose new technologies that utilize fiber space more efficiently. In this regard, few-mode fiber (FMF)-based transmission systems have attracted considerable attention toward achieving high-speed communications utilizing mode division multiplexing (MDM). It takes advantage of the extra degree of freedom provided by several orthogonal modes, and therefore considering each mode as an independent communication channel. Nevertheless, increasing the capacity of optical transmission systems by incorporating MDM will require new network functions and components such as reconfigurable optical add and drop multiplexers (ROADMs) [2]. This causes optical network architectures to be more complex and adaptive [3]. Thus, transmitted optical signals are highly exposed to dynamic optical impairments such as additive spontaneous emission (ASE) noise, chromatic dispersion (CD), etc. In this context, optical performance monitoring (OPM) is found to be a common need in future MDM optical networks. This is to track the amount of signal distortion, hence applying suitable impairment mitigation techniques, allocating resources more efficiently, and optimizing the routing and switching functions [3], [4].

The recent developments in digital coherent receivers and signal processing methods enable OPM of the received signal. Optical signal-to-noise ratio (OSNR) and CD are considered the most crucial parameters that affect the performance of optical fiber communication systems [5]–[8]. OSNR is directly related to the bit-error-rate (BER). Whereas CD arises in each SMF span and has a significant influence on the quality of optical signals [5], [9]. Thus, real-time monitoring is essential to combat CD and maintain optimal channel conditions [5]. Though CD-uncompensated links are becoming more common, the deployment of these fiber types are still in its fancy. Moreover, off-the-shelf FMF are suffering form CD impairments [10]. Hence, monitoring CD is required in the current and near future optical networks.

Various approaches of OSNR estimation have been proposed, which can be classified into two types: data-aided (DA) [11], [12] and non-data-aided (NDA) approaches [13]–[15]. The former approach needs synchronization in order to extract training sequences. Besides, it sacrifices the spectral efficiency (SE) owing to data overhead. On the other side, NDA methods have the benefit of in-service estimation without reducing the SE. These include: (i) statistical moments based approaches [13], [14], which experience degradation of performance when considered for multilevel constellations [15], (ii) empirical cumulative distribution function based approaches [15], which only suit multilevel modulation schemes, and (iii) correlation-based OSNR estimation methods [16], which fail to distinguish the noise level in the presence of fiber nonlinearity [17]. The next important aspect is CD monitoring. To achieve this, training sequence-based approaches have been proposed [18]–[20]. In [18] a constant

Manuscript received June 23, 2020; revised September 8, 2020; accepted September 25, 2020. Date of publication September 29, 2020; date of current version January 15, 2021. This work was supported by the Deputyship for Research & Innovation, Ministry of Education in Saudi Arabia, under Grant DRI-675KSU-1088. *(Corresponding author: Waddah S. Saif.)*

Waddah S. Saif, Tariq A. Alshawi, and Saleh A. Alshebeili are with the Electrical Engineering Department, King Saud University, Riyadh 11421, Saudi Arabia, and also with KACST-TIC in Radio Frequency, and Photonics for the e-Society, King Saud University, Riyadh 11421, Saudi Arabia (e-mail: wsaif@ksu.edu.sa; talshawi@ksu.edu.sa; dsaleh@ksu.edu.sa).

Amr M. Ragheb is with KACST-TIC in Radio Frequency, and Photonics for the e-Society, King Saud University, Riyadh 11421, Saudi Arabia (e-mail: aragheb@ksu.edu.sa).

Color versions of one or more of the figures in this article are available online at https://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JLT.2020.3027725

amplitude zero-autocorrelation (CAZAC) sequence is used. This method is robust to ASE noise. However, it requires high-speed analog to digital converters (ADCs). A fractional Fourier transform (FrFT) based CD estimation method is introduced in [19]. While this method decreases the required sampling rates of the ADCs, the need to transform the signals into the fractional domain and find the optimal order of the energy concentration function makes it of high computational complexity [20]. A CD estimation based delay-tap sampling (DTS) approach is proposed in [20]. This technique can be applied under the ultra-low sampling rate conditions (i.e. about 1/30 of the Nyquist sampling rate). All of the above methods need training sequences in the time domain, which reduces system SE. A blind CD estimation method using spectral cyclostationarity has been proposed in [21]. This approach is fast and does not need a training sequence, however, it requires high-speed ADCs.

On the other hand, machine learning algorithms have been applied to many disciplines and shown to provide superior results compared to traditional digital signal processing (DSP)-based approaches [22]. In the field of optical communications, ML algorithms have been investigated for modulation format identification (MFI) [23], [24] and/or OPM [25]–[27]. The use of ML-based algorithms has shown several advantages compared to traditional OPM approaches [27]. One advantage is in the ability to simultaneously deal with multiple transmission impairments, while conventional OPM techniques have shown limited success [28]. ML-based algorithms were able to provide better results when dealing with non-linear behavior. Also, most of the DSP-based techniques are based on the training sequences that sacrifice SE [29], which is not the case with ML-based approaches. Additionally, recent developments in ML technology have provided new techniques (e.g. deep learning), which can help achieving superior performance. Deep learning techniques have the capability to solve complex models without the need to build exhaustive analytical models; this makes them well-suited to face the challenge imposed by the increased complexity and dynamism of next-generation optical communication [30], [31]. Such advantages have motivated researchers to extensively use ML-based algorithms to monitor and alleviate fiber channel impairments.

Over the last two decades, many ML-based OPM techniques have been proposed. For instance, a deep learning in conjunction with eye diagram is proposed to monitor modulation format, OSNR, roll-off factor (ROF), and timing skew for QAM signals [32]. Although this technique provides high accuracy results, it requires precise time clock recovery. The authors in [33] proposed simultaneous MFI and OSNR monitoring using amplitude histogram with multi-task learning based artificial neural network. This approach is advantageous since it does not need timing recovery and the associated hardware. However, when the received signal is heavily impaired by CD, the distinction between different impairments becomes difficult. Convolutional neural network (CNN) based asynchronous DTS (ADTS) is proposed for OPM in [34]. This method increases the monitoring range owing to its ability to capture information about the detected signal's slope. The downside, however, is that the technique requires two sampling clocks, which increases the cost. The authors in [5] proposed simultaneous OSNR and CD monitoring algorithm using long short-term memory neural network (LSTM-NN). This technique has the advantage that it does not need handcrafted features extraction. However, LSTM models require large amounts of memory. Transfer learning assisted deep neural network (DNN) algorithm is proposed in [35] to perform OSNR monitoring. Transfer learning relies on previous knowledge instead of random initialization, which leads to faster training. This technique provides results with high accuracy over a wide OSNR range. However, it's only used for OSNR monitoring.

So far, the reported OPM methods focus on the use of various ML-based techniques to monitor and estimate optical channel impairments in SMF-based networks. In this work, we aim to investigate ML-based OPM algorithms in the FMF-based optical networks. In contrast to SMF, FMF has other new impairments such as mode coupling (MC), which makes OPM a challenging task. Therefore, the following is considered in this study.

1) We present investigation considering, for the first time, OPM in FMF-based network employing five spatial orthogonal modes (i.e. LP_{01}, LP_{11a}, LP_{11b}, LP_{21a}, and LP_{21b}). We use support vector machine (SVM) trained with one-dimensional (1D) features vector extracted by projecting the two-dimensional (2D) in-phase quadrature histogram (IQH), which combines both the in-phase (I) and quadrature (Q) information, to monitor OSNR, CD, and MC for a 10-Gbaud dual polarization-quadrature phase shift keying (DP-QPSK) system.
2) We consider two types of 1D features vectors. The first type is obtained by taking the horizontal projection of 2D IQH, while the second type is obtained based on the diagonal projection. Then, we compare the monitoring accuracy of both types of 1D features vectors with that of amplitude histogram (AAH) using SVM regressor.
3) We present OPM based on 2D IQH, and compare the monitoring accuracy when three ML-based regressors are used, including the convolutional neural network (CNN), random forest, and SVM algorithms.
4) We assess the effectiveness of the proposed OPM method by testing its performance under different values of phase noise, first-order polarization mode dispersion (PMD), symbol rates, and different modulation schemes.
5) We study the impact of fiber nonlinearity in terms of mode groups self-phase modulation (SPM) and cross-phase modulation (XPM). The OPM performance is investigated under different transmitted optical power.
6) We carry out a proof of concept experiment for OPM using 2D IQH with CNN.

The structure of this paper is organized as follows: Section II provides a background about FMF, feature extraction, and ML algorithms used in this work. The simulation setup is described in Section III. In Section IV, we discuss the obtained results. The proof of concept experiment is presented in Section V. Finally, we conclude the paper in Section VI.

II. OPM in FMF-Based Network

FMF allows the propagation of a finite number of guided propagation modes. These modes are the transverse electric

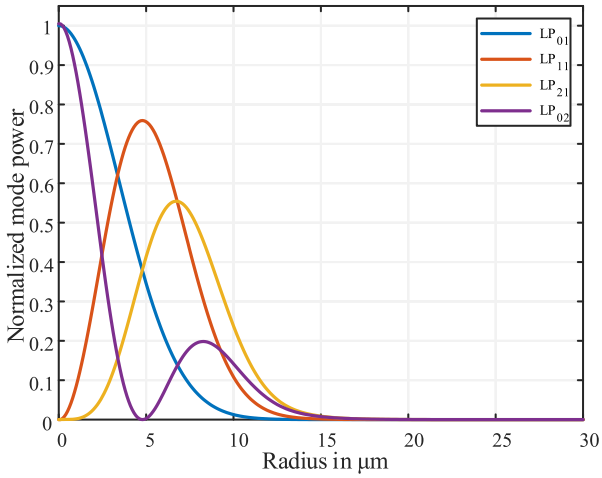

Fig. 1. Normalized Mode Power As Function of the Radius in the Fiber.

field distributions propagating along the fiber axis. The optical fiber modes are described as linearly polarized LP_{lp} waves, where l and p are integer numbers describing the radial and azimuthal indices. The mode intensity for a given LP_{lp} mode can be expressed as [36], [37]

$$\psi(r,\phi) = \begin{cases} F_{lp}(r)\cos(l\phi) \\ F_{lp}(r)\sin(l\phi) \end{cases} \quad (1)$$

where $\cos(.)$ and $\sin(.)$ reflect the choice between two spatial configurations of the mode and F is the mode field, which is inserted into the wave equation, to give [38]

$$F''_{lp}(r) + \frac{F'_{lp}(r)}{r} + \left(n^2(r)k^2 - \frac{l^2}{r^2} - \beta_{lp}^2\right) F_{lp}(r) = 0 \quad (2)$$

where $k = 2\pi/\lambda$, λ is the wavelength, β is the propagation constant, r represents the distance from the center core (i.e $r = 0$ at the center core), and $n(r)$ is the refractive index profile given by

$$n(r) = \begin{cases} n_1 & r < a \\ n_2 & r \geq a \end{cases} \quad (3)$$

where a is the core radius, n_1 is the core refractive index, and n_2 is the cladding refractive index. For example, in Fig. 1, we show the normalized mode intensity of five propagating modes, along the fiber cross-section. The intensity of LP_{01} and LP_{02} shows a peak value at the fiber center. However, the LP_{11} and LP_{21} modes have a maximum value at 4.8 μm and 6.7 μm away from the center and zero value at the fiber center.

FMF-based optical network exploiting LP modes has been considered thoroughly in literature to scale the capacity of optical systems. It can be smoothly integrated with wavelength division multiplexing (WDM) network, as shown in Fig. 2, where different WDM signals can be transmitted at the same band using different modes [39]. This combination of WDM and FMF supports high speed transmission that can reach up to hundreds of Tera bit/s [40], [41]. For instance, three modes graded-index FMF transmission over C and L bands have been demonstrated to achieve data rates of 159 Tbit/s [40]

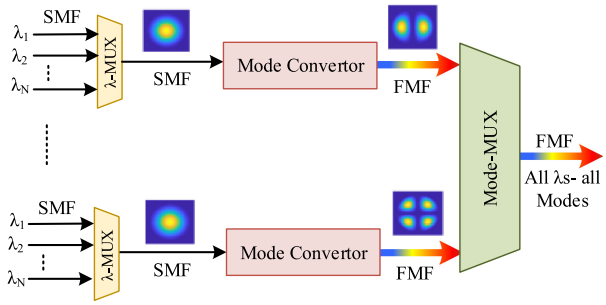

Fig. 2. Hybrid WDM-MDM Architecture..

and 280 Tbit/s [41]. Furthermore, SDM networks based on multi-core fibers and FMF increase the transmission capacity dramatically using a single optical fiber cable [42], [43]. By using 6-mode and 19-core fiber, the authors in [42] and [43] demonstrated 2.05 and 10.16 Pbit/s transmission over the C-band and C+L bands, respectively. Additionally, FMF-based transmission is well-matched with passive optical network (PON), where the different spatial modes can be used for both upstream and downstream communications [39]. In this regard, the authors in [44] proposed a bidirectional MDM-PON architecture and experimentally demonstrated a bidirectional PON transmission over 10-km FMF.

Nevertheless, in addition to the existing impairments of the traditional WDM network, the FMF-based network has additional impairments such as mode coupling (MC), which arises from the crosstalk between the different propagating modes. This results in transmission performance deterioration [45], [46]. Such impairments will impose an additional burden on the next generation fiber networks, where OPM becomes an important tool for system management.

The procedure to perform ML-based OPM, after the signal acquisition, includes two main stages: In the first stage, specific features from the signal is extracted. These features contain information about the signal impairments. In the second stage, the extracted features are applied to an ML algorithm (i.e. regressor) which is used to estimate the signal impairment. More details about the features types and ML algorithms exploited in this work are given in the following subsections.

A. *Features Extraction*

This is the process of choosing the most useful parameters that help reducing the dimensionality of the dataset; thus, enabling the classification/regression algorithms to operate faster and more effective. AAH is one of the most common features in OPM and MFI that is based on asynchronous sampling, where the received signal is sampled at a rate lower than the signal symbol rate. The amplitude of both I and Q sample's components is used to build the histogram. AAH is sensitive to some types of optical impairments, as changing the impairment level causes AAH profiles change. Thus, it can be used to differentiate between different values of the same impairment. However, the presence of multiple signal impairments complicates the monitoring process, when AAH is being used [47].

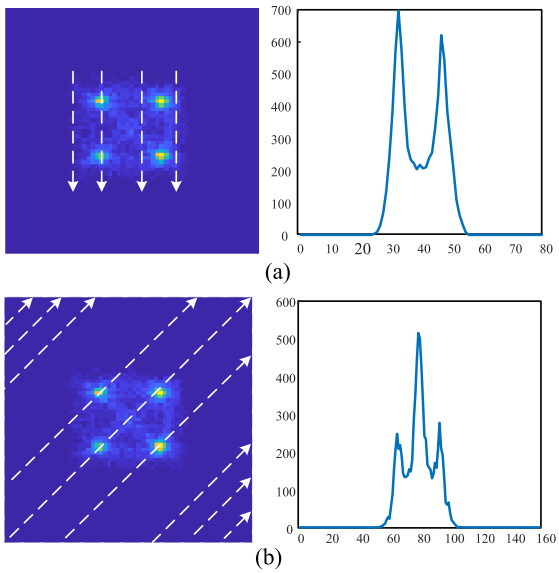

Fig. 3. An Illustration of the Projection Process (a) 2 d IQH (left) and the Corresponding $\mathrm{IQH_H}$ (right), and (b) 2 d IQH (left) and the Corresponding $\mathrm{IQH_D}$ (right).

To extract more details of the statistical properties of the monitored signal, 2D IQH is proposed in [46]–[48]. In 2D IQH, the asynchronous samples of I and Q are utilized to build a 2D histogram.

IQH features are closely related to constellation diagram. As majority of the impairment types are affecting the constellation shape in one way or another, hence IQH features will also be affected (i.e., IQH features will sense the changes; thereby they can be used for monitoring).

In this work, we propose two types of 1D features vectors that can be constructed from projecting a 2D IQH onto two different axes. The main motivation behind the use of 1D features vectors is twofold:

1) to reduce the dimensionality of the problem, hence improve its computational requirements;
2) to allow the use of ML algorithms well-suited for processing 1D features vectors.

The first features vector is obtained by projecting the 2D IQH over either vertical or horizontal axes. In our development, we consider the horizontal axis and we call it $\mathrm{IQH_H}$ (see Fig. 3(a)). On the other hand, the second features vector is obtained by projecting the 2D IQH over the diagonal axis and we call it $\mathrm{IQH_D}$ (see Fig. 3(b)).

Let 2D $\mathrm{IQH}(k,i)$, where $k,i=1,2,\ldots,n$, represent the IQH matrix ($n=80$ in our case). The $\mathrm{IQH_H}$ and $\mathrm{IQH_D}$ vectors can be represented mathematically as

$$\mathrm{IQH_H}(k)=\sum_{i=1}^{n}\mathrm{IQH}(k,i),\quad k=1,2,\ldots,n \tag{4}$$

$$\mathrm{IQH_D}(k)=\begin{cases}\displaystyle\sum_{i=1}^{k}\mathrm{IQH}(k-i+1,i), & 1\le k\le n\\ \displaystyle\sum_{i=k+1-n}^{n}\mathrm{IQH}(k-i+1,i), & n<k\le 2n-1\end{cases} \tag{5}$$

The two features $\mathrm{IQH_H}$ and $\mathrm{IQH_D}$ are, in fact, the inverse Fourier transform (FT) of the 1D slices of the 2D FT of IQH, computed at ($\omega_1=0$ and $\omega_2=\omega$) and ($\omega_1=\omega_2=\omega$), respectively. Specifically, let $C(\omega_1,\omega_2)$ be the 2D FT of IQH; that is,

$$C(\omega_1,\omega_2)=\sum_{k=0}^{n-1}\sum_{i=0}^{n-1}\mathrm{IQH}(k+1,i+1)\exp(-j\omega_1 k-j\omega_2 i) \tag{6}$$

By setting $\omega_1=\omega$ and $\omega_2=0$ in (6), we have

$$C(\omega,0)=\sum_{k=0}^{n-1}\left\{\sum_{i=0}^{n-1}\mathrm{IQH}(k+1,i+1)\right\}\exp(-j\omega k) \tag{7}$$

Equation (7) reveals that the 1D slice $C(\omega,0)$ is the FT of the 1D projection $\mathrm{IQH_H}(k)$. Similarly, it can be shown that the 1D slice $C(\omega,\omega)$ is the FT of the 1D projection $\mathrm{IQH_D}(k)$. It is of interest here to note that the required number of features points has been reduced by using 1D projections from n^2 for the 2D IQH to n points for the $\mathrm{IQH_H}$ and to $2n-1$ points for the $\mathrm{IQH_D}$.

Both $\mathrm{IQH_H}$, $\mathrm{IQH_D}$ features convey information about both amplitude and phase of signal samples. Also, $\mathrm{IQH_D}$ extracts more information from 2D IQH compared to $\mathrm{IQH_H}$, hence, it has better performance as will be shown in the next section. On the other hand, the AAH extracts information from the amplitude only. Hence, IQH achieves better performance than that of AAH. However, compared with coherent detection-based IQH, direct detection-based AAH has the competitive advantage in terms of implementation cost.

These extracted features are affected by signal degradation; thereby provide information about the degradation level for different impairments. Figure 4(a) shows the effect of OSNR on $\mathrm{IQH_H}$ and $\mathrm{IQH_D}$ features in comparison with the traditional AAH features, in the presence of low MC and CD= 0 ps/nm. It is noted that as the OSNR value decreases, the peaks values reduce. Also, the AAH features are more noisier than those of $\mathrm{IQH_H}$ and $\mathrm{IQH_D}$. This is intuitively not surprising because there is more averaging in the computation of $\mathrm{IQH_H}$ and $\mathrm{IQH_D}$ than that in the AAH. Similar conclusions can be drawn from Fig. 4(b), which shows the effect of other impairments (high MC and CD=1160 ps/nm) on the shape of three features.

B. ML-based Regression Techniques

ML regression is a kind of algorithms that aims to build a predictive model that maps continuous inputs to continuous outputs (i.e. targets). In the following subsections, we discuss the ML algorithms used in this paper.

1) Support Vector Machine: Support vector machine (SVM) is one of the most widely used ML-based algorithms. It can be used for both classification and regression problems [49], [50]. To use SVM in FMF-based OPM, the mathematical model is constructed from a training dataset that involves AAH, 2D IQH, $\mathrm{IQH_H}$, and $\mathrm{IQH_D}$. In this work, a linear kernel is used. The sequential minimal optimization (SMO) is utilized for solving the regression optimization problem.

2) Random Forest (RF) Regression: Random forest, as its name implies, comprises several numbers of individual decision trees that work as an ensemble [51]. The random forest offers relatively accurate results because it allows each tree to be

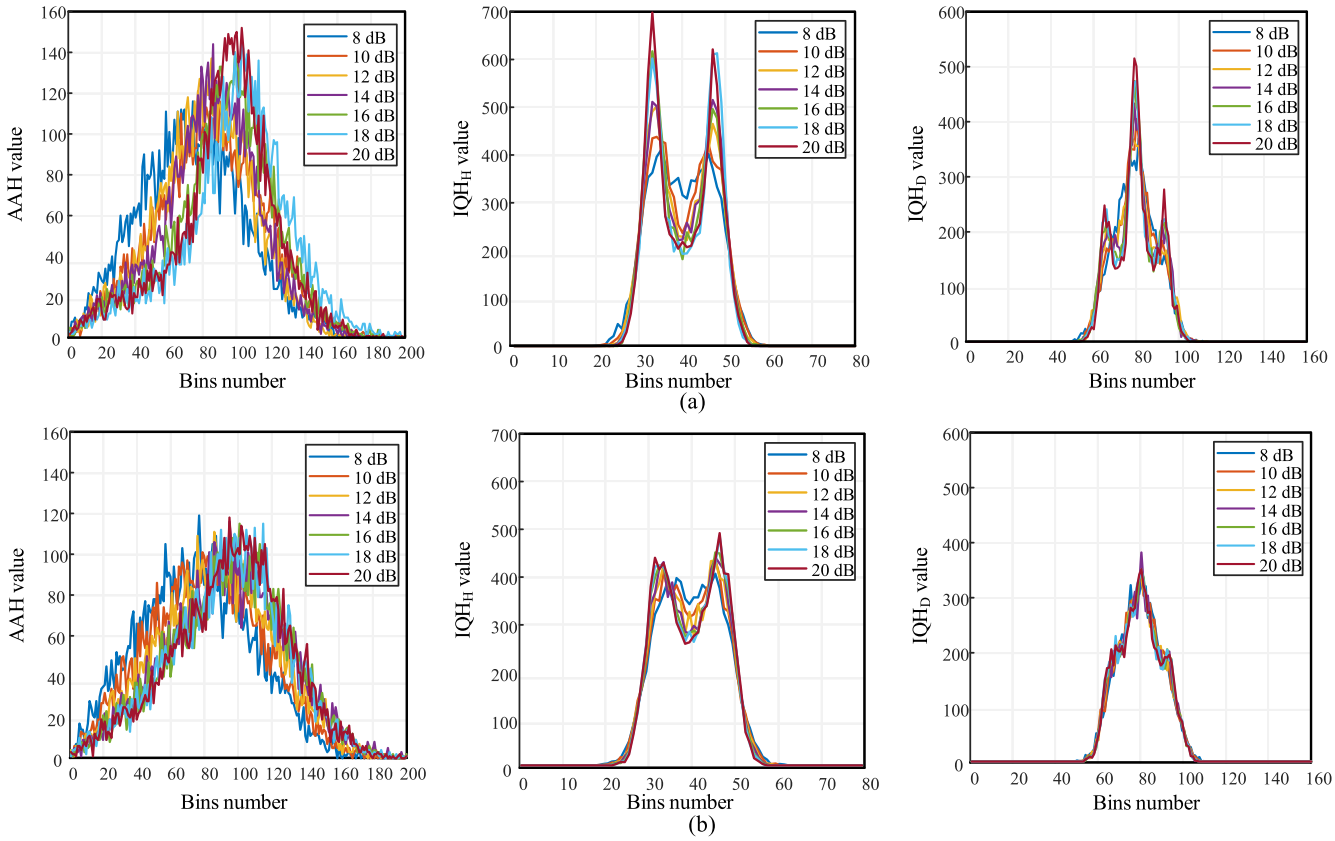

Fig. 4. OSNR Effect on the Features, AAH (left), IQH$_H$ (middle), and IQH$_D$ (right) (a) At Low MC and 0 Ps/nm and (b) At High MC and 1160 Ps/nm.

trained on different conditions, i.e. extract different features, hence, enhance the prediction result [52].

3) Convolutional Neural Network (CNN) Regression: CNN is a kind of DNN algorithms that can be used for handling data that have a grid-like topology, such as image data [53]. It decreases the number of trainable parameters by sharing the weight over the input data. Basically, CNN consists of three types of layers; convolutional, pooling, and fully connected layers, as shown in Fig. 5. The input layer accepts the extracted features, i.e. IQHs, and passes them to the first convolution layer. In the convolutional layer, convolution is performed with kernel filters (of size 3×3) that are initialized randomly to produce some feature maps. The dimensions of the first convolution layer are $80 \times 80 \times 16$, which represent an input of 80×80 pixels and 16 convolution kernels. All the convolution layers are associated to activation functions (i.e. rectified linear unit (ReLU) function). Then, the pooling layer reduces the dimension of each feature map using a down sampling filter of size 2×2. The dimensions of the second, third, and forth convolution layers are $40 \times 40 \times 32, 20 \times 20 \times 64$, and $10 \times 10 \times 64$, respectively. The output of the previous layer is converted to a 1D vector and passes through a connected layer that produces the desired output. CNN is trained, first, to minimize the mean square error (MSE) between network output and target value. More details about the architecture of employed CNN are given in the caption of Fig. 5.

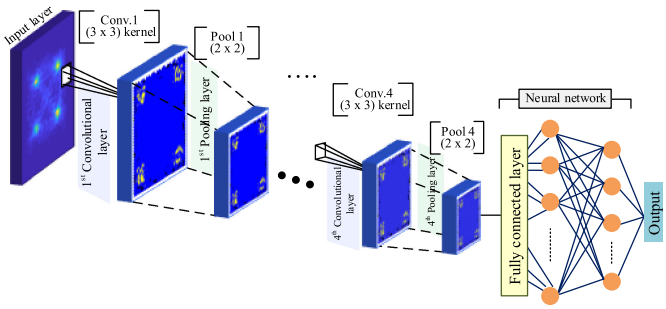

Fig. 5. Schematic Diagram of CNN Algorithm. the Network Hyper Parameters Include; Input Layer: 2 d IQH of 80×80. Convolutional 1 Layer: Sixteen 80×80 Feature Maps Generated Using Sixteen 3×3 Kernels. Pooling 1 Layer: Sixteen 40×40 Feature Maps Obtained After 2×2 Downsampling. Convolutional 2 Layer: Thirty-two 40×40 Feature Maps Generated Using Thirty-two 3×3 Kernels. Pooling 2 Layer: Thirty-two 20×20 Feature Maps Obtained After 2×2 Downsampling. Convolutional 3 Layer: Sixty-four 20×20 Feature Maps Generated Using Sixty-four 3×3 Kernels. Pooling 3 Layer: Sixty-four 10×10 Feature Maps Obtained After 2×2 Downsampling. Convolutional 4 Layer: Sixty-four 10×10 Feature Maps Generated Using Sixty-four 3×3 Kernels. Pooling 4 Layer: Sixty-four 5×5 Feature Maps Obtained After 2×2 Downsampling.the Used Activation Function is the Rectified Linear Unit (ReLU).

III. Simulation Setup

Fig. 6 shows the system configuration used in the simulation setup, developed using *VPItransmissionMaker* ver. 10. It comprises five optical DP transmitters operating at 1550 nm,

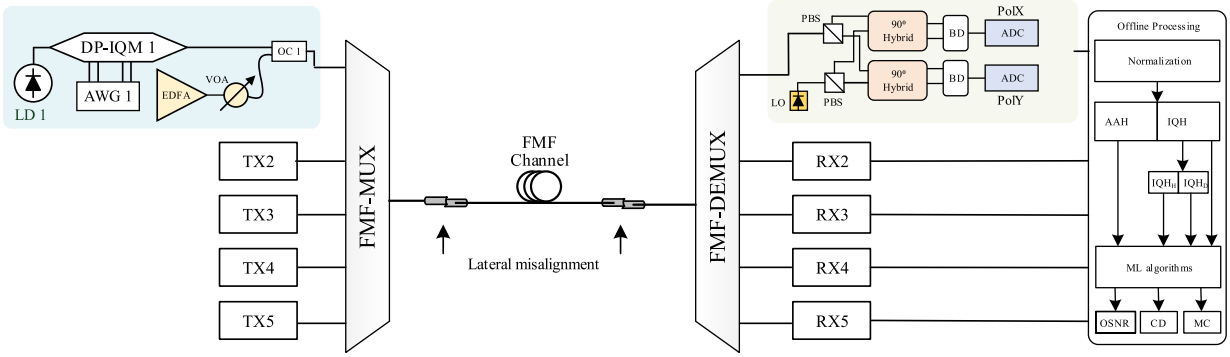

Fig. 6. Simulation Setup of the Proposed FMF-OPM System. LD: Laser Diode, EDFA: Erbium-doped Fiber Amplifier, VOA: Variable Optical Attenuator, MUX: FMF Multiplexer, DEMUX: FMF De-multiplexer, FMF: Few Mode Fiber, PBS: Polarization Beam Splitter, ADC: Analog to Digital Convertor, LO: Local Oscillator, BD: Balanced Detector.

TABLE I
PHYSICAL PROPERTIES OF FMF CHANNEL

Name	Value
Dispersion	LP_{01}: 21.1 ps/(nm·km) LP_{11}: 22.0 ps/(nm·km) LP_{21}: 21.4 ps/(nm·km)
Dispersion slope	LP_{01}: 0.066 ps/(nm^2·km) LP_{11}: 0.064 ps/(nm^2·km) LP_{21}: 0.056 ps/(nm^2·km)
Differential group delay	LP_{11}-LP_{01}: 2 ps/m LP_{21}-LP_{01}: 3.9 ps/m

FMF multiplexer and de-multiplexer, FMF cable, and five digital coherent receivers. At each transmitter block, a DP-QPSK signal is generated at a system speed of 10 Gbaud. The 1550 nm laser source has a linewidth of 10 kHz. An Erbium-doped fiber amplifier (EDFA) with variable optical attenuator (VOA) is employed to emulate different OSNR values. Five spatial modes including LP_{01}, LP_{11a}, LP_{11b}, LP_{21a} and LP_{21b} are considered, where a spatial multiplexer with five SMF inputs is exploited to generate a linear combination of the spatial modes into the FMF channel. The FMF cable has a step-index (SI) profile and physical properties such as dispersion, dispersion slop, and differential group delay (DGD). In Table I, we summarize the properties of a commercial FMF (OFS company) that is used in our system. The deviation from the ideal group delays in the degenerate groups of the LP modes (i.e. intra mode group) is 5×10^{-15} s/m. Also, for nonlinear calculations, the effective area of the fundamental mode (LP_{01}) is 80×10^{-12}m^2 and the non-linear refractive index is 2.6×10^{-20}m^2/w.

To emulate the influence of MC, two non-ideal connectors with lateral misalignment are assumed, as in [45]. Additionally, the length of FMF is adjusted in order to provide different values of CD. At the receiver side, we used modes de-multiplexer to separate the five spatial modes and fed them into five optical receivers. A digital coherent receiver is utilized at each receiver block which converts the received optical signal, of each mode, into an electrical signal. For each polarization, the detected signal's components (i.e. I and Q) are digitized by two ADCs working at 500 Msamples/s. The obtained samples (i.e. 8192 amplitude samples for each component) are captured directly after the ADCs as in [5], [54] and stored for offline processing. First, we applied sample normalization by division over sample's standard deviation. Then, we built the 2D IQH with size (80×80) and AAH with length 200 [46], [55]. Finally, the IQH_H and IQH_D are computed by applying the projection process to obtain features vectors with lengths 80 and 159, respectively, as discussed in Section II. These features are used as inputs to the different ML algorithms.

IV. RESULTS AND DISCUSSION

MC is a key impairment that affects transmission over FMF cables. It arises from the misalignment in optical connectors/splices and is controlled by the displacement parameter δD [45]. This impairment destroys the orthogonality among fiber modes and results in crosstalk between different modes expressed by the MC matrices shown in Fig. 7. These matrices are obtained according to [45], where the diagonal values denote the self-coupling coefficients while the remaining values indicate the cross-coupling coefficients. In this study, two misaligned connectors/splices are considered along the FMF cable. The displacement parameters for both connectors vary from 0.05 to 0.25 μm with a step of 0.05 μm.

To assess OPM in FMF-based network, we built an impairments dataset that comprises: OSNR values ranging from 8 to 20 dB in a step of 2 dB (note that changing the step of OSNR to 1 or 0.5 dB has marginal effect on the monitoring accuracy); CD values ranging from 160 to 1120 ps/nm in a step of 160 ps/nm; and a list of self-coupling values of 0.96, 0.88, 0.81, 0.74, 0.67, 0.62, 0.53, and 0.47. We generated 200 realizations for each impairment value. For instance, for OSNR monitoring, the total number of realizations is 7 (OSNR values) $\times$ 200 $=$ 1400 realizations. The dataset is split into 70% for training and 30% for testing. In the next subsections, we present results considering the following cases.

1) We fix the regressor (to be the SVM) and evaluate the monitoring performance utilizing the three 1D features

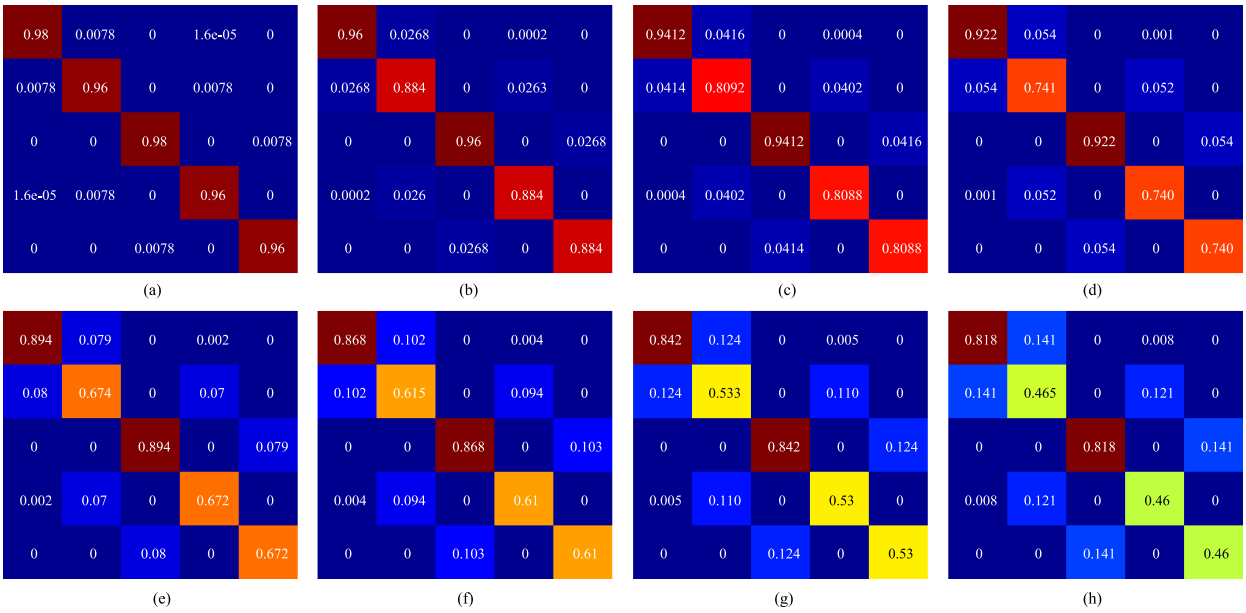

Fig. 7. MC Matrices of the Five Spatial Modes, (a) At ($\delta D1 = 0.05$ and $\delta D2 = 0.05$), (b) At ($\delta D1 = 0.1$ and $\delta D2 = 0.1$), (c) At ($\delta D1 = 0.1$ and $\delta D2 = 0.15$), (d) At ($\delta D1 = 0.15$ and $\delta D2 = 0.15$), (e) At ($\delta D1 = 0.15$ and $\delta D2 = 0.2$), (f) At ($\delta D1 = 0.2$ and $\delta D2 = 0.2$), (g) At ($\delta D1 = 0.2$ and $\delta D2 = 0.25$), (h) At ($\delta D1 = 0.25$ and $\delta D2 = 0.25$).

vectors (AAH, IQH_H, and IQH_D). The output of this part of study reveals that IQH_D has the best performance.

2) We fix the input features to be the 2D IQH and evaluate the monitoring performance when three ML regressors (CNN, RF, and SVM) are being used. Note that the columns of 2D IQH features matrix is concatenated in the form of 1D features vector when applied as input to RF and SVM algorithms. In this study, it has been found that the monitoring performance of CNN outperforms that of other two ML algorithms.
3) We fix the regressor (to be the SVM) and evaluate the monitoring performance utilizing the 2D IQH features and the 1D features vector IQH_D. The output of this part of study reveals that the 2D IQH has better performance.
4) We fix the regressor and features to be the CNN and 2D IQH, respectively, and evaluate the monitoring performance under different values of phase noise, first-order PMD, symbol rates, and modulation schemes.

It is worth noting that the performance of a single spatial mode (i.e. LP_{11a}) is considered, to avoid redundancy in results.

A. Results of OPM Using 1D Projections

In this subsection, we discuss the monitoring results of 1D features in conjunction with SVM regressor. We considered two types of projections including IQH_H and IQH_D, then we compared their performances with the commonly used AAH features. The effectiveness of monitoring is assessed using graphical inspection and quantitative measure. The graphical representation is performed using boxplot. A boxplot is a standardized way to display data distribution by using five statistical measures, which are the minimum, first quartile (Q_1), median, third quartile (Q_3), and maximum of dataset [56]. It tells about the skewness of data distribution and presence of outliers. In this context, the minimum and maximum of a dataset are defined as $Q_1 - 1.5 \times IQR$ and $Q_3 + 1.5 \times IQR$, respectively. Here, IQR is the difference between Q_3 and Q_1. Therefore, any sample of value less than the minimum or greater than the maximum is considered an outlier.

Also, the coefficient of determination is used as a quantitative measure, which is expressed as [57]

$$\rho = 1 - \frac{\sum_{n=1}^{N} (y_n - \hat{y_n})^2}{\sum_{n=1}^{N} (y_n - \bar{y})^2} \tag{8}$$

where y_n is the actual data, $\hat{y}$ is the estimated data, N is the total number of test samples, and $\bar{y}$ is the sample mean. This measure gives some information about the goodness-of-fit of a model. In particular, ρ tells how well the regression results approximate the true target values. When $\rho = 1$, this means the model's output exactly match the target (ground truth) values. While $\rho = 0$ means that the model cannot predict the true target values.

The regression results of OSNR, CD, and MC are shown in Fig. 8. The results show that the IQR range for both IQH_H, and IQH_D is relatively less than AAH which provides initial indication that the IQH_H and IQH_D produce better results than AAH. The results is confirmed using the coefficient of determination where its values for OSNR monitoring in case of AAH, IQH_H and IQH_D features are 0.90, 0.94, and 0.97, respectively. Similarly, for CD and MC, IQH_D provides the best prediction results with coefficient of determination values of 0.86 and 0.77, respectively.

The OPM performance is also assessed in the coexistence of multiple impairments. Nine monitoring cases have been considered for each impairment, as illustrated in Fig. 9. For instance,

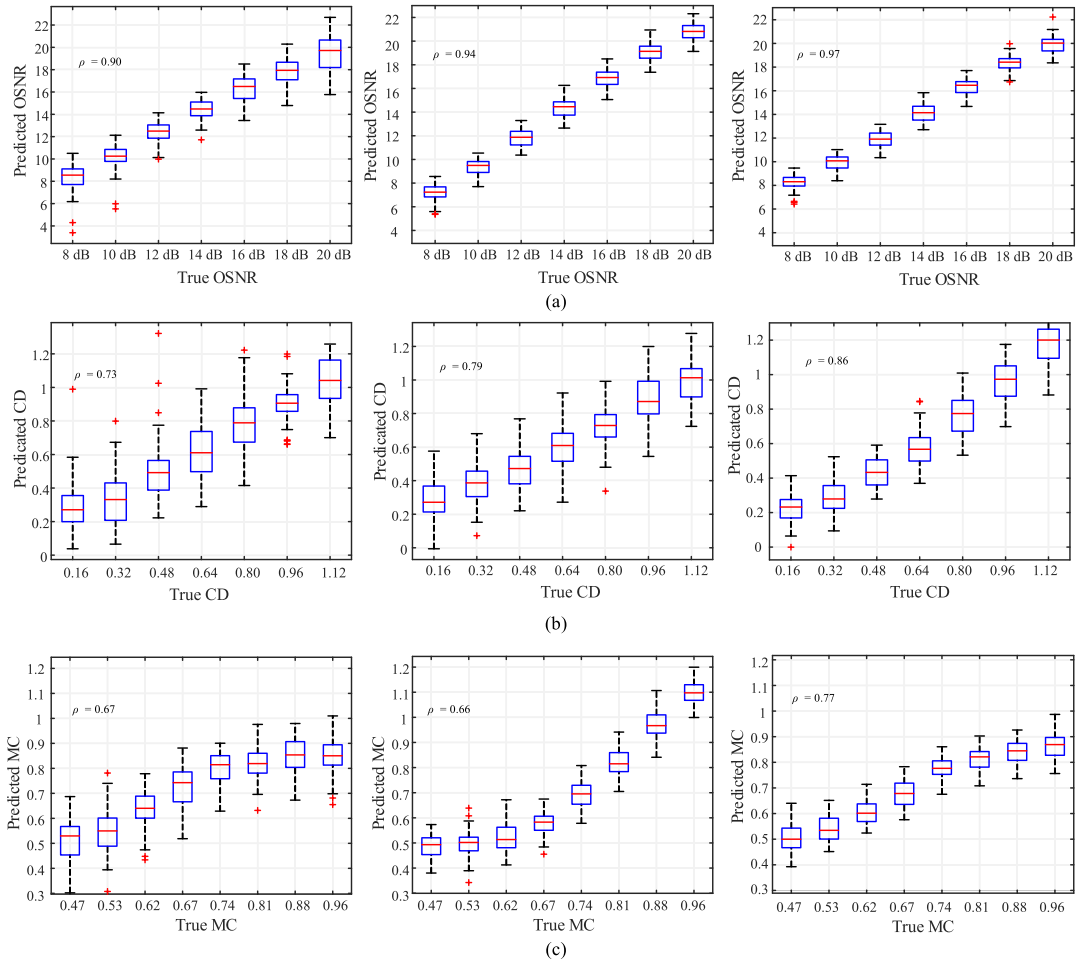

Fig. 8. OPM Using SVM With AAH (left), IQH_H(Middle), and IQH_D (right) for (a) OSNR, (b) CD, and (c) MC.

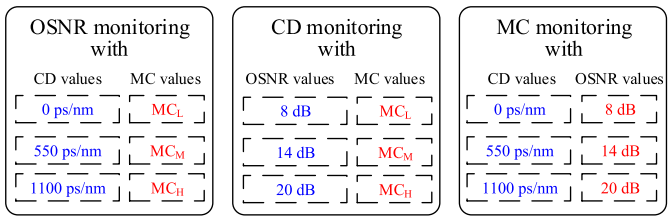

Fig. 9. Monitoring Strategy of the Different Impairments.

to monitor OSNR, we choose MC values of 0.97 (MC_L), 0.74 (MC_M), and 0.47 (MC_H) in the coexistence of 0, 550, and 1100 ps/nm CD values. This generates 9 monitoring cases. The monitoring results are demonstrated according to ρ (z-axis) versus the impairments type (x- and y-axis), as shown Fig. 10. The OSNR monitoring accuracies at CD = 0 ps/nm, decreased from 90%, 94%, and 97% (at MC_L) to 70%, 84%, and 85% (at MC_H) for AAH, IQH_H, and IQH_D, respectively. It is worth noting that the prediction accuracy of OSNR outperforms that of CD and MC, since the OSNR values have relatively more noticeable effect on the constellation diagrams than those of CD and MC, which reflects on the extracted features. And, in all cases, the performance of proposed 1D projection features outperforms that obtained by the widely used AAH features. Further, the IQH_D is the best among all other features vectors.

B. Results of OPM Using 2D IQH

To improve the monitoring accuracy, CNN in conjunction with the 2D IQH are employed. CNN has the ability to extract abstractive information from high-dimensional input. Besides, it is robust with respect to over-fitting [58]. In order to demonstrate the effectiveness of CNN, SVM and random forest algorithms are considered. Fig. 11 illustrates the monitoring results for separate impairments (i.e. OSNR, CD, or MC) using 2D IQH with different regressors including CNN, random forest, and SVM. It is clear that the prediction accuracy for OSNR and CD, in terms of the coefficient of determination, shows quit similar performance for the three regressors. However, for MC, the SVM produces the worst accuracy. In Fig. 12, we show the coefficient of determination of OSNR, CD, and MC in the coexistence of multiple impairments, as described in Fig. 9. The results indicate that monitoring accuracies of CNN are greater than SVM and random forest. However, random forest provides relatively better accuracy results than those of SVM.

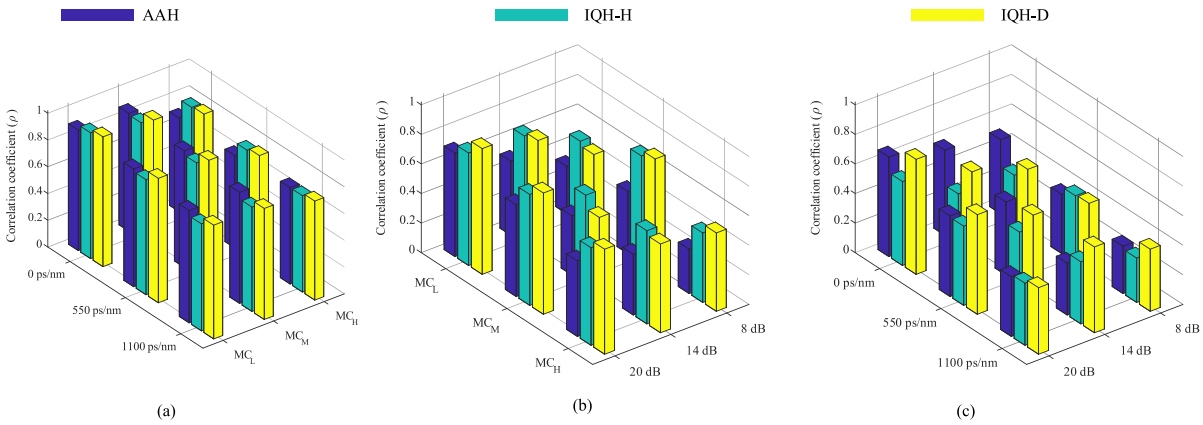

Fig. 10. Monitoring Results Using AAH, IQH_H, and IQH_D Features and SVM Regressor, (a) OSNR, (b) CD, and (c) MC.

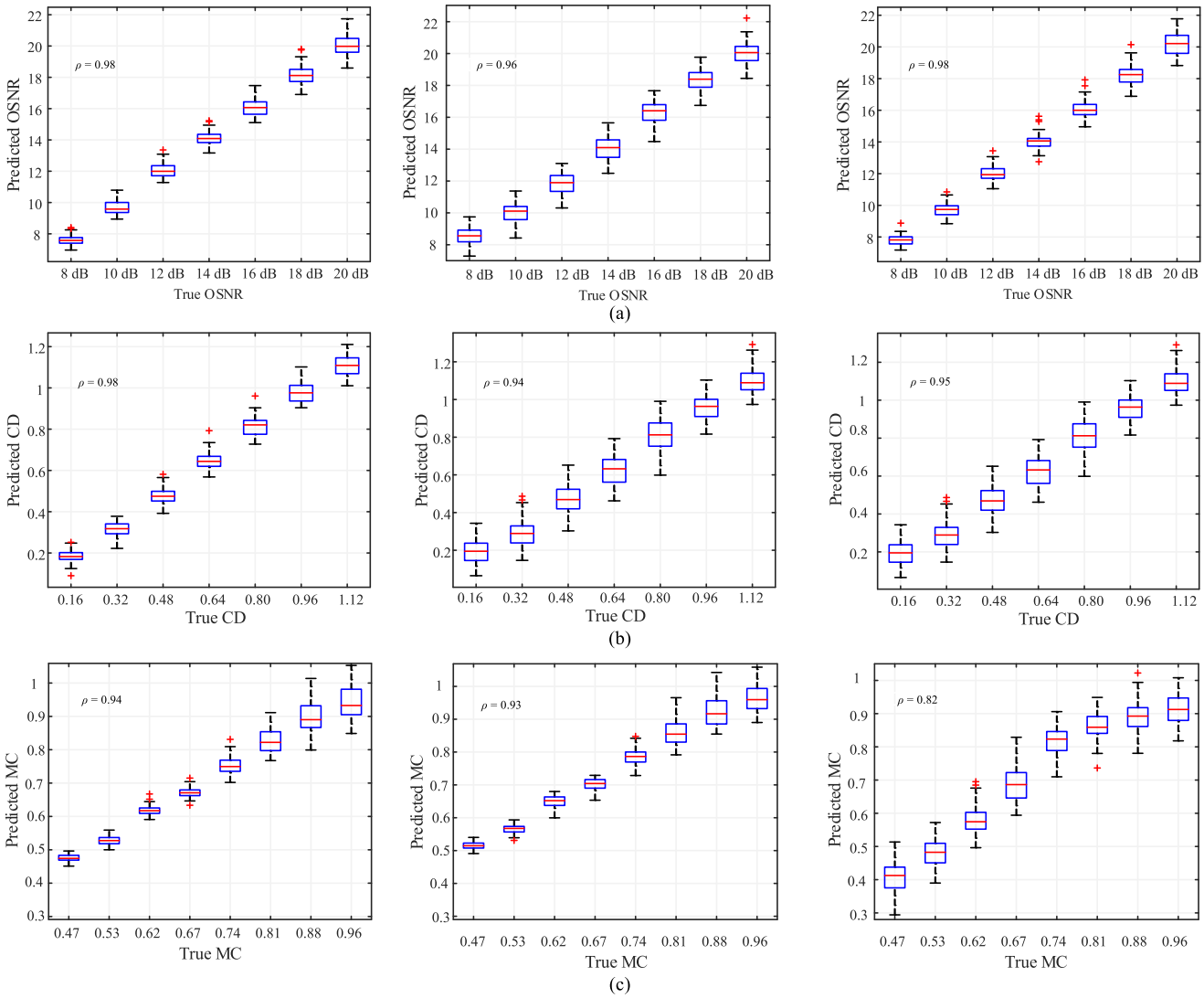

Fig. 11. OPM Using 2 d IQH With CNN (left), Random Forest (Middle), and SVM (right) (a) OSNR (b) CD, and (c) MC.

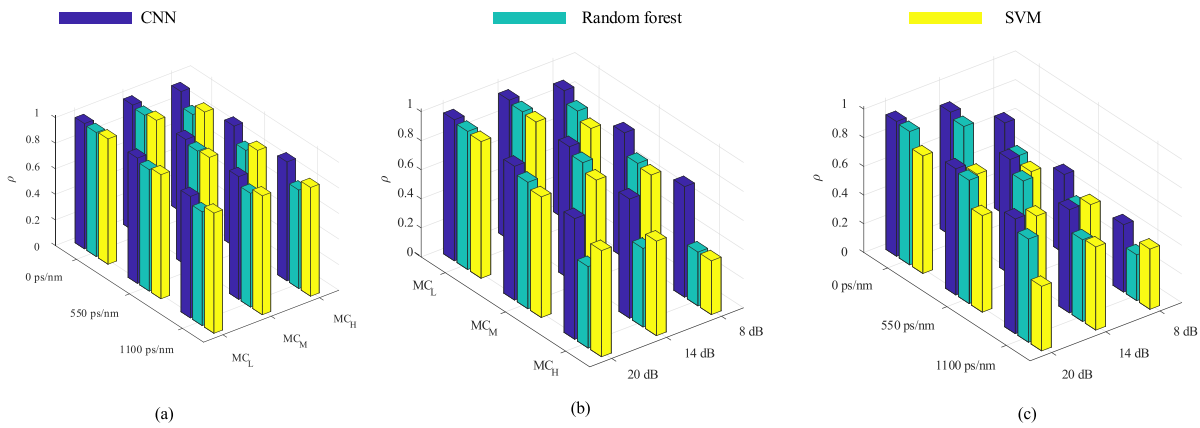

Fig. 12. Monitoring Results Using 2 d IQH Features With CNN, SVM, and Random Forest Regressors for (a) OSNR, (b) CD, and (c) MC.

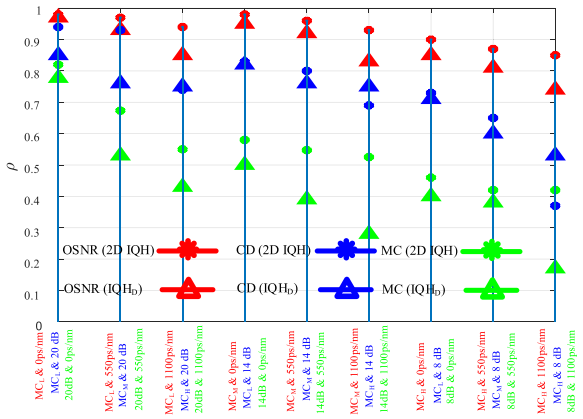

Fig. 13. Comparisons Between Monitoring Results of 2 d IQH and IQH_D.

Also, in Fig. 13, we compare the performance of 2D IQH using SVM with the results obtained in the previous section for IQH_D using the same regressor. It can be observed that the OSNR monitoring results for both features are quite similar, except at high CD value (i.e. 1100 ps/nm), where the performance of 2D IQH outperforms that of IQH_D. Similarly, for CD monitoring, both features provide the same performance, except at high OSNR value (i.e. 20 dB), the 2D IQH performance is better than IQH_D. However, for MC monitoring, 2D IQH provides the best performance.

As 2D IQH with CNN shows better performance than other algorithms, we focus only on CNN's regressor in the following analysis. We assess the FMF-based OPM for various conditions including PN, PMD, and different modulation schemes and baud rates. This is achieved in the coexistence of multiple impairments, in particular, monitoring OSNR (when CD = 550 ps/nm and MC_M), CD (when OSNR = 14 dB and MC_M), and MC (when OSNR = 14 dB and CD = 550 ps/nm). First, we investigate the OPM performance in case of PN ranging from 0 to 1 MHz and 1^{st} PMD with a DGD ranging from 5 to 25 ps, and a principal states of polarization (PSP) angle of 45^o. It is observed from Fig. 14(a) that the OSNR monitoring accuracy maintains greater than 0.97 even when PN reaches

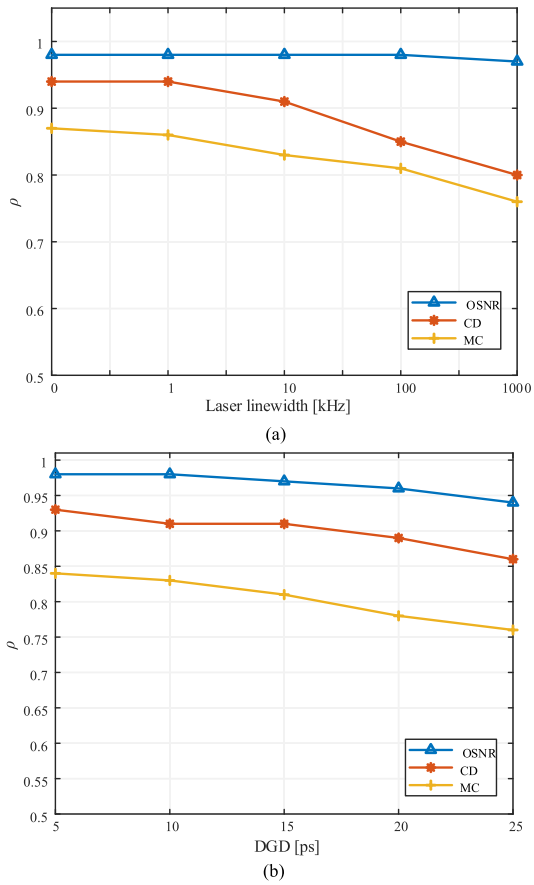

Fig. 14. OPM Results Using 2 d IQH Features and CNN Regressor At (a) Different PN Values, and (b) Different PMD Values.

1 MHz, while the CD and MC accuracies decrease to 0.8 and 0.77, respectively. Similarly, in case of PMD, the accuracy is still greater than 0.94 when DGD reaches 25 ps. However, CD and MC accuracy reduces to 0.85 and 0.76, respectively, as shown in Fig. 14(b).

Second, we investigated OPM at 10, 14, 20, and 28 Gbaud system speeds. It is clear from Fig. 15 that the OSNR accuracy approximately remains unchanged up to 20 Gbaud and decreases

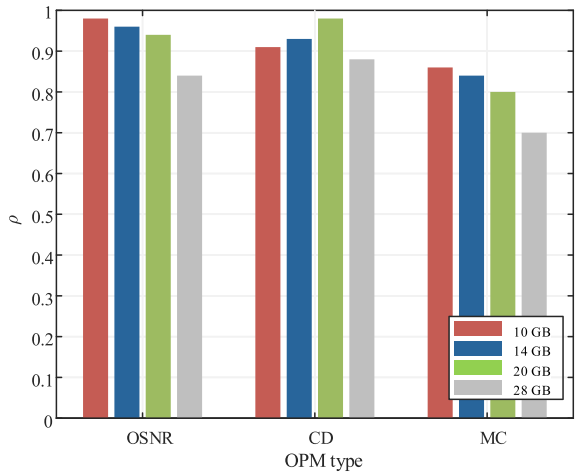

Fig. 15. OPM Results Using 2 d IQH Features and CNN Regressor At Different Data Rates.

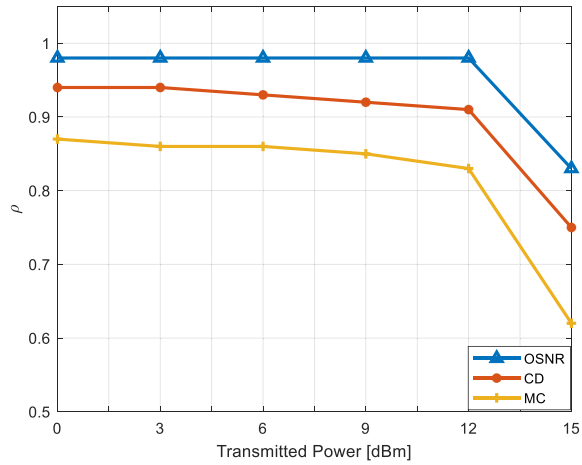

Fig. 17. OPM Results Using 2 d IQH Features and CNN Regressor At Different Transmitted Power Values in the Presence of Fiber Nonlinearity.

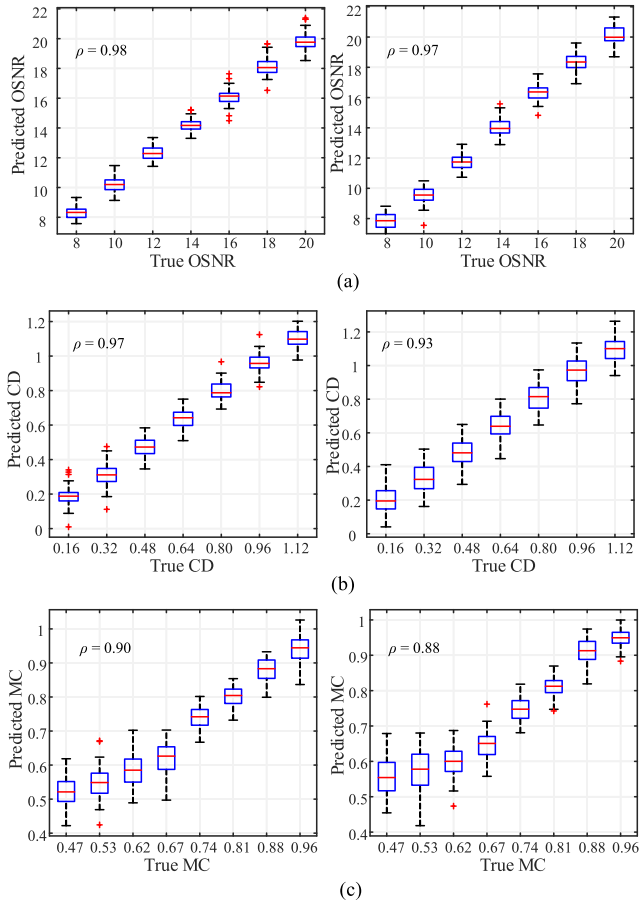

Fig. 16. OPM Results Using 2 d IQH Features and CNN Regressor When Using DP-8QAM (left) and DP-16QAM (right) (a) OSNR Monitoring, (b) CD Monitoring, and (c) MC Monitoring.

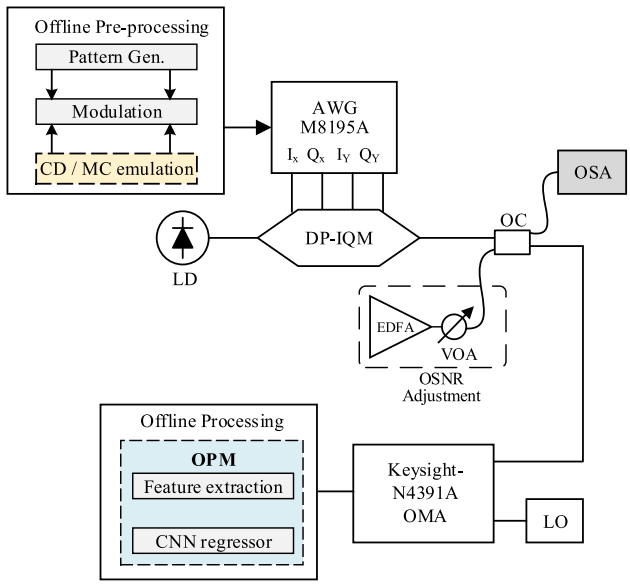

Fig. 18. Experimental Setup of the Proposed FMF-OPM System. CD: Chromatic Dispersion, MC: Mode Coupling, LO: Local Oscillator, LD: Laser Diode, EDFA: Erbium-doped Fiber Amplifier, VOA: Variable Optical Attenuator, OC: Optical Coupler, OSA, Optical Signal Analyzer, OMA, Optical Modulation Analyzer, CNN, Convolutional Neural Network.

to 0.85 when the symbol rate reaches 28 Gbaud. However, the CD monitoring accuracy gradually increases from 0.91 at 10 Gbaud to 0.98 at 20 Gbaud then decreases to 0.88 at 28 Gbaud. The reason for this is that in the case of 10 Gbaud, there is a very slight effect on the 2D IQH for the different CD values, especially at small values such as 160 and 320 ps/nm. On the other hand, as the symbol rate increases to 14 and 20 Gbaud, the disparity between the different CD values becomes more pronounced, which increases the monitoring accuracy. However, as the symbol rate increases to 32 Gbaud, the 2D IQH becomes noisy especially in high CD values, which leads to a decrease in monitoring accuracy. Besides, the MC accuracy reduces from 0.86 at 10 Gbaud to 0.7 at 28 Gbaud.

Third, the OPM for DP-M-QAM (M = 8 and 16) is addressed. It can be seen from Fig. 16 that there are no major differences in the accuracies of the two schemes. The prediction accuracies for DP-8-QAM(DP-16-QAM) are 0.98(0.97), 0.97(0.93), and 0.90(0.88) for OSNR, CD, and MC monitoring, respectively.

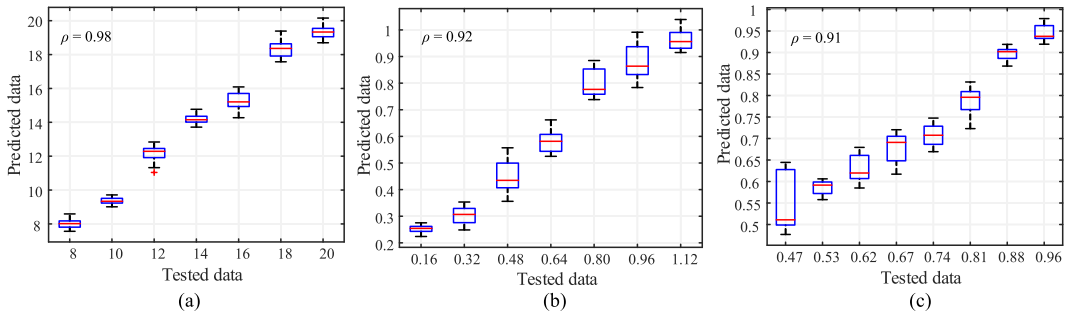

Fig. 19. Experimental Results of OPM Using 2 d IQH With CNN (a) OSNR Monitoring, (b) CD Monitoring, and (c) MC Monitoring.

This performance makes the 2D IQH a viable candidate for OPM in FMF-based optical networks.

Finally, to investigate the effect of fiber nonlinearity on the OPM accuracy, we studied the fiber nonlinear effect in terms of the nonlinear factor k_{uv}, which takes into account the SPM and XPM coefficients [59]. k_{uv} describes the nonlinear factor of the two spatial modes u, v. For instance, if G_u, defined as mode group, contains only a single spatial mode with two orthogonal polarizations, the k_{uu} factor will be given by [60] $k_{uu} = 8/9$. Also, we considered $k_{uv} = 0.895$ [59]. The OPM is investigated under different transmitted optical power as shown in Fig. 17. The transmitted power ranges from 0 to 15 dBm. It can be observed that the OSNR, CD, and MC accuracies remain almost unchanged up to 12 dBm and decrease to 0.83, 0.75, and 0.62, respectively, at 15 dBm optical transmitted power.

V. Experiment Verification

A proof-of-concept experiment is performed to verify the feasibility of the proposed OPM scheme. The experimental setup is illustrated in Fig. 18. At the transmitter side, 10 different sets of pseudo-random binary sequence (PRBS) of length 2^{11}-1 are generated. These data are mapped into five 10 Gbaud DP-QPSK signals that represent five spatial modes (i.e. LP_{01}, LP_{11a}, LP_{11b}, LP_{21a} and LP_{21b}). For each polarization, the effect of crosstalk between the different modes (i.e. MC) and CD are introduced digitally in the transmitter DSP. The MC effect can be implemented using the following equation [46].

$$\mathbf{a_{out}} = \mathbf{C}\,\mathbf{a_{in}} \tag{9}$$

where $\mathbf{a}_{\text{in}}$ and $\mathbf{a}_{\text{out}}$ are the input and output mode vectors, respectively, and $\mathbf{C}$ is the coupling matrix given by

$$\mathbf{C} = \begin{bmatrix} c_{01-01} & c_{11a-01} & c_{11b-01} & c_{21a-01} & c_{21b-01} \\ c_{01-11a} & c_{11a-11a} & c_{11b-11a} & c_{21a-11a} & c_{21b-11a} \\ c_{01-11b} & c_{11a-11b} & c_{11b-11b} & c_{21a-11b} & c_{21b-11b} \\ c_{01-21a} & c_{11a-21a} & c_{11b-21a} & c_{21a-21a} & c_{21b-21a} \\ c_{01-21b} & c_{11a-21b} & c_{11b-21b} & c_{21a-21b} & c_{21b-21b} \end{bmatrix}$$

where the matrix element c_{i-j} is the coupling coefficient between the i^{th} and j^{th} modes, as in Fig. 7.

Similarly, the CD effect is introduced by multiplying its frequency response $e^{-j\lambda^2 w^2 DL}/4\pi c$ by the frequency domain

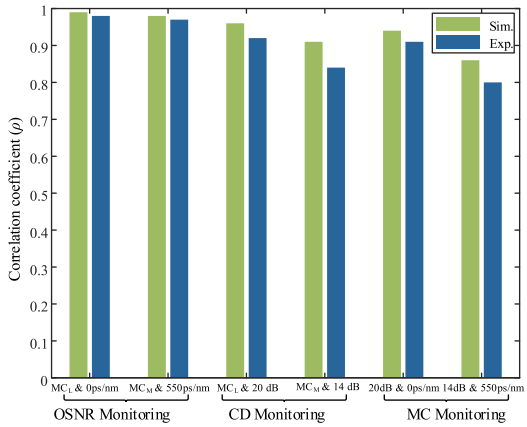

Fig. 20. Simulation vs Experimental Results of OPM Using 2 d IQH With CNN.

of each mode, where λ is the signal wavelength, w is the angular frequency, L is the fiber length, c is the speed of light, and D is the dispersion parameter which is given in Table I. The MC is precisely adjusted according to Fig. 7 and the CD value is varied from 160 ps/nm to 1100 ps/nm. The OPM performance based on LP_{11a} mode is investigated, as in the simulation section. Four electrical multilevel driving signals (I and Q for individual polarization) are generated using Keysight M8195 A 64 GSa/s arbitrary waveform generator and applied into a DP-IQ Mach-Zehnder modulator (DP-IQM). A narrow-linewidth distributed-feedback fiber laser (NKT-Photonics) centered at 1550 nm is used as a carrier signal and modulated by the DP-IQM. In order to address the effect of OSNR, an ASE noise source is added to the optical signal using an EDFA and an optical attenuator, which is used to control the OSNR between 8 - 20 dB. The generated optical signal is combined with ASE noise source via a 50:50 optical coupler (OC). Then, the OC's output is split into two branches. One branch is used to adjust the OSNR value via an optical spectrum analyzer (OSA) of 0.06 nm resolution bandwidth while the other is connected to an optical coherent receiver (N4391 A Keysight optical modulation analyzer (OMA)). Lastly, the offline processing for OPM is performed using the 2D IQH features and CNN regressor. It is worth noting that the obtained samples are taken, after the sampling stage, before demodulation stage. We studied the OPM for two cases. In the first case, we monitored OSNR (when CD = 0 ps/nm and MC_L),

CD (when OSNR = 20 dB and MC_L), and MC (when OSNR = 20 dB and CD = 0 ps/nm). Whereas, in the second case, we monitored OSNR (when CD = 550 ps/nm and MC_M), CD (when OSNR = 14 dB and MC_M), and MC (when OSNR = 14 dB and CD = 550 ps/nm). Fig. 19 shows the OPM monitoring accuracies results for the first case. As shown in this figure, the coefficient of determination for OSNR, CD, and MC are 0.98, 0.92, and 0.91, respectively. Besides, in Fig. 20, we compare the simulation and experimental results for both monitoring cases where a good agreement between them can be observed.

VI. Conclusion

In this paper, we investigated OPM in FMF-based optical networks by proposing 1D features (the IQH_H and IQH_D) and the 2D IQH features. We evaluated the performance of OPM, via simulation and experiment, using 10 Gbuad-DP-QPSK signal. The results showed that we can accurately estimate OSNR values in the range of 8 to 20 dB, CD values ranging from 160 to 1100 ps/nm, and different mode coupling coefficient values. However, the presence of impairment(s) with level(s), making IQH features almost indistinguishable for different cases, will negatively impact the monitoring performance. Also, we showed that both IQH_D and IQH_H are providing better performance than the conventional AAH, besides, IQH_D outperforms IQH_H. Also, there is a trade-off between IQH_D and 2D IQH, where the 2D IQH provides better accuracy, but it requires larger number of input features samples to be applied to ML regressors. Furthermore, the results show that the proposed CNN-based OPM can tolerate laser PN and PMD, as well as, it has slight variations due to change in data rate and modulation format. This shows the viability of the proposed 2D IQH features to be considered for OPM in FMF-based coherent optical networks.

References

[1] W. Klaus, J. Sakaguchi, B. J. Puttnam, Y. Awaji, and N. Wada, "Optical technologies for space division multiplexing," in *Proc. 13th Workshop Inf. Opt.*, Neuchatel, 2014, pp. 1–3.

[2] G. M. Saridis, D. Alexandropoulos, G. Zervas, and D. Simeonidou, "Survey and evaluation of space division multiplexing: From technologies to optical networks," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 4, pp. 2136–2156, Oct.–Dec. 2015.

[3] Z. Dong, F. N. Khan, Q. Sui, K. Zhong, C. Lu, and A. P. T. Lau, "Optical performance monitoring: A review of current and future technologies," *J. Lightw. Technol.*, vol. 34, no. 2, pp. 525–543, 2015.

[4] A. E. Willner, Z. Pan, and C. Yu, "Optical performance monitoring," in *Proc. Opt. Fiber Telecommun. VB*. Elsevier, 2008, pp. 233–292.

[5] C. Wang *et al.*, "Joint OSNR and CD monitoring in digital coherent receiver using long short-term memory neural network," *Opt. Express*, vol. 27, no. 5, pp. 6936–6945, 2019.

[6] F. N. Khan, A. P. T. Lau, T. B. Anderson, J. C. Li, C. Lu, and P. K. A. Wai, "Simultaneous and independent OSNR and chromatic dispersion monitoring using empirical moments of asynchronously sampled signal amplitudes," *IEEE Photon. J.*, vol. 4, no. 5, pp. 1340–1350, Oct. 2012.

[7] F. Wu, A. Yang, P. Guo, Y. Qiao, L. Zhuang, and S. Guo, "QPSK training sequence-based both OSNR and chromatic dispersion monitoring in DWDM systems," *IEEE Photon. J.*, vol. 10, no. 4, pp. 1–10, Aug. 2018.

[8] C. Courvoisier, J. Fatome, and C. Finot, "Measurement of residual chromatic dispersion or OSNR via nonlinear spectral evolution," *IEEE Photon. Technol. Lett.*, vol. 23, no. 9, pp. 537–539, May 2011.

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

[10] P. Sillard, M. Bigot-Astruc, and D. Molin, "Few-mode fibers for mode-division-multiplexed systems," *J. Lightw. Technol.*, vol. 32, no. 16, pp. 2824–2829, 2014.

[11] C. Do, A. V. Tran, C. Zhu, D. Hewitt, and E. Skafidas, "Data-aided OSNR estimation for QPSK and 16-QAM coherent optical system," *IEEE Photon. J.*, vol. 5, no. 5, pp. 6 601 609–6 601 609, Oct. 2013.

[12] C. C. Do, C. Zhu, and A. V. Tran, "Data-aided OSNR estimation using low-bandwidth coherent receivers," *IEEE Photon. Technol. Lett.*, vol. 26, no. 13, pp. 1291–1294, Jul. 2014.

[13] M. S. Faruk, Y. Mori, and K. Kikuchi, "Estimation of OSNR for Nyquist-WDM transmission systems using statistical moments of equalized signals in digital coherent receivers," in *Proc. Opt. Fiber Commun. Conf. Exhibition*, San Francisco, CA, 2014, pp. 1–3.

[14] Y. Ma, M. Gao, L. Wang, Y. Sha, W. Shao, and G. Shen, "Accuracy enhancement of moments-based OSNR monitoring in QAM coherent optical communication," *IEEE Commun. Lett.*, vol. 24, no. 4, pp. 821–824, Apr. 2020.

[15] X. Lin, O. A. Dobre, T. M. Ngatched, and C. Li, "A non-data-aided OSNR estimation algorithm for coherent optical fiber communication systems employing multilevel constellations," *J. Lightw. Technol.*, vol. 37, no. 15, pp. 3815–3825, 2019.

[16] W. Moench and E. Loecklin, "Measurement of optical signal-to-noise-ratio in coherent systems using polarization multiplexed transmission," in *Proc. Opt. Fiber Commun. Conf. Exhib.*, Los Angeles, CA, 2017, pp. 1–3.

[17] D. Gariépy, S. Searcy, G. He, S. Tibuleac, M. Leclerc, and P. Gosselin-Badaroudine, "Novel OSNR measurement techniques based on optical spectrum analysis and their application to coherent-detection systems," *J. Lightw. Technol.*, vol. 37, no. 2, pp. 562–570, 2019.

[18] F. Wu, P. Guo, A. Yang, and Y. Qiao, "Chromatic dispersion estimation based on CAZAC sequence for optical fiber communication systems," *IEEE Access*, vol. 7, pp. 139 388–139 393, 2019.

[19] Y. Ma *et al.*, "Training sequence-based chromatic dispersion estimation with ultra-low sampling rate for optical fiber communication systems," *IEEE Photon. J.*, vol. 10, no. 6, pp. 1–9, Dec. 2018.

[20] D. Tang, X. Wang, L. Zhuang, P. Guo, A. Yang, and Y. Qiao, "Delay-tap-sampling-based chromatic dispersion estimation method with ultra-Low sampling rate for optical fiber communication systems," *IEEE Access*, vol. 8, pp. 101 004–101 013, 2020.

[21] K. Horikoshi, A. Matsushita, S. Okamoto, and M. Nakamura, "Fast blind chromatic-dispersion esitimation for small-rolloff Nyquist pulse-shaped signal using spectral cyclostationarity," in *Proc. 45th Eur. Conf. Opt. Commun.*, Dublin, Ireland, 2019, pp. 1–3.

[22] W. G. Hatcher and W. Yu, "A survey of deep learning: Platforms, applications and emerging research trends," *IEEE Access*, vol. 6, pp. 24 411–24 432, 2018.

[23] R. A. Eltaieb *et al.*, "Efficient classification of optical modulation formats based on singular value decomposition and Radon transformation," *J. Lightw. Technol.*, vol. 38, no. 3, pp. 619–631, 2020.

[24] F. N. Khan, Y. Zhou, A. P. T. Lau, and C. Lu, "Modulation format identification in heterogeneous fiber-optic networks using artificial neural networks," *Opt. Express*, vol. 20, no. 11, pp. 12 422–12 431, 2012.

[25] T. S. R. Shen, K. Meng, A. P. T. Lau, and Z. Y. Dong, "Optical performance monitoring using artificial neural network trained with asynchronous amplitude histograms," *IEEE Photon. Technol. Lett.*, vol. 22, no. 22, pp. 1665–1667, Nov. 2010.

[26] J. Thrane, J. Wass, M. Piels, J. C. Diniz, R. Jones, and D. Zibar, "Machine learning techniques for optical performance monitoring from directly detected PDM-QAM signals," *J. Lightw. Technol.*, vol. 35, no. 4, pp. 868–875, 2016.

[27] W. Saif, M. A. Esmail, A. Ragheb, T. Alshawi, and S. Alshebeili, "Machine learning techniques for optical performance monitoring and modulation format identification: A Survey," *IEEE Commun. Surveys Tuts.*, pp. 1–1, Sep. 2020.

[28] F. N. Khan, C. Lu, and A. P. T. Lau, "Optical performance monitoring in fiber-optic networks enabled by machine learning techniques," in *Proc. Opt. Fiber Commun. Conf. Expo.*, San Diego, CA, 2018, pp. 1–3.

[29] Q. Xiang, Y. Yang, Q. Zhang, and Y. Yao, "Joint and accurate OSNR estimation and modulation format identification scheme using the feature-based ANN," *IEEE Photon. J.*, vol. 11, no. 4, pp. 1–11, Aug. 2019.

[30] J. Mata *et al.*, "Artificial intelligence (AI) methods in optical networks: A comprehensive survey," *Opt. Switching Netw.*, vol. 28, pp. 43–57, 2018.

[31] F. N. Khan, Q. Fan, C. Lu, and A. P. T. Lau, "An optical communication's perspective on machine learning and its applications," *J. Lightw. Technol.*, vol. 37, no. 2, pp. 493–516, 2019.

[32] Y. Zhang *et al.*, "Eye diagram measurement-based joint modulation format, OSNR, ROF, and skew monitoring of coherent channel using deep learning," *J. Lightw. Technol.*, vol. 37, no. 23, pp. 5907–5913, 2019.
[33] Z. Wan, Z. Yu, L. Shu, Y. Zhao, H. Zhang, and K. Xu, "Intelligent optical performance monitor using multi-task learning based artificial neural network," *Opt. Express*, vol. 27, no. 8, pp. 11 281–11 291, 2019.
[34] D. Wang *et al.*, "Cost-effective and data size–adaptive OPM at intermediated node using convolutional neural network-based image processor," *Opt. Express*, vol. 27, no. 7, pp. 9403–9419, 2019.
[35] L. Xia, J. Zhang, S. Hu, M. Zhu, Y. Song, and K. Qiu, "Transfer learning assisted deep neural network for osnr estimation," *Opt. Express*, vol. 27, no. 14, pp. 19 398–19 406, 2019.
[36] A. W. Snyder and J. Love, *Optical Waveguide Theory*, New York, NY, USA: Springer Science & Business Media, 2012.
[37] S. J. Garth *et al.*, "Few-mode optical waveguides and their study by the four-photon mixing process," Ph.D. dissertation, Dept. Appl. Math., The Australian National University, 1987.
[38] R. Paschotta, *Field Guide to Optical Fiber Technology*, Bellingham, WA, USA: SPIE, 2010.
[39] Y. Weng, X. He, and Z. Pan, "Space division multiplexing optical communication using few-mode fibers," *Opt. Fiber Technol.*, vol. 36, pp. 155–180, 2017.
[40] G. Rademacher *et al.*, "159 Tbit/s C+ L band transmission over 1045 km 3-mode graded-index few-mode fiber," in *Proc. Opt. Fiber Commun. Conf.*, San Diego, CA, 2018, Paper Th4C–4.
[41] G. Rademacher *et al.*, "93.34 Tbit/s/mode (280 Tbit/s) transmission in a 3-mode graded-index few-mode fiber," in *Proc. Opt. Fiber Commun. Conf.*, San Diego, CA, 2018, Paper. W4C–3.
[42] D. Soma *et al.*, "2.05 Peta-bit/s super-Nyquist-WDM SDM transmission using 9.8-km 6-mode 19-core fiber in full C band," in *Proc. Eur. Conf. Opt. Commun.*, Valencia, 2015, pp. 1–3.
[43] D. Soma, T. Tsuritani, and I. Morita, "10 Pbit/s SDM/WDM transmission," in *Proc. IEEE Photon. Conf.*, Reston, 2018, pp. 1–2.
[44] T. Hu *et al.*, "Demonstration of bidirectional PON based on mode division multiplexing," *IEEE Photon. Technol. Lett.*, vol. 28, no. 11, pp. 1201–1204, Jun. 2016.
[45] J. Vuong *et al.*, "Mode coupling at connectors in mode-division multiplexed transmission over few-mode fiber," *Opt. Express*, vol. 23, no. 2, pp. 1438–1455, 2015.
[46] W. S. Saif, A. M. Ragheb, H. E. Seleem, T. A. Alshawi, and S. A. Alshebeili, "Modulation format identification in mode division multiplexed optical networks," *IEEE Access*, vol. 7, pp. 156 207–156 216, 2019.
[47] W. S. Saif, T. Alshawi, M. A. Esmail, A. Ragheb, and S. Alshebeili, "Separability of histogram based features for optical performance monitoring: An investigation using t-SNE technique," *IEEE Photon. J.*, vol. 11, no. 3, pp. 1–12, Jun. 2019.
[48] T. A. Almohamad *et al.*, "Automatic modulation recognition in wireless communication systems using feature-based approach," in *Proc. 10th Int. Conf. Robot., Vis., Signal Process. Power Appl.*, Penang, Malaysia, 2018, pp. 403–409.
[49] V. Vapnik, *The Nature Statistical Learning Theory*, New York, NY, USA: Springer Science & Business Media, 2013.
[50] D. Basak, S. Pal, and D. C. Patranabis, "Support vector regression," *Neural Inf. Process.-Lett. Rev.*, vol. 11, no. 10, pp. 203–224, 2007.
[51] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
[52] Y. Zhao *et al.*, "Low-complexity and nonlinearity-tolerant modulation format identification using random forest," *IEEE Photon. Technol. Lett.*, vol. 31, no. 11, pp. 853–856, Jun. 2019.
[53] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Advances Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
[54] T. Tanimura *et al.*, "Deep learning based OSNR monitoring independent of modulation format, symbol rate, and chromatic dispersion," in *Proc. 42nd Eur. Conf. Opt. Commun.*, Dusseldorf, Germany, 2016, pp. 1–3.
[55] L. Guesmi, A. M. Ragheb, H. Fathallah, and M. Menif, "Experimental demonstration of simultaneous modulation format/symbol rate identification and optical performance monitoring for coherent optical systems," *J. Lightw. Technol.*, vol. 36, no. 11, pp. 2230–2239, 2017.
[56] J. M. Chambers, *Graphical Methods Data Analysis*, New York, NY, USA: CRC Press, 2018.
[57] T. O. Kva *et al.*, "Note on the R^2 measure of goodness of fit for nonlinear models," *Bulletin Psychonomic Soc.*, vol. 21, no. 1, pp. 79–80, 1983.
[58] D. Wang *et al.*, "Intelligent constellation diagram analyzer using convolutional neural network-based deep learning," *Opt. Express*, vol. 25, no. 15, pp. 17 150–17 166, 2017.
[59] A. Mecozzi, C. Antonelli, and M. Shtaif, "Coupled Manakov equations in multimode fibers with strongly coupled groups of modes," *Opt. Express*, vol. 20, no. 21, pp. 23 436–23 441, 2012.
[60] D. Marcuse, C. Manyuk, and P. Wai, "Application of the Manakov-PMD equation to studies of signal propagation in optical fibers with randomly varying birefringence," *J. Lightw. Technol.*, vol. 15, no. 9, pp. 1735–1746, 1997.

Waddah S. Saif received the B.Sc. degree in communication engineering from Hadhramout University, Yemen, in 2003 and the M.S. degree in electrical engineering from King Saud University (KSU), Riyadh, Saudi Arabia, in 2015. He is currently working toward the Ph.D. degree with the electrical engineering department, KSU. Since 2010, he has been with the Electrical Engineering Department, KSU, as a Graduate Researcher. His research interests include optical CDMA, coherent optical receivers, performance monitoring of optical networks, and machine learning.

Amr M. Ragheb received the B.S. (Hons.) and M.Sc. degrees from Tanta University, Egypt, in 2001 and 2007, respectively, and the Ph.D. degree from King Saud University, Riyadh, Saudi Arabia, in 2015, all in electrical engineering. He was a Teaching Assistant (TA) at Tanta University from 2003 to 2008. He was a TA at King Saud University from 2010 to 2015. He has over seven years of experience at the Photonics Telecommunication Laboratory. He is currently an Assistant Professor with King Saud University. He has contributed to the research areas, such as photonic-microwave integration, quantum dash-based lasers, free-space optical communication, optical modulation format identification, coherent optical receivers, multi-format highspeed optical transmitter, and passive optical networks.

Tariq A. Alshawi received the M.S. degree in electrical engineering from the University of Michigan, Ann Arbor, USA and the Ph.D. degree from the Georgia Institute of Technology, Atlanta, USA, in 2013 and 2018, respectively. He is currently an Assistant Professor with the electrical engineering department, King Saud University, Riyadh. His research interests include signal and image processing, learning-based models, bio-medical imaging, communication, and networks.

Saleh A. Alshebeili was the Chairman at the electrical engineering department, King Saud University, from 2001 to 2005. He has over 27 years of teaching and research experience in the area of communications and signal processing. He was a member of the Board of Directors with the King Abdullah Institute for Research and Consulting Studies, from 2007 to 2009, a member of the Board of Directors with the Prince Sultan Advanced Technologies Research Institute, from 2008 to 2017, where he was the Managing Director from 2008 to 2011, and the Director of the Saudi-Telecom Research Chair from 2008 to 2012. He has been the Director of the Technology Innovation Center, RF and Photonics in the e-Society, funded by the King Abdulaziz City for Science and Technology (KACST), since 2011. He is currently a Professor with the electrical engineering department, King Saud University. He has been on the Editorial Board of the Journal of Engineering Sciences of King Saud University from 2009 to 2012. He has also an active involvement in the review process of a number of research journals, KACST general directorate grants programs, and national and international symposiums and conferences.