Machine Learning Techniques for Optical Performance Monitoring and Modulation Format Identification: A Survey

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

***Abstract*—The trade-off between more user bandwidth and quality of service requirements introduces unprecedented challenges to the next generation smart optical networks. In this regard, the use of optical performance monitoring (OPM) and modulation format identification (MFI) techniques becomes a common need to enable the development of next-generation autonomous optical networks, with ultra-low latency and self-adaptability. Recently, machine learning (ML)-based techniques have emerged as a vital solution to many challenging aspects of OPM and MFI in terms of reliability, quality, and implementation efficiency. This article surveys ML-based OPM and MFI techniques proposed in the literature. First, we address the key advantages of employing ML algorithms in optical networks. Then, we review the main optical impairments and modulation formats being monitored and classified, respectively, using ML algorithms. Additionally, we discuss the current status of optical networks in terms of MFI and OPM. This includes standards, monitoring parameters, and the available commercial products with their limitations. Second, we provide a comprehensive review of the available ML-based techniques for MFI, OPM, and joint MFI/OPM, describing their performance, advantages, and limitations. Third, we give an overview of the exiting ML-based OPM and MFI techniques for the emerging optical networks such as the new fiber-based networks that use future space division multiplexing techniques (e.g., few-mode fiber), the hybrid radio-over-fiber networks, and the free space optical networks. Finally, we discuss the open issues, potential future research directions, and recommendations for the potential implementation of ML-based OPM and MFI techniques. Some lessons learned are presented after each section throughout the paper to help the reader identifying the gaps, weaknesses, and strengths in this field.**

***Index Terms*—Optical performance monitoring, modulation format identification, optical networks, machine learning.**

Manuscript received December 6, 2019; revised May 11, 2020 and July 6, 2020; accepted August 13, 2020. Date of publication September 1, 2020; date of current version November 20, 2020. This work was supported by the Research and Development Program (Research Pooling Initiative), Ministry of Education, Saudi Arabia, (RPI-KSU). *(Corresponding author: Tariq Alshawi.)*

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 the 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).

Maged A. Esmail is with the Communications and Networks Engineering Department, Prince Sultan University, Riyadh 12435, Saudi Arabia, and also with the Smart Systems Engineering Laboratory, Prince Sultan University, Riyadh 12435, Saudi Arabia (e-mail: mesmail@psu.edu.sa).

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

Digital Object Identifier 10.1109/COMST.2020.3018494

I. Introduction

OPTICAL networks are evolving to provide candidate solutions that can cope with the required data traffic. The capacity of such networks outperforms the radio frequency (RF) and copper cable networks. Therefore, fiber optics are key drivers for future emerging technologies and networks such as smart cities, Internet of Things (IoT), data centers, and 5G [1], [2]. However, the rapid advances in information technology and large data usage creat new challenges and limitations on optical networks in terms of bandwidth, latency, and reliability. To deal with such challenges, there has been an evolution/revolution in network transmission systems and architectures such as the utilization of advanced modulation formats, new multiplexing techniques, flex-grid transmission, and reconfigurable optical add-drop multiplexer (OADM). Nonetheless, these advances come at the cost of increased network complexity and create more challenges to the operation and management of optical networks.

On the other side, the current optical networks are static, where the physical channel path from the transmitter to the receiver is fixed. This network architecture reduces the complexity and requirements of the network nodes and terminals. However, the future optical networks, such as the elastic [3] and cognitive networks [4]–[6] are expected to be dynamic, spectrum grid-free, modulation format-free, and reconfigurable [3], [7], [8]. These features improve the overall network performance, flexibility, and efficiency, requiring the upgrade of the current optical nodes to be intelligent. Part of this intelligence is monitoring the signal performance and identifying its type to enable the network to determine the degradation source and initiate precautionary procedures to improve network reliability. Moreover, the ability of signal receivers to identify the signal type known as modulation format identification (MFI) enables building adaptive, efficient, and flexible networks where the signal type and bandwidth is determined based on the network conditions. Additionally, the ability to monitor the signal quality, i.e., optical performance monitoring (OPM), at network nodes enables the network

operator to make a tradeoff between spectral efficiency, signal quality, and reach distance. In such a case, higher-order modulation formats will be transmitted over good condition optical channels while lower-order modulation formats will be used in bad condition optical channels.

Machine learning (ML) techniques are promising solutions to add intelligence to nodes in the optical network. ML techniques can help the network nodes to learn from the conditions of the network and use this knowledge in future to optimize the network resources. Recently, there has been extensive research on the use of ML in optical networks such as fiber nonlinearity compensation [9], [10], nonlinear phase noise compensation [11], [12], and nonlinear equalizers design [13]–[15], and that by learning the behavior of the impairment or channel from the observed data. Then, a model can be built to mitigate the channel or the impairment effect. In addition, ML can be utilized for optical channel modeling [16], [17], especially for cases where theoretical modeling is not feasible. Recently, ML is proposed in the literature for performing OPM and MFI. This enables the use of adaptive modulation formats according to the transmission conditions, which in turn requires the utilization of suitable MFI techniques to identify the modulation format type at the receiver. Table I presents the definitions of the used acronyms in this article.

Figure 1 illustrates the architecture of future optical networks that include access, metro, and core networks. Over the past two decades, new technologies have been introduced to improve the network capacity, such as free space optical (FSO) communication in data centers and mobile access networks, in addition to the integration between space and wavelength division multiplexing (SDM and WDM) fiber networks in the metro segment. These new advances in fiber networks introduce tremendous challenges, in terms of latency, reliability, availability, adaptability, and heterogeneity. ML technologies can play an important role in network management, organization, and optimization. However, to achieve these benefits, a suitable training dataset, collected from the network elements, is required to reach an acceptable level of generalization that will perform well when deployed. Then, the resultant model can be used to control and optimize the network resources by performing remote OPM, MFI, and routing.

TABLE I
MAIN ACRONYMS

Acronym	Definition
AAH	Asynchronous amplitude histograms
ADTS	Asynchronous delay-tap sampling
AH	Amplitude histogram
ANN	Artificial neural network
ASE	Amplified spontaneous emission
AT	Atmospheric turbulence
BER	Bit error rate
BR	Bit rate
BRI	Bit rate identification
BVT	Bandwidth-variable transmitter
CA	Crossing amplitude
CART	classification and regression tree
CCA	Connected component algorithm
CD	Chromatic dispersion
CDF	Cumulative distribution function
CFSFDP	Clustering by fast search and find of density peaks
CL	Closure
CMA	Constant modules algorithm
CNN	Convolutional Neural Network
CS	compressed sensing
CZT-SS	Chirp-z transform software synchronization algorithm
DBSCAN	Density-based spatial clustering of applications with noise
DCF	Dispersion compensating fiber
DGD	Differential group delay
DI	Delay interferometer
DML	Direct modulated laser
DNN	Deep neural network
DP	Dual polarization
DPSK	Differential phase shift keying
DQPSK	Differential quadrature phase shift keying
DR	Dimensionality reduction
DSDM	Dense Space division multiplexing
DSP	Digital signal processing
DT	Decision tree
DWDM	Dense wavelength division multiplexing
EH	Eye height
EM	Expectation-Maximization
FB	Feature-based
FD	Fractal dimension
FFT	Fast Fourier transform
FM	Frequency modulation
FMF	Few mode fiber
FM-MCF	Few mode multi-core fiber
FSO	Free space optical communication
FWM	four-wave mixing
GA	Genetic algorithms
GMM	Gaussian mixture model
GPU	Graphics processing units
GVD	Group velocity dispersion
HOC	High-order cumulant
ICA	Independent component analysis
ID3	Iterative dichotomiser 3
IF	Intermediate frequency
IM-DD	Intensity modulation –direct detection
IoT	Internet of things
IQH	In-phase and quadrature histogram
ISI	Inter-symbol interference
JRMS	Root-mean-square jitter
k-NN	k-nearest neighbor
LB	Likelihood-based
LF	Low frequency
LO	Local oscillator
LSTM	Long-short term memory
MAE	Mean absolute error
MC	Mode coupling
MCF	Multi-core fiber
MDS	Multidimensional scaling
MFI	Modulation format identification

A. Advantages of Using ML in OPM and MFI for Optical Networks

MFI and OPM can be performed using two main approaches: classical and ML approaches. In classical approaches, we mainly have the Likelihood-based (LB) and features-based (FB) methods. The LB methods are optimum with respect to parameters estimation and classification accuracy. However, this optimality comes at the cost of requiring a prior knowledge of channel parameters and/or a comprehensible mathematical model describing the channel under consideration [18], [19]. In FB methods, MFI and OPM are performed by making use of hand-crafted features. These features are selected in an ad-hoc manner involving a long trial and error process to decide which features best describe different channel impairments and/or modulation formats. Furthermore, the actual implementation of a FB classifier often requires manual construction of a decision tree with a set of pre-determined thresholds. These thresholds are often computed using a theoretical noise-free signal or optimized with given channel conditions.

TABLE I
MAIN ACRONYMS (CONTINUED)

MIMO	Multi-input-multi-output
ML	Machine learning
ML-ANN	Multi-layers artificial neural network
MLP3	Multi-layer perceptron 3
MMF	Multi-mode fiber
MMS	Modulus mean square
M-PAM	M-ary pulse amplitude modulation
M-PSK	M-ary phase shift keying
M-QAM	M-ary quadrature phase shift
MSE	Mean square error
MTL	Multi-task learning
MZM	Mach-Zehnder modulator
NL-NP	nonlinear noise power
NRZ	Non-return to zero
OADM	Optical add-drop multiplexer
OAM	Orbital angular momentum
ODB	Optical duobinary
OFDM	Orthogonal frequency-division multiplexing
OOK	On-Off keying
OPM	Optical performance monitoring
OPTICS	Ordering points to identify the clustering structure
OSA	Optical spectrum analyzer
OSNR	Optical signal-to-noise ratio
PAED	Parametric asynchronous eye diagram
PBC	Polarization beam combiner
PBS	Polarization beam splitter
PC	Pulse carver
PCA	Principle component analysis
PDL	Polarization dependent loss
PMD	Polarization mode dispersion
PM-DD	Phase modulation-direct detection
PN	Phase noise
PNN	Probabilistic neural network
PoCR	Probability of correct recognition
PSP	Principle state of polarization
QoS	Quality of service
RF	Radio frequency
RL	Reinforcement learning
RMS	Root-mean-square
RMSE	Root mean square error
RNN	Recurrent neural network
RoF	Radio over fiber
RoFSO	Radio over free space optical communication
RT	Radon transform
ROP	Received optical power
RZ	Return to zero
SBS	Stimulated Brillouin scattering
SDM	Space division multiplexing
SDN	Software defined network
SE	Spectral efficiency
SLA	Service level agreement
SMF	Single mode fiber
SP	Single polarization
SPE	Stochastic proximity embedding
SPM	Self-phase modulation
SRS	Stimulated Raman scattering
SSB	Single side band
SVD	Singular value decomposition
SVM	Support vector machine
t-SNE	t-distribution stochastic neighbor embedding
VBEM	Variational Bayesian expectation maximization
VLC	Visible light communication
WDM	Wavelength division multiplexing
XPM	Cross-phase modulation

ML algorithms, on the other hand, can build predictive and classification models in an automated manner and without a prior knowledge of channel model or channel parameters. Indeed, ML algorithms based on the concept of end-to-end learning replace extracting hand-crafted features by iterating through deep learning architectures to automatically learn rich features directly from raw data; see Section II. For this reason, recently, ML-based algorithms have been used extensively in diverse fields of optical communication systems [20]–[24]. The use of ML techniques to perform OPM and MFI can provide many benefits either to the current optical networks or for the future adaptive and autonomous optical networks. Here, we discuss the main benefits of using ML for OPM and MFI.

1) Real-Time Adaptability Using Online Learning Procedures: Using ML in OPM of optical networks helps in utilizing information about the network status in real time [25]–[27]. OPM based on ML helps in building proactive networks by relying on constantly-adapting models. These models can predict the possibility of fault occurrence and recommend suitable solutions even when operation parameters are changing, thereby guaranteeing stable and reliable network operation. Hence, the optical network becomes able to monitor its functions, detect performance changes, and provide feedback to the network management to improve the operational performance. Improving network performance in an automatic manner requires ML-based MFI techniques to adapt the transmission speed based on OPM feedback. ML-based OPM and MFI techniques ensure reducing downtime and increase network availability [26]. Note that network reliability is very important for some applications as in medicine, where short diagnosis and treatment times are highly required.

2) Superior Flexibility and Reconfigurability: Future optical networks will be dynamic, flexible, and adaptive, where data rate and modulation formats can change according to the customer needs and physical link status. Leaning-based models can help in this regard by building OPM and MFI techniques that are data-driven and adaptable to the variety of operation conditions [25], [28]. This can be achieved using unified learning algorithm that works across the spectrum.

3) Improved Network Security: Similar to other data transmission networks, optical networks are vulnerable to unpredictable and detrimental attacks targeting service disruption, or unauthorized acquiring of transmitted data. As OPM provides continuous information about the optical parameters, any attack causes changes in the relationship between these parameters [29]. Since accurate models of physical-layer impairments under attacks do not exist, ML greatly helps for recognizing and detecting these security breaches. Furthermore, ML for OPM and MFI helps in securing the network by utilizing techniques such as incremental learning [30], [31]. This technique improves in-use models or reinforce their ability, by continually adapting the models based on a constantly arriving data stream, to detect the presence of an unauthorized data acquirement.

4) Advanced Network Operation Features at Reduced Cost: OPM and MFI techniques introduce some cost when built and integrated in the network. However, their ability to re-learning OPM and MFI models improves their cost-effectiveness compared to non-ML learning techniques. Additionally, an ML-based technique can be adapted to a variety of conditions to

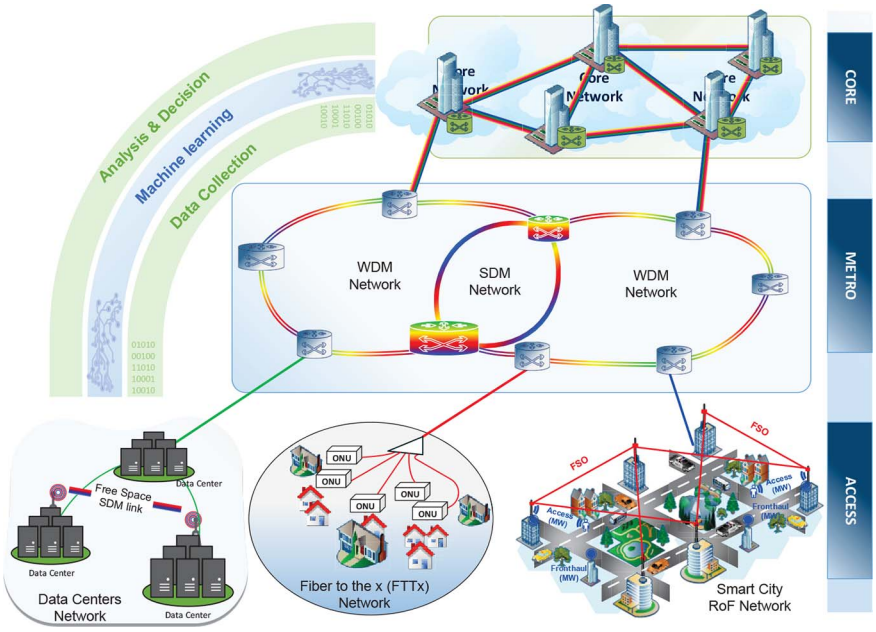

Fig. 1. Next generation heterogeneous optical network.

satisfy multi-tasks objectives leading to more efficient usage, as non-ML-based techniques are typically designed to tackle each task objective separately [32]. Therefore, ML adds to the reduction of the overall network operational costs by providing proactive monitoring of the network impairments. This, in turn, improves the quality of service (QoS) to respect the service level agreement (SLA) of customers.

5) Improved Network Efficiency: Building intelligent nodes with OPM and MFI functions helps the usage of network resources in an efficient way [25]. However, unless the monitoring functions can improve their performance and efficiency over time, as in ML-based OPM and MFI, there will be a gap between maximum efficiency of network resources and deployed solutions to achieve said efficiency. ML-based OPM and MFI functions can minimize waste in the network resources especially in the next generation optical networks; e.g., elastic optical networks. These networks are expected to handle large number of tunable parameters such as modulation format, signal power, symbol rate, adaptive coding rate, and adaptive channel bandwidth, etc. In such situations, it is very difficult to model a system through non-ML closed-form formulas that relies on approximations and sacrifices resource utilization. In contrast, ML algorithms can easily handle complex non-linear behavior, which leads to better resources utilization [33].

B. Review of Relevant Survey Articles

For many decades, MFI has been a hot topic of research in RF digital communication systems, where many approaches have been proposed in literature [18]. Similarly, optical networks are gaining more attention owing to the advances in photonics technologies and the interest in MFI and OPM for these networks is increasing dramatically. In the literature, few technical reviews have discussed the importance of MFI and OPM for current and future optical networks. The work in [34]–[36] addressed the development of different OPM techniques in optical communication networks such as the optical filtering, interpolation, and polarization nulling. Although many of these techniques are proposed to monitor different performance and impairment parameters, they did not address ML techniques that are considered effective solutions for OPM. In [37], the authors discussed the application of ML techniques in software defined network (SDN). Their contribution is focused on traffic classification, routing optimization, resource management, QoS and security. However, ML for OPM and MFI of optical networks is not covered. Recently, a review work pertaining to ML and its applications in optical communication networks is reported in [25]. In this work, the authors mainly focused on the issues related to the physical layer. Specifically, the aim was to describe the mathematical foundations of basic ML techniques and the benefits of using ML for optical networks. In that context, this work reviewed briefly ML-based OPM techniques proposed for optical networks. Similarly, the authors in [33], [38] reviewed ML techniques when applied to different areas in optical networks. One of these areas is the utilization of ML in OPM and MFI, which was covered briefly.

C. Summary of Paper's Contributions

In this work, our goal is to conduct an extensive review that covers ML-based OPM and MFI techniques in optical networks. Our contributions are summarized as follows:

- Review and discuss the current situation of optical networks in terms of MFI and OPM: standards, monitoring parameters, availability of commercial products and their limitations.
- Review, extensively the proposed ML techniques for MFI, OPM, and joint MFI/OPM for direct and coherent optical networks during the last two decades.

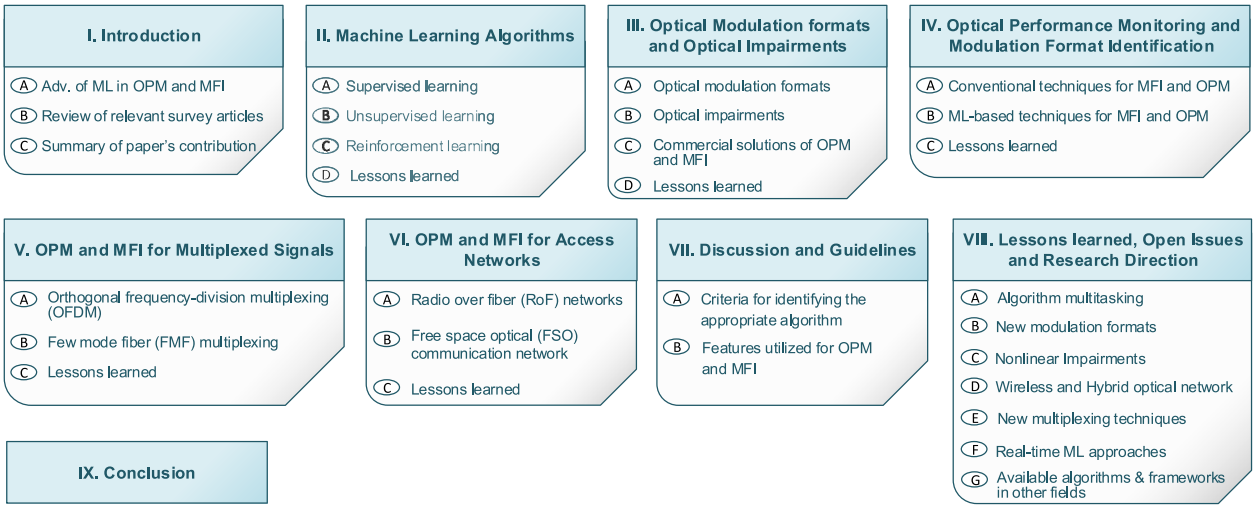

Fig. 2. Paper organization.

- Evaluate and compare the different proposed techniques in a tabular manner according to different algorithmic aspects.
- Review the available ML-based OPM and MFI techniques for other types of optical networks such as the hybrid radio-over-fiber (RoF) networks, the orbital angular momentum (OAM), or other fiber-based networks that use future SDM multiplexing techniques such as the few mode fiber (FMF).

D. Paper Organization

The rest of this survey is organized as illustrated in Fig. 2. In Section II, we describe the different ML algorithms used to build OPM and MFI functions, which include supervised ML, un-supervised ML, and reinforcement learning. Section III discusses the different types of modulation formats and optical impairments that are classified and monitored, respectively, in the proposed MFI and OPM techniques for optical networks. Moreover, it discusses the available OPM commercial products and their limitations. In Section IV, we first address some conventional techniques for OPM and MFI that are not ML-based. Then, we comprehensively discuss the ML-based techniques for OPM, MFI, and joint OPM/MFI in traditional fiber-based optical network. OPM and MFI for multiplexed signals are presented in Section V. In Section VI, we discuss the OPM and MFI for RoF and FSO networks. Further discussions and guidelines are presented in Section VII. Open issues in current proposed OPM and MFI techniques and recommended research directions are discussed in Section VIII. Finally, concluding remarks are shown in Section IX. For each section in the paper, some lessons learned are given in order to deepen the readers' understanding of the discussed topics in this survey.

II. Machine Learning Algorithms

ML is a branch of artificial intelligence that focuses on developing algorithms that, when given access to adequate amount of training data, can learn the relationship between inputs and outputs, without explicitly articulating the relationship. ML has been used extensively in image processing, medical applications and recently exploited in optical networks for MFI and OPM.

In general, ML algorithms aim to estimate an unknown function that maps inputs, representing the parameters of a given problem, to outputs representing solutions to said problem. When both input-output pairs are available for the learning algorithm, such a learning task is called "Supervised Learning." However, when only the inputs (or outputs) are available for the learning algorithm, the learning process is called "Unsupervised Learning." A special class of problems that requires a series of decisions or functions, to reach the final solution, can be learned using methods fall under "Reinforcement Learning (RL)." Additionally, the output type can affect the learning process. For instance, when the outputs represent a generic description of the class of data, e.g., "tree" versus "bird," the problem is known to be a classification problem. MFI in optical networks is a classification problem where supervised and unsupervised learning are utilized to identify the different modulation formats. On the other hand, when the outputs represent a precise numerical value for each given input, the problem is known as a regression problem. It is worth noting that the majority of the reported learning-based OPM techniques fall under supervised learning because OPM estimation problems typically require some form of regression. Furthermore, joint OPM and MFI is proposed but with individual training [39]. This kind of learning is called single task learning. Recently multi-task learning (MTL) [40], [41] is proposed to perform classification and regression simultaneously with joint training.

In this section, we review the commonly used learning-based algorithms and focus exclusively on the ones used for OPM and MFI in optical networks. The different types of ML algorithms are illustrated in Fig. 3.

A. Supervised Learning

Supervised learning is achieved by analyzing input-output pairs of examples (instances) of an unknown function that

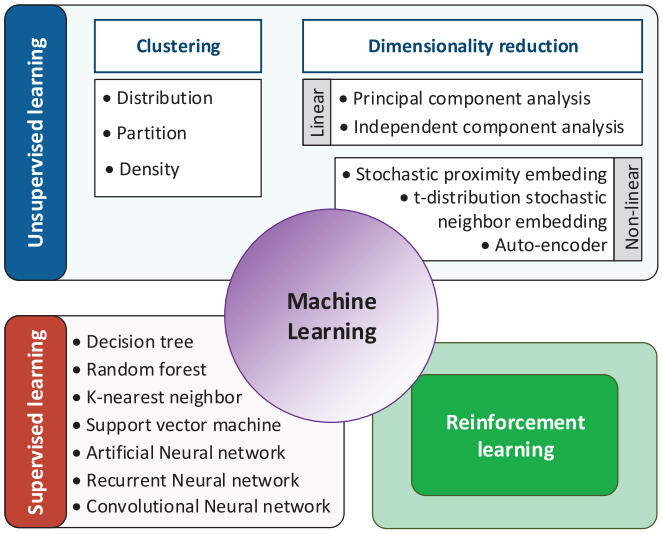

Fig. 3. Different types of ML algorithms used for MFI and OPM in optical networks.

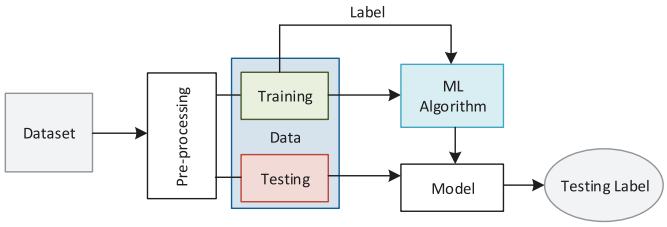

Fig. 4. A typical pipeline for learning a supervised model.

maps inputs to outputs. The learned model will approximate the function based on few assumptions, which gives rise to multiple supervised learning algorithms each is suited for a specific set of assumptions. To learn a model, supervised learning algorithms are fed a large number of training examples containing the input data and their corresponding labeled outputs (ground truth) to estimate the parameters of the learned model. The output during training phase is iteratively compared with the ground truth to minimize the error, known as the training error. Once a model is found, the model is tested using unknown examples during the training phase, typically called testing instances. The performance of the model is estimated by calculating the accuracy precision of the model when estimating the output of testing instances. A typical pipeline for learning a supervised model is illustrated in Fig. 4. Supervised ML algorithms can provide accurate results because the outputs are known during training phase and the parameters of the learning model are tweaked to minimize the error in estimating (predicting) such outputs. In the following discussion we list a few supervised learning algorithms that are widely used in learning-based OPM and MFI techniques.

1) K-Nearest Neighbor (K-NN): k-nearest neighbor (k-NN) is one of the simplest supervised ML algorithm that belongs to a subclass called nonparametric models. The idea behind k-NN is to estimate (predict) an unknown output from a given input by using "close; input-output pairs that are known during training. The closeness notion is dictated by the problem

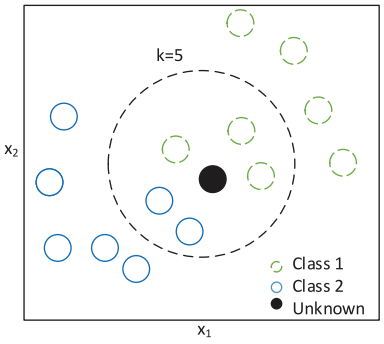

Fig. 5. Illustration of ML classification using k-NN algorithm with $k = 5$.

settings and the characteristics of the input data, nevertheless, Euclidean, Manhattan, and Hamming distance metrics are commonly used [42]. In the training phase, the dataset is sorted and indexed for easy and fast future retrieval. During testing, a majority vote of the k-nearest neighbors is used to estimate (predict) the unknown output of a given input. Figure 5 illustrates an example of classification problem using k-NN algorithm. The dashed-green and solid-blue circles represent class 1 and 2, respectively, while the filled-black circle represents the unknown (unlabeled) point. For $k = 5$, among the five closest neighbors shown in Fig. 5, three belong to class 1 and two belong to class 2. Thus, the unknown circle belongs to class 1. k-NN algorithm can also be used to solve regression problems by calculating the average of numerical target of the k-nearest neighbor. There are numerous proposals to estimate the optimum value of k, where cross validation is commonly used. The cross validation is performed by testing independent data at different values of k and select the value that provides the best accuracy. The k-NN algorithm is easy to implement and well suited for large datasets and highly non-linear mapping functions, however, storage requirements and retrieval time can limit its applications in real-time systems [43].

2) Support Vector Machine (SVM): This is a widely used supervised learning algorithm, which can be interpreted in the so-called kernel methods framework. This framework is based on mapping a low dimensional input feature space into a higher dimensional kernel feature space, and then solving a linear problem in that kernel feature space [44]. SVM is an extension of support vector classifier; however, its family of algorithms can tackle both classification and regression problems. The SVM algorithm seeks finding a hyperplane (a plane in the n-dimensional kernel feature space) that maximizes the margin, i.e., the separation distance between two classes within a dataset, as shown in Fig. 6; thus, SVM is a maximal margin classifier. Therefore, when data points of two classes are non-linearly intertwined, the data points are, first, linearized by projecting them into a high dimensional space using kernel functions such as the sigmoid, polynomial, and radial basis functions. In Fig. 6, the support vectors are simply the vectors defined in terms of the co-ordinates of individual observations (mapped data points), which are closest to the boundary between the classes. These vectors are used to estimate the parameters of the hyperplane. Their importance

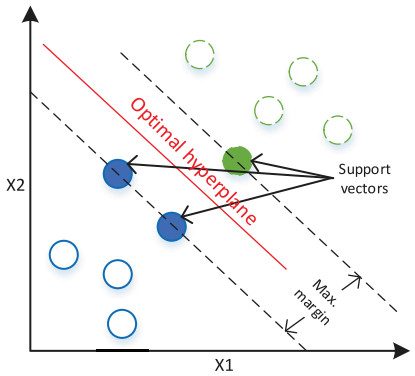

Fig. 6. Classification of a dataset consists of two classes using SVM.

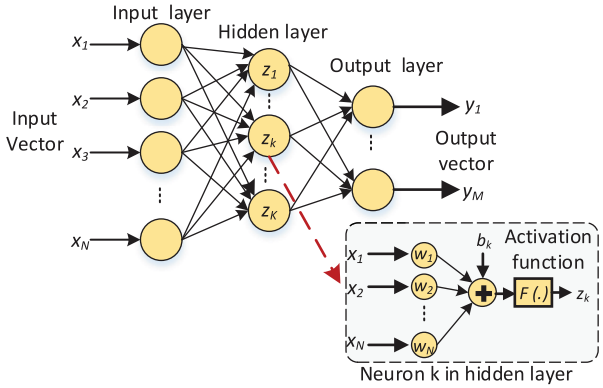

Fig. 7. ANN algorithm architecture (MLP3).

comes from their influence on the cost minimization procedure to identify the hyperplane parameters.

The SVM algorithms can be extended to classify between more than two classes by utilizing either "1-against-1" or "1-against the rest" methods. In 1-against-1 technique, multiple SVMs are used on parallel where each classifies between two classes and the final decision is based on majority vote. On the other hand, 1-against the rest technique redefines the problem as multiple independent problems where each classifies between a class and the rest of the classes combined in a "Super" class, and the final decision is taken based on majority vote [45]. The SVM algorithm can be modified to solve regression problems by utilizing the hyperplane as the estimated function with the addition of tolerance range. This ensures the existence of global minima in the optimization of loss function [43]. In practice, SVM can provide superior performance compared to *k*-NN and artificial neural networks (ANN) algorithms both in accuracy and training speed, however, its meta-parameters such as the kernel function, tolerance margin, soft margin, etc. can be difficult to optimize and depend on the data topology as well as domain knowledge [46].

3) Artificial Neural Networks (ANNs): Inspired by the biological structure of the neural cells in human brains, artificial neural networks present a computational model for the learning and decision making in biological entities [47]. However, state-of-the-art neural networks have diverged significantly from biological models. The most basic building block of an ANN is called a neuron [47]. In ANN, layers of neurons are connected in cascade and information propagates through the network where it goes under various transformation until it reaches the end and produces outputs. The goal of ANN algorithm is to change the parameters of each neuron in the network so that the output yields the desired values. Each neuron is modeled as a non-linear activation function whose inputs (coming from proceeding layers) are multiplied by weights (w_k) and shifted by bias coefficients (b_k) giving the overall ANN algorithm non-linear properties and making it possible to learn virtually any function. The basic ANN algorithm architecture contains three layers of neurons; an input layer (x), one hidden layer (z) and an output layer (y); this architecture is called multi-layer perceptron 3 (MLP3), as shown in Fig. 7. The input layer accepts an input vector, then transfers the vector samples to all neurons of the hidden layer. The connection between the layers is done by the network weights. In the training phase, the weight and bias coefficients are optimized. The training phase stops when it reaches its specified margin, i.e., the error between the output and the target [48]. The advantage of the ANN algorithm is that the testing phase is fast once training is completed. However, it is hard to interpret the trained ANN model. Additionally, during training, the optimization process may fail to reach a global minimum especially when the number of layers and/or neurons cannot model the complexity of the learning task.

There are many types of ANN, which have the same concept but differ in the architecture. For example, Probabilistic Neural Network (PNN) is a kind of feed-forward artificial neural network, which can approach a Bayes-optimal solution [49]. This solution chooses the class that has the maximum a posteriori probability of occurrence. PNN consists of input layer, pattern layer, summation layer, and output layer. Similar to ANN, the input layer accepts an input vector, which gets transferred to the pattern layer. This second layer calculates the Euclidean distance between the input vector and the vectors of all classes in a reference dataset. Therefore, the number of nodes in this layer is equal to the number of patterns in the reference dataset, with each node has a radial basis activation function. This nonlinear function accepts at its input the measured distance between the input vector and a reference pattern, and produces at its output a probability value. The reference dataset contains certain number of patterns for each class. Therefore, the summation layer performs an averaging over the outputs of the pattern layer. The averaging is performed for each class alone to produce a vector whose entries are of values ≤ 1, representing the probabilities of classes. Finally, the class with the maximum probability is identified as the true class.

When ANN algorithm contains a large number of hidden layers (typically 2 or more hidden layers), it is called deep neural network (DNN). DNN helps in complex non-linear modeling problems. However, it requires a large dataset for training, which is more time consuming than training an ANN.

4) Convolutional Neural Networks (CNN): Similar to multi-layer ANN, convolutional neural network (CNN) algorithm is a kind of DNN algorithms where multi-hidden layers are used, however, in CNN inputs are typically

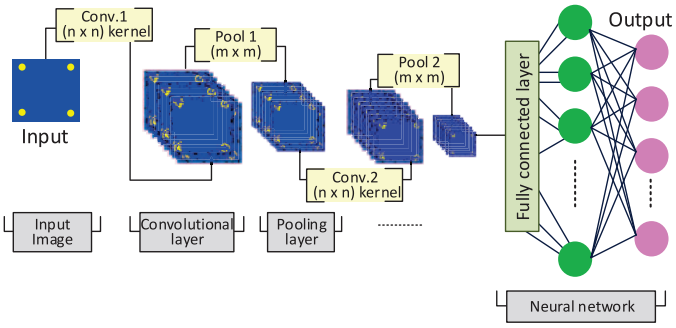

Fig. 8. Schematic diagram of CNN algorithm, where $n \times n$ is kernel filter size and $m \times m$ is down-sampling ratio.

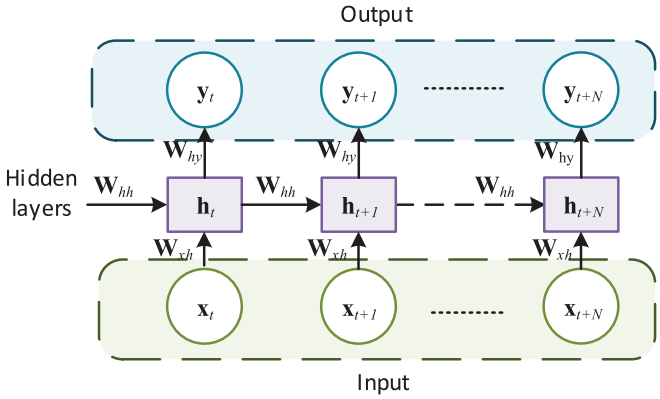

Fig. 9. RNN algorithm's schematic diagram. $\mathbf{x}_t$: input at time step t. $\mathbf{h}_t$: hidden state at time step t, $\mathbf{y}_t$: output at time step t, $\mathbf{W}_{xh}$: weights between input layer and hidden layer, $\mathbf{W}_{hh}$: weights between hidden layers, $\mathbf{W}_{hy}$: weights between hidden layer and output layer.

multidimensional such as images, i.e., inputs are highly correlated. Architecturally, CNN consists of three types of layers; convolutional layer, pooling layer, and fully connected layer, as shown in Fig. 8. In the convolutional layer, each point in the original multidimensional input is convolved with kernel filters that are initialized randomly to produce some feature maps. Then, the pooling layer reduces the dimension of each feature map using a filter such as average and max filters. The output of the previous layers is flattened and feeds into a fully connected neural network which perform the classification or regression tasks [50]. CNN algorithms are useful for both classification and regression problems, however, its training computational time and cost are high compared with other ML algorithms.

5) Recurrent Neural Network (RNN): Recurrent neural networks (RNNs) are a special class of DNNs, where the output of the network is fed back to the input making it especially useful when handling sequential input data, i.e., data with time-dependency, by utilizing internal state (memory). A commonly-used simple structure of RNN is an MLP with hidden layers that contain feedback loop to provide some information about the previous states, as shown in Fig. 9. The hidden layer is sharing the same weights (i.e., $\mathbf{W}_{xh}$, $\mathbf{W}_{hh}$ and $\mathbf{W}_{hy}$, which has the advantage of reducing the training parameters compared to other NNs [51]. RNN algorithm provides better performance by taking advantage of ability to

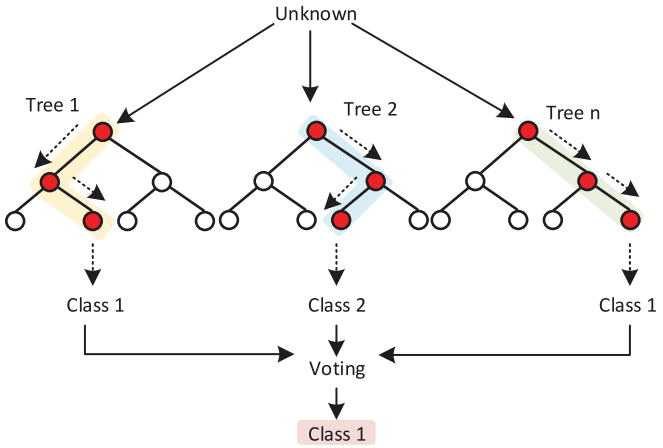

Fig. 10. Random forest schematic diagram.

memorize information and build complexity through recursive processing. However, their advantage can be limitation at the same time since optimization algorithms can be vulnerable to exploding or vanishing gradients in training phase, consequently, training might fail for long term sequential data. This problem is tackled in long-short term memory (LSTM) RNN, where the hidden layers are controlled using gates that pass information to the hidden layers and maintain extracted features from previous time steps [52].

6) Decision Tree (DT): This algorithm uses a tree-like graph to adopt both classification and regression. The tree structure includes a root node (the root of a tree) that connects to internal nodes, through links (branches), which in turn connect to leaf nodes (leaves of a tree). Each node represents a feature (attribute) and each leaf represents an outcome (target value). A DT model consists of an ensemble of (binary) decisions arranged in a hierarchical manner [53]. Each pathway from the root node through internal nodes to a leaf node represents a decision rule. Decision rules can be constructed from a training dataset using different algorithms such as the iterative dichotomiser 3 (ID3) and classification and regression tree (CART). The collection of all such rules is the corresponding DT model. Once a decision tree model is constructed, it can be utilized for classification or prediction of a new case. This algorithm is simple and fast, however, it is also critically sensitive to the features and thresholds used in making routing decisions. To combat such a disadvantage, multiple trees are trained and processed on parallel in what is known as the random forest algorithm.

7) Random Forest: This algorithm consists of multiple DTs which provide diversity across feature space. Each DT is constructed randomly from the original dataset during the training phase giving rise to a different subset of features for each tree. The output of random forest is determined by plurality voting for classification problems. In regression problems, the output will be the mean for all the decision tree's results [54]. Figure 10 shows an illustration of using a decision forest in classification problems. In comparison with single decision tree, random forest provides accurate results because it relies on majority voting of multiple decision trees. Additionally,

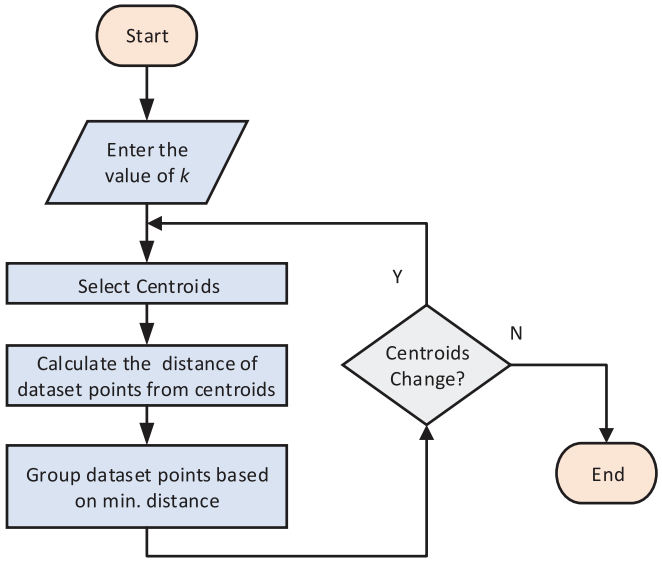

Fig. 11. Flow chart of k-means clustering algorithm.

they are less prone to overfitting that plagues most other ML algorithms.

B. Unsupervised Learning

Unlike supervised learning, unsupervised learning can process and extract information from data without a clear input-output pairs. Unsupervised models are developed by analyzing patterns common between dataset points in order to recognize (cluster) similar data points into classes, as the case in clustering and dimensionality reduction (DR) algorithms.

1) Clustering Algorithms: Dataset clustering aims to partition a dataset points into several groups or clusters according to their similarity. Clustering algorithms are divided into three types based on partition, distribution, and density.

a) Clustering algorithm based on partition: In partitioned clustering, the dataset points are initially partitioned into a set of clusters. Then, the cluster algorithm uses iterative procedures to update the partitioning by moving objects from one cluster to another. k-means [55] and k-medoids [56] are the two most prominent examples of partitioning algorithms. The principle of k-means algorithm is relying on selecting randomly k points as initial centroids (a center of a cluster). Then, each data point is assigned to the nearest centroid based on a distance metric. By iteratively updating centroids and moving points from one cluster to another, the clustering algorithm can reach equilibrium. The flow chart in Fig. 11 illustrates the procedure followed by the k-means clustering algorithm to perform dataset partitioning. k-means algorithm is easy to implement and fast, however, choosing the distance metric and the number of clusters k can be tricky. Additionally, equilibrium is not guaranteed for all partitioning problems.

b) Clustering algorithm based on distribution: Distribution-based clustering, such as Gaussian mixture model (GMM), is based on the assumption that points within a cluster are likely to be drawn from the same distribution [57]. GMM algorithm groups observations into k multi-dimensional Gaussian distribution and compute an initial estimate of distribution parameters such as the mean, or the covariance, and mixing coefficient (mixture weights) for each distribution. These parameters are then updated for each iteration. The limitation of GMM algorithm is its sensitivity to the initial estimates which sometimes leads to solution divergence. To overcome this problem, expectation-maximization (EM) techniques [58] are used, which iteratively find the maximum-likelihood estimates for model parameters. Because EM algorithm is complex and cannot be implemented directly, an approximation is often invoked by using, e.g., the variational Bayesian expectation maximization (VBEM) algorithm [59] to estimate the model parameters.

c) Clustering algorithm based on density: Density-based spatial clustering establishes a cluster of data points, in a high-density region, and regards neighbors that belong to a lower density region as outliers. By differentiating between core data points in a cluster and its outliers, density-based clustering can successfully partition challenging datasets where other clustering algorithms cannot achieve. Density-based spatial clustering of applications with noise (DBSCAN) algorithm [60] and ordering points to identify the clustering structure (OPTICS) algorithm [61] are examples of this type of clustering. DBSCAN applies threshold to decide on noisy points (outliers), which makes the algorithm sensitive to threshold setting. This is overcome using OPTICS algorithm which takes into consideration both the density and spatial closeness of data points within a cluster. More recently, clustering by fast search and find of density peaks (CFSFDP) algorithm, which is an updated version of DBSCAN algorithm, clusters data points by detecting 1) density peaks (representing cluster centroids), and 2) nearest neighbor for each cluster [62].

2) Dimensionality Reduction (DR): DR aims to transform a high dimensional dataset into a lower dimensional space while trying to preserve the overall structure and properties of the original dataset. There is a lot of research related to this area in literature. Some are concerned with preserving the global structure, i.e., the distance between the dataset points in higher and lower dimensions [63]. Other algorithms are concerned with preserving the local structure, i.e., dataset topology and continuity [64]. Recently, researchers showed interest in developing algorithms that preserve the global and local structures at the same time [65]. In general, DR algorithms can be categorized into linear and nonlinear. In this subsection, we discuss some algorithms that were used in OPM and MFI of optical networks.

a) Linear dimensionality reduction: In linear DR, the transformation from a higher dimension into a lower dimension is performed as linear mapping. This means the lower dimension data can be obtained by a linear combination of the original dataset points, for instance, by applying some weights or finding the projections of the original dataset points. Several algorithms that depend on linear transformation have been studied in literature such as the principal component analysis (PCA) [66]. PCA transforms a high dimensional correlated data into a lower dimensional uncorrelated data, which can be obtained by finding the orthogonal linear combination of the dataset points (PCA space). To find the PCA space, we

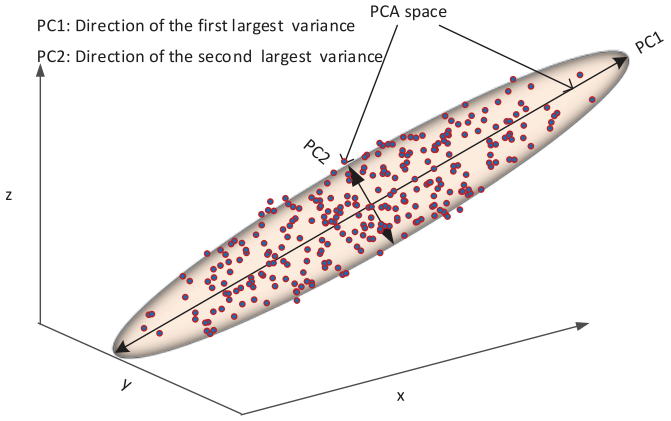

Fig. 12. Illustration of DR using PCA.

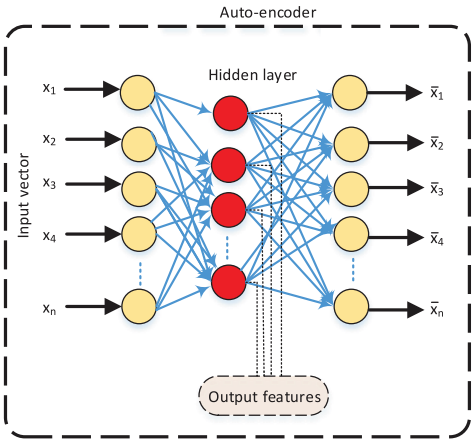

Fig. 13. Schematic diagram of auto-encoder.

evaluate the direction of the largest variance, i.e., the largest eigenvalue of a covariance matrix. Figure 12 illustrates how PCA can transform a dataset from 3D to 2D. Since the PCA is based on finding the maximum deviation, it eliminates the noise effect. However, it is hard to evaluate the covariance matrix accurately [67].

Another linear DR technique, closely related to PCA, is the independent component analysis (ICA) [68]. It is based on extracting the linear independent components of high dimensional data. Both PCA and ICA techniques preserve the global structure. These techniques are simple in implementation but their performance degrades when used on nonlinear datasets.

b) Non-linear dimensionality reduction: The nonlinear algorithms accomplish the reduction by applying a kernel function to the original data, where some can maintain the global structure such as the multidimensional scaling (MDS) algorithm [69] and stochastic proximity embedding (SPE) algorithm [70], while others attempt to preserve local structure such as the auto-encoder [71]. MDS performs a nonlinear transformation while maintaining the pairwise distance. The quality of transformation is evaluated by stress function, such as Kruskal's stress function, which calculates the error between the data in high and low dimensions. Similar to MDS, SPE maintains pairwise distances but it utilizes an iterative process to minimize the stress function which improves the performance. On the other hand, an auto-encoder is similar to an ANN algorithm, with the number of input and output layers are equal, as shown in Fig. 13. The auto-encoder is trained in a way such that its output matches the input. Hence, the hidden layers represent the input and are used as features extractor or data compressor.

Preserving both global and local structure can be achieved using the t-distribution stochastic neighbor embedding (t-SNE) algorithm [65]. In an SNE algorithm, the similarities of pairwise Euclidean distances in high and low dimensional spaces are expressed as joint probabilities. The initial values of lower dimensional space are chosen randomly, from a Gaussian distribution. Figure 14 illustrates reducing a dataset from 2D to 1D using SNE algorithm. Both data in lower and higher dimensions are projected into a Gaussian distribution. The similarity between high and low dimensional spaces is optimized using

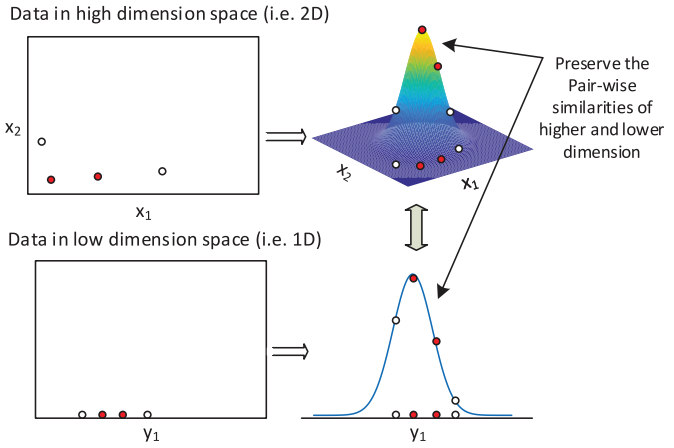

Fig. 14. Illustration of DR using SNE algorithm.

an optimization algorithm such as the gradient descent. In SNE, dataset points in the lower dimension are crowded. This makes it difficult to maintain the overall structure. An improved SNE algorithm (t-SNE) has solved this problem by projecting the data in lower dimensional onto t-distribution.

C. Reinforcement Learning (RL)

RL is a part of ML that is concerned with a sequence of actions. In applications where it is not feasible to obtain an accurate model or the complexity of the problem is significant, RL, also known as approximate dynamic programming, can produce an autonomous agent that can navigate the search space and provide solutions. This is achieved by iteratively taking actions, assessing state, and computing reward function whose value increases as it gets closer to the desired solution. While such a learning approach has produced impressive results [72], RL is critically sensitive to all three aspects of the algorithm (action, state, reward) making its training a challenging task. Recently, RL is applied in some applications in optical networks such as solving decision making problems where RL provides potential solutions regarding routing tables optimization. Moreover, it can be used for dynamic resources allocation [73], [74]. To the best or our knowledge,

TABLE II
HIGHLIGHTS ON MACHINE LEARNING TECHNIQUES

ML algorithm			Advantages	Dis-advantages	Ref.
Supervised ML Algorithms	*k*-NN		• Easy to implement • Supports highly non-linear mapping functions	• Sensitive to data quality and outliers • Requires large storage memory	[42], [43]
	SVM		• Effective in high dimensional spaces • Memory efficient owing to using a subset of training points (support vectors)	• Kernel function selection is heuristic • Inability to utilize data far away from class border	[44]–[46]
	ANN		• Supports a broad spectrum of function representation • Able to handle noisy data	• Hard to interpret the trained model • Limited to few layers due to vanishing gradients	[47], [48]
	CNN		• Provides state-of-the-art performance in dealing with two dimensional inputs	• High computational complexity	[50]
	RNN		• Suitable for handling sequential data (i.e. suitable for time dependent tasks)	• Recursive processing can be difficult to optimize • Vulnerable to exploding or vanishing gradients	[51], [52]
	DT		• Fast and easy to implement • Supports highly non-linear mapping functions	• Critically sensitive to features and thresholds	[53]
	Random forest		• Supports highly non-linear mapping functions • Less prone to overfitting problems	• Complex and time consuming, compared to DT	[54]
Un-upervised ML Algorithms	Clustering algorithms	Clustering based on partition	• Fast and easy to implement	• Choosing the number of clusters *k* is typically heuristic • Equilibrium among clusters is not guaranteed for partitioning problems	[55], [56]
		Clustering based on distribution	• Handle uncertain data that are geometrically indistinguishable • Soft clustering	• Complex to implement	[57]–[59]
		Clustering based on density	• Works well in the presence of noise	• Requires large storage memory	[60]–[62]
	DR algorithms	Linear DR	• Easy to implement	• Unfit for nonlinear datasets	[66]–[68]
		Non-linear DR	• Preserves the local neighborhoods (manifold learning)	• Complex to implement	[65], [69]–[71]
RL			• Applicable to complex optical networks	• Complex to implement • Equilibrium is not guaranteed	[72], [73]

RL has not been applied to OPM and MFI problems in optical networks. Nevertheless, it has been suggested that RL can be applied to real and complex optical networks because it can achieve fast convergence, meanwhile minimize the influence of non-optimal actions taken during the pre-trained-based algorithms [25].

Table II summarizes the different types of ML techniques that are used for OPM and MFI in optical networks. Also, it provides highlights on their advantages and disadvantages, which may help facilitating the selection of appropriate ML algorithm for a particular application.

D. Lessons Learned

In this section, we discussed several common ML techniques used in OPM and MFI for optical networks. Few lessons are to follow.

- The ML techniques that are proposed for OPM and MFI in optical networks are mainly of two types: supervised and un-supervised ML algorithms. In OPM, ML regressors are used because the impairment is a continuous parameter that has infinite number of values within a range. On the other hand, for MFI, classifiers are used to select one among different modulation formats. DR techniques are also of value in both OPM/MFI as they provide the tools to obtain features of reduced dimensionality and facilitate dataset visualization by reducing dataset dimensionality into order 2 or 3.
- Each ML technique has its own benefits and drawbacks. Some have high accuracy while others have lower computational complexity. Therefore, ML algorithm selection is subject to the problem at hand. For example, the *k*-NN algorithm is the simplest algorithm in the supervised category and few parameters are required for its implementation. However, it is not recommended for high dimensional data. The SVM is more appropriate for high dimensional space and linear separable problems. It is worth noting that the usage of kernel in SVMs made them non-linear learning algorithms but selecting the suitable kernel function is not a trivial task. DT algorithm is preferred when the examined dataset is small. However, it is susceptible to overfitting, especially when the tree is particularly deep. On the other side, the random forest algorithm aggregates many DTs to limit overfitting at the expense of its computational complexity. Deep learning algorithms are performing outstanding capabilities in

complex and nonlinear models. In addition, they deal directly with raw data; that is, there is no need for hand-crafted features extraction. However, they require a large dataset for training which may not be always available. Also, the training phase is often time-consuming. CNN is well-suited for two-dimensional input data. For sequential (i.e., time dependent) datasets, RNN is more appropriate than other supervised algorithms. Moreover, MTL algorithms are recommended in simultaneous classification and regression problems.

- Unsupervised ML-based algorithms are preferably used for clustering and DR purposes. For clustering, the k-means algorithm is fast and easy to implement. However, it employs hard-decision clustering with the requirement of a pre-selection of the number of clusters. The GMM-EM technique performs a more flexible (soft-decision) clustering, where it relies on the posterior probabilities. However, the requirement of optimizing many parameters makes its implementation costly. Other clustering algorithms such as OPTICS, DBSCAN, and CFSFDP are recommended for noisy datasets, as the clustering is established based on local densities. On the other hand, DR techniques are commonly used as a preprocessing step to improve learning and avoid curse of dimensionality. PCA and auto-encoders are examples of DR techniques. PCA finds the main components of the dataset that corresponds to the direction of maximum variance, while the auto-encoder is trained in a way such that its output matches the input. It is worth noting that some DR techniques can be used within the context of learning a model in a supervised manner, but their main original formulation is to be unsupervised.

III. Optical Modulation Formats Generation and Optical Impairments

In this section, we review different types of modulation formats where ML-based MFI techniques are proposed for their classification. This includes common modulation formats used in direct and coherent detection optical systems. Also, we show the representations of optical modulation schemes in different domains so that the readers can have better understanding of features extracted from each domain. Additionally, we discuss the general types of linear and nonlinear impairments that affect the OPM and MFI algorithms. Then, we address the current OPM and MFI in the deployed optical networks where international standards and commercial products are discussed.

A. Optical Modulation Formats

Many techniques have been proposed to classify different types of modulation formats [75]–[77]. In particular, some modulation formats are sensitive to specific impairments while others are more tolerant to the same impairments. Understanding the properties and techniques of the generation and detection of various modulation formats helps in identifying the challenges and selecting the appropriate algorithms for MFI and OPM. In this subsection, we review the various modulation formats where ML-based MFI techniques play an important role in their classification.

1) Intensity Modulation–Direct Detection (IM-DD):

a) On-off keying (OOK): On-Off keying (OOK) modulation format has dominated optical communication field for long period due to its simple transmitter and receiver structures. However, the advances in communication technology and the requirement for high spectral efficiency (SE) networks make OOK not a favored option for ultra-high data transmission. It provides acceptable performance for data rates up to 10 Gbps. However, it becomes more susceptible to noise distortions and channel impairments at higher network speeds. OOK can be generated as non-return-to-zero (NRZ) and return-to-zero (RZ) formats. In NRZ-OOK, the bit "1" is represented by a light pulse and no light for a bit "0." The RZ-OOK scheme is similar to the NRZ-OOK with a difference in the optical pulse width. The "1" bit has a different pulse width, compared to the NRZ-OOK, according to the required optical pulse duty cycle (defined as the ratio of the optical pulse width to the total signal period), such as 33% RZ, 50% RZ, and 67% RZ. Either direct modulated lasers (DMLs) or external modulators, such as Mach-Zehnder modulators (MZMs), are exploited to generate NRZ-OOK and RZ-OOK formats, while the later requires an additional modulator called pulse carver (PC) to control the laser pulse width. At the receiver side, a photodiode (PD) is used to convert the optical power into an electrical current. In general, RZ formats require more transmitter complexity and wider transmission bandwidth (i.e., less tolerant to channel dispersion) than NRZ formats. However, they are more tolerant to optical noise (i.e., require less optical signal to noise ratio (OSNR) for a given bit error rate (BER)) than NRZ schemes owing to the less impact of inter-symbol-interference (ISI) on RZ formats. Figure 15 (a) and (b) show simulated optical intensity time domain, optical intensity eye diagram, phase constellation, and optical spectrum at 10 Gbps NRZ-OOK and RZ-OOK, respectively. Besides, Table III shows and compares the transmitter and receiver hardware complexity of various modulation formats.

b) Optical duobinary (ODB): This modulation is similar to OOK in varying the light source amplitude based on the data, and using a simple direct detection at the receiver (i.e., similar hardware complexity as NRZ-OOK, see Table III). However, it requires a pre-defined phase relation between successive bits to reduce the effect of fiber channel dispersion. A pre-encoder generates three signal levels (i.e., "-1," "0" and "1") to represent data bits. Bit "1" is represented by either "1" or "-1" level while bit "0" is represented by "0" level. If two successive "1" bits are separated by an odd number of zeros, then the corresponding pre-coded signals will be 180° out of phase. This has the effect of reducing transmitted signal bandwidth and increasing the tolerance to fiber channel dispersion compared to OOK schemes. Figure 15 (c) shows the Optical duobinary (ODB) simulated optical intensity time domain, optical intensity eye diagram, phase constellation, and optical spectrum at 10 Gbps data rate.

2) Phase Modulation–Direct Detection (PM-DD):

a) Differential binary phase shift keying (DBPSK): Differential binary phase shift keying (DBPSK) or simply

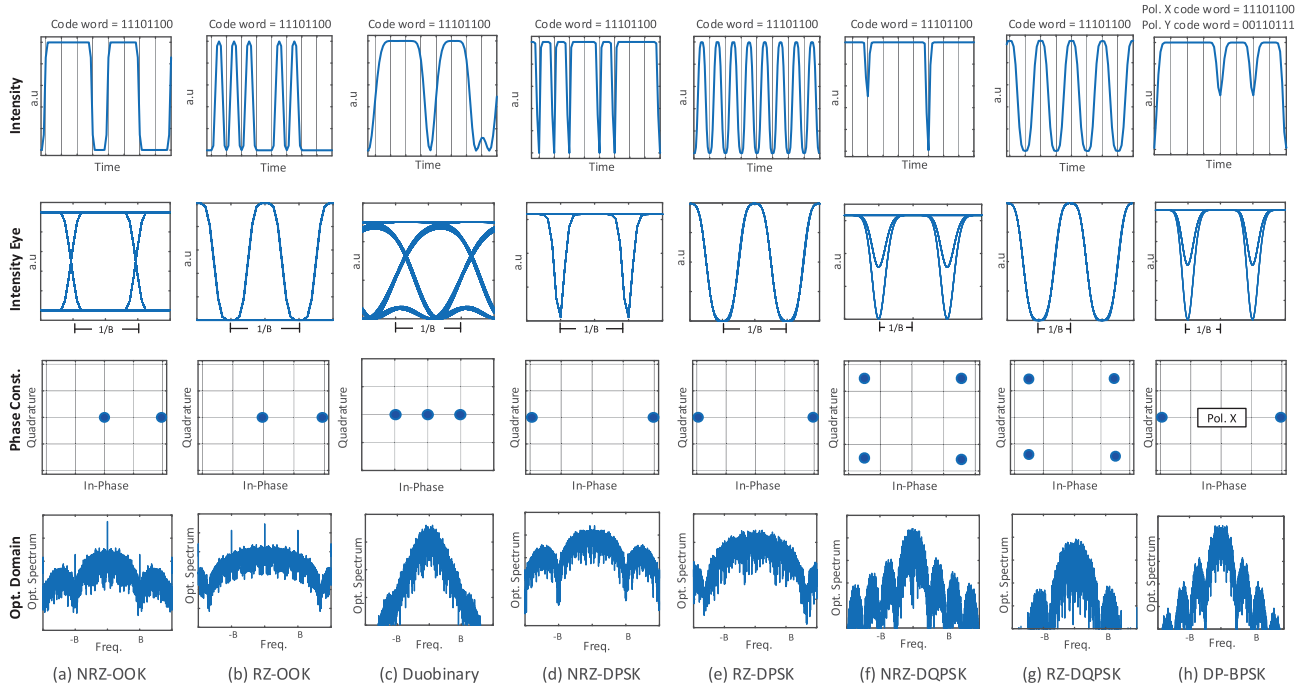

Fig. 15. Intensity time domain, intensity eye diagram, phase constellation, and optical spectrum of different modulation formats. The code word "11101100" is used to generate the intensity time domain signal for NRZ-OOK, RZ-OOK, Duobinary, NRZ-DPSK, RZ-DPSK, NRZ-DQPSK, and RZ-DQPSK modulations. For DP-BPSK modulation, the code words "11101100" and "00110111" are used for Pol. X and Pol. Y, respectively.

TABLE III
HARDWARE COMPLEXITY OF VARIOUS MODULATION FORMATS

Modulation Format	Transmitter	Receiver
OOK-NRZ	1 MZM	1 PD
OOK-RZ	1 MZM + 1 PC	1 PD
Duobinary	1 MZM	1 PD
DPSK-NRZ	1 MZM	1 DI + 2 PDs
DPSK-RZ	1 MZM + 1 PC	1 DI + 2 PDs
DQPSK-NRZ	2 MZMs + 1 PM	2 DIs + 4 PDs
DQPSK-RZ	2 MZMs + 1 PM + 1 PC	2 DIs + 4 PDs
M-PSK/M-QAM (SP)	2 MZMs + 1 PM	LO + 1 90^o hybrid + 4 PDs
M-PSK/M-QAM (DP)	4 MZMs + 2 PMs or 2 MZMs + 1 PM + DP emulator (PBS/PBC)	LO + 2 PBS + 2 90^o hybrid + 8 PDs

DI: Delay interferometer, PC: pulse carver, PM: phase modulator, PBS: polarization beam splitter, PBC: polarization beam combiner, LO: local oscillator, MZM: Mach-Zehnder modulator.

DPSK systems that employing delay-demodulation or differential detection, at the receiver side, are known as DPSK systems. DPSK formats encode the data using the difference in optical carrier phase between adjacent bits, such that a π phase change in the carrier's phase, between the successive bits, represents the bit "1" and absence of phase change represents the bit "0." Hence, signal power is almost constant over the transmitted data. This makes DPSK less prone to nonlinear and dispersion effects [78]. Besides, the π phase shift between transmitted symbols improves the receiver sensitivity with respect to OOK formats; see DPSK constellation diagram in Fig. 15 (d) and (e). DPSK can be generated as NRZ or RZ formats with the later requires more hardware complexity in signal generation, see Table III. Figure 15 (d) and (e) show the NRZ- and RZ-DPSK simulated optical intensity time domain, optical intensity eye diagram, phase constellation, and optical spectrum at 10 Gbps data rate.

b) Differential quadrature phase shift keying (DQPSK): In differential quadrature phase shift keying (DQPSK), each transmitted symbol carries two bits where the phase of the optical carrier hops between 0, $+\pi/2$, $-\pi/2$, and π. Hence, for the same data rate, information can be transmitted over less signal bandwidth (i.e., more SE) than single bit per symbol formats (i.e., OOK and PSK). This increases the tolerance to channel dispersion. The DQPSK spectrum shape is similar to DPSK scheme, however, its transmitter is more complex than DPSK, as it is implemented using two parallel MZMs and a phase modulator to generate the quadrature signals; see Table III. Figure 15 (f) and (g) shows the optical intensity time domain, optical intensity eye diagram, phase constellation, and optical spectrum for NRZ-DQPSK and RZ-DQPSK, respectively, at 10 Gbps data rate.

3) Advanced Modulation Formats With Coherent Optical Detection: Unlike the previously discussed modulation formats which use the direct detection techniques to recover the transmitted information, coherent-based modulation formats use the so-called coherent detectors for data recovery. Starting from 1980s, coherent optical communication has gained much attention owing to its advantages over traditional direct detection systems. Modulation formats that implement phase and amplitude modulation reduce the system cost and devices

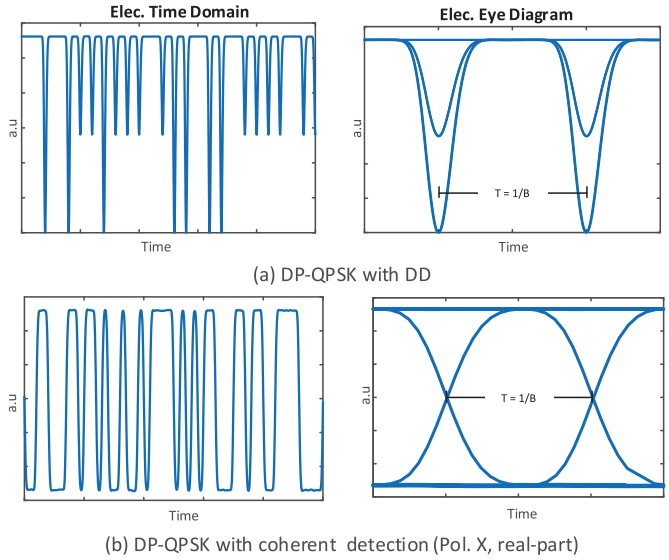

Fig. 16. Recovered DP-QPSK signal using (a) direct and (b) coherent detection.

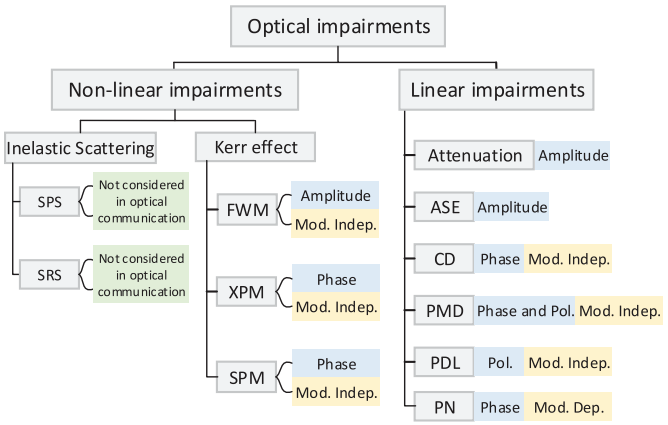

Fig. 17. Optical impairments classification. The blue shaded boxes indicate the optical signal parameter affected by the impairment. The yellow shaded boxes show the dependence of impairment's mitigation algorithm on the modulation format.

requirements in high-speed applications. The generated symbols may differ in phase only such as M-ary phase shift keying (M-PSK) or in combinations of amplitude and phase as in M-QAM. In this technique, a local oscillator laser is beaten with the modulated signal to improve the receiver sensitivity and allow the detection of both signal's intensity and phase. Besides, two orthogonal polarizations of the optical signal can be generated and detected in a dual polarization (DP) system to duplicate the SE. Moreover, the progress in digital signal processing (DSP) algorithms opens the door to build a digital coherent receiver where channel impairments can be digitally mitigated using post and pre-processing algorithms. On the other hand, coherent detection suffers from the high receiver cost that is at least 10 times that of direct detection system, for the same bandwidth. In Table III, we show the complexity of single polarization (SP) and DP coherent optical systems. Also, the simulated optical intensity time domain, optical intensity eye diagram, phase constellation, and optical spectrum for DP-BPSK signal at 10 Gbps data rate are illustrated in Fig. 15 (h).

So far, we show the various optical signatures, of common modulation formats, at different domains (i.e., time, eye, spectrum, and constellation). However, for the purpose of MFI, the optical signals are converted to the electrical domain with new signatures according to the type of optical receiver (i.e., simple PD or complex coherent detector). In Fig. 16 (a) and (b), we show the recovered DP-QPSK signal using direct detection and the coherent detection (for the real part of one polarization), respectively. It is clear that each detection method produces its own signature which affects the corresponding MFI process. Hence, an adaptive modulation format identifier will be of a potential value in heterogeneous optical networks.

B. Optical Impairments

As any communication medium, fiber cable introduces impairments that affect either the amplitude, phase, and/or the polarization of the optical signal. These propagation impairments can be divided into linear and nonlinear types. Such impairments degrade the system capacity to a limit that no useful data can be obtained at the receiver. In this subsection, among many impairments that affect the optical signal, we focus on the common impairments where ML-based techniques are reported in literature either to predict their levels or to identify the modulation format type in their presence. In specific, our focus here is on the linear impairments which are more dominant in current optical networks. The different types of optical impairments (gray shaded boxes), the effect on optical signal parameters (blue shaded boxes), and the dependency of impairment's mitigation algorithm on modulation type (yellow shaded boxes) are summarized in Fig. 17. Moreover, Fig. 18 illustrates the individual effect of the common impairments using different signal representations. The illustrated diagrams (ideal in red and impaired in blue) are generated using a 10 Gbps DP-QPSK optical system with a roll-off factor of 0.18.

1) Linear Impairments:

a) Signal attenuation: Attenuation causes decay in the power of light signal as it propagates through fiber channel. Thus, it affects the amplitude of the optical signal. The loss in optical power is owing to fiber-based intrinsic factors including scattering and absorption or extrinsic factors including environment, physical bending, and installation process stress. It has been shown that the light's power is decreasing exponentially with distance. Current standard single mode fiber (SSMF) cable introduces $\sim$ 0.2 dB/km attenuation factor at 1550 nm window, according to ITU-T G.652.B standardization [79]. It is worth to note that fiber attenuation can be mitigated using inline optical amplifiers (OAs).

b) Amplified spontaneous emission (ASE) noise: To extend the reach distance of a transmitted signal in fiber, optical amplifiers (OAs) are widely used. These amplifiers enhance the optical signal power at the expense of introducing some undesirable signal called amplified spontaneous emission (ASE) noise. The amount of ASE noise in the optical signal is defined by the OSNR parameter, which is the ratio of the

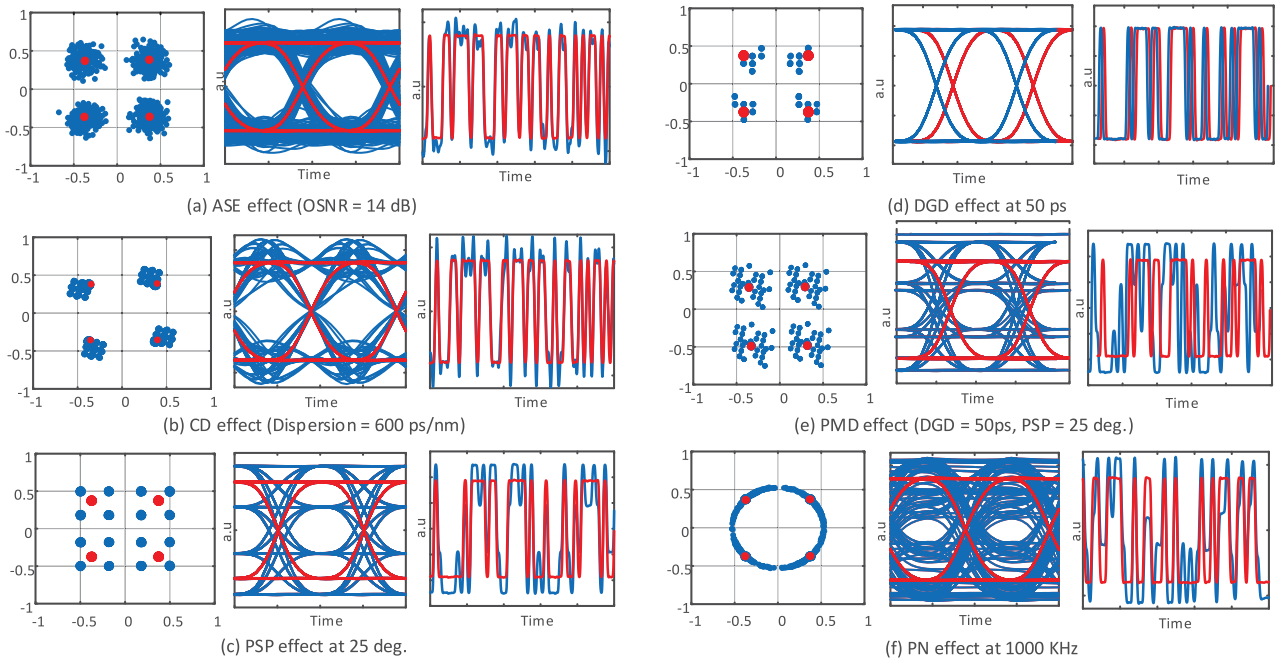

Fig. 18. The effect of the different impairments on the phase constellation, eye diagram, and time-domain for 10 Gbps DP-QPSK signal (ideal in red and impaired in blue). Note that the subfigures are only for Pol. X to avoid redundancy, where the eye-diagrams and time-domain pulses are plotted for the real-part of the signal.

optical signal power to noise power. OSNR is an important quality metric in optical networks, similar to SNR in wireless communications. In Fig. 18 (a), we show the effect of ASE with OSNR of value 14 dB on the phase constellation, eye diagram, and time-domain of 10 Gbps QPSK signal. This effect degrades the recognition accuracy of different classifiers and calls for the implementation of robust monitoring algorithms.

c) Chromatic dispersion (CD): Chromatic dispersion (CD) is a linear impairment that affects the optical signal phase. CD results from both waveguide and material dispersions which cause optical pulse broadening. Waveguide dispersion depends on the fabrication process; however, material dispersion arises from the frequency dependent of the propagated signal on the refractive index of the fiber. In high-speed transmission systems, this causes different spectral components of optical signal to travel through fiber channel with different speeds which leads to broadening the transmitted pulses. Hence, it introduces ISI between adjacent pulses and its severity depends on the fiber channel length and system bitrate. This phenomenon is commonly referred to as group velocity dispersion (GVD) or simply fiber dispersion. CD is a static dispersion as it changes very slowly with time and with small amount due to variation in fiber temperature. CD can be mitigated either optically using dispersion compensating fiber (DCF) modules or electronically using digital filters [80]. Note that the mitigation techniques of CD do not depend on the modulation format. Hence, the MFI algorithms can be applied after removing this effect from the received data. In Fig. 18 (b), we show the effect of CD with fiber dispersion of 600 ps/nm on the phase constellation, eye diagram, and time-domain for 10Gbps QPSK signal. This impairment affirms the need for building intelligent nodes in the future self-reconfigurable heterogeneous optical networks, where signal routing depends on network conditions. This stimulates the development of fast and reliable OPM algorithms that can cope with the dramatic changes in CD amount.

d) Polarization mode dispersion (PMD): Polarization mode dispersion (PMD) affects both the phase and polarization of optical signals. PMD has its origin from optical birefringence. In ideal cases, fiber has cylindrical geometry, so the two orthogonal polarizations of the optical signal travel with the same group delay. However, in real systems, the two-polarization components travel through the fiber with different group delays, due to manufacturing imperfection or fiber tension. The fiber asymmetry breaks the degeneracy of the orthogonal polarized modes resulting in birefringence, in addition to a difference in the group velocity of the two orthogonal modes, known as the differential group delay (DGD). The instantaneous value of DGD varies along the fiber and follows Maxwellian distribution [80]. The mean DGD value is known as the fiber PMD and measured in ps/$\sqrt{\text{km}}$. In modern optical systems, the mean DGD value is in the range of 0.2 ps/$\sqrt{\text{km}}$. This value has a major effect on optical systems operating at data rates greater than 40 Gbps because it can be a significant fraction from the symbol period. Associated with fiber's DGD, fiber's principle state of polarization (PSP) rotates along fiber channel [81]. These two phenomena together form the 1^{st} order PMD effect. Note that PMD is a dynamic dispersion since it changes on a time scale of milliseconds. The effect of PMD can be mitigated using an adaptive multi-input-multi-output (MIMO) equalizer [80]. Figure 18 (c), (d), and (e) show the effect of DGD (50 ps), PSP rotation (25 deg.), and 1^{st} order PMD (DGD $=$ 50 ps, PSP $=$ 25 deg.) on the phase constellation for 10 Gbps DP-QPSK signal (Pol. X). The eye

diagram and time-domain representations are also plotted for the same signal but only for the real part of Pol. X to avoid redundancy.

e) Polarization dependent loss (PDL): Polarization dependent loss (PDL) is another impairment that affects polarization multiplexed systems. It results from the polarization dependence of optical components which causes the two polarization components of an optical signal to reach the destination with different OSNR values. The most common sources of PDL include passive couplers, multiplexers, and photodetectors. The PDL can be expressed as the loss difference peak-to-peak of various states of polarization.

f) Phase noise (PN): Phase noise (PN) originates from the laser at the transmitter side or the local oscillator (LO) at the receiver side. This impairment can be modeled by a Wiener process and it depends proportionally on the laser (transmitter laser and LO) linewidth and symbol duration. PN has no effect on the performance of non-coherent optical systems since the decision on those receivers depends on the signal intensity. However, it has a sever effect on coherent optical systems because the data is modulated in the optical signal phase. The optical carrier's phase recovery, due to PN, depends on the modulation type [82], [83]. Thus, the MFI will be applied before the phase recovery algorithms. Figure 18 (f) shows the effect of laser PN for linewidth of 1000 KHz on the phase constellation, eye diagram, and time-domain for 10 Gbps QPSK signal.

2) Non-Linear Impairments: Since optical fiber is a nonlinear medium, several nonlinear effects start to appear as the optical power level increases. These nonlinearities result from the coexistence of high strength optical fields simultaneously in optical channel or the interaction of high intensity optical fields with the acoustic waves and/or molecular vibrations. This causes power gain or loss at different wavelengths, wavelength conversion, and crosstalk between wavelength channels. Optical nonlinearities can be classified into two general categories: nonlinear inelastic scattering process which includes the stimulated Raman scattering (SRS) and stimulated Brillouin scattering (SBS) and Kerr effect (i.e., the dependence of refractive index on light intensity) which includes the self-phase modulation (SPM), cross-phase modulation (XPM), and four-wave mixing (FWM). SBS, SRS, and FWM result in power gain or loss in wavelength channels by adding gain to some channels or depleting power from others. SPM and XPM affect only the phase of the signals which causes frequency chirping in digital pulses. Frequency chirping worsens the pulse broadening due to dispersion. It worths to note that in other contexts, some nonlinear effects can be very useful to perform important functions. This includes wavelength conversion, in WDM networks, using FWM and optical wave amplification using SRS and SBS.

C. Current Available Commercial Solutions and Standards/Recommendations for OPM and MFI in Optical Networks

Owing to the need of OPM for signals in the current optical networks, different commercial products and standards/recommendations have been released. However, to the best of the authors' knowledge, there are no commercial solutions that use ML for OPM. Moreover, there are no commercial solutions to MFI for the existing optical networks, regardless built using ML or not. This is because of to the fact that MFI application will be employed in the next generation of optical networks. These networks are going to be adaptive, where nodes and receivers will be capable of identifying the modulation type to facilitate signal demodulation. Therefore, in this subsection, we will discuss the commercial products and standards/recommendations currently available for OPM.

OPM can be achieved electrically by end-to-end monitoring of some digital parameters such as the BER [84]; here BER provides accurate measurements of signal quality. However, it requires complete demodulation at higher network layers instead of performing this task at the physical layer. This reduces impairments monitoring speed. In addition, it is difficult to determine the root causes of signal degradation because BER value involves the accumulated effects of all impairments in the network. Furthermore, the technology is replacing electronic devices in optical nodes with optical devices to improve the speed. Hence, monitoring BER electrically at higher layers is not possible in all network nodes.

The first optical based commercial approaches for monitoring optical networks include optical power meter and optical spectrum analyzer (OSA) [85]. Optical power meter can be utilized to monitor the aggregate power transmitted through a fiber. It is a simple and low-cost widely used device. However, it does not work with multichannel networks such as WDM. To measure the power of a specific channel in WDM networks, OSA-based device is used which, in addition to power measurement, provides information about the channel's wavelength drift.

The noise power in optical networks is one of the main limiting impairments. OSNR is an important parameter in analyzing the signal quality in the optical domain. It defines the signal quality by measuring the signal power to the noise power. The available solutions in the market include those for in-service monitoring of wavelength, optical power, and OSNR of each channel in dense WDM (DWDM) networks in C and L bands; examples of which are provided by Optoplex and Lightwave2020 Companies [45], [47].

In addition to the noise power impairment, two other important impairments that affect optical signals are fiber CD and PMD. Recommended test and measurement methods of CD and PMD impairments are reported in the International Telecommunication Union-Telecommunications (ITU-T) G.650.1 [86] and ITU-T G.650.2 [87] recommendations, respectively. For CD measurement, pulse delay technique and phase shift technique can be used among other techniques [86]. For PMD measurements, Stokes parameters evaluation method, state of polarization method, and interferometric method are recommended [87]. Some commercial devices are available for CD and PMD impairments monitoring [88]–[91]. For instance, PE.fiberoptics Company has developed a product to measure out-of-service CD and PMD in DWDM networks supporting up to 40 Gbps data speed [89]. For high speed DWDM networks, VIVA Company provided

out-of-service solution for CD and PMD monitoring with speeds up to 100 Gbps [91].

Field trials for in-service monitoring of CD and PMD for WDM networks have been reported in literature [92], [93]. In [92], Stokes parameters based technique is used for predicting the CD and PMD values. The trial was performed for 820 km fiber link between Stockholm to Hudiksvall in Sweden. However, this technique introduces modifications to the transmitter, which is not appropriate as it increases the overall cost of the network. Another field trail was conducted in [93] for monitoring CD and PMD for a WDM network over 140 km fiber cable length and 10 Gbps data rate speed. In this work, part of the signal is taped, sampled asynchronously using two samplers at rate lower than the data rate. The two sampled signals are plotted in a 2D histogram, which is used to monitor the CD and PMD impairments.

The utilization of ML for monitoring purposes in current optical networks is reported in [94], where a field trial was conducted for monitoring OSNR in SDN-based optical network of 436.4 km fiber length in U.K. The obtained measured values of OSNR are used to improve the spectral efficiency by utilizing probabilistic-shaping based bandwidth-variable transmitter (BVT). To the best of authors knowledge, this is the first time ML is being used in real optical networks for monitoring signal impairments.

D. Lessons Learned

In this section, we reviewed the widely used types of optical modulation formats and impairments that have been considered in the optical communication field. Few lessons can be drawn as follows.

- The traditional optical modulation formats (i.e., intensity, phase, and intensity and phase schemes) enjoy different signatures in different signal representation domains, which could be exploited for automatic MFI and/or OPM. However, the presence of impairments may cause considerable distortions for these signatures, which make the task of MFI and OPM difficult, especially when multiple types of distortions are present.
- The choice of the optical receiver type (i.e., direct or coherent receiver) is a key factor in building optical systems. The former requires low-cost devices, at the expense of having low SE. The later doubles the system SE, but mainly suffers from the ultra-high cost of its components. These components are required to increase the receiver sensitivity and allow real-time implementation of advanced digital processing of received samples.
- The presence of optical impairments harden the reconstruction process of information bits; therefore, optical impairments need to be mitigated for proper demodulation. Fortunately, the majority of impairment mitigation algorithms are modulation independent. However, the mitigation techniques for carrier phase and frequency offsets are often modulation dependent, which require a prior knowledge of the modulation format [95]. Additionally, SBS and SRS impairments can be neglected in optical communications, as these impairments are stimulated by the presence of high power requirement (>& 20 dBm), which is not applicable in fiber-based communications [96].
- There are no commercial products or standards/recommendations for MFI. However, the commercial OPM devices are only available to monitor signal noise, CD, and PMD. The currently available monitoring devices require service-cut to monitor specific impairments. Therefore, a non-intrusive solution that can simultaneously monitor all these types of impairments is mandatory for future optical networks.

IV. Optical Performance Monitoring and Modulation Format Identification

The advances in optical networks have established an increase in the data rate using advanced modulation formats and introduced different access techniques in the physical layer. This, in turn, has increased the network complexity and effect of channel impairments. Moreover, future optical networks will be adaptive where the network resources are allocated based on the link condition and customer requirements. Therefore, the need for OPM in the optical network is becoming mandatory. The task of OPM is to estimate optical signal impairments such as ASE noise (defined in terms of OSNR parameter), CD, PMD, and PN at the network nodes without disturbing the traffic. OPM reduces the network downtime and increases its availability and reliability. The predicted values enable the network elements to compensate for these impairments and optimize the adaptive network resources. In addition, development of optical nodes with built-in signal format classification and baud rate estimation will pave the road to the development of autonomous optical networks. Such networks have the capability to identify the signal type and perform signal processing tasks such as de-modulation, equalization, filtering, etc. without the need for prior signal information.

In this section, we first discuss briefly some conventional OPM and MFI techniques for optical networks. Then, we comprehensively discuss the proposed ML-based techniques for OPM and MFI.

A. Conventional OPM and MFI Techniques

1) Conventional OPM Techniques: Over the last two decades, many OPM techniques have been proposed. The early approaches to perform OSNR monitoring were relying on out-of-signal band noise power measurement [43]. This technique was not effective in WDM networks due to the existence of the OADMs, which remove major part of ASE out of band noise. Thus, the in-band monitoring provides best and effective solution especially in WDM networks. Some techniques for measuring the in-band noise include wavelength down-converters [97], polarization nulling [98]–[105], delay interferometer (DI) [106]–[113], and electrical sampling [114], [115] which have some limitations. Wavelength down-converters based solutions require additional hardware to select the single sideband (SSB) signal and down-convert it to an intermediate frequency (IF). In polarization nulling

based solutions, two polarizations are used where only one polarization carries the signal and the other is used for noise measurement. Then, the received signal is split into two orthogonal polarization components to measure the OSNR. If depolarization such as PMD occurs, the polarization that carries the noise will be mixed with the signal which leads to inaccurate OSNR measurement. The use of DI for OSNR monitoring does not depend on either PMD or CD. Specifically, the received optical signal (signal and ASE noise) is measured by constructive interference, while the ASE noise is measured by excluding the optical signal using destructive interference. This method is insensitive to CD or polarization effects. However, it needs precise wavelength control. In the electrical sampling technique, the received signal is sampled using two samplers with a short delay between them. The obtained sampled signals are displayed in a 2D plot. Then, OSNR can be measured by analyzing the statistical mean and variance of the 2-D phase portrait. This technique can monitor a wide dynamic range of OSNR. However, the phase portrait is sensitive to PMD and CD impairments, which results in inaccurate OSNR measurement. In addition to the previously reported techniques, the authors in [116]–[118] proposed to analyze the noise produced by beating the optical signal with ASE noise at the square-law receiver (i.e., direct detection receiver). This approach, however, requires expensive RF components.

In [119]–[123], OSNR monitoring is proposed to be integrated in coherent digital receivers. In [119]–[121], the second- and fourth-order moments of the received signal are used for OSNR monitoring. Although this technique is insensitive to frequency offset and PN, its performance is limited by the performance of the used equalizer [124]. Exploiting the signal's cumulative distribution function (CDF) for OSNR monitoring is reported in [122], [123]. In [123], a reference CDF of a candidate OSNR is compared with the CDF of the normalized amplitude of the received signal, where the candidate OSNR that has the closest average distance to the signal CDF is selected. To reduce the computational complexity of this technique, the authors in [122] proposed a non-aided OSNR estimation technique that exploits the empirical CDF to extract the information of noise variance.

For CD impairment monitoring, the early approaches were based on detecting the phase of the optical signal [125]. This is achieved by inserting subcarriers in the transmitter side and observing the RF tones in the receiver. Another technique is to apply frequency modulation (FM) on a pilot subcarrier in the transmitter side. In the receiver, the clock phase of the signal is detected and observed to monitor any deviation [126]. All previous methods require modification in the transmitter, which increases the system's cost. Vestigial side band filtering and clock phase detection [127] have also been used for CD monitoring, which do not require any modification in the transmitter side. However, it has limited monitoring range.

For PMD impairment monitoring, the usage of eye diagram has been proposed in literature, but it is costly due to its high-speed clock requirement [128]. Alternatively, methods based on RF power spectrum [129] can be used, however, their performance is influenced by the presence of severe CD impairment.

2) Conventional MFI Techniques: In general, MFI techniques can be divided into two types: LB and FB schemes. LB schemes are related to the probabilistic model of the modulation to be classified. They require formulating the exact probabilistic model. Therefore, the computational complexity increases. On the other hand, FB schemes only require extracting some features from the received signal. Therefore, they are simple and effective at the same time, and widely used in optical communications [130]–[138].

MFI-based normalized power distribution was proposed in [130], where the empirical probability distribution of received signal powers is used to distinguish between the different modulation formats. The proposed technique is independent of PN and offset frequency. However, it is not adequate for some modulation formats such as M-PSK signals because they have the same power distribution. Therefore, this method is limited only for multi-level modulation formats. Besides, it requires a high OSNR to achieve better MFI. On the other hand, for signals suffering low OSNR values, MFI-based compressed sensing (CS) and higher-order cyclic cumulants were proposed in [131]. Since various modulation formats have different cumulants (i.e., fourth-order cumulants), these cumulants can be used for classification purposes. This technique utilizes CS before the classification stage to enable low-cost classification. It is noise-tolerant and can be used for both M-PSK and MQAM signals; however, only the effect of ASE impairment is considered.

Features extracted from the signal's amplitude deviation were exploited in [132]. The ratio of amplitude deviation for two modulations (i.e., ideal DP-QPSK and DP-16-QAM) was utilized as a reference feature. Similarly, MFI based on parameters extracted from the signal's phase and amplitude distributions were reported in [133]. These parameters include the differential phase and amplitude ratio of adjacent symbols. The product of these two parameters can be used to discriminate the various modulation formats. Both techniques do not require a large number of samples. However, they are sensitive to decision thresholds. In [134], the entropy of the amplitude histogram was used for MFI. The entropy of amplitude histogram with a different number of bins (i.e., 3, 5, 7, and 9) was used to identify different modulation formats. This technique is tolerant to fiber nonlinearity. However, it can only distinguish M-QAM signals. In addition, it is sensitive to predefined threshold values. Exploiting the DC component of the received signal for MFI is reported in [135]. In this technique, the ratio of the 4^{th} power (exponent) of the DC component to that of the received signal without phase rotation is utilized to distinguish the different modulation formats. This technique can tolerate PN and fiber nonlinearity. However, it is limited only for three modulation formats.

Using normalized received signal's amplitude distribution for MFI is proposed in [136]. This technique is based on finding the CDF of the received signal's amplitude and compare it with reference CDFs of all possible candidate modulations. The decision is taken by measuring the similarity (i.e., minimum average distance) between the received signal's CDF and reference CDFs. This technique is insensitive to PN and frequency offset. Moreover, it requires a low number of

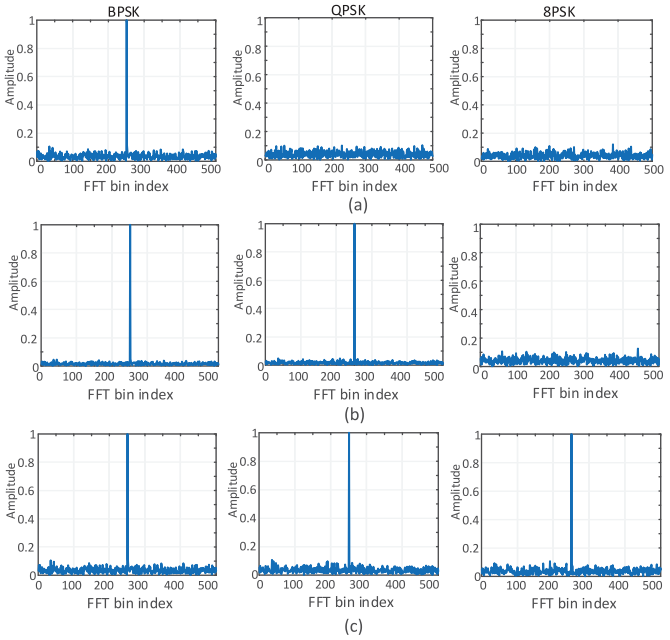

Fig. 19. Different modulation formats and their FFT after nonlinear power, (a) FFT after $(I+jQ)^2$, (b) FFT after $(I+jQ)^4$, and (c) FFT after $(I+jQ)^8$.

samples and does not require carrier recovery. However, it is not appropriate for M-PSK classification.

Instead of time-domain features, the authors in [137], [138] proposed frequency domain features for MFI. The authors in [137] used nonlinear power transformation for MFI. In this method, the fast Fourier transform (FFT) was used to extract information about the modulation type of the signal. The different powers of the input signal (i.e., $(I + jQ)^2$, $(I + jQ)^4$, and $(I + jQ)^8$) help determining its modulation type at the FFT output, as illustrated in Fig. 19. Peak detection is used to distinguish between the different modulation formats. This technique is robust to ASE noise. However, it is hard to identify modulation formats beyond 16-QAM. MFI-based amplitude variance and the 4^{th} order FFT (i.e., FFT after applying power 4 for the received signal) were proposed in [138]. In this method, the incoming signals' amplitude variance is used for the separation between M-PSK (M $=$ 4 and 8) and M-QAM (M $=$ 16 and 32) signals. The 4^{th} order FFT was used for sub-categories classification for M-PSK and M-QAM signals. This technique has ASE noise and nonlinearity tolerance. However, it is sensitive to pre-defined threshold values.

B. ML-Based Techniques for OPM and MFI

The general procedure to achieve ML-based OPM and MFI is illustrated in Fig. 20. There are three steps to develop the model before using it for impairment estimation or modulation classification. In the first step, the optical signal is converted into an electrical signal and sampled to build a dataset. The second step is to extract specific features of the signal that contain some information about the impairments or modulation formats. Examples of such features include the amplitude histogram, CDF, and eye diagram statistics. These features are

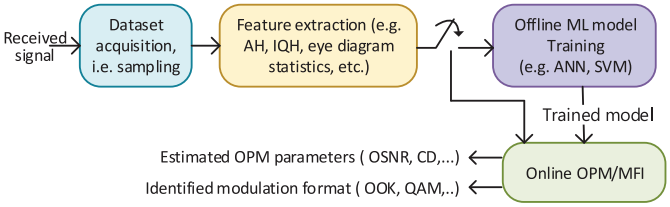

Fig. 20. The general procedure for performing OPM and MFI using ML techniques, where AH: amplitude histogram, IQH: in-phase quadrature histogram.

then used in the last step to offline train a certain ML algorithm to perform impairment estimation or modulation classification. Once the offline training is complete, the developed model along with the features extraction step are then used for either online estimation of a specific impairment such as OSNR or classification of modulation type. Note that in DNN, there is no hand-crafted features extraction step as it is performed by the algorithm itself [139], [140]

In the following, we first discuss the proposed ML-based techniques for OPM and then the proposed techniques for MFI. Finally, we discuss the joint OPM/MFI techniques.

1) ML-Based Techniques for OPM:

a) ML for OPM using direct detection: Several ML techniques have been presented in the last decade for OPM. Most of these techniques are exploiting ANN as a ML algorithm to monitor the optical network impairments. They differ in the type of features that are used to train the ANN algorithm, and the way of extracting these features.

The first attempt was in [141]. In this work, the authors proposed identifying the impairments' types using SVM in conjunction with features extracted from the eye diagram image. These features contain 23 low-order Zernike moments [142], which include a set of orthogonal polynomials. These polynomials represent the image properties with no redundancy for the purpose of discriminating whether the signal is normal or contains impairments like CD, PMD, and crosstalk. This technique has been verified experimentally for 10 Gbps OOK signal. It provides accuracy greater than 95%. However, it is only limited for determining the type of impairments.

The multi-impairments monitoring using ANN algorithm that is trained with features derived from the eye diagram, after a synchronous sampling (see Fig. 21 (a)), was proposed in [143], [144]. The extracted features to train the ANN algorithm include: Q-factor, closure (CL), root-mean-square (RMS), jitter and crossing amplitude (CA) [143]. This method can be deployed to monitor OSNR, CD and DGD. Its performance was verified using simulation for 10-Gbps NRZ-OOK and 40-Gbps RZ-DPSK signals at OSNR range of 18 to 30 dB, CD range of 100 to 700 ps/nm, and DGD range of 0 and 35 ps. The correlation coefficient was 0.91 and 0.96 for NRZ-OOK and RZ-DPSK, respectively. This technique can be used to monitor multi-impairments. However, it requires precise timing/clock recovery. Therefore, its use especially in the intermediate nodes is costly and even impossible, especially for high-speed transmission.

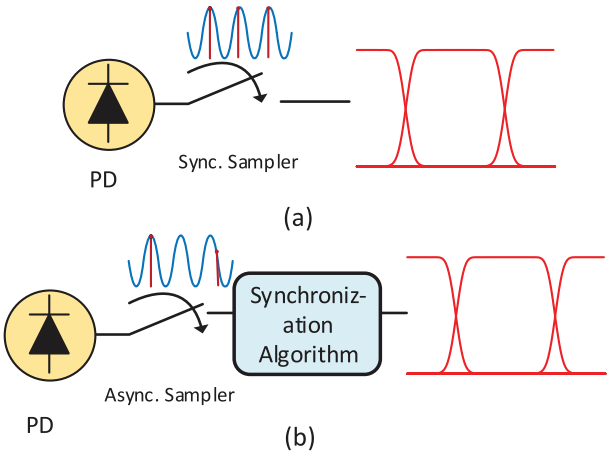

Fig. 21. Generation of eye diagrams using (a) synchronous sampling and (b) asynchronous sampling with a synchronization algorithm.

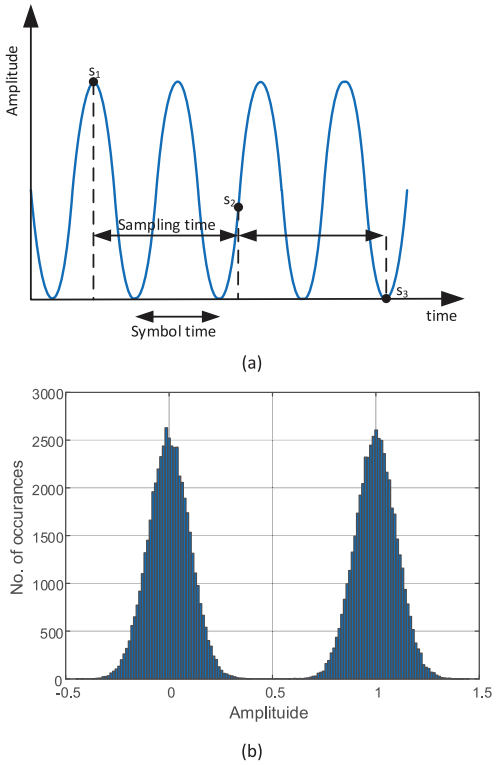

Fig. 22. Amplitude sampled signal for AAH generation, and (b) AAH for OOK NRZ signal.

To overcome the need for a precise timing/clock recovery, the authors in [145] proposed extracting the required features to train the ANN algorithm using asynchronous sampling, and then reconstruct the eye diagram. In this method, a 40-Gbps RZ-OOK signal is sampled asynchronously (at a rate lower than the signal symbol rate). Then, the eye diagram is reconstructed using chirp-z transform software-synchronization algorithm (CZT-SS) (see Fig. 21 (b)) [146]. The extracted parameters include Q-factor, eye height (EH), CA, and root-mean-square jitter (JRMS). This approach is verified by simulation to monitor OSNR, CD and DGD at OSNR range of 22.5 to 37.5 dB, CD range of 4 to 28 ps/nm, and DGD range of 1 to 7 ps. The results showed that correlation coefficient of 0.97 was achieved with root mean square error (RMSE) for OSNR, CD, and DGD of 0.69 dB, 1.05 ps/nm, and 0.38 ps, respectively. Although this method does not necessitate timing recovery, it requires additional circuity, i.e., software synchronization to reconstruct the eye diagram.

A multi-layer ANN (ML-ANN) algorithm, using the asynchronous sampling technique, has been applied to monitor a 56-Gbps 4-PAM signal in terms of OSNR, CD and DGD [147]. The training parameters were extracted from reconstructed eye diagrams. These parameters include eye level values, eye heights, crossing amplitudes, and jitter values. This method was verified by simulation at OSNR range of 26 to 42 dB, CD range of 0 to 400 ps/nm, and DGD range of 0 to 8 ps. The results showed that the RMSE for OSNR, CD, and DGD were 0.21 dB, 6.79 ps/nm, and 0.8 ps, respectively. This work deals with multi-level modulation formats, and does not require timing recovery. However, it needs additional circuity, i.e., software synchronization.

The asynchronous amplitude histograms (AAHs) method was exploited in [148], as a feature to train ANN, and then to monitor the impairments. In this method, the detected signal is sampled at a rate lower than the signal symbol rate, as illustrated in Fig. 22 (a). The samples are divided into uniformly spaced levels to form a histogram for OOK NRZ signal, as illustrated in Fig. 22 (b). The performance has been verified by simulating 40-Gbps RZ-DQPSK and NRZ-16-QAM signals at OSNR range of 10 to 30 dB, CD range 0 to 400 ps/nm, and DGD range of 0 to 10 ps. The results showed that the RMSE of RZ-DQPSK(16-QAM) signals were 0.43(0.2) dB, 9.82(9.66) ps/nm, 0.92(0.65) ps for OSNR, CD and DGD, respectively. The AAH provides information about the statistical properties of the sampled signal. It does not require timing recovery and additional hardware/circuits. However, when the received signal is heavily impaired by CD and DGD, the distinction between different impairments becomes difficult.

Features defined in terms of empirical moments have been used in [149] to train ANN algorithm. In this work, the detected signal was sampled asynchronously followed by moments computation. The trained ANN was used to monitor OSNR, CD and DGD impairments. The performance was verified by simulation for 40/56 Gbps RZ-DQPSK and 40-Gbps RZ-DPSK systems at OSNR range of 10 to 26 dB, CD range of -500 to 500 ps/nm, and DGD range of 0 to 14 ps. The results showed that RMSEs of 40 (56) Gbps RZ-DQPSK were 0.1(0.1) dB, 27.3(29) ps/nm, 0.94(1.3) ps for OSNR, CD and DGD, respectively, while RMSEs of 40 Gbps RZ-DPSK were 0.1 dB for OSNR, 17 ps/nm for CD, and 1 ps for DGD. This technique can be utilized for monitoring both magnitude and sign of accumulated CD. However, it is not suitable for high dispersive channels. This is because the extracted moments from the signal's amplitude samples lack slope information that is affected by some impairments like CD.

To extract more details of the statistical properties of the monitored signal, the 2D histogram has been proposed as a feature to train ANN algorithm instead of the 1D

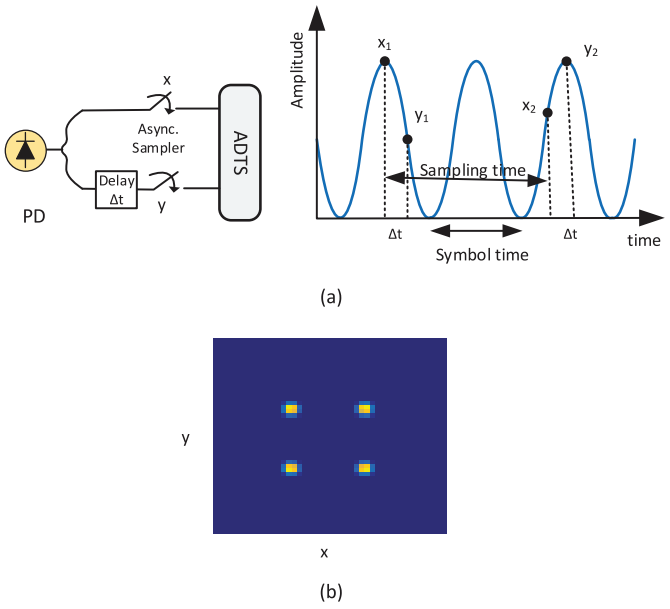

Fig. 23. (a) ADTS block diagram (left) and the amplitude sampled signal (right), and (b) generated 2D ADTS plot.

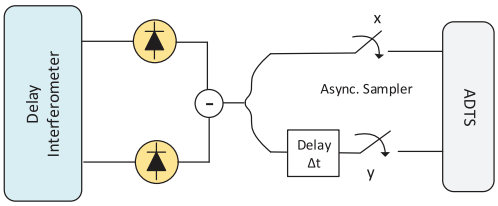

Fig. 24. Generation of ADTS using balanced detection.

histogram [150]. The 2D histogram is achieved using asynchronous delay-tap sampling (ADTS). In ADTS, the amplitude of the signal is sampled using two clocks. A constant electrical delay time called delay tap between these two clocks is used, as illustrated in Fig. 23 (a) (left). The two samplers generate two sequences x and y that are separated by this time delay, as shown in Fig. 23(a) (right). The output is a 2D ADTS plot, as shown in Fig. 23 (b). This technique was tested by subjecting a 10-Gbps NRZ-OOK signal to OSNR, CD and DGD, and then the received signal is monitored in terms of these impairments. The performance was evaluated by simulation at OSNR range of 18 to 30 dB, CD range of 100 to 500 ps/nm, and DGD of range 5 to 35 ps. The results showed correlation coefficient of value 0.97. Although this technique extends the monitoring range due to its capability of capturing information about the slope of the detected signal, the cost of implementing ADTS is high due to the use of pair of sampling clocks. Figure 23 (b). shows the ADTS of an NRZ-OOK signal.

In [151], the authors proposed to use an optical interferometer followed by two balanced-detectors before generating ADTS, as shown in Fig. 24. This allows capturing the transition of phase changes which provides better results especially for higher-order modulation formats. The validity of this approach was confirmed by monitoring OSNR, CD and DGD impairments in a high speed 100-Gbps QPSK signal. The effectiveness of this method was verified experimentally at OSNR range of 16 to 28 dB, CD range of 0 to 50 ps/nm, and DGD range of 0 to 10 ps. The results showed RMSE values of 1.27 dB, 2.22 ps/nm, and 0.91 ps for the three impairments under study. This technique achieved accurate results at high speed transmission. However, its drawback is in the requirement for additional hardware such as interferometer and two balanced detectors.

Similar to the ADTS approach where the 2D histogram is computed to act as a signal feature, the authors in [153] proposed a different technique, called parametric asynchronous eye diagram (PAED). In their work, the detected signal is passed through two branches, the first of which has a sampler while the second has a differentiator followed by a sampler. The differentiator is exploited to capture information about pulse broadening, which is affected by CD and DGD. The output of two samplers is represented in a 2D plot. This 2D plot looks like an eye diagram. The eye is divided into six quadrants. For each division, different parameters can be extracted such as the signal mean and standard deviation. The performance was verified by simulation for monitoring multiple impairments of different modulation formats including RZ/NRZ-OOK and QPSK, and monitoring different bit rates including 10 Gbps, 20 Gbps, and 40 Gbps. Although shown good performance, this monitoring approach requires differentiator and two samplers which increase the overall system's cost.

Because RF spectrum is insensitive to CD effect, the authors in [154] proposed exploiting the low-frequency (LF) components of the RF spectrum as a feature in monitoring OSNR parameter. Simulation has been conducted for monitoring 112 Gbps DP-RZ-QPSK at OSNR range from 10 to 24 dB under large amounts of CD up to 27,000 ps/nm impaired signal. The obtained results showed RMSE value of 0.84 dB. This technique provides accurate results at high CD values. However, it is not appropriate for CD and DGD monitoring because the RF spectrum is insensitive to these impairments.

The previously mentioned techniques in this subsection make use of ANN/SVM for impairment monitoring. In [152], the authors proposed the utilization of CNN for joint OSNR and CD monitoring for 10 Gbps, NRZ-OOK signal. The input to the CNN was images representing the 2-D histogram, which were extracted using the ADTS approach. Simulation results for simultaneous monitoring of OSNR and CD in the range of 10 to 40 dB and 0 to 2000 ps/nm, respectively, showed correlation coefficient of value 0.995. This method achieved good performance for joint monitoring of OSNR and CD at a wide range of impairments values. However, it handles images which increases the system's complexity.

Table IV summarizes the ML-based OPM techniques for direct detection. Specifically, we list for each reference the type of extracted features, modulation formats under consideration, data rate, the utilized ML technique, type of impairments and their range of values, and monitoring accuracy.

b) ML for OPM using coherent detection: Coherent detection is used to demodulate complex signals that include amplitude and phase information represented by the components I and Q. In [155], the received I and Q components (we call them here x and y) are sampled asynchronously to

TABLE IV
SUMMARY OF ML-BASED OPM TECHNIQUES FOR DIRECT DETECTION

Features domain	ML tech.	Impairment type*	impairment range	Data rate (Gbps)	Modulation formats	Sim. / Exp.	Monitoring accuracy	Ref.
Synchronous eye diagram features (Q, CL, RMS, J, CA)	ANN	OSNR CD DGD	18 – 30 100 – 700 0 – 35	10 40	NRZ-OOK RZ-DPSK	Sim.	$\rho = 0.91$** $\rho = 0.96$	[143]
Asynchronous eye diagram features (Q, EH, CA, JRMS)	ANN	OSNR CD DGD	22.5 – 37.5 4 – 28 1 – 7	40	RZ-OOK	Sim,	$\rho = 0.97$	[145]
Asynchronous eye diagram features (Amp. levels, EHs, CAs, Js)	DNN (ML-ANN)	OSNR CD DGD	26 – 42 0 – 400 0 – 8	56	4-PAM	Sim.	RMSE = 0.21 dB RMSE = 6.79 ps/nm RMSE = 0.8 ps	[147]
AAH	ANN	OSNR CD DGD	10 – 30 0 – 400 0 – 10	40 40	RZ-DQPSK 16-QAM	Sim.	RMSE <0.43 dB RMSE <9.8 ps/nm RMSE <0.92 ps	[148]
Empirical moments	ANN	OSNR CD DGD	10 – 26 -500 – 500 0 – 14	40 56 40	RZ-DQPSK RZ-DQPSK RZ-DPSK	Sim.	RMSE <0.1 dB RMSE <29 ps/nm RMSE <1.3 ps	[149]
ADTS	ANN	OSNR CD DGD	18 – 30 100 – 400 5 – 35	10	NRZ-OOK	Sim.	$\rho = 0.97$	[150]
ADTS after two balanced-detectors	ANN	OSNR CD DGD	16 – 28 0 – 50 0 – 10	100	QPSK	Exp.	RMSE = 1.27 dB RMSE = 2.22 ps/nm RMSE = 0.91 ps	[151]
ADTS	CNN	OSNR CD	18 – 30 0 – 2000	10	NRZ-OOK	Sim.	$\rho = 0.995$	[152]
PAED features	ANN	OSNR CD DGD	10 – 40 0 – 800 0 – 20	10 20 10 20	RZ-OOK NRZ-OOK RZ-OOK NRZ-OOK	Exp.	RMSE <0.75 dB RMSE <47.4 ps/nm RMSE <2.3 ps	[153]
LF components of RF spectrum	ANN	OSNR	10 – 24	112	DP-QPSK	Sim.	RMSE = 0.84 dB.	[154]

*OSNR in dB, CD in ps/nm, DGD in ps.
Sim: simulation, Exp.: experimental, RMSE: root mean square error, ρ: correlation coefficient, CL:closure, RMSE: root-mean-square error, J: jitter, CA: crossing amplitude, EH: eye height, CA: crossing point amplitude, JRMS: jitter root-mean-square, Amps: amplitude values.
** The correlation coefficient for all impairments.

construct a 2D constellation diagram. The constellation diagram is divided into four quadrants (i.e., (x_i, y_i) quadrant1 (Q_1) if $x_i < 0$ and $y_i < 0$). Two quadrants, Q_1 and Q_3, are considered. The extracted features include μ and σ of Q_1 and Q_3 amplitudes, maximum and minimum values of the y's at the $x = 0$ axis and the Q-factor that is given by $(\mu_3-\mu_1)/(\sigma_1+\sigma_3)$. These features were used to develop a regression model using ANN algorithm. The validity of this approach was examined by monitoring OSNR, CD, and DGD for an impaired 40-Gbps RZ-QPSK signal. This technique was verified by simulation at OSNR range of 14 to 30 dB, CD range of 20 to 180 ps/nm, DGD range of 2 to 18 ps. The results showed RMSE value of 0.77 dB for OSNR, 18.7 ps/nm for CD, and 1.17 for DGD. Note that the utilization of asynchronous sampling eliminates the need for timing recovery. However, the performance was evaluated in the existence of small amounts of DGD, CD, and high OSNR values.

The insensitivity of AAH technique to PN has been mitigated in [156] by asynchronously sampling the received signal and building a 2D I and Q histogram (IQH), as shown in Fig. 25. In this work, the IQH histogram was used to develop an SVM-based regressor to monitor OSNR, CD, DGD and PN (i.e., laser linewidth) impairments. Moreover, the separability of impairments was investigated using the t-SNE algorithm for

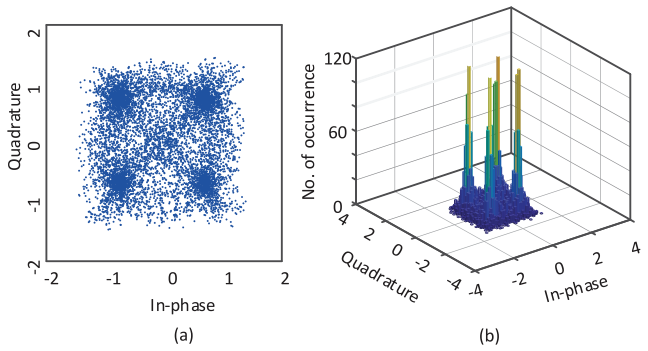

Fig. 25. Asynchronous QPSK constellation diagram, and (b) corresponding IQH plot.

both single and simultaneous multiple impairments. The t-SNE is a nonlinear DR technique that facilitates visualization of complex high-dimensional signals; thereby helps investigating the conditions under which the optical channel impairments can be monitored. This monitoring method was verified by simulation, where a 12.5-Gbps DP-QPSK signal was monitored in terms of OSNR of (9 to 19 dB), CD (200 to 1600 ps/nm), DGD (10 to 70 ps), and PN (10 KHz to 1 MHz). The

TABLE V
SUMMARY OF ML-BASED OPM TECHNIQUES FOR COHERENT DETECTION

Features domain	ML tech.	Impairment type*	Impairment range	Data range (Gbps)	Modulation formats	Sim./ Exp.	Monitoring accuracy	Ref.
Asynchronous constellation diagram features (Q, max., min.)	ANN	OSNR CD DGD	14 – 30 20 – 180 2 – 18	40	RZ-QPSK	Sim.	RMSE = 0.77 dB RMSE = 18.7 ps/nm RMSE = 1.17 ps	[155]
IQH	SVM	OSNR CD DGD PN	9 – 19 200 – 1600 10 – 70 0.01 – 1	12.5	DP-QPSK	Sim.	$\hat{\rho} = 0.97$**	[156]
Asynchronous raw data (IH, QH, IV, QV)	DNN ML-ANN	OSNR	7.5 – 27	14/16 14/16	DP-QPSK DP-16-QAM	Exp.	Error <0.5 dB	[139]
Asynchronous raw data (IH, QH, IV, QV)	LSTM-RNN	OSNR CD	15 – 30 1360 – 2040	28/35 28/35	DP-16-QAM DP-64-QAM	Sim.	MAE <0.1 dB MAE <0.64 ps/nm	[140]
		OSNR CD	15 – 30 1360 - 2040	5/10 5/10	DP-16-QAM DP-64-QAM	Exp.	MAE <0.12 dB MAE <1.09 ps/nm	
FFT domain	LSTM-RNN	OSNR NL-NP***	15 – 30 -40 –15	28 28 28	QPSK 16-QAM 64-QAM	Sim.	Error <1 dB Error <1 dB	[157]
AH	DNN (ML-ANN)	OSNR	5 – 30	56/28	QPSK	Exp.	RMSE <0.1 dB	[158]

Sim: simulation, Exp.: experimental, RMSE: root mean square error, MAE: mean absolute error.
* OSNR in dB, CD in ps/nm, DGD in ps, non-linear noise power (NL-NP) in dBm, phase noise (PN) in MHz.
** Coefficient of determination $(\hat{\rho}) = 1 - \sum_{i=1}^{N}(t_i - \hat{t}_i)^2 / \sum_{i=1}^{N}(t_i - \bar{t})^2$ where t_i is the actual data, $\hat{t}_i$ is the estimated data and $\bar{t}$ is the sample mean.
***All modulation formats are affected by all these impairments (OSNR and NL-NP).

results showed high accuracy compared with AAH. However, both I and Q samples are needed to build IQH.

To deal with the raw data directly without feature pre-engineering, like building histograms, the authors in [139] proposed a technique using DNN (i.e., ML-ANN) algorithm to monitor OSNR at range from 7.5 to 27 dB. The DNN was trained using asynchronous raw data for both vertical and horizontal polarizations (i.e., IH, QH, IV and QV). The proposed technique was verified experimentally using 14/16 Gbaud DP-QPSK and 14/16 Gbaud 64-QAM signals with different symbol rates and multi-modulation formats. The results showed mean absolute error (MAE) less than 0.5 dB. This technique performs OSNR monitoring without manual feature engineering. However, it requires the utilization of complex data models. Moreover, it requires expensive graphics processing units (GPUs). Similarly in [140], the authors utilized asynchronous raw data to train LSTM-RNN for simultaneous OSNR and CD monitoring at 28/35 Gbaud DP-16-QAM and 28/35 Gbaud 64-QAM signals. The performance of this method was evaluated by simulation for 28/35 Gbaud DP-16-QAM/64-QAM and experimentally for 5/10 Gbaud DP-16-QAM/64-QAM at OSNR range from 15 to 30 dB, and CD range from 1360 to 2040 ps/nm. The results showed RMSE value less than 0.12 dB for OSNR and 1.09 ps/nm for CD. In this technique, the simultaneous monitoring of OSNR and CD at high CD range was performed without a need for preprocessing to extract training features. However, the requirement for a memory in the LSTM model makes it of high cost [51].

The traditional DNN requires retraining when the parameters to be monitored get changed; thus time consuming. The authors in [158] proposed OSNR monitoring for 56/28 Gbps QPSK signals using transfer learning assisted DNN algorithm, utilizing amplitude histogram (AH). Transfer learning relies on adjusting neuron weights based on previous knowledge rather than random initialization, which in turn greatly accelerates the training procedure. This technique was verified experimentally at OSNR range from 5 to 30 dB, residual CD range from 0 to 600 ps/nm, and optical launched power range from -6 to 8 dBm. The results showed RMSE value of less than 0.1 dB. This technique provides high accuracy results in wide OSNR range. However, it is limited just for OSNR monitoring only.

LSTM-RNN algorithm was also used in [157] to monitor simultaneous OSNR and nonlinear noise power (NL-NP) caused by fiber nonlinearity. The fiber nonlinearity is due to transmitting high signal power through the fiber. The LSTM-NN algorithm was trained using FFT with length 1024. In this work, simulations were used for verification puposes by building a setup consisting of five channels with 50 GHz channel spacing. Different modulation formats were considered (QPSK, 16-QAM, and 64-QAM) at 28 Gbaud symbol rate at OSNR range from 15 to 30 dB and at optical power range from -3 dBm to 3 dBm. The transmission length was varied from 100 to 1000 km. The obtained results showed that OSNR monitoring is tolerant to fiber nonlinearity. However, since the LSTM-based classification/monitoring deals with sequential data in the time domain, the LSTM-based features extracted in the frequency domain do not take full benefit of LSTM.

Table V summarizes the literature pertaining to the ML-based OPM techniques for coherent detection. Specifically, we list for each reference the type of extracted features, modulation formats under consideration, data rate, the utilized ML technique, type of impairments and their range of values, and monitoring accuracy.

2) ML-Based Techniques for MFI: In this section, we review the proposed techniques for MFI. These techniques are classified based on the type of signal detection, whether it is direct or coherent. Although direct detection has an advantage in that it reduces the cost of proposed solution, we

TABLE VI
SUMMARY OF ML-BASED MFI TECHNIQUES FOR DIRECT DETECTION

Features domain	ML tech.	Modulation formats	Data rate (Gbps)	Impairment type*	Impairment range	Sim. / Exp.	Classification accuracy	Ref.
AAH	ANN	RZ-OOK	10	OSNR CD DGD	12 - 26 -500 - 500 0 - 10	Sim.	100%	[159]
		NRZ-DPSK	40				99.81%	
		ODB	40				99.9%	
		RZ-DQPSK	40				97.98%	
		DP-RZ-QPSK	100				97.34%	
		DP-NRZ-16-QAM	200				99.35%	
AAH	ANN + GA	RZ-OOK	10	OSNR CD DGD	12 - 26 -500 - 500 0 - 10	Sim.	99.84%	[160]
		NRZ-DPSK	40				99.20%	
		ODB	40				98.56%	
		RZ-DQPSK	40				99.68%	
		DP-RZ-QPSK	100				98.40%	
		DP-NRZ-16-QAM	200				97.76%	
AAH	DT+SVM	OOK	10	OSNR CD DGD	5 - 30 0 - 4000 0 - 10	Exp.	100%	[161]
		DPSK	40				100%	
		QPSK	100				100%	
		OQPSK	100				100%	
		64-QAM	200				100%	
		16-QAM	200				100%	

* OSNR in dB, CD in ps/nm, DGD in ps.
Sim: simulation, Exp.: experimental.

notice that most of the proposed MFI techniques are using coherent receivers, because of the difficulty to recover phase information in direct detection receivers.

a) ML for MFI using direct detection: The literature in this category includes [159], where the AAH was utilized to train ANN. The validly of this technique has been verified by simulation, for six modulation formats with different data rates, including 10 Gbps RZ-OOK, 40 Gbps NRZ-DPSK, 40 Gbps ODB, 40 Gbps RZ-DQPSK, 100 Gbps DP-RZ-QPSK, and 200 Gbps DP-NRZ-16-QAM. The MFI was performed at OSNR in range of 12 to 26 dB, CD in range of -500 to 500 ps/nm, and DGD in the range of 0 to 10 ps. The AAH was utilized after detecting the received signal using a photodetector (direct detection). The results showed MFI accuracy greater than 97% for all modulation formats. In order to reduce the ambiguity in MFI, the authors proposed splitting the optical received signal into two polarizations. Each polarization was directly detected and then sampled asynchronously. The samples for both polarizations were considered and the AAH of the resulting samples were generated. This modification increased the MFI accuracy by more than 99%. Similarly, AAH has been used in [160], [162] for MFI with ANN classifier optimized by genetic algorithm (GA). The results showed same MFI accuracy as obtained in [159], with few number of neurons and hidden layers. The MFI-based AAH provides high accuracy. However, it cannot be utilized to identify phase modulation formats such as the M-PSK.

For low OSNR and different modulation formats, the authors in [161] proposed MFI technique using DT twin support vector where feature extraction is exploited using higher-order cumulants (HOC) (i.e., set of quantities describing a probability distribution). Fourth-order and eighth-order cumulants and cyclic spectrum were utilized to identify six modulation formats including 10 Gbps OOK, 40 Gbps DPSK, 100 Gbps QPSK, 100 Gbps OQPSK, 200 Gbps 16-QAM and 200 Gbps 64-QAM. This technique was verified experimentally at OSNR range of -10 to 30 dB, DGD range of 0 and 10 ps, and different fiber lengths to produce CD varying from 0 to 4000 ps/nm. The achieved MFI accuracy for all modulation formats was upto 100% when OSNR equal 5 dB. This technique provides accurate results at very low OSNR values. However, it is sensitive to pre-defined threshold values.

Table VI summarizes the literature of ML-based MFI techniques for direct detection. Specifically, we list for each reference the type of extracted features, modulation formats under consideration, data rate, the utilized ML technique, type of impairments and their range of values, and monitoring accuracy.

b) ML for MFI using coherent detection: In the next generation fiber-optic networks, there will be a need for flexible transceivers that support multiple data rates and multiple modulation formats [163]–[165]. Because of the flexible transceivers, it is no longer assured that signals arriving at the receiver side will have the same data rate and modulation format. Thus, the receiver requires some techniques to adjust these changes. MFI is of great importance for future networks as it makes the network autonomous and flexible. Therefore, coherent receiver must be able to recognize the modulation format of arriving signals to guarantee proper demodulation. Figure 26 shows an adaptive coherent receiver with DSP architecture. Fortunately, there are some DSP algorithm independent of modulation formats such as IQ skew removal, CD compensation timing phase recovery, and constant modules algorithm (CMA) equalization that are built in the data receivers and can be used to improve the signal quality before performing MFI.

There are many MFI techniques proposed in literature for coherent detection. In the following, we discuss these

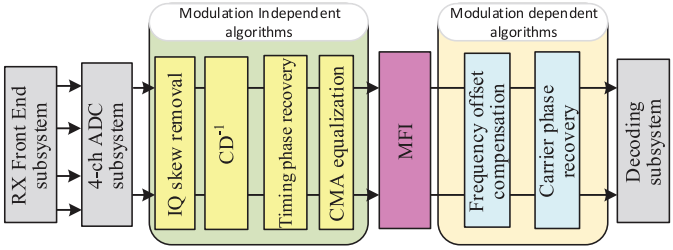

Fig. 26. Coherent receiver architecture with MFI.

techniques in details. These techniques are based on extracting different time domain features from the received signal. In addition, there are other classification techniques such as those relying on image processing. A summary of these techniques is listed in Table VII.

i) MFI-Based Time Domain Features Extraction: In this subsection, we review the ML algorithms that exploit the time-domain to extract features such as the Stokes space representation, AH, IQH, etc.

1. MFI-based Stokes Space Representation: For dual polarization transmission, the received signal can be represented by Stokes parameters. The Stokes parameters can be obtained as:

$$\mathbf{s} = \begin{bmatrix} s_0[k] \\ s_1[k] \\ s_2[k] \\ s_3[k] \end{bmatrix} = \begin{bmatrix} x^2[k] + y^2[k] \\ x^2[k] - y^2[k] \\ 2Re(xy^*) \\ 2Im(xy^*) \end{bmatrix} \tag{1}$$

where $x[k]$ and $y[k]$ are the amplitudes of the complex signal sample k in the V and H polarizations, respectively. s_0 parameter is the signal power, and $(s_1, s_2, s_3)^T$ parameters represent the 3-D Stokes space constellations. It can be observed from (1) that the phase information is removed after transformation to Stokes space. Hence, Stokes parameters are completely independent of the laser frequency offset and phase noise as well as the polarization rotation of the signal. Other information inferred from (1) is that the M-PSK signal only exists in the $(s_1, s_2, s_3)^T$ plane because $s_1 = 0$. However, the M-QAM signal is more complex because it exists in the plane $(s_1, s_2, s_3)^T$. Figure 27 (a) shows the Stokes space representation of ideal noiseless QPSK, 8-PSK, 8-QAM, and 16-QAM modulation formats in Poincare sphere. The projections of modulation formats on (s_1, s_2), (s_2, s_3), and (s_1, s_3) planes are shown in Fig. 27 (b-d), respectively. It is clear from Fig. 27 that QPSK and 8-PSK are allocated on the plane $s_1 = 0$, while 8-QAM, and 16-QAM are allocated on several planes that are parallel to the plane $s_1 = 0$.

• *MFI-based Stokes space representation with non-supervised ML:* Using Stokes space, the different modulation formats are represented as clusters in 3D space. Then, it is helpful to use clustering algorithms, i.e., unsupervised ML, to classify the different types of modulation formats in this space. This is the reason why most of the proposed Stokes space-based ML algorithms are unsupervised.

The authors in [166], [167] proposed an MFI technique based on Stokes space. The modulation formats are represented in the 3D Stoke space as clusters (point clouds). Applying VBEM with GMM algorithms makes it possible to determine the number of clusters (i.e., count clusters number) that represents the modulation format type. This technique was verified by simulation for DP-BPSK, DP-QPSK, DP-8-PSK, DP-8-QAM, DP-12QAM, and DP-16-QAM modulation formats at 10 Gbaud transmission speed, 30 dB OSNR, and back to back configuration. Then, the performance was further verified by proof-of-concept experiments using DP-QPSK and DP-16-QAM at 19 dB and 27 dB OSNR, respectively. The experiments were conducted at 10 Gbaud transmission speed, 100 kHz-linewidth laser, several hundred MHz offset frequency, and back to back transmission. The obtained results revealed that the proposed technique can be used to separate the aforementioned modulation formats without prior training. In addition, it is robust to polarization rotation and offset frequency. However, this work did not consider the effect of residual CD or low OSNR values.

To study the effect of different OSNR values and residual CD, the authors in [168], [169] proposed MFI scheme based on Stokes space analysis followed by HOC (i.e., fourth-order cumulants) and spatial cross-correlation. In this work, different types of amplitude and phase modulation formats were considered. The principle of this technique is as follow. First, Stokes space is used to separate the modulation formats represented by a 2D Stokes space (e.g., OOK, M-ary pulse amplitude modulation (M-PAM) and M-PSK modulation formats) from the 3D modulation formats such as M-QAM modulation formats. For the 2D modulation formats, HOC is utilized to distinguish between them while for 3D modulation formats, spatial cross-correlation is utilized. This technique was verified experimentally to identify four modulation formats including 32 Gbaud OOK, 32 Gbaud BPSK, 32 Gbaud QPSK, and 16-32-Gbaud 16-QAM within 11 to 34 dB OSNR range. The experimental results showed the possibility of achieving more than 60% probability of correct recognition (PoCR) in back to back configuration. In case of introducing residual CD by 1056 km transmission fiber, the PoCR was reduced to 50%. In general, the results showed poor performance in identifying wide range of modulation formats.

To improve the MFI accuracy performance, the authors in [170] exploited Stokes space and maximum-likelihood clustering algorithms to identify different types of phase modulation formats. In addition, the performance of these techniques was compared with other proposed techniques including *k*-means, EM, DBSCAN, OPTICS, and spectral clustering. Simulations were used to identify five modulation formats including DP-BPSK, DP-QPSK, DP-8-PSK, DP-8-QAM, and DP-16-QAM, each was transmitted at 28 Gbaud speed with OSNR in the range of 5 to 30 dB (0 to 15 dB for BPSK) and back to back configuration. The simulation results showed that under the same conditions, the proposed technique achieves more than 95% MFI accuracy which is better than that of other techniques. However, its computational time is higher than that of DBSCAN and OPTICS algorithms.

To decrease the computational time, MFI-based non-iterative algorithms was proposed in [171], [172]. The proposed method is relying on Stokes space and connected component algorithm (CCA). It is based on taking the projection of the 3D Stokes space on (s_2, s_3) plane. To reduce the

TABLE VII
SUMMARY OF ML-BASED MFI TECHNIQUES FOR COHERENT DETECTION

Features domain	ML tech.	Modulation formats*	Data rate (Gbaud)	OSNR range (dB)	Sim./Exp.	Min. OSNR (dB)**	Ref.
Stokes space	VBEM- GMM	BPSK QPSK 8-PSK 8-QAM 12QAM 16-QAM	10	30	Sim.	30	[166]
		QPSK 16-QAM	10 10	19 27	Exp.	19 27	
Stokes space	Maximum likelihood	BPSK QPSK 8-PSK D8-QAM 16-QAM	28	0 - 30	Sim.	0 9.5 18 18 18	[170]
Stokes space	CCA	QPSK 8-PSK 16-QAM	32	10 - 30	Sim.	15 22 22	[171]
		QPSK 16-QAM	32	12 - 30	Exp.	–	
Stokes space	Adaptive CFSFDP	QPSK 8-QAM 16-QAM 32-QAM 64-QAM	28	8 - 28	Sim.	12 15 19 22 23	[175]
		QPSK 16-QAM 64-QAM	28	12 - 30	Exp.	12 19 26	
Stokes space	Fuzzy mean	BPSK QPSK 8-PSK 8-QAM 16-QAM	28	7 - 28	Sim.	11 14 14 15 10	[177]
		QPSK 16-QAM	28	10 - 28	Exp.	15 18	
Stokes space	CFSFDP	BPSK QPSK 8-PSK 16-PSK 8-QAM 16-QAM	28	7 - 24	Sim.	9 11.5 13 16.5 17.5 19.5	[178]
		QPSK 8-PSK 8-QAM 16-QAM	28	7 - 24	Exp.	11.5 13.5 17.5 20	
Stokes space (AH + constellation) Images	PNN	QPSK 8-PSK 16-QAM 64-QAM	28	7 - 35	Sim.	9 16 17 25	[49]
		QPSK 16-QAM 64-QAM	28	10 - 28	Exp.	12 16 19	
Stokes space	CNN	BPSK QPSK 8-PSK 16-QAM 32-QAM 64-QAM	28	9 - 35	Sim.	9 9 9 15 12 12	[179]
Intensity fluctuation	SVM	QPSK 8-QAM 16-QAM 64-QAM	28	9 - 16.5 12.5 - 21 19 - 26.5 19 - 28.5	Exp.	9 12.5 19.5 20	[180]
AH	Random forest	QPSK 8-QAM 16-QAM 32-QAM 64-QAM	16	5 - 32	Sim.	5 12 16 20 24	[181]
		16-QAM 32-QAM 64-QAM	16	– – –	Exp.	– – –	

* All modulation formats are dual polarization, ** Minimum OSNR (dB) required for accuracy>99%, Sim.: simulation, Exp.: experimental.

effect of ASE on the constellation points in (s_2, s_3) plane, Voronoi polygon filtering method [173] was utilized. The remaining constellation points were converted into a binary image. Then, the CCA [174] was used to count the number of clusters, i.e., constellation points. Simulation was conducted to identify three different modulation formats including DP-QPSK, DP-8-PSK, and DP-16-QAM, all transmitted at 32 Gbaud transmission speed, with OSNR in the range of 10 to 30 dB. The simulation results showed the possibility of achieving high MFI accuracy of more than 99%. Furthermore, this method was verified experimentally through building autonomous receiver able to identify and detect two

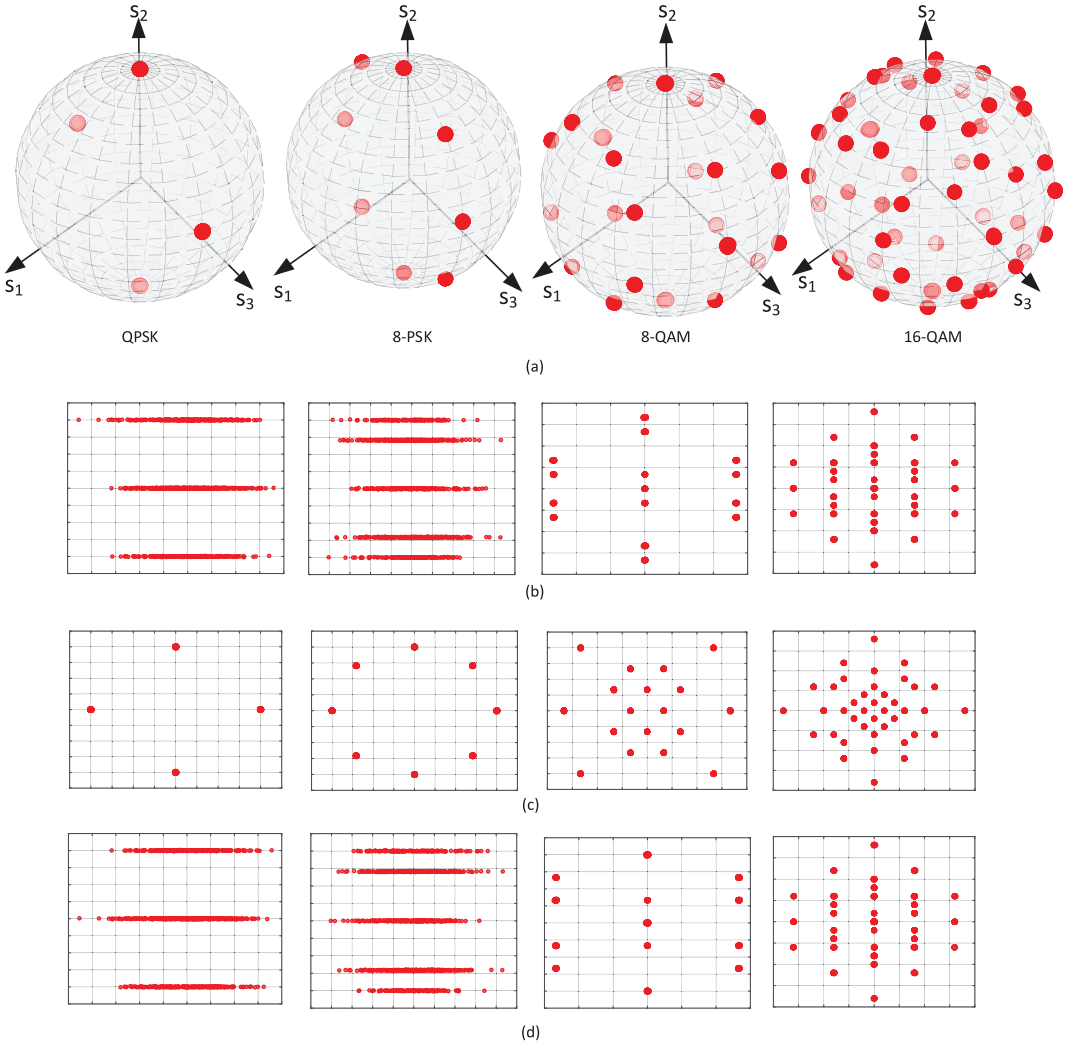

Fig. 27. (a) Stokes space representation of QPSK, 8-PSK, 8-QAM, and 16-QAM, (b) their projection on (s_1, s_2) plane, (c) their projection on (s_2, s_3) plane, and (d) their projection on (s_1, s_3) plane.

modulation formats including DP-QPSK and DP-16-QAM at 32Gbaud transmission speed. The MFI in this proposed work was achieved with low computational time, but only considered for few modulation formats.

Another MFI technique based on non-iterative algorithm was proposed in [175]. This technique relies on Stokes space and adaptive CFSFDP algorithm. In this technique, (s_2, s_3) plane is produced and exploited to discriminate between the various modulation formats according to the density of the clusters which is achieved using adaptive CFSFDP algorithm. Simulation was performed to identify five different modulation formats including DP-QPSK, DP-8-QAM, DP-16-QAM, DP-32-QAM, and DP-64-QAM, all are transmitted at 28 Gbaud transmission speed, with OSNR in the range of 8 to 28 dB. The minimum OSNR required to achieve 100% MFI accuracy was 12 dB for DP-QPSK dB, and 23 dB for DP-64-QAM. Furthermore, this technique was validated experimentally for three different modulation formats including DP-QPSK DP-16-QAM and DP-64-QAM, with 28 Gbaud transmission speed, laser linewidth 100 kHz, OSNR with range of 12 to 30 dB, and back to back transmission. The minimum OSNR required to achieve 100% MFI accuracy for DP-QPSK DP-16-QAM and DP-64-QAM was 12 dB, 19 dB, and 26 dB, respectively. Furthermore, the effect of fiber nonlinearity and CD were investigated where 100% MFI accuracy can be achieved for all modulation formats when the transmitted power ranges from 5 to 6 dBm. Although this technique is non-iterative which requires low computational time, it is difficult to determine the number of clusters for low OSNR values.

MFI based on subtraction (fuzzy mean) clustering [176] was reported in [177]. In this technique, first, the 3D Stokes vector is projected into the 2D planes (s_2, s_3) and (s_1, s_3). Then, the fuzzy mean algorithm is applied to determine clusters density center. Different modulation formats can be distinguished according to their statistical parameters (i.e., forth order cumulant) of the cluster centers distributed on (s_2, s_3) and (s_1, s_3) planes. The performance of this technique was verified by simulation to identify five modulation formats including DP-BPSK, DP-QPSK, DP-8-PSK, DP-8-QAM, and DP-16-QAM, with 28 Gbaud transmission speed, and OSNR in range of 7 dB to 28 dB. The simulation results showed the possibility of achieving 99% MFI accuracy. Moreover,

this technique was validated experimentally for DP-QPSK and DP-16-QAM, with 28 Gbaud transmission speed, OSNR in range of 10 dB to 28 dB and back to back transmission configuration. The results showed the possibility of achieving 99% MFI accuracy. Furthermore, the MFI accuracy of this technique was studied under the effect of CD and transmission power. In comparison to the algorithms reported in [170], this technique provides accuracy quit similar to maximum-likelihood and better than DBSCAN. However, its computational time is relatively similar to DBSCAN and better than maximum-likelihood.

The utilization of CFSFDP algorithm with Stokes space was reported in [178] for MFI of different M-PSK (M = 2, 4, 8, and 16) and M-QAM (M = 8 and 16) modulation formats. The principle of this technique is as follows. First, the M-PSK modulations set is separated from M-QAM modulations set according to the value of s_1 (i.e., $s_1 = 0$ in case of M-PSK signals). Then, the CFSFDP algorithm is applied to (s_2, s_3) and (s_1, s_3) planes to identify the M-PSK and M-QAM signals, respectively. This technique was validated by simulations for DP-BPSK, DP-QPSK, DP-8-PSK, DP-16-PSK, DP-8-QAM, and DP-16-QAM modulation formats, with 32 Gbaud transmission speed, 100 kHz laser linewidth, and OSNR in the range of 7 to 24 dB. The simulation results showed the possibility of achieving more than 95% MFI accuracy. The effect of other impairments such as CD, PMD, and PDL were also considered in this work. Furthermore, this technique was verified experimentally for the same modulation formats, where more than 95% MFI accuracy was achieved. The proposed technique has the potential to identify DP-M-PSK and M-QAM modulation formats.

• *MFI-based Stokes space representation with supervised ML:* Instead of exploiting unsupervised ML in conjunction with Stokes space, the authors in [49] and [179] proposed using supervised ML algorithms. In [49], Stokes space in conjunction with two consecutive PNN algorithms was proposed. In this technique, the AH of projected parameters on (s_1, s_3) plane are generated first. The AH is utilized to train PNN1 in order to separate the modulation formats into different sets according to energy level in (s_1, s_3) plane. Since the AH is limited for distinguishing the multi-level signals (i.e., M-QAM), the rest of modulation formats (i.e., M-PSK) are identified using the constellation images of data on (s_2, s_3) plane that are used to train PNN2. This technique was verified by simulation and experiments to identify DP-QPSK, DP-8-PSK, DP-16-QAM, and DP-64-QAM modulation formats. These modulation formats were transmitted at 28 Gbaud transmission speed with OSNR in the range of 7 to 35 dB, and back to back configuration. The DP-M-PSK, DP-16-QAM and DP-64-QAM signals were separated using PNN1 while the M-PSK signals were identified by PNN2. The achieved MFI accuracy was more than 95%. Furthermore, the effect of fiber nonlinearity and CD effect were investigated. Compared with the traditional supervised ML (i.e., ANN), MFI-based PNN is faster and more accurate. However, it needs extra memory space to store the model.

DNN as a supervised ML algorithm was proposed in [179] in conjunction with Stokes space for MFI. The constellation in 3D Stoke space is projected onto three planes; (s_1, s_2), (s_2, s_3,) and (s_1, s_3) to obtain three images for each examined modulation format. These images are used as inputs to three-channels pre-trained CNN algorithm called MobileNet. This proposed method was verified by simulation to identify six modulation formats including DP-BPSK, DP-QPSK, DP-8-PSK, DP-16-QAM DP-32-QAM, and DP-64-QAM, all transmitted at 28 Gbaud speed with OSNR in the range of 9 to 35 dB. In addition, a frequency offset of 1 GHz was introduced. An MFI accuracy more than 95% was achieved with OSNR greater than 8 dB. In addition, the effect of residual CD was investigated. This technique provides accurate results. However, it deals with images which increase the computational cost.

2. Other MFI-based time domain features extraction techniques: DNN was exploited in [95] for MFI. The features were extracted using AH and auto-encoder with Softmax classifier. First, the AH of I and Q signal's components are obtained. Then, an auto-encoder is used to extract a few representative features of the signal. These features are used as input for the Softmax to classify various modulation formats. This technique was verified experimentally for DP-QPSK, DP-16-QAM, and DP-64-QAM modulation formats transmitted at 28, 14, and 20 Gbaud symbol rate, respectively. Signal noise and PN impairments were considered. In addition, 1 GHz frequency offset was introduced. This technique has the advantage of providing accurate results and robustness to PN and offset frequency. However, it is restricted only for three modulation formats and cannot be used to discriminate the M-PSK modulation formats.

To perform MFI in the co-existence of different impairments, MFI-based fractal dimension (FD) [182], [183] and variance of the incoming signals' amplitude was proposed in [184]. The FD technique tries to find the slope of the received signal. The scatter plot of FD-variance brings together each modulation on one cluster with few overlaps. SVM was used to separate the different clusters. The validity of this technique was tested by simulation for six modulation formats include BPSK, QPSK, 8-PSK, 16-QAM, 32-QAM, and 64-QAM, transmitted at 20 Gbaud speed. In addition, 1 GHz frequency offset, residual CD and DGD impairments were considered. The results showed that MFI accuracy greater than 98.05% can be achieved. The proposed technique considered both M-PSK and M-QAM in the presence of different channel conditions such as ASE noise, residual CD, DGD, and frequency offset. However, FD requires complex calculations.

MFI-based SVM with intensity fluctuation features was proposed in [180]. These features include Godard's criterion error [185], [186] and intensity variance. The Godard's criterion error is given by

$$\left(\sum_{n=1}^{N} |\Big(D(n)|^2 - \Big(E|D(n)|^4 / E|D(n)|^2\Big)\Big)\right) \tag{2}$$

where $|D(n)|^2$ is intensity of the received signal $D(n)$, N is the number of samples, and $E|.|$ is the mean. Different modulation format can be separated by plotting these features in 2D plane, before applying SVM. This technique was verified experimentally for different modulation formats including

28 Gbaud DP-QPSK, DP-8-QAM, and DP-16-QAM, and 20 Gbaud DP-32-QAM. The results showed that MFI accuracy greater than 99.8% can be achieved. In addition, the nonlinearity effect was investigated by launching variable signal's power. This technique can be used for both M-PSK and M-QAM signals. However, it requires a large number of symbols to get accurate results.

CFSFDP algorithm in conjunction with the signal amplitude was proposed in [187] for MFI. The amplitude for both and I and Q signal's components is obtained first. Then, every amplitude level in M-QAM signal is separated using CFSFDP algorithm. This technique was verified by simulation to identify six modulation formats including DP-QPSK, DP-8-QAM, DP-16-QAM, DP-32-QAM, DP-64-QAM, and DP-256QAM, all transmitted at 12.5 Gbaud speed. The achieved MFI accuracy is 95% under ASE impairment effect. To proof the concept, this technique was validated experimentally for DP-QPSK, DP-16-QAM, and DP-64-QAM, with 12.5 Gbaud, transmission speed, where the minimum achieved MFI accuracy is 95%. The proposed method is noise tolerant. However, it is limited for M-QAM signals.

Random forest and AH were exploited in [181] for MFI. After AH extraction, random forest was used to identify the different modulation formats. Simulation was conducted to identify five modulation formats in WDM transmission system with three channels. The modulation formats include DP-QPSK, DP-8-QAM, DP-16-QAM DP-32-QAM, and DP-64-QAM, all transmitted at 16 Gbaud speed and subjected to ASE. The minimum OSNR required to achieve 100% MFI accuracy was between 5 and 24 dB. The effect of non-linearity and frequency offset were considered too. Furthermore, this technique was verified experimentally for DP-16-QAM DP-32-QAM, and DP-64-QAM modulation formats. In addition, the accuracy and complexity of the random forest is compared with other techniques including *k*-NN, SVM and DNN. The results showed that the MFI accuracy for random forest is better than SVM and *k*-NN and the computational complexity for random forest is less than DNN. However, this work is limited only for M-QAM modulation formats.

ii) MFI using image processing techniques: Instead of depending on time/frequency domain features, another MFI technique that considers the pixel points of an image as the data was described in [188]. In this work, images of constellation diagram with CNN were used to classify different types of modulation formats. The validly of this technique was tested by simulation for six modulation formats including QPSK, 8-QAM, 16-QAM, 32-QAM, and 64-QAM, all subjected to ASE. The results showed that 100% MFI accuracy can be obtained. However, its computational complexity is relatively high.

Similarly, the authors in [189] proposed two techniques for MFI based on the singular value decomposition (SVD) and Radon transform (RT) of the constellation diagrams. The RTs for different modulation formats are shown in Fig. 28. Different classifiers including SVM, *k*-NN, and DT were used. This technique was verified by simulation to identify 4-, 16-, 64-, and 256-QAM and 2-, 4-, 8-, and 16-PSK at 10 Gbaud transmission speed, OSNR values range from 2 to 30 dB, PN range from 1KHz to 10 MHz, and state of polarization (SoP) range from 5° to 45°. To proof the concept, this technique was also validated experimentally for DP-QPSK, DP-16-QAM and DP-64-QAM at 10 Gbaud transmission speed and back to back transmission. The results showed that the proposed technique provides accuracy up to 100% MFI even at low OSNR values of 10 dB. This technique covers wide range of modulation formats. However, its computational complexity is relatively high.

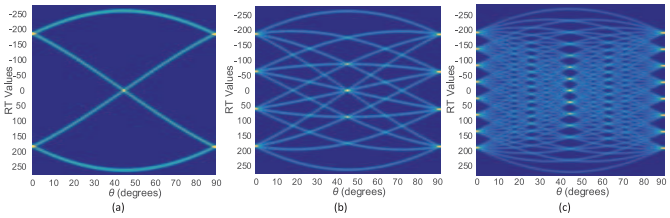

Fig. 28. RTs of different modulation formats, (a) QPSK, (b) 16-QAM, and (d) 64-QAM.

3) ML-Based Joint MFI and OPM Techniques: In this section, we review the proposed techniques for joint MFI/OPM in both direct and coherent systems. It is observed that most of the proposed techniques, especially in coherent systems, focused on joint OSNR monitoring and MFI because of the direct relationship to BER and the existence of algorithms which are able to compensate some other impairments (e.g., the CD) before MFI.

a) ML for joint MFI/OPM in direct detection systems: In [190], SVM and ANN algorithms were used for MFI and OSNR monitoring, respectively. This technique was verified experimentally for 32 Gbaud 4, 8, 16, and 64 DP-QAM signals. In this method, the power eye-diagram is produced after up-sampling the detected signal in 10 samples/symbol. Statistical features are extracted from the eye diagram including mean, variance, etc. The disadvantage of this method is using the up-sampler ADC which makes it cost in-effective.

In [191], the authors proposed simultaneous bit rate identification (BRI), MFI and OPM based on PCA algorithm in conjunction with ATDS. The principle of this method is as follows. The ATDS with $N \times N$ dimensions is converted into one dimensional vector with length N^2. Then, the PCA algorithm is used to reduce the dimension (vector length). The Euclidean distance between the vectors in lower dimension was used to distinguish the different modulation formats. This technique was verified by simulation of 10/20 Gbps RZ-OOK,40/100 Gbps DP-RZ-QPSK, 100/200 Gbps DP-NRZ-16-QAM at OSNR range from 14 to 28 dB, CD range from -500 to 500 ps/nm, and DGD range from 0 to 10 ps. This method provides information about the slope, so it extends the impairments range. However, it requires two samplers.

Instead of using two samplers, as in [191] which increases the cost, the authors in [192] proposed MFI and OSNR monitoring technique using PCA in conjunction with the asynchronous single channel sampling (ASCS). ASCS is similar to ADTS but it uses single sampler rather than two samplers. This technique was validated experimentally to identify RZ-OOK,

NRZ-DPSK, and RZ-DPSK, all transmitted at 10 Gbps transmission speed at OSNR range from 10 to 25 dB and CD range from 0 to 700 ps/nm. Compared with ADTS, ASCS uses one sampler, hence, the implementation complexity and cost of the system is reduced. However, it requires additional circuit (i.e., software synchronization).

The authors in [193] proposed joint OSNR and MFI using CNN. Synchronous eye diagram was used to train CNN. This technique was verified by simulation for DP-BPSK, DP-QPSK, DP-8-PSK, DP-8-QAM, DP-12QAM, and DP-16-QAM modulation formats at 10 Gbaud transmission speed, 30 dB OSNR, and back to back configuration. This technique does not need manual intervention, i.e., it does not require extracting features from the eye diagram. The limitations of this method are that first, it requires timing recovery leading to high cost. Second, the monitoring is performed using a classifier, not a regressor.

In [194], the authors proposed simultaneous MFI and OSNR monitoring for 28 Gbaud NRZ-OOK, 4-PAM and 8-PAM, using AH with MTL-based ANN. In MTL, the network (i.e., ANN) performs multiple tasks, such as monitoring and classification, simultaneously. This technique was verified by simulation and experiments at OSNR range from 10 to 35 dB under residual CD effect ranging from -100 to 100 ps/nm. The results showed 100% MFI accuracy and OSNR monitoring mean square error (MSE) below 0.12 dB. The use of MTL improves the performance compared with traditional training that is based on single-task learning. However, this work is limited only for intensity modulation formats.

Similarly, the authors in [41] considered simultaneous BRI, MFI and OSNR monitoring for higher-order modulation using MTL-based ANN in conjunction with AAH. This technique was verified by simulation and experiments for DP-QPSK, DP-8-QAM, and DP-16-QAM at OSNR range from 10 to 26 dB and CD value range from 0 to 1600 ps/nm. The monitoring is asynchronous as it uses AAH, hence no time recovery is required and low cost/low complexity is achieved. The limitations of this method are that first, the performance decreases as CD increases owing to the fact that the AHH offers only information about the amplitude. Second, the simulation results do not match with the experimental results.

In [195], the authors proposed MFI and OSNR monitoring using CNN in conjunction with ADTS. The ADTP images are used as input for the CNN for joint OSNR and MFI. This technique was verified by simulation for DP-16-QAM, DP-16-QAM, and DP-64-QAM modulation formats, all transmitted at 28 Gbaud symbol and subjected to ASE and CD impairments. 100% MFI and OSNR accuracies are obtained. Furthermore, it was verified experimentally for 16-QAM and 64-QAM signals with achieved accuracy more than 96%. The advantage of this technique is providing accurate results in the existence of CD. However, its cost is relatively high because of using two samplers and the need for more computational time due to processing images.

The authors in [40] proposed BRI, MFI and OPM including OSNR, CD and DGD using MTL-based CNN in conjunction with ATDS. This technique was verified by simulation for NRZ-OOK, RZ-OOK, and NRZ-DPSK for 10/20 Gbps at OSNR values range from 10 to 28 dB, CD range from 0 to 450 ps/nm, and DGD range 0 to 10 ps. MTL improves the performance due to the simultaneous training of BRI, MFI and OPM. However, this technique is computational expensive because it deals with ADTS as images.

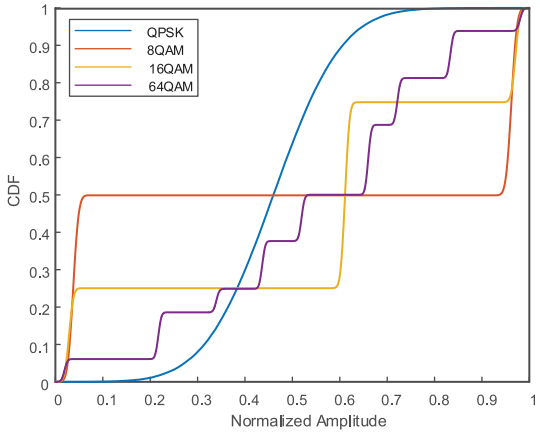

Fig. 29. CDFs of different modulation formats.

Table VIII summarizes the literature of ML-based joint OPM-MFI techniques for direct detection.

b) ML for joint MFI/OPM in coherent detection systems: Among the different types of impairments that affect optical signals in coherent systems, OSNR is the only parameter considered in literature in joint MFI and OPM. This is because OSNR is easier to estimate compared to other impairments when MFI is required.

The authors in [39] proposed a technique based on AH in conjunction with DNN (i.e., ML-ANN) to identify MFI as well as joint OSNR monitoring. This technique was verified experimentally for identifying three optical signals 56 baud QPSK, 28 Gbaud 16-QAM, and 60 Gbaud 64-QAM. The results showed the possibility of monitoring OSNR with less than 1.2 dB estimation error and 100% MFI accuracy. The proposed technique is sensitive to PN and offset frequency. However, it is limited for M-QAM modulation.

Joint BRI, MFI and OPM based on ANN algorithm that exploits AH features was reported in [196]. The proposed technique was verified experimentally for DP QPSK and 16-QAM signals with 12.5, 14, and 16 Gbaud symbol rates, at OSNR range from 9 to 19 dB, CD range from 200 to 1600 ps/nm, DGD range from 10 to 70 ps, and PN range from 1 kHz to 1 MHz. This work considered BRI and MFI. However, the influence of simultaneous co-existence of multi-impairments was not taken into consideration.

In addition to exploiting AAH features, CDF was exploited in [197] to train SVM algorithm for joint OSNR monitoring and MFI. The CDFs for different modulation formats are shown in Fig. 29. This technique was verified numerically and experimentally for three modulation formats; 4QAM, 16-QAM, and 64-QAM with 12.5 Gbaud speed and back-to-back setup configuration. The results showed high accuracy of identifying the transmitted modulation formats. In addition, they showed the possibility of estimating OSNR with less than

TABLE VIII
SUMMARY OF ML-BASED JOINT MFI-OPM TECHNIQUES FOR DIRECT DETECTION

Features domain	ML tech.	Modulation formats	BR GBaud	OPM*	OSNR monitoring range (dB)	Sim./ Exp.	Classification accuracy	Monitoring accuracy	Ref.
Synchronous eye diagram features	SVM + ANN	DP-QPSK DP-8-QAM DP-16-QAM DP-64-QAM	32 32 32 32	OSNR	7 - 30	Exp.	96% 100% 87% 92%	MSE <3.5 dB	[190]
ADTS	PCA	RZ-OOK DP-RZ-QPSK DP-NRZ-16-QAM	10/20 40/100 100/200	OSNR CD DGD BRI	14 - 28 -500 - 500 0 - 10	Sim.	100% 100% 100%	MSE = 1 dB MSE = 4 ps/nm MSE = 1.6 ps BRI = 100%	[191]
ASCS	PCA	RZ-OOK NRZ-DPSK RZ-DPSK	10 10 10	OSNR	10 - 25	Sim.	100% 100% 100%	MSE <1 dB	[192]
Eye diagram Images	CNN	RZ-OOK NRZ-OOK RZ-DPSK 4-PAM	25 25 25 25	OSNR	10 - 25	Sim.	100% 100% 100% 100%	Accuracy = 100%	[193]
AH	DNN (MTL-ANN)	NRZ-OOK 4-PAM 8-PAM	28 28 28	OSNR	10 - 25 15 - 30 20 - 35	Sim.	100% 100% 100%	MSE <0.12 dB	[194]
		NRZ-OOK 4-PAM 8-PAM	20 20 20	OSNR	14 - 29 17 - 32 22 - 38	Exp.	100% 100% 100%	MSE <0.12 dB	
AAH	DNN (MTL-ANN)	DP-QPSK DP-8-QAM DP-16-QAM	20/30 20/30 20/30	OSNR BRI	10 - 22 14 - 24 17 - 26	Sim.	100% 100% 100%	RMSE <0.97 dB BRI >95.9%	[41]
		DP-QPSK DP-8-QAM DP-16-QAM	2.8/9.8 2.8/9.8 2.8/9.8	OSNR BRI	10 - 22 14 - 24 17 - 26	Exp.	100% 100% 100%	RMSE <0.76 dB BRI >92.4%	
ADTS	CNN	DP-16-QAM DP-32-QAM DP-64-QAM	28 28 28	OSNR	15 - 30 20 - 35 25 - 40	Sim.	100% 100% 100%	Accuracy = 100%	[195]
		DP-16-QAM DP-64-QAM	28 28	OSNR	15 - 30 25 - 40	Exp.	99.68% 100%	Accuracy $\geq$ 96.56%	
ADTS	MTL-CNN	NRZ-OOK RZ-OOK NRZ-DPSK	10/28 40/100 100/200	OSNR CD DGD BRI	14 - 28 0 - 450 0 - 10	Sim.	100% 100% 100%	RMSE = 0.73 dB RMSE = 1.34 ps/nm RMSE = 0.47 ps BRI = 100%	[40]

* OSNR in dB, CD in ps/nm, DGD in ps.
Sim: simulation, Exp.: experimental, BRI: bit rate identification, RMSE: root mean square error, MSE: means square error.

1 dB estimation error. This technique requires few number of samples and is robust to PN and offset frequency. However, it is limited for M-QAM signals.

Stokes space with DNN (ML-ANN) algorithm was used in [198] for joint monitoring of OSNR and MFI. The first order derivations of (s_2, s_3) and (s_1, s_3) planes were utilized as features. Two cascaded DNN algorithms were utilized. The first was used for MFI while the second was utilized for OSNR estimation. Four signals with PMD were used to evaluate the proposed technique performance. The signals under consideration were QPSK, 8-QAM, 16-QAM at 28 Gbaud symbol rate and 32-QAM at 21.5 Gbaud symbol rate. The results showed high accuracy of identifying the different transmitted modulation formats. In addition, OSNR estimation error was found to be less than 0.5 dB. This technique is basically insensitive to carrier PN, frequency offset and polarization mixing. However, it contains multiple hidden layers with multiple neurons leading to an increase in the computational complexity.

CNN algorithm was used in [199] for MFI and OSNR estimation. Images of constellation diagram of received signals were used as a training dataset for the algorithm. The validly of this technique was verified by simulation for six modulation formats: QPSK, 8-PSK, 8-QAM, 16-QAM, 32-QAM, and 64-QAM. The results showed an accuracy larger than 95% for identifying the different modulation formats, with less than 0.7 estimating error for the OSNR. However, the authors dealt with the OSNR monitoring as a classification problem, which is only suitable for discrete OSNR values, as well the computational complexity is relatively high. Moreover, it is difficult or even impossible to recover the constellations without prior information about the modulation formats [179].

Table IX summarizes the literature of ML-based joint OPM-MFI techniques for coherent detection.

C. Lessons Learned

In this section, we discussed the conventional and ML-based techniques for OPM and MFI. Few lessons are given below.

- The conventional OPM techniques have been mainly applied to either single impairment or few impairments

TABLE IX
SUMMARY OF ML-BASED JOINT MFI-OPM TECHNIQUES FOR COHERENT DETECTION

Features domain	ML tech.	Modulation formats	BR GBaud	OPM	OSNR monitoring range (dB)	Sim./ Exp.	Classification accuracy	Min. OSNR in (dB)*	Monitoring accuracy	Ref.
AH	DNN (ML-ANN)	DP-QPSK	28	OSNR	10 - 23	Exp.	100%	–	MAE = 1.2 dB	[39]
		DP-16-QAM	14		17 - 26		100%	–	MAE = 0.4 dB	
		DP-64-QAM	20		25 - 37		100%	–	MAE = 1 dB	
CDF	SVM	DP-QPSK	28	OSNR	5 - 30	Sim.	–	5	MAE = 0.24 dB	[197]
		DP-16-QAM	28		5 - 30		–	17	MAE = 0.8 dB	
		DP-64-QAM	28		5 - 30		–	5	MAE = 0.8 dB	
		DP-QPSK	12.5	OSNR	10 - 25	Exp.	100%	–	MAE = 0.38 dB	
		DP-16-QAM	12.5		14 - 25		97%	–	MAE = 0.68 dB	
		DP-64-QAM	12.5		16 - 25		100%	–	MAE = 0.62 dB	
Stokes space	DNN (ML-ANN)	DP-QPSK	28	OSNR	9.3 - 16.8	Exp.	–	10	MSE = 0.21 dB	[198]
		DP-8-QAM	28		12.5 - 22.5		–	13	MSE = 0.48 dB	
		DP-16-QAM	28		17 - 27		–	18	MSE = 0.35 dB	
		DP-32-QAM	21.5		18 - 30		–	22	MSE = 0.44 dB	
Constellation images	CNN	QPSK	25	OSNR	15 - 30	Sim	100%	–	Accuracy >95%	[199]
		8-PSK	25		15 - 30		100%	–		
		8-QAM	25		15 - 30		100%	–		
		16-QAM	25		15 - 30		100%	–		
		32-QAM	25		15 - 30		100%	–		
		64-QAM	25		20 - 35		100%	–		
		QPSK	25	OSNR	15 - 30	Exp.	100%	–	Error <0.7 dB	
		16-QAM	25		15 - 30		100%	–		

* Minimum OSNR (dB) required for accuracy>99%.
BR: Baud rate, Sim.: simulation, Exp.: experimental, MAE: mean absolute error, MSE: mean square error.

with a small range of monitoring values. Besides, the conventional FB-based MFI algorithms are either sensitive to pre-defined threshold values or only limited to classifying few modulation formats.

- The time domain-based AH features extracted for MFI can cope with the existence of PN. However, its accuracy is affected by ASE noise. On the other hand, features extracted from the frequency domain are robust to ASE noise. Therefore, it would be useful to have features that are extracted from both time and frequency domains. Such features are already exploited in RF communication systems and can be adopted for optical communication, as well.
- The majority of ML-based OPM techniques adopt supervised learning, especially ANN. Additionally, OPM techniques are envisioned to be deployed in a large number of intermediate network nodes to provide comprehensive surveillance. Hence, inexpensive acquisition systems should be employed. Therefore, receivers with asynchronous sampling and direct detection are preferred over expensive coherent systems. Moreover, the vast majority of OPM techniques mainly consider monitoring OSNR parameter, and CD and DGD impairments in transmission systems employing OOK, DPSK, and DQPSK direct detection modulation formats. However, it is more challenging in direct detection acquisition to monitor impairments when using higher order modulation formats such as M-QAM (M = 16, 32, and 64) signals. Therefore, more studies are needed in this field.
- Most of the ML-based MFI techniques are mainly proposed for coherent adaptive receivers where the received modulation format changes according to the network status. Such receivers are using coherent detection to benefit from the built-in DSP algorithms for mitigating some signal impairments such as CD. Note that some algorithms developed for impairments mitigation are modulation dependent which require identifying the modulation format first. Therefore, it is advisable that ML algorithms are built around features that are impairment independent so that the MFI is achieved with high accuracy. Stokes space features are a viable option, which allows MFI with relatively high accuracy under stressed optical signal.
- The intermediate nodes and/or optical receivers are envisioned to have the capability of performing both OPM and MFI functions. However, achieving such simultaneous tasks is not an easy task. ML-based techniques for joint OPM/MFI are available. But, these techniques are mainly limited to monitoring only a single parameter (OSNR) during the process of performing MFI for either direct or coherent detection acquisition systems. Therefore, this calls for conducting more studies to develop joint OPM/MFI techniques capable of incorporating more impairments to monitor.
- A closer look at the majority of ML-based OPM and MFI techniques reveals that these algorithms have been developed for a set of modulation formats/impairments, which may not cover the whole set of formats of practical interest. Therefore, it is necessary to develop ML-based

OPM/MFI techniques that consider a wider range of modulation formats and channel impairments, without compromising the classification/monitoring accuracy or adding impractical computational complexity.

- The proposed supervised ML techniques for OPM and MFI are often trained off-line using a specific dataset that might not be comprehensive to include all key network situations. Hence, the applicability of these techniques in practical systems remains questionable, and self-learning approaches could play a vital role in such cases.

V. OPM AND MFI FOR MULTIPLEXED SIGNALS

In Section IV, we have discussed ML for OPM and MFI for fiber-based optical networks. However, the advances in optical communications opened new applications that require introducing new types of optical networks or using more efficient multiplexing techniques. In this section, we discuss the ML-based OPM and MFI methods that are proposed in literature for multiplexed signals.

In general, multiplexing techniques are useful to increase network data rate, hence improve bandwidth efficiency. However, they come with some challenges that require adequate OPM and/or MFI techniques. In the next subsections, we review such multiplexing techniques that are proposed in optical networks.

A. Orthogonal Frequency-Division Multiplexing (OFDM)

Orthogonal frequency-division multiplexing (OFDM) is one of the multiplexing techniques that can be used in next generation optical networks. It provides some advantages such as its resistance to CD effect. During the last two years, few MFI and OPM techniques have been proposed for OFDM signals using direct detection [200], [201] and coherent detection [202] receivers. In [200], an MFI technique that is relying on ANN algorithm was proposed. The ANN is trained using AH features extracted from the real part of data after FFT stage. This technique was verified experimentally to identify five modulation formats including 4-, 16-, 32-, 64- and 128-QAM, transmitted over 25 km fiber link. The proposed method achieved 100% identification accuracy for received optical power (ROP) greater than -10 dBm. Later, the authors extended their work to perform joint OSNR monitoring and MFI using *k*-NN algorithm which was trained using AH [203]. To reduce the computational complexity, the authors applied some pre-processing to reduce the features size. The results showed reduction in the computational complexity of 1.4% compared with ANN-based technique. Direct detection was used in this work, which reduces the cost so that MFI can be used for intermediate nodes. However, MFI-based AH cannot be used for phase modulation formats (e.g., M-PSK) because these types of modulations have constant amplitude, whereas AH deals with changes in amplitude.

Another MFI technique based on CFSFDP algorithms that does not require prior training was presented in [204]. In this technique, the peak density and distance between data points are utilized to define the number of clusters, based on which the MFI is achieved. The proposed technique was verified experimentally for 4-, 8-, 16-, 32-, and 64-QAM modulation formats. The results showed that this technique can achieve 100% classification accuracy but only for high OSNR greater than 21 dB. This work was extended in [201] by utilizing *k*-NN regression method to count the number of clusters. The simulation results showed that this technique can achieve 100% identification accuracy for 8-PSK, QPSK, 8, 16, 64, and 128-QAM modulation formats transmitted at 12.5 Gbaud symbol rate with OSNR greater than 19.5 dB.

In [202], the authors introduced MFI based modulus mean square (MMS) features. In this method, different modulation formats are transmitted over each subcarrier. The mean square of the I and Q points for each modulation is calculated and then an appropriate threshold is applied for each subcarrier. The proposed method was verified through simulation and experiments to identify different QPSK and M-QAM (M = 8, 16, 32, and 64) signals. The maximum OSNR required to achieved 100% identification rate is 25 dB. This work can be used to identify hybrid modulation formats. However, it is based on coherent detection to extract the dataset which in turn increases the cost.

B. Few Mode Fiber (FMF) Multiplexing

FMF is a type of SDM where more than one mode is exploited for data transmission. Recently, there has been much interest to exploit FMF in future elastic optical networks to reduce the overall cost while improving the network capacity. Using MFI for SMF-based networks has been widely studied in literature. However, applying MFI for FMF-based networks is still in its infancy. In contrast to SMF, the optical signal in FMF is subject to some additional impairments such as the introduced cross-talk between the modes (mode coupling (MC)). The authors in [205] investigated by simulation identifying six types of modulation formats, including DP-BPSK, DP-QPSK, DP-8-QAM, DP-16-QAM, DP-32-QAM, and DP-64-QAM, under the effect of ASE and CD impairments, besides, the MC. ANN classifier was used which exploits the sampled received signal's IQH features to train the network. The average identification accuracy was found to be 98% in the presence of low MC. However, this accuracy reduces to 90% under the effect of high MC and CD. Furthermore, the average identification accuracy was investigated under different symbol rates including 14 and 20 Gbaud.

C. Lessons Learned

The ML-based techniques of OPM and MFI for multiplexed signals are still in its infancy and only few studies are reported in the literature. For the OFDM signal, the reported monitoring techniques considered the traditional impairments of the optical signal and ignored the impairments related to OFDM signal such as inter-carrier-interference. Similarly, reported techniques for FMF signals did not consider all the impairments related to FMF signals such as PMD, mode-dependent loss (MDL), and frequency offset. Since FMF technology is a promising solution for future optical networks, more research is needed in this area. Moreover, developing adaptive techniques that can work for single and multi-carrier

(i.e., super-channel and WDM) networks is of interest to next-generation high-speed networks.

VI. MFI and OPM for Access Networks

Here, we survey OPM and MFI for access networks. In particular, we consider two networks where MFI and/or OPM are used to improve their performance. These are RoF and free space optical (FSO) communication networks.

A. Radio Over Fiber (RoF) Network

RoF is a hybrid network, where the RF wireless signals' transportation and distribution is achieved through optical networks. This technology is gaining more interest especially in future wireless networks such as the 5G communication, where RF transmission distance is getting shorter due to exploiting the high frequencies. Hybrid RoF network can solve the distance reach issue by extending the RF transmission distances using fiber. Using ML for MFI in RF wireless networks is well investigated in literature [206]–[208]. However, MFI in hybrid RoF technology is still in its early stages. In [209], the authors proposed using ANN algorithm to classify four types of modulation formats; BPSK, QPSK, 16-QAM, and OFDM-QPSK. The extracted AAH features of the sampled received signal are used to train the ANN. Different impairments were considered, including CD, DGD, and ASE. Using 60 GHz RF carrier, the simulation results showed that an accuracy more than 99% can be achieved regardless of the type of modulation formats at data rate less than 2.6 Gbps.

Another technique for MFI in hybrid RoF was proposed in [210], which exploits an auto-encoder preceded by a pre-processing step. The pre-processing step involves sorting the values of the in-phase and quadrature samples of a received signal. In this work, a 28 GHz RF signal and six different modulation formats; BPSK, QPSK, 8-PSK, 16-, 64-, and 256-QAM were considered. The performance of the classifier was investigated experimentally and by simulation under the effect of CD and ASE impairments. The results showed good agreement between the simulation and experiments. A classification accuracy of 98% was achieved for OSNR greater than 10 dB and fiber length less than 60 km. For fiber length longer than 70 km, the performance starts degrading more because of CD effect.

B. Free Space Optical (FSO) Communication Network

In FSO communications, OAM is a new multiplexing technique that has gained interest during the last years due to its capability to provide new freedom of signal's carriers and hence double the data rate of optical communication systems [211]. When the orthogonal OAM modes are transmitted over turbulent free space channel, the wave-front phase is perturbed and hence cross-talk is introduced at the receiver. Monitoring the atmospheric turbulence (AT) impairment helps in exploiting adaptive modulation techniques or even correcting AT impairment. In [212], a CNN algorithm was used to determine the severity of AT and simultaneously detect the OAM modes, in an M-ary pattern coding system. The proposed algorithm was verified in simulation by considering six values for AT covering weak to strong turbulence for 4-OAM, 8-OAM, and 16-OAM. The detecting accuracy of AT types was found to be 95% on average. Instead of monitoring specific AT types, the authors in [213] used CNN to build a system that is capable of providing feedback to the transmitter to correct the OAM's transmitted mode, which was disturbed by random AT. The results showed that by using such a technique, the received OAM's mode has been found to be close to the desired profile.

C. Lessons Learned

Few lessons can be extracted from this section pertaining to MFI and OPM for access networks.

- The reach distance in future RF wireless networks such as 5G/6G is very short. Fiber cables are then viable solutions to extend the reach distance. The transmitted signal over this hybrid RF/fiber link is subject to different types of impairments, some of which occur in the optical domain and the others in the electrical domain. Therefore, the development of future OPM/MFI techniques should consider both types of impairments. Additionally, fiber impairment, particularly CD, shows a different behavior in RoF channels, owing to the double-sideband transmission, as compared to the optical baseband distortion such as the repetitive signal fading and time-shifting effects. This new behavior deserves special attention from researchers to investigate its effect on the performance of ML classifiers.
- Fiber installation is sometimes difficult or even impossible. Therefore, FSO technology is proposed to replace the fiber in such cases. However, FSO technology has its own impairments which require designing suitable MFI and OPM techniques. The work in this area is still in its infancy and more research is required to consider the different types of impairments, such as signal scattering, turbulence, pointing errors, and phase distortion, in the development of OPM/MFI techniques.

VII. Discussions and Guidelines

Since there are many ML-based techniques proposed for OPM and MFI, some criteria are needed to help identifying the appropriate algorithms for specific applications. Moreover, feature selection is a primary element in determining the algorithm performance. In this section, we first list some criteria that can be used for ML algorithm selection. Then, we discuss and compare the different types of features that are proposed in the literature for OPM and MFI in optical networks.

A. Criteria for Identifying the Appropriate Algorithm

1) Accuracy: The accuracy is an important metric that identifies how much the proposed technique is accurate and sensitive in predicting the amount of a specific impairment or identifying a type of a modulation format. High accuracy is required to provide suitable decisions accordingly.

2) Multitasking: Multitasking in OPM is defined as the capability of the proposed technique to monitor multiple impairments simultaneously because the optical signal is

subjected to different simultaneous impairments while propagating in the channel. In MFI, the proposed techniques are developed to identify multiple modulation formats, as in adaptive optical networks where the modulation formats change according to the channel condition. Therefore, multitask solutions can be utilized to perform joint OPM and MFI; thereby implementing one system/algorithm instead of many.

3) Cost of Data Acquisition Hardware: The first step in OPM and MFI is data acquisition. The hardware required for signal acquisition needs to be simple and inexpensive. This is because OPM and MFI functions need to be installed in many nodes in the network, therefore low-cost solutions help in reducing the overall network cost.

4) Implementation: The proposed techniques need to be easy to implement and integrate with the network equipment. A technique that only requires installing of the OPM/MFI algorithm on the receiver memory is easy to implement. On the other side, a technique that requires modifying the receiver somehow to implement the algorithm is harder to implement.

5) Computational Complexity: Computational complexity means the amount of time, storage, and other resources that are needed to execute a particular algorithm. MFI and OPM techniques that exploit algorithms with low computational complexity are preferred. The input data size to the algorithm needs to be small. Furthermore, predicting the impairment or identifying the modulation format needs to be fast to ensure building proactive optical networks.

6) Impairments Range: The severity of an impairment depends on the conditions of the impairment's source. For example, low OSNR values exist in case of fiber damage or power eavesdropping. In such cases, OSNR can drop to low values in range of few decibels. Therefore, an OPM technique should be able to monitor a specific impairment over wide range extending from low values to high values. Similarly, an MFI technique should be able to identify the type of a modulation format even under harsh conditions such as low signal power, large noise, and large CD.

B. Features Utilized for OPM and MFI

The performance of ML algorithms is heavily based on the utilized features and classifiers/regressors. In Section II, we discussed in details the different classifiers/regressors employed in literature. Therefore, to complete the picture about the surveyed algorithms, we consider here the most commonly used features in literature for OPM and MFI.

Each OPM or MFI technique reported in this survey is using certain type of features for training the ML algorithm. Therefore, these features play an important role in determining the effectiveness of the ML-based OPM and MFI techniques. In this section, we further discuss these features with emphasis on their pros and cons.

By virtue of the discussion previously presented in Sections IV and V, we observe that most of the features extracted from the time domain signals can be classified according to their sampling technique, either synchronous or asynchronous. Asynchronous features are often used such as the reconstruction eye diagram using the chirp-z conversion software synchronization algorithm [145], AAHs [148], [159], [162], ADTSs [40], [150], [152], [191], [195], ASCS [192], IQH [156], [205], and asynchronous constellation diagram [155]. Most of these features have been extracted using low cost direct detection acquisition systems, making them attractive for intermediate nodes in the optical networks.

In heterogeneous fiber optic networks, direct detection acquisition systems are not appropriate for long fiber transmission links due to the accumulated linear impairments such as the CD. In contrast, in coherent detection acquisition systems, MFI becomes more reliable because DSP algorithms for CD and some other impairments compensation are often employed. Features that have been considered from such acquisition systems include AH [39], [49], [95], [181], [196], [200], [203], CDF [197], and power distribution [130]. These features are insensitive to PN and frequency offset. In addition, they are suitable for multi-level modulation formats such M-QAM (M = 2, 4, and 8) because of their dependency on the signal amplitude. However, for phase modulation formats such as the M-PSK (M = 4, 16, 32, and 64), they are not appropriate. Moreover, they cannot be used for PN impairment monitoring because they are insensitive to it. In dual polarization modulations, the signals are prone to polarization rotation as well as PN and frequency offset. The received signal that is acquired coherently is often mapped into Stokes space constellation to produce features that are completely independent of these impairments [49], [166]–[172], [175], [177]–[179], [198]. However, Stokes space constellation features are sensitive to ASE noise, because the mapping distorts the ASE noise probability density function. Moreover, Stokes space constellation features are distorted by PMD and PDL.

Besides, HOC features and some other features extracted from the signal amplitude and phase that are acquired coherently have been exploited in [132], [133], [135]. In contrast to Stokes space constellation features, these features are insensitive to ASE noise. However, they are relying on DT that is sensitive to pre-defined threshold values.

Features extracted from the frequency domain have been considered as well. LF components of RF spectrum have been used as features for OSNR monitoring [154]. These features showed accurate results under high CD impairment values. However, they are not suitable for CD and PMD monitoring. On the other hand, FFT-based features extracted after taking different powers of the time-domain received signal [137], [138] introduce computational complexity. In addition, these features are only robust at low values of OSNR due to their association with thresholding.

Features extracted from images, e.g., the eye diagram images [193], constellation images [199], and RT images [189] require long processing time and large memory. Similarly, deep learning algorithms, which extract features directly from raw data require long training time and large data size to get acceptable results.

Figure 30 presents pictorial classification for the commonly used features in the literature for OPM and MFI, while Table X summarizes the pros and cons of these features. Moreover, Table XI lists some reported comparisons in the literature for different ML techniques using specific features.

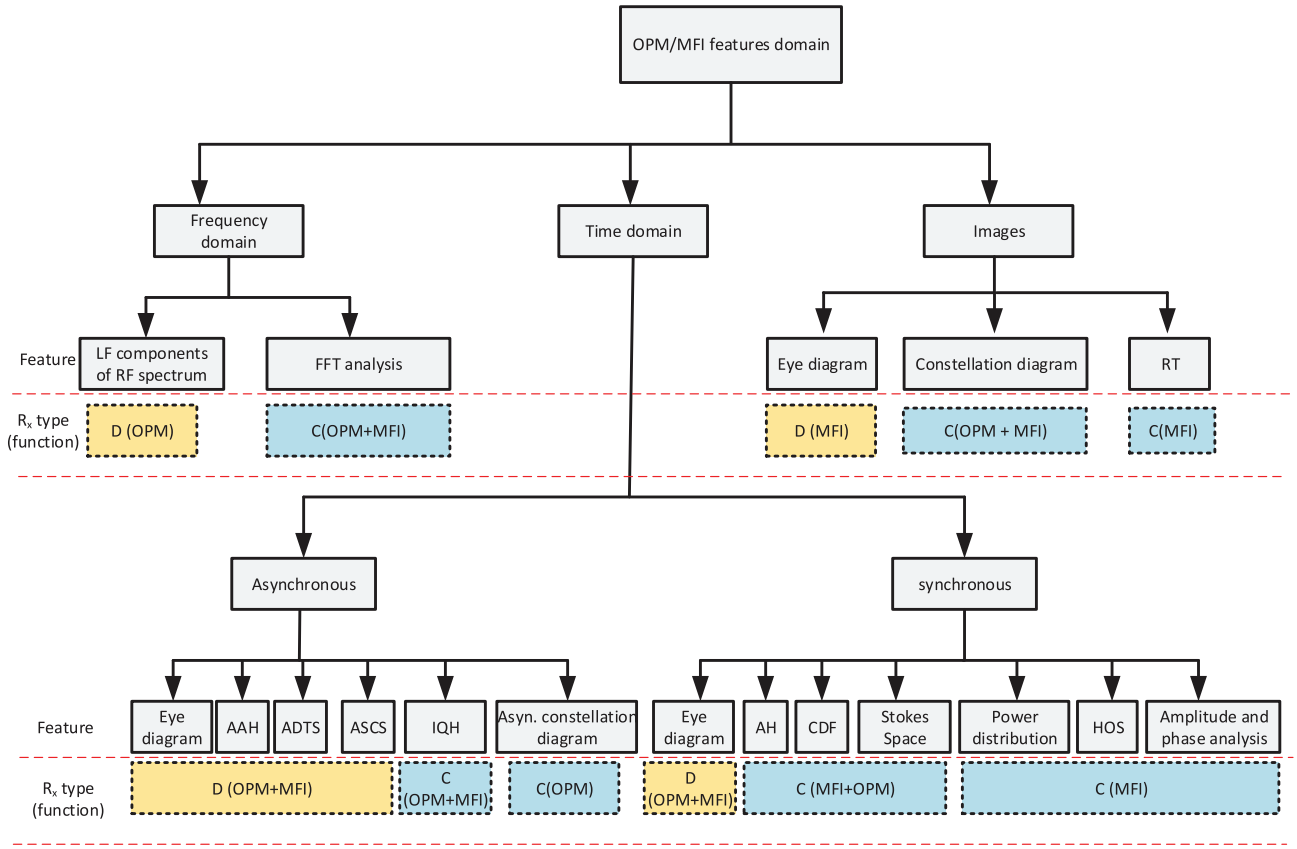

Fig. 30. Features used in the literature. C: coherent detection and D: direct detection.

VIII. Lessons Learned, Open Issues and Research Direction

ML-based OPM and MFI for optical networks have received considerable attentions over the last decade. However, there are some challenges that still require more investigations. In addition, there are some other optical networks that require the development of MFI and OPM techniques adequate for their nature. In this section, we discuss the learned lessons, highlight the main gaps in the currently proposed ML-based techniques, and present our vision to improve their performance. Furthermore, we discuss the research directions pertaining to OPM and MFI in the coming few years.

A. Algorithm Multitasking

ML-based OPM and MFI techniques in the current existing works are limited to monitoring/identifying specific impairment/modulation formats, which do not support future optical networks that are subject to different types of impairments and support variety of modulation formats. For example, the algorithms that exploit AH features can be used to identify M-QAM (M = 4, 16, 32, and 64) modulation formats but not M-PSK (M = 2, 4, and 8) modulation formats. Similarly, the usage of AH features in OPM algorithms does not support monitoring PN impairment. Therefore, there is a need to propose more comprehensive features so that OPM and MFI techniques have the capability to monitor a wide range of impairments and identify a large number of modulation formats. The ability to develop a multitasking algorithm that performs joint OPM/MFI is the foundation for future intelligent optical networks.

B. New Modulation Formats

So far, the surveyed literature is limited to the recognition of the traditional modulation formats such as M-PSK (M = 2, 4, and 8) and M-QAM (M = 8, 16, 32, and 64). Apart from BPSK and QPSK, M-PSK ($M \geq 8$) systems are not recommended for future optical fiber networks (i.e., 400 Gbps and 1Tbps) [214]. Besides, for the M-QAM schemes, every point in the constellation diagram is located on a uniform Cartesian grid and transmitted with equal probability. Recently, new more efficient modulation formats have been proposed to improve the capacity and achieve better power efficiency such as the probabilistic constellation shaping modulation and geometric constellation shaping formats [215]–[217]. The former relies on transmitting the constellation points at different probabilities while the latter is based on optimizing the OSNR. The introduction of such new modulation formats in optical communication requires new MFI algorithms at the receiver not only to determine the different modulation orders but also the signal type such as OOK, QAM, probabilistic, geometric, etc. to facilitate proper demodulation. A recent study for MFI of probabilistic shaping modulation formats showed that for some special cases the probabilistic shaping 16-QAM format might have very small differences in the Jones space with the

TABLE X
SUMMARY OF PROS AND CONS OF THE FEATURES UTILIZED IN THE LITERATURE

Acquisition system	Features domain		Pros	Cons	Ref
Direct detection	Eye diagram	Synchronous	Simple (no pre-processing)	High speed sampling	[143], [144], [190]
		Asynchronous	Low speed sampler (cost effective)	Requires software synchronization algorithms	[145]
	AAH		Low speed sampler (cost effective)	Highly affected by CD and PMD	[148], [159], [162]
	ADTS		Effective for multiple impairments monitoring and formats classification	High-cost (two sampler)	[40], [150], [152], [191], [195]
	ASCS		Similar to ADTS but uses single samplers	Requires software synchronization algorithms	[193]
	RF spectrum		Effective for OSNR monitoring under large CD and PMD impairments	Limited to OSNR monitoring	[154]
	Eye diagram Images		No requirement for Pre-processing	-Require time recovery -High computational complexity	[193]
Coherent detection	AH		Insensitive to PN and offset frequency	Not appropriate for M-PSK classification and PN monitoring	[39], [49], [95], [181] [196], [200], [203]
	CDF		-Insensitive to PN and offset frequency - Uses less sample.	Not appropriate for M-PSK classification and PN monitoring	[136], [197]
	Power distribution		-Insensitive to PN and offset frequency	Requires prior information of OSNR -Not suitable for M-PSK	[130]
	IQH		Appropriate for phase and amplitude modulation formats	Requires high number of samples.	[156], [205]
	Stokes space		Insensitive to polarization rotation, frequency offset, and PN	Sensitive to ASE noise, CD, and PMD	[49], [166]–[172], [175], [177]–[179] [198]
	Amplitude and phase analysis		No need extensive for training datasets	Sensitive to pre-defined threshold values	[132], [133]
	FFT		Insensitive to ASE noise	Sensitive to pre-defined threshold values	[137], [138]
	Constellation and RT Images		Appropriate for phase and amplitude modulation formats	High computational complexity	[189], [199]

standard 4-QAM format. Similarly, it would be the case with the probabilistic shaping 64-QAM format and the standard 16-QAM format. Therefore, current MFI techniques used for standard M-QAM modulation formats may produce low classification accuracy when employed for the classification of probabilistic shaping modulation formats, and hence new MFI techniques are required for such new formats [218].

C. Nonlinear Impairments

The vast majority of research has focused on single carrier optical systems, making these studies limited only for linear effects. However, in WDM networks, beside the linear effects, there are some other nonlinear impairments such as FWM, SPM, and XPM that require OPM. Non-linear impairments limit the maximum signal power in optical channel which reduce the transmission distance. Therefore, it is necessary to accurately measure these effects to enhance the quality of transmission. According to ITU-T G.697 recommendations [84], in 100 Gbps NRZ/RZ line coding high speed coherent networks, nonlinear impairments have medium effect which requires adequate OPM solutions. ML techniques can help in providing such solutions with high performance and reasonable cost. In fact, ML-based OPM is an attractive solution because it does not require exact knowledge of the mathematical modeling of impairment's nonlinearity.

D. Wireless and Hybrid Optical Networks

Similar to fiber-based optical networks, some other optical networks such as RoF, free space optical communication (FSO), radio over FSO (RoFSO), and visible light communication (VLC) are gaining more interest as counterparts to fiber-based optical networks. They are useful when fiber installation is impossible or costly. Such networks require the development of suitable MFI and OPM techniques that consider additional types of impairments such as RF noise, path loss, shadowing and fading effect in RoF and RoFSO networks. In addition, the VLC and FSO networks have their own impairments such as turbulence, free space signal

TABLE XI
SUMMARY OF REPORTED COMPARISON IN LITERATURE OF ML ALGORITHMS IN TERMS OF COMPUTATIONAL COMPLEXITY AND PERFORMANCE (ACCURACY)

Employed ML algorithms	MFI/OPM	Utilized Features	Accuracy	Computational complexity	Ref.
ANN, *k*-NN, DT, SVM, and CNN	MFI	Constellation images	• CNN has the best accuracy • DT has the worst accuracy	• DT has the lowest computational time • *k*-NN has the largest computational time	[199]
ANN, *k*-NN, DT, SVM, and CNN	MFI	Eye diagram images	• CNN has the best accuracy • DT has the worst accuracy	—	[193]
Random forest, *k*-NN, SVM and DNN	MFI	AH	• Random forest achieves the highest accuracy	• Random forest achieves the lowest complexity	[181]
DNN and CNN	OPM (OSNR)	AAH and ADTS	• ADTS with CNN achieves better accuracy than AHH with DNN under strong CDs' values	—	[195]
SVM and DNN	Joint OPM OSNR/MFI	CDF	• SVM attains better accuracy	• SVM and DNN have comparable time complexity	[197]
MTL-CNN and PCA	Joint OPM/MFI	ADTS	• CNN-based MTL has higher accuracy	• CNN-based MTL has better computational time	[40]
EM, *k*-means, DBSCAN, OPTICS, Spectral clustering, and Maximum-likelihood	MFI	Stoke Space	• Maximum-likelihood requires the lowest OSNR to achieve accuracy 95%	• OPTICS & DBSCAN have the lowest computational time	[170]
DBSCAN, Maximum-likelihood, and Subtraction Clustering	MFI	Stoke Space	• Subtraction clustering and Maximum-likelihood require the lowest OSNR level to achieve accuracy 95%	• Subtraction clustering and DBSCAN have the lowest computational time	[177]
CFSFDP, DBSCAN, and OPTICS	MFI	Stoke Space	• CFSFDP requires the lowest OSNR level to achieve accuracy 95%	—	[178]
Adaptive CFSFDP, Maximum-likelihood, and DBSCAN	MFI	Stoke Space	• Adaptive CFSFDP and Maximum-likelihood require the lowest OSNR level to achieve accuracy 95%	• DBSCAN has the lowest computational time	[175]
Adaptive CFSFDP and PNN	MFI	Stoke Space	• PNN requires the lowest OSNR level to achieve accuracy 95%	—	[49]

scattering, light reflection, and pointing errors. As the hybrid networks combine different types of channels, the task of designing suitable OPM and MFI becomes more complicated.

E. New Multiplexing Techniques

Besides the traditional multiplexing technique, SDM is a large player in next generation of optical networks to overcome the capacity limit of the existing optical communication systems. SDM is realized by multi-core fiber (MCF), multi-mode fiber (MMF), and FMF. Furthermore, the dense SDM (DSDM) systems, MCFs with FMFs, known as few-mode multicore fibers (FM-MCFs) are also proposed [219]. Such multiplexing techniques have their own new impairments that are not yet considered in the literature such as mode coupling and mode dependent loss which may require the development of new ML-based OPM and MFI algorithms.

F. Real-Time ML Approaches

In many ML-based algorithms, real-time processing of samples and number of samples required to achieve certain accuracy are major concerns especially in time-sensitive applications such as optical communication. Therefore, the processing efficiency is an important factor in developing ML-based solutions. On the other hand, it is noted that all ML-based OPM and MFI techniques utilize offline training data. However, in real dynamic optical network, the traffic data as well as the optical components behavior may change over time. It is envisaged that the optical network would be capable to perform self-learning, self-adapting, and self-optimizing. Therefore, ML-based OPM and MFI techniques need to perform training and fast convergence in real time. Therefore, optical components could be considered to build such algorithms, which will lead to a quantum leap because of their high speed.

G. Available Algorithms and Frameworks in Other Fields

Since using ML for OPM and MFI in optical networks is recent, researchers can benefit from the proposed techniques and used features in other fields such as RF communication, bioinformatics and image processing. MATLAB software is common among researchers in optical communication to develop ML algorithms for OPM and MFI, which is not popular in ML community. Other software such as Python might be more useful for developing ML algorithms because of the availability of many open-source frameworks such as TensorFlow, Pytorch, and Caffe.

IX. Conclusion

OPM and MFI are expected to be an essential part of the next generation optical networks by enabling autonomous optical nodes and receivers which provide increased stability, adaptability, and efficient utilization of network resources. Machine learning has emerged as a reliable solution to build estimation and classification models for OPM and MFI, respectively, due to their ability to provide data-drive solutions that are efficient and accurate. This article provided a comprehensive survey of ML-based OPM, MFI, and joint OPM/MFI techniques for both direct and coherent systems. It also provided comparisons between these proposed techniques and addressed future research directions and open issues.

Proposed algorithms in the literature span a wide spectrum of techniques but can be categorized in terms of feature extraction and training method. The majority of algorithms use supervised approaches to train OPM and MFI models. However, most of the proposed OPM algorithms have the advantages of exploiting low-cost direct detection acquisition systems with simple asynchronous features. On the other hand, most of the MFI algorithms are using coherent receivers for data acquisition equipped with impairments' mitigation DSP algorithms (e.g., CD compensation algorithms) as preprocessing step. Features used in most MFI algorithms are handcrafted to be transparent to certain types of impairments by utilizing different domains such as Stokes space, CDF, and AH domains. Some MFI algorithms have been proposed using unsupervised learning techniques to cluster data samples from different modulations, and then identify new data samples by finding common features with existing clusters. For joint OPM and MFI, most of the proposed algorithms are limited to joint monitoring of OSNR with MFI.

Unlike OPM/MFI algorithms that uses traditional ML techniques, recent algorithms propose using deep CNNs and LSTMs models to automatically identify features and perform classification or estimation using the same network. This raises the issue of lack of experimental datasets that are truly representative of the different settings in optical networks. One approach to resolve this issue is the use of generative adversarial neural networks to expand existing experimental measurements data. Additionally, transfer learning can be used to efficiently train deep neural networks on new and emerging technologies such as free space optics, few mode fiber, or orbital angular momentum. Other trends in machine learning field such as adaptive learning and online learning can be enabling technologies for OPM and MFI to build the next generation optical networks.

Acknowledgment

The authors acknowledge the funding from the Research and Development (R&D) Program (Research Pooling Initiative), Ministry of Education, Saudi Arabia, (RPI-KSU).

References

[1] I. Yaqoob, I. A. T. Hashem, Y. Mehmood, A. Gani, S. Mokhtar, and S. Guizani, "Enabling communication technologies for smart cities," *IEEE Commun. Mag.*, vol. 55, no. 1, pp. 112–120, Jan. 2017.

[2] M. Jaber, M. A. Imran, R. Tafazolli, and A. Tukmanov, "5G backhaul challenges and emerging research directions: A survey," *IEEE Access*, vol. 4, pp. 1743–1766, 2016.

[3] V. López and L. Velasco, *Elastic Optical Networks, Architectures, Technologies, and Control*. Cham, Switzerland: Springer, 2016.

[4] R. Borkowski *et al.*, "Cognitive optical network testbed: EU project CHRON," *IEEE/OSA J. Opt. Commun. Netw.*, vol. 7, no. 2, pp. A344–A355, Feb. 2015.

[5] E. Palkopoulou *et al.*, "Cognitive heterogeneous reconfigurable optical network: A techno-economic evaluation," in *Proc. Future Netw. Mobile Summit*, 2013, pp. 1–10.

[6] W. Wei, C. Wang, and J. Yu, "Cognitive optical networks: Key drivers, enabling techniques, and adaptive bandwidth services," *IEEE Commun. Mag.*, vol. 50, no. 1, pp. 106–113, Jan. 2012.

[7] M. J. O'Mahony, C. Politi, D. Klonidis, R. Nejabati, and D. Simeonidou, "Future optical networks," *J. Lightw. Technol.*, vol. 24, no. 12, pp. 4684–4696, Dec. 9, 2006.

[8] A. Nag, M. Tornatore, and B. Mukherjee, "Optical network design with mixed line rates and multiple modulation formats," *J. Lightw. Technol.*, vol. 28, no. 4, pp. 466–475, Feb. 15, 2010.

[9] T. S. R. Shen and A. P. T. Lau, "Fiber nonlinearity compensation using extreme learning machine for DSP-based coherent communication systems," in *Proc. 16th Opto Electron. Commun. Conf.*, 2011, pp. 816–817.

[10] A. Amari, X. Lin, O. A. Dobre, R. Venkatesan, and A. Alvarado, "A machine learning-based detection technique for optical fiber nonlinearity mitigation," *IEEE Photon. Technol. Lett.*, vol. 31, no. 8, pp. 627–630, Apr. 27, 2019.

[11] D. Wang *et al.*, "Nonlinear decision boundary created by a machine learning-based classifier to mitigate nonlinear phase noise," in *Proc. Eur. Conf. Opt. Commun. (ECOC)*, 2015, pp. 1–3.

[12] M. Li, S. Yu, J. Yang, Z. Chen, Y. Han, and W. Gu, "Nonparameter nonlinear phase noise mitigation by using M-ary support vector machine for coherent optical systems," *IEEE Photon. J.*, vol. 5, no. 6, pp. 1–13, Dec. 2013.

[13] A. G. Reza and J. K. Rhee, "Nonlinear equalizer based on neural networks for PAM-4 signal transmission using DML," *IEEE Photon. Technol. Lett.*, vol. 30, no. 15, pp. 1416–1419, Aug. 1, 2018.

[14] Z. Wan *et al.*, "Nonlinear equalization based on pruned artificial neural networks for 112-Gb/s SSB-PAM4 transmission over 80-km SSMF," *Opt. Exp.*, vol. 26, no. 8, pp. 10631–10642, Apr. 2018.

[15] E. Giacoumidis *et al.*, "Blind nonlinearity equalization by machine-learning-based clustering for single- and multichannel coherent optical OFDM," *J. Lightw. Technol.*, vol. 36, no. 3, pp. 721–727, Feb. 1, 2018.

[16] C. Häger and H. D. Pfister, "Deep learning of the nonlinear Schrödinger equation in fiber-optic communications," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, 2018, pp. 1590–1594.

[17] C. Monterola and C. Saloma, "Solving the nonlinear schrodinger equation with an unsupervised neural network," *Opt. Exp.*, vol. 9, no. 2, pp. 72–84, Jul. 2001.

[18] O. A. Dobre, A. Abdi, Y. Bar-Ness, and W. Su, "Survey of automatic modulation classification techniques: Classical approaches and new trends," *IET Commun.*, vol. 1, no. 2, pp. 137–156, Apr. 2007.

[19] Z. Zhu and A. K. Nandi, *Automatic Modulation Classification: Principles, Algorithms and Applications*. New York, NY, USA: Wiley, 2015.

[20] S. Troia *et al.*, "Machine-learning-assisted routing in SDN-based optical networks," in *Proc. Eur. Conf. Opt. Commun.*, 2018, pp. 1–3.

[21] M. Klinkowski, P. Ksieniewicz, M. Jaworski, G. Zalewski, and K. Walkowiak, "Machine learning assisted optimization of dynamic crosstalk-aware spectrally-spatially flexible optical networks," *J. Lightw. Technol.*, vol. 38, no. 7, pp. 1625–1635, Apr. 1, 2020.

[22] T. Panayiotou, G. Savva, I. Tomkos, and G. Ellinas, "Decentralizing machine-learning-based QoT estimation for sliceable optical networks," *IEEE/OSA J. Opt. Commun. Netw.*, vol. 12, no. 7, pp. 146–162, Apr. 2020.

[23] Y. Zhao, B. Yan, D. Liu, Y. He, D. Wang, and J. Zhang, "SOON: Self-optimizing optical networks with machine learning," *Opt. Exp.*, vol. 26, no. 22, pp. 28713–28726, Oct. 2018.

[24] A. Ragheb *et al.*, "Identifying structured light modes in a desert environment using machine learning algorithms," *Opt. Exp.*, vol. 28, no. 7, pp. 9753–9763, Mar. 2020.

[25] 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, Jan. 15, 2019.

[26] D. Rafique, T. Szyrkowiec, H. Grießer, A. Autenrieth, and J. Elbers, "Cognitive assurance architecture for optical network fault management," *J. Lightw. Technol.*, vol. 36, no. 7, pp. 1443–1450, Apr. 1, 2018.

[27] X. Liu *et al.*, "AI-based modeling and monitoring techniques for future intelligent elastic optical networks," *Appl. Sci.*, vol. 10, no. 1, p. 363, Jan. 2020.

[28] M. Bouda *et al.*, "Accurate prediction of quality of transmission based on a dynamically configurable optical impairment model," *IEEE/OSA J. Opt. Commun. Netw.*, vol. 10, no. 1, pp. A102–A109, Jan. 2018.

[29] M. Furdek and C. Natalino, "Machine learning for optical network security management," in *Proc. Opt. Fiber Commun. Conf. Exhibit. (OFC)*, 2020, pp. 1–3.

[30] A. Gepperth and B. Hammer, "Incremental learning algorithms and applications," in *Proc. Eur. Symp. Artif. Neural Netw.*, Bruges, Belgium, 2016, pp. 357–368.

[31] J. Li, D. Xue, W. Wu, and J. Wang, "Incremental learning for malware classification in small datasets," *Security Commun. Netw.*, vol. 20, no. 20, pp. 1–12, Feb. 2020.

[32] 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. (OFC)*, San Diego, CA, USA, 2018, pp. 1–3.

[33] F. Musumeci *et al.*, "An overview on application of machine learning techniques in optical networks," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 2, pp. 1383–1408, 2nd Quart., 2019.

[34] Z. Dong *et al.*, "Optical performance monitoring: A review of current and future technologies," *J. Lightw. Technol.*, vol. 34, no. 2, pp. 525–543, Jan. 15, 2016.

[35] Z. Pan, C. Yu, and A. E. Willner, "Optical performance monitoring for the next generation optical communication networks," *Opt. Fiber Technol.*, vol. 16, no. 1, pp. 20–45, 2010.

[36] D. Dahan, U. Mahlab, A. Teixeira, I. Zacharopoulos, and I. Tomkos, "Optical performance monitoring for translucent/transparent optical networks," *IET Optoelectron.*, vol. 5, no. 1, pp. 1–18, Feb. 2011.

[37] J. Xie *et al.*, "A survey of machine learning techniques applied to software defined networking (SDN): Research issues and challenges," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 1, pp. 393–430, 1st Quart., 2019.

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

[39] F. N. Khan *et al.*, "Joint OSNR monitoring and modulation format identification in digital coherent receivers using deep neural networks," *Opt. Exp.*, vol. 25, no. 15, pp. 17767–17776, Jul. 2017.

[40] X. Fan *et al.*, "Joint optical performance monitoring and modulation format/bit-rate identification by CNN-based multi-task learning," *IEEE Photon. J.*, vol. 10, no. 5, pp. 1–12, Oct. 2018.

[41] Y. Cheng, S. Fu, M. Tang, and D. Liu, "Multi-task deep neural network (MT-DNN) enabled optical performance monitoring from directly detected PDM-QAM signals," *Opt. Exp.*, vol. 27, no. 13, pp. 19062–19074, Jun. 2019.

[42] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Trans. Inf. Theory*, vol. IT-13, no. 1, pp. 21–27, Jan. 1967.

[43] H. Suzuki and N. Takachio, "Optical signal quality monitor built into WDM linear repeaters using semiconductor arrayed waveguide grating filter monolithically integrated with eight photodiodes," *Electron. Lett.*, vol. 35, no. 10, pp. 836–837, 1999.

[44] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf, "Support vector machines," *IEEE Intell. Syst. Appl.*, vol. 13, no. 4, pp. 18–28, Jul./Aug. 1998.

[45] *Optical Performance Monitor*, Optoplex Corporat., Fremont, CA, USA, Nov. 2019.

[46] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sep. 1995.

[47] *1310 1550nm WDM Filter Integrated Optical Power Monitor*, Lightwaves2020, Milpitas, CA, USA, Nov. 2019.

[48] S. Haykin, *Neural Networks: A Comprehensive Foundation*. Upper Saddle River, NJ, USA: Prentice-Hall, 1994.

[49] M. Hao *et al.*, "Stokes space modulation format identification for optical signals using probabilistic neural network," *IEEE Photon. J.*, vol. 10, no. 3, pp. 1–13, Jun. 2018.

[50] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst. Conf.*, 2012, pp. 1097–1105.

[51] H. Salehinejad, S. Sankar, J. Barfett, E. Colak, and S. Valaee. (2017). *Recent Advances in Recurrent Neural Networks*. [Online]. Available: https://arxiv.org/abs/1801.01078.

[52] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[53] J. R. Quinlan, "Induction of decision trees," *Mach. Learn.*, vol. 1, no. 1, pp. 81–106, Mar. 1986.

[54] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001.

[55] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proc. 5th Berkeley Symp. Math. Stat. Probab.*, Berkeley, CA, USA, 1967, pp. 281–297.

[56] H.-S. Park and C.-H. Jun, "A simple and fast algorithm for k-medoids clustering," *Expert Syst. Appl.*, vol. 36, no. 2, pp. 3336–3341, 2009.

[57] C. E. Rasmussen, "The infinite Gaussian mixture model," in *Proc. 12th Int. Conf. Neural Inf. Process. Syst.*, Cambridge, MA, USA, 1999, pp. 554–560.

[58] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Stat. Soc. B Methodol.*, vol. 39, no. 1, pp. 1–22, 1977.

[59] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer, 2006.

[60] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, *A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases With Noise*. New York, NY, USA: AAAI, 1996, pp. 226–231.

[61] M. Ankerst, M. M. Breunig, H. P. Kriegel, and J. Sander, "Optics: Ordering points to identify the clustering structure," *ACM SIGMOD Rec.*, vol. 28, no. 2, pp. 49–60, 1999.

[62] A. Rodriguez and A. Laio, "Clustering by fast search and find of density peaks," *Science*, vol. 344, no. 6191, pp. 1492–1496, 2014.

[63] L. V. D. Maaten, E. Postma, and J. V. D. Herik, "Dimensionality reduction: A comparative," *J. Mach. Learn. Res.*, vol. 10, nos. 66–71, pp. 1–35, 2009.

[64] S. Schaal, S. Vijayakumar, and C. G. Atkeson, "Local dimensionality reduction," in *Proc. Adv. Neural Inf. Process. Syst.*, 1998, pp. 633–639.

[65] L. V. D. Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.

[66] H. Hotelling, "Analysis of a complex of statistical variables into principal components," *J. Educ. Psychol.*, vol. 24, no. 6, p. 417, 1933.

[67] S. Karamizadeh, S. M. Abdullah, A. A. Manaf, M. Zamani, and A. Hooman, "An overview of principal component analysis," *J. Signal Inf. Process.*, vol. 4, no. 3, pp. 173–175, 2013.

[68] A. Hyvärinen and E. Oja, "Independent component analysis: Algorithms and applications," *Neural Netw.*, vol. 13, nos. 4–5, pp. 411–430, 2000.

[69] W. S. Torgerson, "Multidimensional scaling: I. Theory and method," *Psychometrika*, vol. 17, no. 4, pp. 401–419, 1952.

[70] D. K. Agrafiotis, "Stochastic proximity embedding," *J. Comput. Chem.*, vol. 24, no. 10, pp. 1215–1221, 2003.

[71] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.

[72] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: MIT Press, 2018.

[73] X. Chen, B. Li, R. Proietti, H. Lu, Z. Zhu, and S. J. B. Yoo, "DeepRMSA: A deep reinforcement learning framework for routing, modulation and spectrum assignment in elastic optical networks," *J. Lightw. Technol.*, vol. 37, no. 16, pp. 4155–4163, Aug. 15, 2019.

[74] J. Suárez-Varela *et al.*, "Routing based on deep reinforcement learning in optical transport networks," in *Proc. Opt. Fiber Commun. Conf. Exhibit.*, San Diego, CA, USA, 2019, pp. 1–3.

[75] P. J. Winzer and R. Essiambre, "Advanced optical modulation formats," *Proc. IEEE*, vol. 94, no. 5, pp. 952–985, May 2006.

[76] L. N. Binh, *Optical Modulation: Advanced Techniques and Applications in Transmission Systems and Networks*. Hoboken, NJ, USA: CRC Press, 2017.

[77] *Optical System Design and Engineering Considerations*. ITU, Geneva, Switzerland, 2016.

[78] I. Kaminow, T. Li, and A. E. Willner, *Optical Fiber Telecommunications VB: Systems and Networks*. Amsterdam, The Netherlands: Elsevier, 2010.

[79] "Characteristics of a single-mode optical fibre and cable," IETF, Fremont, CA, USA, ITU Recommendations G.652, 2016.

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

[81] J. Li *et al.*, "A self-coherent receiver for detection of PolMUX coherent signals," *Opt. Exp.*, vol. 20, no. 19, pp. 21413–21433, Sep. 2012.

[82] P. J. Winzer *et al.*, "Spectrally efficient long-haul optical networking using 112-Gb/s polarization-multiplexed 16-QAM," *J. Lightw. Technol.*, vol. 28, no. 4, pp. 547–556, Feb. 30, 2010.

[83] C. Xie and G. Raybon, "Digital PLL based frequency offset compensation and carrier phase estimation for 16-QAM coherent optical communication systems," in *Proc. 38th Eur. Conf. Exhibit. Opt. Commun.*, Amsterdam, The Netherlands, 2012, pp. 1–3.

[84] "Optical monitoring for dense wavelength division multiplexing systems," IETF, Fremont, CA, USA, ITU Recommendations G.697, 2016.

[85] S. Stanic, S. Subramaniam, H. Choi, G. Sahin, and H.-A. Choi, "On monitoring transparent optical networks," in *Proc. Int. Conf. Parallel Process. Workshop*, Vancouver, BC, Canada, 2002, pp. 217–223.

[86] "Definitions and test methods for linear, deterministic attributes of single-mode fibre and cable," IETF, Fremont, CA, USA, ITU Recommendations G.650.1, 2018.

[87] "Definitions and test methods for statistical and non-linear related attributes of single-mode fibre and cable," IETF, Fremont, CA, USA, ITU Recommendations G.650.2, 2015.

[88] C. D. Poole and D. L. Favin, "Polarization-mode dispersion measurements based on transmission spectra through a polarizer," *J. Lightw. Technol.*, vol. 12, no. 6, pp. 917–929, Jun. 6, 1994.

[89] *ONA600 Optical Network Analyser*, PE.Fiberoptics, Wokingham, U.K., Sep. 2008.

[90] *CHROMOS11 Optical Network Chromatic Dispersion and PMD Test Set*, PE.fiberoptics, Wokingham, U.K., Sep. 2005.

[91] *VIAVI T-BERD/MTS High-Resolution Dispersion Test Solution*, VIAVI Solutions, San Jose, CA, USA, 2018.

[92] M. Skold *et al.*, "Field trial of polarization-assisted optical performance monitoring operating in an 820 km WDM system," in *Proc. Eur. Conf. Opt. Commun.*, 2006, pp. 1–2.

[93] T. Morgan, Y. R. Zhou, A. Lord, and T. Anderson, "Non-intrusive simultaneous measurement of OSNR, CD and PMD on live WDM system," in *Proc. OFC/NFOEC*, Los Angeles, CA, USA, 2012, pp. 1–3.

[94] S. Yan *et al.*, "Field trial of machine-learning-assisted and SDN-based optical network planning with network-scale monitoring database," in *Proc. Eur. Conf. Opt. Commun. (ECOC)*, 2017, pp. 1–3.

[95] F. N. Khan, K. Zhong, W. H. Al-Arashi, C. Yu, C. Lu, and A. P. T. Lau, "Modulation format identification in coherent receivers using deep machine learning," *IEEE Photon. Technol. Lett.*, vol. 28, no. 17, pp. 1886–1889, Sep. 1, 2016.

[96] A. Amari *et al.*, "A survey on fiber nonlinearity compensation for 400 Gb/s and beyond optical communication systems," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 4, pp. 3097–3113, 4th Quart., 2017.

[97] L. Baker-Meflah, S. Savory, B. Thomsen, J. Mitchell, and P. Bayvel, "In-band OSNR monitoring using spectral analysis after frequency down-conversion," *IEEE Photon. Technol. Lett.*, vol. 19, no. 2, pp. 115–117, Jan. 15, 2007.

[98] M. Rasztovits-Wiech, M. Danner, and W. R. Leeb, "Optical signal-to-noise ratio measurement in WDM networks using polarization extinction," in *Proc. 24th Eur. Conf. Opt. Commun.*, Madrid, Spain, 1998, pp. 549–550.

[99] J. H. Lee and Y. C. Chung, "An improved OSNR monitoring technique based on polarization-nulling method," in *Proc. Opt. Fiber Commun. Conf. Exhibit*, Anaheim, CA, USA, 2001, pp. 1–3.

[100] M.-H. Cheung, L.-K. Chen, and C.-K. Chan, "A PMD-insensitive OSNR monitoring scheme based on polarization-nulling with off-center narrowband filtering," in *Proc. Opt. Fiber Commun. Conf.*, Los Angeles, CA, USA, 2004, pp. 2562–2564.

[101] M.-H. Cheung, L.-K. Chen, and C.-K. Chan, "PMD-insensitive OSNR monitoring based on polarization-nulling with off-center narrow-band filtering," *IEEE Photon. Technol. Lett.*, vol. 16, no. 11, pp. 2562–2564, Nov. 11, 2004.

[102] H. Y. Choi *et al.*, "Improved polarization-nulling technique for monitoring OSNR in WDM network," in *Proc. Opt. Fiber Commun. Conf. Nat. Fiber Opt. Eng. Conf.*, Anaheim, CA, USA, 2006, pp. 1–3.

[103] M. D. Feuer, "Measurement of OSNR in the presence of partially polarized ase," *IEEE Photon. Technol. Lett.*, vol. 17, no. 2, pp. 435–437, Feb. 24, 2005.

[104] J. H. Lee, H. Y. Choi, S. K. Shin, and Y. C. Chung, "A review of the polarization-nulling technique for monitoring optical-signal-to-noise ratio in dynamic WDM networks," *J. Lightw. Technol.*, vol. 24, no. 11, pp. 4162–4171, Nov. 13, 2006.

[105] J. H. Lee, D. K. Jung, C. H. Kim, and Y. C. Chung, "OSNR monitoring technique using polarization-nulling method," *IEEE Photon. Technol. Lett.*, vol. 13, no. 1, pp. 88–90, Jan. 2001.

[106] M. Brodsky and M. D. Feuer, "Interferometric method and apparatus for measuring optical signal quality in optical communications system," U.S. Patent 7 499 173, Mar. 2009.

[107] X. Liu *et al.*, "OSNR monitoring method for OOK and DPSK based on optical delay interferometer," *IEEE Photon. Technol. Lett.*, vol. 19, no. 15, pp. 1172–1174, Aug. 1, 2007.

[108] E. Flood *et al.*, "In-band OSNR monitoring using a pair of Michelson fiber interferometers," *Opt. Exp.*, vol. 18, no. 4, pp. 3618–3625, Feb. 2010.

[109] A. Annoni and F. Morichetti, "Enhancing the sensitivity of interferometer based in-band OSNR monitoring by narrow band filtering," *J. Lightw. Technol.*, vol. 31, no. 9, pp. 1447–1453, May 7, 2013.

[110] J. M. Oh, M. Brodsky, L. E. Nelson, G. Cadena, and M. D. Feuer, "Interferometric optical signal-to-noise ratio measurements of telecom signals with degraded extinction ratio," *Opt. Lett.*, vol. 33, no. 18, pp. 2065–2067, Sep. 2008.

[111] E. Flood *et al.*, "Interferometer based in-band OSNR monitoring of single and dual polarisation QPSK signals," in *Proc. 36th Eur. Conf. Exhibit. Opt. Commun.*, 2010, pp. 1–3.

[112] J. Qiu, Z. Huang, B. Yuan, N. An, D. Kong, and J. Wu, "Multi-wavelength in-band OSNR monitor based on Lyot–Sagnac interferometer," *Opt. Exp.*, vol. 23, no. 16, pp. 20257–20266, Aug. 2015.

[113] Z. Tao, Z. Chen, L. Fu, D. Wu, and A. Xu, "Monitoring of OSNR by using a Mach–Zehnder interferometer," *Microw. Opt. Technol. Lett.*, vol. 30, no. 1, pp. 63–65, 2001.

[114] F. Khan, A. P. T. Lau, Z. Li, C. Lu, and P. Wai, "OSNR monitoring for RZ-DQPSK systems using half-symbol delay-tap sampling technique," *IEEE Photon. Technol. Lett.*, vol. 22, no. 11, pp. 823–825, Jun. 5, 2010.

[115] F. N. Khan, A. P. Lau, C. Lu, and P. K. Wai, "Wide dynamic range OSNR monitoring for RZ-DQPSK systems using delay-tap sampling technique," in *Proc. Opt. Fiber Commun. Conf.*, San Diego, CA, USA, 2010, pp. 1–3.

[116] S. K. Shin, K. J. Park, and Y. C. Chung, "A novel optical signal-to-noise ratio monitoring technique for WDM networks," in *Proc. Opt. Fiber Commun. Conf.*, Baltimore, MD, USA, 2000, pp. 182–184.

[117] W. Chen, R. S. Tucker, X. Yi, W. Shieh, and J. S. Evans, "Optical signal-to-noise ratio monitoring using uncorrelated beat noise," *IEEE Photon. Technol. Lett.*, vol. 17, no. 11, pp. 2484–2486, Nov. 2005.

[118] M. Bakaul, "Low-cost PMD-insensitive and dispersion tolerant in-band OSNR monitor based on uncorrelated beat noise measurement," *IEEE Photon. Technol. Lett.*, vol. 20, no. 11, pp. 906–908, Jun. 1, 2008.

[119] 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.*, San Francisco, CA, USA, 2014, pp. 1–3.

[120] M. S. Faruk, Y. Mori, and K. Kikuchi, "In-band estimation of optical signal-to-noise ratio from equalized signals in digital coherent receivers," *IEEE Photon. J.*, vol. 6, no. 1, pp. 1–9, Feb. 2014.

[121] C. Zhu *et al.*, "Statistical moments-based OSNR monitoring for coherent optical systems," *Opt. Exp.*, vol. 20, no. 16, pp. 17711–17721, Jul. 2012.

[122] X. Lin, O. A. Dobre, T. M. N. 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, Aug. 1, 2019.

[123] X. Lin, O. A. Dobre, O. Omomukuyo, Y. Eldemerdash, and C. Li, "OSNR estimation algorithm for higher-order modulation formats in coherent optical systems," in *Proc. Asia Commun. Photon. Conf. (ACP)*, Guangzhou, China, 2017, pp. 1–3.

[124] D. Zhao, L. Xi, X. Tang, W. Zhang, Y. Qiao, and X. Zhang, "Periodic training sequence aided in-band OSNR monitoring in digital coherent receiver," *IEEE Photon. J.*, vol. 6, no. 4, pp. 1–8, Aug. 2014.

[125] M. Tomizawa, Y. Yamabayashi, I. Sato, and T. Kataoka, "Nonlinear influence on PM-AM conversion measurement of group velocity dispersion in optical fibres," *Electron. Lett.*, vol. 30, no. 17, pp. 1434–1435, Aug. 1994.

[126] Q. Yu, Z. Pan, L-S. Yan, and A. E. Willner, "Chromatic dispersion monitoring technique using sideband optical filtering and clock phase-shift detection," *J. Lightw. Technol.*, vol. 20, no. 12, pp. 2267–2271, Dec. 2002.

[127] Y. Takushima, H. Yoshimi, Y. Ozeki, K. Kikuchi, H. Yamauchi, and H. Taga, "In-service dispersion monitoring in 32×10.7 Gbps WDM transmission system over transatlantic distance using optical frequency-modulation method," *J. Lightw. Technol.*, vol. 22, no. 1, pp. 257–265, Jan. 2004.

[128] F. Buchali, S. Lanne, J. Thiery, W. Baumert, and H. Bulow, "Fast eye monitor for 10 Gbit/s and its application for optical PMD compensation," in *Proc. Opt. Fiber Commun. Conf. Exhibit. (OFC)*, Anaheim, CA, USA, 2001, pp. 1–3.

[129] M. Bohn and W. Rosenkranz, "Experimental verification of combined adaptive PMD and GVD compensation in a 40 Gb/s transmission using integrated optical FIR-filters and spectrum monitoring," in *Proc. Opt. Fiber Commun. Conf. Exhibit. (OFC)*, vol. 1. Los Angeles, CA, USA, 2004, pp. 1–3.

[130] J. Liu, Z. Dong, K. Zhong, A. P. T. Lau, C. Lu, and Y. Lu, "Modulation format identification based on received signal power distributions for digital coherent receivers," in *Proc. Opt. Fiber Commun. Conf. (OFC)*, San Francisco, CA, USA, 2014, pp. 1–3.

[131] B. Hui, X. Tang, N. Gao, W. Zhang, and X. Zhang, "High order modulation format identification based on compressed sensing in optical fiber communication system," *Chin. Opt. Lett.*, vol. 14, no. 11, pp. 1–5, Nov. 2016.

[132] Z. Zhao, A. Yang, P. Guo, and W. Tang, "A modulation format identification method based on amplitude deviation analysis of received optical communication signal," *IEEE Photon. J.*, vol. 11, no. 1, pp. 1–7, Feb. 2019.

[133] Q. Tan, A. Yang, and P. Guo, "Blind modulation format identification using differential phase and amplitude ratio," *IEEE Photon. J.*, vol. 11, no. 1, pp. 1–12, Feb. 2019.

[134] Z. Zhao, A. Yang, and P. Guo, "A modulation format identification method based on information entropy analysis of received optical communication signal," *IEEE Access*, vol. 7, pp. 41492–41497, 2019.

[135] Q. Tan, A. Yang, and P. Guo, "Blind modulation format identification using the DC component," *IEEE Photon. J.*, vol. 11, no. 2, pp. 1–10, Apr. 2019.

[136] X. Lin, Y. A. Eldemerdash, O. A. Dobre, S. Zhang, and C. Li, "Modulation classification using received signal's amplitude distribution for coherent receivers," *IEEE Photon. Technol. Lett.*, vol. 29, no. 21, pp. 1872–1875, Sep. 20, 2017.

[137] G. Liu, R. Proietti, K. Zhang, H. Lu, and S. J. B. Yoo, "Blind modulation format identification using nonlinear power transformation," *Opt. Exp.*, vol. 25, no. 25, pp. 30895–30904, Dec. 2017.

[138] A. Yi, H. Liu, L. Yan, L. Jiang, Y. Pan, and B. Luo, "Amplitude variance and 4th power transformation based modulation format identification for digital coherent receiver," *Opt. Commun.*, vol. 452, pp. 109–115, Dec. 2019.

[139] 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. (ECOC)*, 2016, pp. 1–3.

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

[141] R. A. Skoog *et al.*, "Automatic identification of impairments using support vector machine pattern classification on eye diagrams," *IEEE Photon. Technol. Lett.*, vol. 18, no. 22, pp. 2398–2400, Nov. 13, 2006.

[142] A. Khotanzad and Y. H. Hong, "Invariant image recognition by Zernike moments," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 5, pp. 489–497, May 1990.

[143] J. A. Jargon, X. Wu, and A. E. Willner, "Optical performance monitoring using artificial neural networks trained with eye-diagram parameters," *IEEE Photon. Technol. Lett.*, vol. 21, no. 1, pp. 54–56, Jan. 1, 2009.

[144] X. Wu, J. A. Jargon, R. A. Skoog, L. Paraschis, and A. E. Willner, "Applications of artificial neural networks in optical performance monitoring," *J. Lightw. Technol.*, vol. 27, no. 16, pp. 3580–3589, Aug. 15, 2009.

[145] J. Lai, A. Yang, and Y. Sun, "Multiple-impairment monitoring for 40-Gbps RZ-OOK using artificial neural networks trained with reconstructed eye diagram parameters," in *Proc. Int. Quant. Electron. Conf. (IQEC) Lasers Electro Opt. (CLEO)*, 2011, pp. 563–565.

[146] J.-S. Lai, A.-Y. Yang, and Y.-N. Sun, *Application of CHIRP-Z Transform to the Optical-Sampling Optical Performance Monitoring*, Trans. Beijing Inst. Technol., Beijing, China, 2011.

[147] Y. Huang, Y. Chen, and J. Yu, "Optical performance monitoring of 56Gbps optical PAM4 signal using artificial neural networks," in *Proc. Asia Commun. Photon. Conf. (ACP)*, Guangzhou, China, 2017, pp. 1–3.

[148] 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. 27, 2010.

[149] F. N. Khan, T. S. R. Shen, Y. Zhou, A. P. T. Lau, and C. Lu, "Optical performance monitoring using artificial neural networks trained with empirical moments of asynchronously sampled signal amplitudes," *IEEE Photon. Technol. Lett.*, vol. 24, no. 12, pp. 982–984, Jun. 13, 2012.

[150] J. A. Jargon, X. Wu, and A. E. Willner, "Optical performance monitoring by use of artificial neural networks trained with parameters derived from delay-tap asynchronous sampling," in *Proc. Conf. Opt. Fiber Commun. (OFC)*, San Diego, CA, USA, 2009, pp. 1–3.

[151] X. Wu, J. A. Jargon, L. Paraschis, and A. E. Willner, "ANN-based optical performance monitoring of QPSK signals using parameters derived from balanced-detected asynchronous diagrams," *IEEE Photon. Technol. Lett.*, vol. 23, no. 4, pp. 248–250, Feb. 10, 2011.

[152] T. Mrozek, "Simultaneous monitoring of chromatic dispersion and optical signal to noise ratio in optical network using asynchronous delay tap sampling and convolutional neural network (deep learning)," in *Proc. 20th Int. Conf. Transp. Opt. Netw. (ICTON)*, 2018, pp. 1–4.

[153] V. Ribeiro, L. Costa, M. Lima, and A. L. J. Teixeira, "Optical performance monitoring using the novel parametric asynchronous eye diagram," *Opt. Exp.*, vol. 20, no. 9, pp. 9851–9861, Apr. 2012.

[154] T. S. R. Shen, Q. Sui, and A. P. T. Lau, "OSNR monitoring for PM-QPSK systems with large inline chromatic dispersion using artificial neural network technique," *IEEE Photon. Technol. Lett.*, vol. 24, no. 17, pp. 1564–1567, Sep. 18, 2012.

[155] J. A. Jargon, X. Wu, H. Y. Choi, Y. C. Chung, and A. E. Willner, "Optical performance monitoring of QPSK data channels by use of neural networks trained with parameters derived from asynchronous constellation diagrams," *Opt. Exp.*, vol. 18, no. 5, pp. 4931–4938, Mar. 2010.

[156] 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.

[157] Z. Wang, A. Yang, P. Guo, and P. He, "OSNR and nonlinear noise power estimation for optical fiber communication systems using LSTM based deep learning technique," *Opt. Exp.*, vol. 26, no. 16, pp. 21346–21357, Aug. 2018.

[158] L. Xia, J. Zhang, S. Hu, M. Zhu, Y. Song, and K. Qiu, "Transfer learning assisted deep neural network for OSNR estimation," *Opt. Exp.*, vol. 27, no. 14, pp. 19398–19406, Jul. 2019.

[159] 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. Exp.*, vol. 20, no. 11, pp. 12422–12431, May 2012.

[160] S. Zhang, Y. Peng, Q. Sui, and Z. Li, "Modulation format identification using sparse asynchronous amplitude histograms," in *Proc. Adv. Photon.*, Boston, MA, USA, 2015, pp. 1–3.

[161] X. Sun, S. Su, Z. Huang, Z. Zuo, X. Guo, and J. Wei, "Blind modulation format identification using decision tree twin support vector machine in optical communication system," *Opt. Commun.*, vol. 438, pp. 67–77, May 2019.

[162] S. Zhang, Y. Peng, Q. Sui, J. Li, and Z. Li, "Modulation format identification in heterogeneous fiber-optic networks using artificial neural networks and genetic algorithms," *Photon. Netw. Commun.*, vol. 32, no. 2, pp. 246–252, Oct. 2016.

[163] K. Roberts and C. Laperle, "Flexible transceivers," in *Proc. Eur. Conf. Exhibit. Opt. Commun.*, 2012, pp. 1–3.

[164] D. A. Mello *et al.*, "Optical networking with variable-code-rate transceivers," *J. Lightw. Technol.*, vol. 32, no. 2, pp. 257–266, Jan. 15, 2013.

[165] P. Isautier, K. Mehta, A. J. Stark, and S. E. Ralph, "Robust architecture for autonomous coherent optical receivers," *J. Opt. Commun. Netw.*, vol. 7, no. 9, pp. 864–874, Sep. 2015.

[166] R. Borkowski, D. Zibar, A. Caballero, V. A. Unno, and I. T. Monroy, "Optical modulation format recognition in stokes space for digital coherent receivers," in *Proc. Opt. Fiber Commun. Conf. Expo. Nat. Fiber Opt. Eng. (OFC/NFOEC)*, Anaheim, CA, USA, 2013, pp. 1–3.

[167] R. Borkowski, D. Zibar, A. Caballero, V. Arlunno, and I. T. Monroy, "Stokes space-based optical modulation format recognition for digital coherent receivers," *IEEE Photon. Technol. Lett.*, vol. 25, no. 21, pp. 2129–2132, Nov. 1, 2013.

[168] P. Isautier, J. Pan, and S. E. Ralph, "Robust autonomous software-defined coherent optical receiver," in *Proc. Opt. Fiber Commun. Conf. (OFC)*, San Francisco, CA, USA, 2014, pp. 1–3.

[169] P. Isautier, J. Pan, R. DeSalvo, and S. E. Ralph, "Stokes space-based modulation format recognition for autonomous optical receivers," *J. Lightw. Technol.*, vol. 33, no. 24, pp. 5157–5163, Dec. 15, 2015.

[170] R. Boada, R. Borkowski, and I. T. Monroy, "Clustering algorithms for stokes space modulation format recognition," *Opt. Exp.*, vol. 23, no. 12, pp. 15521–15531, Jun. 2015.

[171] T. Bo, J. Tang, and C. C. Chan, "Blind modulation format recognition for software-defined optical networks using image processing techniques," in *Proc. Opt. Fiber Commun. Conf. Exhibit. (OFC)*, Anaheim, CA, USA, 2016, pp. 1–3.

[172] T. Bo, J. Tang, and C. Chan, "Modulation format recognition for optical signals using connected component analysis," *IEEE Photon. Technol. Lett.*, vol. 29, no. 1, pp. 11–14, Jan. 1, 2017.

[173] S. Fortune, "A sweepline algorithm for Voronoi diagrams," *Algorithmica*, vol. 2, no. 1, pp. 153–157, Nov. 1987.

[174] A. AbuBaker, R. Qahwaji, S. Ipson, and M. Saleh, "One scan connected component labeling technique," in *Proc. IEEE Int. Conf. Signal Process. Commun.*, 2007, pp. 1283–1286.

[175] X. Mai *et al.*, "Stokes space modulation format classification based on non-iterative clustering algorithm for coherent optical receivers," *Opt. Exp.*, vol. 25, no. 3, pp. 2038–2050, Feb. 2017.

[176] S. L. Chiu, "Fuzzy model identification based on cluster estimation," *J. Intell. Fuzzy Syst.*, vol. 2, no. 3, pp. 267–278, May 1994.

[177] P. Chen, J. Liu, X. Wu, K. Zhong, and X. Mai, "Subtraction-clustering-based modulation format identification in stokes space," *IEEE Photon. Technol. Lett.*, vol. 29, no. 17, pp. 1439–1442, Sep. 1, 2017.

[178] L. Jiang *et al.*, "Blind density-peak-based modulation format identification for elastic optical networks," *J. Lightw. Technol.*, vol. 36, no. 14, pp. 2850–2858, Jul. 15, 2018.

[179] W. Zhang *et al.*, "Identifying modulation formats through 2D stokes planes with deep neural networks," *Opt. Exp.*, vol. 26, no. 18, pp. 23507–23517, Sep. 2018.

[180] L. Jiang *et al.*, "Robust and blind modulation format identification for elastic optical networks," in *Proc. Eur. Conf. Opt. Commun. (ECOC)*, 2018, pp. 1–3.

[181] 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. 1, 2019.

[182] C.-T. Shi, "Signal pattern recognition based on fractal features and machine learning," *Appl. Sci.*, vol. 8, no. 8, pp. 1–15, 2018.

[183] B. Mandelbrot, "How long is the coast of britain? Statistical self-similarity and fractional dimension," *Science*, vol. 156, no. 3775, pp. 636–638, 1967.

[184] H. Zhou *et al.*, "Fractal dimension aided modulation formats identification based on support vector machines," in *Proc. Eur. Conf. Opt. Commun. (ECOC)*, 2017, pp. 1–3.

[185] D. Godard, "Self-recovering equalization and carrier tracking in two-dimensional data communication systems," *IEEE Trans. Commun.*, vol. C-28, no. 11, pp. 1867–1875, Nov. 1980.

[186] L. Jiang *et al.*, "Chromatic dispersion, nonlinear parameter, and modulation format monitoring based on Godard's error for coherent optical transmission systems," *IEEE Photon. J.*, vol. 10, no. 1, pp. 1–12, Feb. 2018.

[187] J. Zhang *et al.*, "Blind and noise-tolerant modulation format identification," *IEEE Photon. Technol. Lett.*, vol. 30, no. 21, pp. 1850–1853, Nov. 1, 2018.

[188] J. Zhang *et al.*, "Intelligent adaptive coherent optical receiver based on convolutional neural network and clustering algorithm," *Opt. Exp.*, vol. 26, no. 14, pp. 18684–18698, Jul. 2018.

[189] 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, Feb. 1, 2020.

[190] J. Thrane, J. Wass, M. Piels, J. C. M. 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, Feb. 15, 2017.

[191] M. C. Tan, F. N. Khan, W. H. Al-Arashi, Y. Zhou, and A. P. T. Lau, "Simultaneous optical performance monitoring and modulation format/bit-rate identification using principal component analysis," *IEEE/OSA J. Opt. Commun. Netw.*, vol. 6, no. 5, pp. 441–448, May 2014.

[192] F. N. Khan *et al.*, "Experimental demonstration of joint OSNR monitoring and modulation format identification using asynchronous single channel sampling," *Opt. Exp.*, vol. 23, no. 23, pp. 30337–30346, Nov. 2015.

[193] D. Wang *et al.*, "Modulation format recognition and OSNR estimation using CNN-based deep learning," *IEEE Photon. Technol. Lett.*, vol. 29, no. 19, pp. 1667–1670, Oct. 1, 2017.

[194] 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. Exp.*, vol. 27, no. 8, pp. 11281–11291, Apr. 2019.

[195] D. Wang *et al.*, "Cost-effective and data size-adaptive OPM at intermediated node using convolutional neural network-based image processor," *Opt. Exp.*, vol. 27, no. 7, pp. 9403–9419, Apr. 2019.

[196] 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, Jun. 1, 2018.

[197] X. Lin, O. A. Dobre, T. M. N. Ngatched, Y. A. Eldemerdash, and C. Li, "Joint modulation classification and OSNR estimation enabled by support vector machine," *IEEE Photon. Technol. Lett.*, vol. 30, no. 24, pp. 2127–2130, Dec. 15, 2018.

[198] A. Yi *et al.*, "Modulation format identification and OSNR monitoring using density distributions in stokes axes for digital coherent receivers," *Opt. Exp.*, vol. 27, no. 4, pp. 4471–4479, Feb. 2019.

[199] D. Wang *et al.*, "Intelligent constellation diagram analyzer using convolutional neural network-based deep learning," *Opt. Exp.*, vol. 25, no. 15, pp. 17150–17166, Jul. 2017.

[200] Q. W. Zhang *et al.*, "Artificial neural network based modulation identification for elastic optical networks," in *Proc. 23rd Opto Electron. Commun. Conf. (OECC)*, 2018, pp. 1–2.

[201] Y. Ma *et al.*, "Modulation format identification based on constellation diagrams in adaptive optical OFDM systems," *Opt. Commun.*, vol. 452, pp. 203–210, May 2019.

[202] L. Zhao, H. Xu, S. Bai, and C. Bai, "Modulus mean square-based blind hybrid modulation format recognition for orthogonal frequency division multiplexing-based elastic optical networking," *Opt. Commun.*, vol. 445, pp. 284–290, Aug. 2019.

[203] Q. Zhang *et al.*, "A simple joint modulation format identification and OSNR monitoring scheme for IMDD OOFDM transceivers using k-nearest neighbor algorithm," *Appl. Sci.*, vol. 9, no. 18, pp. 1–12, 2019.

[204] Y. Ma *et al.*, "Modulation format identification for adaptive optical OFDM system," in *Proc. 24th OptoElectron. Commun. Conf. (OECC) Int. Conf. Photon. Switch. Comput. (PSC)*, 2019, pp. 1–3.

[205] 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. 156207–156216, 2019.

[206] M. Chen, U. Challita, W. Saad, C. Yin, and M. Debbah, "Artificial neural networks-based machine learning for wireless networks: A tutorial," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 4, pp. 3039–3071, 4th Quart., 2019.

[207] C. Zhang, P. Patras, and H. Haddadi, "Deep learning in mobile and wireless networking: A survey," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 3, pp. 2224–2287, 3rd Quart., 2019.

[208] M. Mozaffari, W. Saad, M. Bennis, Y. Nam, and M. Debbah, "A tutorial on UAVs for wireless networks: Applications, challenges, and open problems," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 3, pp. 2334–2360, 3rd Quart., 2019.

[209] L. Guesmi and M. Menif, "Modulation formats recognition technique using artificial neural networks for radio over fiber systems," in *Proc. 17th Int. Conf. Transp. Opt. Netw. (ICTON)*, 2015, pp. 1–4.

[210] M. Alharbi, A. Alhuseini, A. Ragheb, M. Altamimi, T. Alshawi, and S. Alshebeili, "Automatic modulation classification: Investigation for millimeter wave over fiber channels," *IEEE Photon. Technol. Lett.*, vol. 31, no. 13, pp. 1092–1095, Jul. 1, 2019.

[211] A. Trichili, K. Park, M. Zghal, B. S. Ooi, and M.-S. Alouini, "Communicating using spatial mode multiplexing: Potentials, challenges, and perspectives," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 4, pp. 3175–3203, 4th Quart., 2019.
[212] J. Li, M. Zhang, D. Wang, S. Wu, and Y. Zhan, "Joint atmospheric turbulence detection and adaptive demodulation technique using the CNN for the OAM-FSO communication," *Opt. Exp.*, vol. 26, no. 8, pp. 10494–10508, Apr. 2018.
[213] S. Lohani and R. T. Glasser, "Turbulence correction with artificial neural networks," *Opt. Lett.*, vol. 43, no. 11, pp. 2611–2614, Jun. 2018.
[214] J. D. Reis, V. Shukla, D. Stauffer, and K. Gass, "Technology options for 400G implementation," Opt. Netw. Forum, Fremont, CA, USA, White Paper, 2015.
[215] Z. Qu and I. B. Djordjevic, "Optimal constellation shaping in optical communication systems," in *Proc. 20th Int. Conf. Transp. Opt. Netw. (ICTON)*, 2018, pp. 1–5.
[216] Z. Qu and I. B. Djordjevic, "Geometrically shaped 16QAM outperforming probabilistically shaped 16QAM," in *Proc. Eur. Conf. Opt. Commun. (ECOC)*, 2017, pp. 1–3.
[217] J. Cho and P. J. Winzer, "Probabilistic constellation shaping for optical fiber communications," *J. Lightw. Technol.*, vol. 37, no. 6, pp. 1590–1607, Mar. 15, 2019.
[218] W. Zhang *et al.*, "Identifying probabilistically shaped modulation formats through 2D stokes planes with two-stage deep neural networks," *IEEE Access*, vol. 8, pp. 6742–6750, 2020.
[219] Y. Sasaki, K. Takenaga, S. Matsuo, K. Aikawa, and K. Saitoh, "Few-mode multicore fibers for long-haul transmission line," *Opt. Fiber Technol.*, vol. 35, pp. 19–27, Oct. 2017.

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) in 2015, where he is currently pursuing the Ph.D. degree with the Electrical Engineering Department. 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.

Maged A. Esmail received the B.E. degree (Hons.) in electronic engineering from IBB University in 2006, and the M.Sc. and Ph.D. degrees (First Hons.) in electrical engineering from King Saud University (KSU) in 2012 and 2017, respectively. Since August 2018, he has been an Assistant Professor with Prince Sultan University. From April 2017 to June 2018, he was a Postdoctoral Researcher with KACST Technology Innovation Center in Radio Frequency and Photonics for the e-Society, KSU. He has over five years of experience with Advanced Optical Communication Laboratory. He has authored/coauthored over 40 journal and conference publications. His current research interests include fiber-optic communications, PON and long-reach PON, network management and protection, free space optics, microwave photonics, visible light communications, and machine learning. He is a recipient of KSU Scientific Excellence Award in 2011 and 2013. Also, he is an Active Reviewer for various IEEE, OSA, and other Journals. He serves as an Associate Editor for IEEE ACCESS journal.

Amr M. Ragheb received the B.S. (Hons.) and M.Sc. degrees in electrical engineering from Tanta University, Egypt, in 2001 and 2007, respectively, and the Ph.D. degree in electrical engineering from King Saud University, Riyadh, Saudi Arabia, in 2015. He was a Teaching Assistant (TA) with Tanta University, from 2003 to 2008. He was a TA with King Saud University, from 2010 to 2015. He has over seven years of experience with 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, multiformat highspeed optical transmitter, and passive optical networks.

Tariq A. Alshawi received the M.S. degree in electrical engineering from the University of Michigan at Ann Arbor, Ann Arbor, and the Ph.D. degree from the Georgia Institute of Technology, Atlanta. He is currently an Assistant Professor with the Electrical Engineering Department, King Saud University, Riyadh. His research interests include signal and image processing and human perception models, with applications in computer vision, bio-medical imaging, communication, and networks.

Saleh A. Alshebeili was the Chairman of 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, and 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.