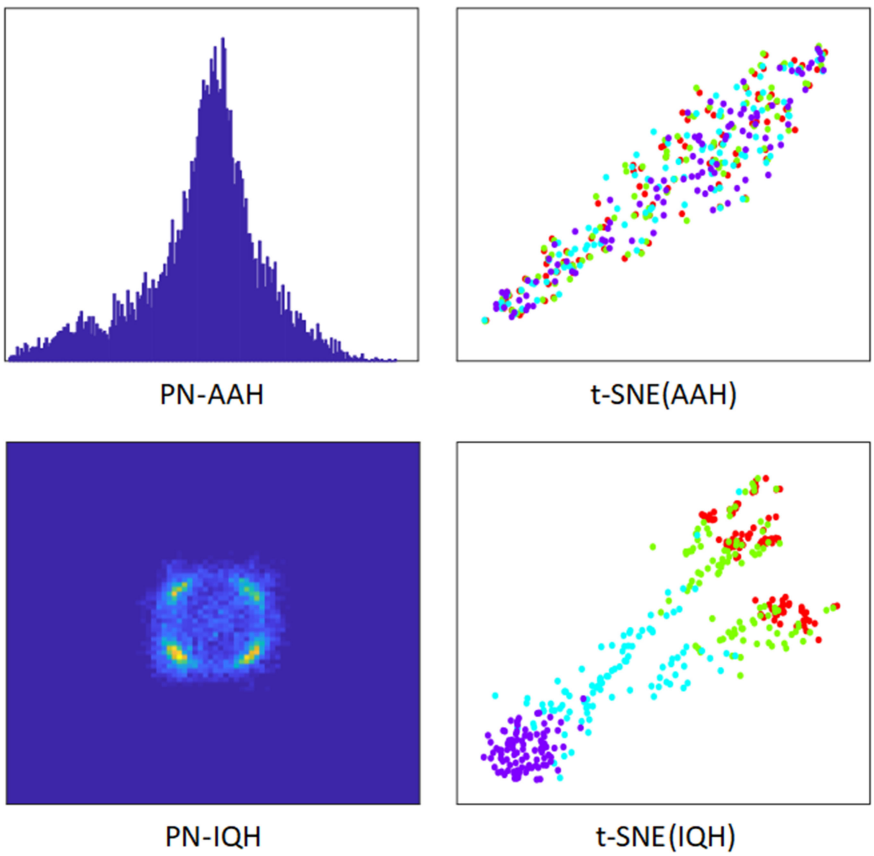
IEEE Photonics Journal
An IEEE Photonics Society Publication



# Separability of Histogram Based Features for Optical Performance Monitoring: An Investigation Using t-SNE Technique

Volume 11, Number 3, June 2019

Waddah S. Saif
Tariq Alshawi
Maged Abdullah Esmail
Amr Ragheb
Saleh Alshebeili




DOI: 10.1109/JPHOT.2019.2913687
1943-0655 © 2019 IEEE

# Separability of Histogram Based Features for Optical Performance Monitoring: An Investigation Using t-SNE Technique

**Waddah S. Saif**,[1,2] **Tariq Alshawi**,[1,2] **Maged Abdullah Esmail**,[3] **Amr Ragheb**,[2] **and Saleh Alshebeili**[1,2]

[1]Department of Electrical Engineering, King Saud University, Riyadh 11421, Saudi Arabia
[2]KACST-TIC in Radio Frequency and Photonics for the e-Society, King Saud University, Riyadh 11451, Saudi Arabia
[3]Communications and Networks Department, Prince Sultan University, Riyadh 12435, Saudi Arabia

*DOI:10.1109/JPHOT.2019.2913687*


Manuscript received March 21, 2019; revised April 22, 2019; accepted April 24, 2019. Date of publication April 29, 2019; date of current version May 8, 2019. This work was supported by Deanship of Scientific Research, King Saud University, under the Research Group Program under Contract RG-1438-092.
Corresponding author: Waddah S. Saif (e-mail: wsaif@ksu.edu.sa).

**Abstract:** In this paper, we study the separability of the commonly used features to monitor the performance of optical signal in coherent optical systems. Specifically, our study focuses on the histogram-based features; the asynchronous amplitude histogram (AAH) and the two-dimensional extension of AAH, which we call IQ histogram (IQH). We investigate the conditions under which the optical channel impairments can be monitored. This study utilizes a dimensionality reduction technique, known as the t-distribution stochastic neighbor embedding (t-SNE). Using t-SNE, we show that under certain conditions the histogram-based features cannot be used to distinguish between the high and low impairment effects, rendering these features useless for certain cases, regardless of the type or complexity of the implemented impairments' estimator/classifier. Extensive simulations results have been conducted to investigate single and multiple impairments conditions and the performance of histogram-based features in each case. The results show that both AAH and IQH can be used for monitoring all types of single Impairment except phase noise in case of AAH. Moreover, under multiple impairments conditions, polarization mode dispersion values can be monitored to some extent, while it is difficult to monitor optical signal-to-noise ratio and chromatic dispersion values especially in the case of concurrent presence of more than two impairments. The results of these investigations are validated by providing a quantitative measure that ties their usefulness to the actual monitoring performance through the estimation of channel impairments under different conditions using linear support vector machine (SVM) regression. It has been found that there is a match to a great extent between the separability investigation using t-SNE and estimation results obtained using SVM.

**Index Terms:** Optical performance monitoring (OPM), histogram-based features, dimensionality reduction, optical coherent receivers.

## 1. Introduction

Nowadays, the optical communication networks are rapidly expanding in its coverage and technology in order to introduce ultra-high data rates and a-much-needed abundance in bandwidth to support future communications. To enable such advancement, efficient optical modulations are

employed to deliver the data rates required by future networks. Additionally, the network topology of future optical networks will comprise elastic architectures that rely on reconfigurable components such as the reconfigurable optical add-drop multiplexer (ROADM). This will improve the efficiency, flexibility, and fault-tolerance of the optical networks [1]. Nonetheless, these enhancements will force the optical channel monitoring techniques to be adaptive in order to cope with the changes in the optical channel and/or the signal impairments [2]–[4].

During the last few years, many optical performance monitoring (OPM) techniques have been proposed. These include asynchronous amplitude histogram (AAH) [5]–[9], which is cost-effective and does not require high speed sampler. However, multiple impairments with similar AAH profiles make the distinction between different impairments difficult. Another approach for OPM is based on the asynchronous delay-tap sampling (ADTS), which relies on the fact that the amplitude of a signal could be sampled using two low speed clocks with a constant delay between them [10]–[13]. Although, this algorithm provides more details about the statistical properties of the signal than AAH, it has ineffective cost in addition to its sensitivity to the delay between the sampling clocks [14]. Besides, OPM techniques based on extracting some features from the signal eye diagram such as Q factor, closure, and jitter, have also been proposed. However, the eye diagram itself is based on synchronous high speed sampling, which is seriously affected by the chromatic dispersion (CD) and polarization mode dispersion (PMD). Therefore, it is not effective for high data rate transmission [15].

So far, the aforementioned techniques are based on studying the statistical features of the signal intensity that are directly extracted from the detected signals using direct detection receiver. However, such systems ignore the phase information, which can be exploited to monitor the distortion in advanced modulation signals. Therefore, coherent optical receivers are used to detect both the intensity and phase of the optical signals. Additionally, these receivers provide the possibility of taking advantage of digital signal processing (DSP) algorithms to compensate for linear impairments such as CD and PMD. Most of the proposed research, in coherent detection systems, has focused on the optical signal to noise ratio (OSNR) monitoring owing to the direct relation with bit-error ratio (BER) metric. In [16], [17], the authors addressed the joint OSNR monitoring and modulation format identification using amplitude histogram [16] and cumulative distribution function (CDF) [17]. In [18], an OPM technique based on statistical parameters derived from constellation diagram was proposed. Its performance is evaluated in the presence of a small amount of differential group delay (DGD), CD, and high OSNR values. In [19], some statistical features were extracted from AAH for OSNR, CD, and DGD monitoring. However, the impact of simultaneous co-existence of multiple impairments was not taken into consideration.

During last years, machine learning techniques were extensively implemented and demonstrated in optical communication networks. For instance, they have been used for optical carrier recovery, fiber nonlinearities mitigation, signal phase tracking, and OPM [20]–[22]. The major advantage of using machine learning techniques is its ability to learn and improve their computational models by handling high volumes of available data.

Note that the majority of research related to OPM focuses on the use of various techniques to monitor and estimate optical channel impairments, before studying the feasibility of performing such techniques to the problem at hand. However, there are fundamental questions which are typically overlooked and need to be addressed before proceeding to the OPM development phase. Examples of such questions include: Under what conditions are the optical channel impairments can be monitored? Can we monitor a channel impairment accurately while other impairments exist? Is the representation of channel impairments equivalent across various signal features? Or, is it the case that some signal features are better in providing information regarding the estimation of channel impairments than others? Therefore, this work is to address these important and fundamental questions to help researchers gain insight before proposing new OPM techniques.

Note that all the previously stated questions revolve around one central idea; the representation of various channel impairments in optical signal features. Therefore, we use a data-driven approach to examine the representation of channel impairments in signal features. In this paper, we use a state-of-the-art technique, called t-distributed stochastic neighbor embedding (t-SNE), which provides a

nonlinear dimensionality reduction that can be used to visualize complex high-dimensional signal pattern, as in image processing, computer security research, medical applications, bioinformatics, etc. [23]–[26]. In our work, we consider the data extracted from the AAH algorithm, due to its wide spread usage in literature, and examine it under the effect of four important optical impairments: OSNR, CD, PMD and PN. Moreover, we propose another feature that combines both the in-phase (I) and quadrature (Q) information on two dimensional histogram, which will be denoted by IQH. Based on this work, we discuss the conditions under which AAH and IQH can provide sufficient information to monitor the channel impairments. Simulation results, for example, show that the IQH is insensitive to the optical phase noise compared to the conventional AAH technique, which only relies on the amplitude information. A clear measure that ties the usefulness of presented investigation to the actual monitoring performance is considered.

The paper is organized as follows. In Section 2, we provide a short description about the dimensionality reduction techniques, then, we focus on the formulation of the t-SNE algorithm. In Section 3, we discuss the effects of impairments on the histogram analysis. The simulation model is described in Section 4. In Section 5, we present and discuss the simulation and validation results. Finally, we present conclusions in Section 6.

## 2. Dimensionality Reduction Techniques

In general, the dimensionality reduction methods transform high dimensional data $\mathbf{X}$ of size (D $\times$ N) into a low dimensional space $\mathbf{Y}$ of size (d $\times$ N), where D $\gg$ d and

$$\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots \mathbf{x}_N\} \subset \mathrm{R}^{\mathrm{D}}. \tag{1}$$

$$\mathbf{Y} = \{\mathbf{y}_1, \mathbf{y}_2, \ldots \mathbf{y}_N\} \subset \mathrm{R}^{\mathrm{d}}. \tag{2}$$

The entries $\mathbf{x}_i$ and $\mathbf{y}_i$(i $= 1, 2, \ldots .$N) are vectors of size (D $\times$ 1), and (d $\times$ 1), respectively. For this transformation to be useful, it must preserve the overall structure and properties of the original data $\mathbf{X}$. There are many research problems conducted in this field [27]. Some of these researches are concerned with preserving the global structure of data (i.e., the distance between the data in high dimension and low dimension), while others are concerned with preserving the local structure of data (i.e., continuity and topology). Recently, literature has shown research that performs dimensionality reduction while preserving both the global and local structures [23], [28].

Many algorithms that rely on linear transformation have been developed in literature. This includes, principle component analysis (PCA), which finds the main components of the dataset that correspond to the direction of maximum variance, and the linear discriminant analysis (LDA), which works by minimizing the similarity and maximizing the separation between different classes. These techniques are easy to implement but they fail at the nonlinear data manifolds.

Other dimensionality reduction techniques have been developed to handle the nonlinear data while preserving the global structure properties, such as, the multidimensional scaling (MDS) and stochastic proximity embedding (SPE). MDS applies a nonlinear transformation that preserves pairwise distances, where it maintains the large distances at the expense of small distances. Quality of transformation is measured by the so-called stress function. It computes the error between the pairwise dataset in high and low dimensions. Similar to MDS, SPE preserves pairwise distances but it uses an iterative process to minimize the stress function. In many applications, preserving the local structure is more important than the global structure, and in some cases, relaxing global structure requirement can lead to better performance. Locally Linear Embedding (LLE), and Laplacian Eigenmaps are two examples of the local-structure-preserving algorithms that have been proposed recently. LLE depends mainly on finding the nearest neighbors without the need to calculate the pairwise distance of widely separated points. Similarly, the Laplacian Eigenmaps maintain the local structure by finding the nearest neighbors, but it uses optimization process to minimize the distances between the data in low dimension and its nearest neighbors [29].

Recently, a nonlinear technique called t-SNE has been proposed with the aim to preserve both the general and local structures. It provides the state-of-the-art performance in many applications

such as computer research, biomedical applications, etc. The similarities of pair-wise Euclidean distances on high dimensional space are expressed as joint probabilities

$$p_{ij} = \frac{p_{i\backslash j} + p_{j\backslash i}}{2N}, \tag{3}$$

where $p_{i\backslash j}$, $i \neq j$ is the conditional probability, which is given as [23]

$$p_{j\backslash i} = \frac{\exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2/2\sigma_i^2)}{\sum_{k\neq i}\exp(-\|\mathbf{x}_i - \mathbf{x}_k\|^2/2\sigma_i^2)}, \tag{4}$$

where $\sigma_i$ is the variance of the Gaussian distribution centered on $\mathbf{x}_i$, and $p_{i\backslash i} = 0$. t-SNE performs a binary search for the value of $\sigma_i$ that produces a probability distribution $P_i$, for a fixed perplexity $Perp(P_i)$ that is specified by the user. Note that the perplexity determines the number of effective neighbor points. This tuning parameter is obtained empirically and is related to the size of the utilized data set [30] .The perplexity is defined as

$$Perp(P_i) = 2^{H(P_i)}, \tag{5}$$

where $H(P_i)$ is the Shannon entropy of $P_i$ measured in bits, which is given by

$$H(P_i) = -\sum_j p_{j\backslash i}\log_2 p_{j\backslash i}. \tag{6}$$

The initial values of lower dimension space $\mathbf{Y} = \{\mathbf{y}_1, \mathbf{y}_2, \ldots \mathbf{y}_N\}$ is selected randomly with Gaussian distribution. The similarities of pair-wise Euclidean distances on low dimensional space are expressed as

$$q_{ij} = \frac{\left(1 + \|y_i - y_j\|^2\right)^{-1}}{\sum_k \sum_{l\neq k}\left(1 + \|y_k - y_l\|^2\right)^{-1}}, \tag{7}$$

where $i \neq j$ and $q_{ii} = 0$. The matching between $p_{ij}$ and $q_{ij}$ is measured using Kullback-Leibler divergence, which is given by

$$C = \sum_i \sum_j p_{ij}\log\frac{p_{ij}}{q_{ij}}. \tag{8}$$

The cost function is minimized by the gradient descent method. The gradient has the following form

$$\frac{\delta C}{\delta \mathbf{y}_i} = 4\sum_j (p_{ij} - q_{ij})(\mathbf{y}_i - \mathbf{y}_j)\left(1 + \|\mathbf{y}_i - \mathbf{y}_j\|^2\right)^{-1}. \tag{9}$$

For each iteration, the lower dimension values are updated using the following formula

$$\mathbf{Y}^m = \mathbf{Y}^{(m-1)} + \eta\frac{\delta C}{\delta \mathbf{Y}} + \alpha(m)\left(\mathbf{Y}^{(m-1)} - \mathbf{Y}^{(m-2)}\right), \tag{10}$$

where $\eta$ is the learning rate and $\alpha(m)$ is the momentum at iteration m ( $= 1, 2, \ldots$ ), which is used to accelerate the optimization process. Figure 1 summarizes the t-SNE algorithm using the flow diagram.

## 3. The Effect of Channel Impairments on Detected Signal Histograms

The AAH is a commonly used feature for OPM, which is based on computing the histogram of the detected signal. In this section, we share insights regarding the effects of channel impairments on histogram shape and readability. Additionally, we show that some of these effects can be mitigated using the proposed IQH. Actually, IQH requires three dimensions (I, Q, and amplitude of each combination), so we use color map to represent the amplitude of IQH.

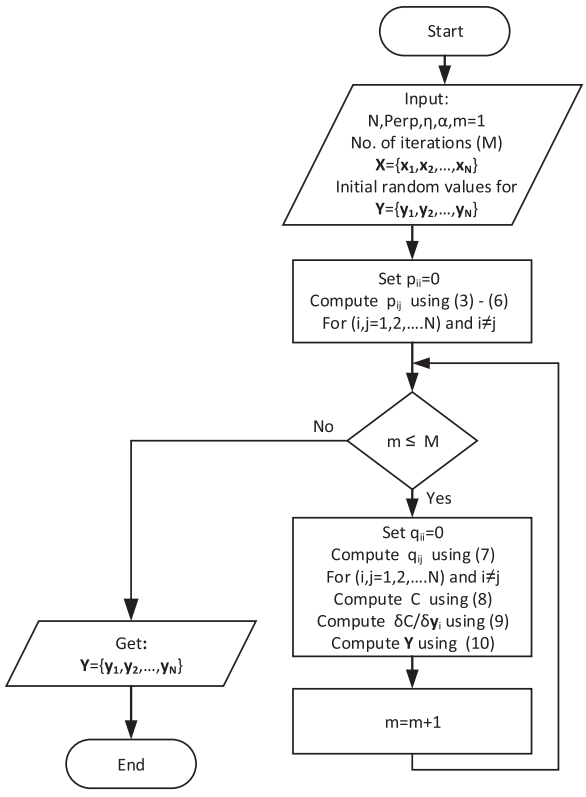



Fig. 1. Flow diagram of the t-SNE algorithm.

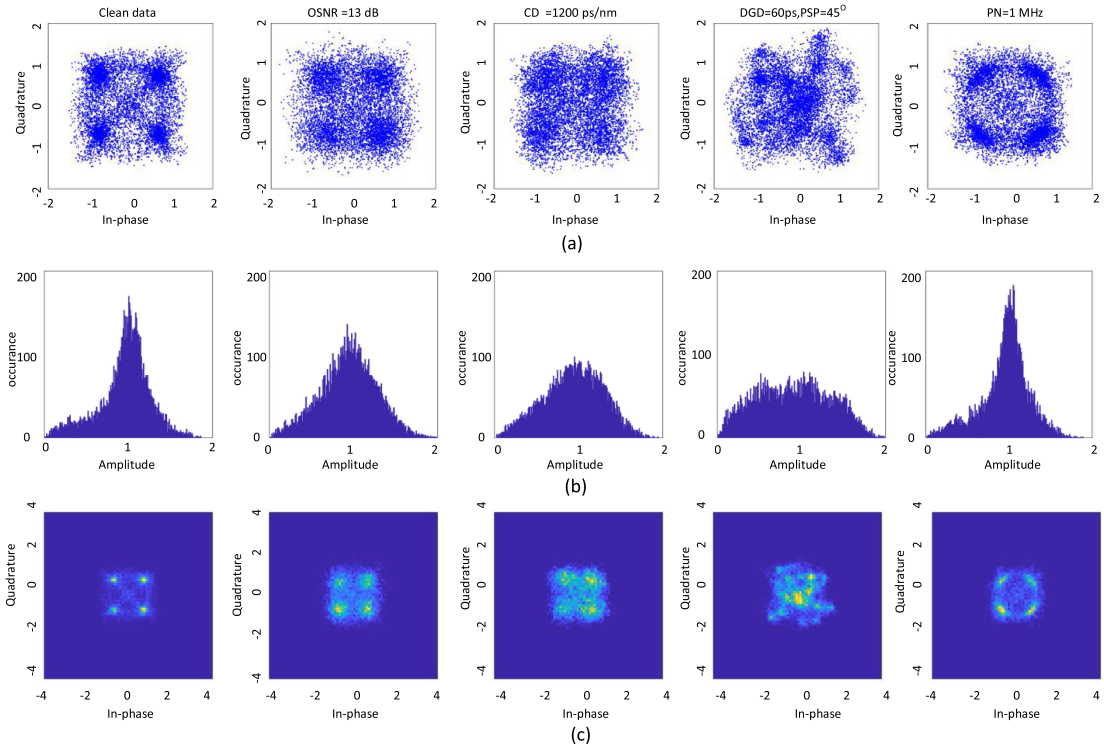



Fig. 2. (a) Impairments effect on constellation diagram, (b) the corresponding AAHs, and (c) the corresponding IQHs.

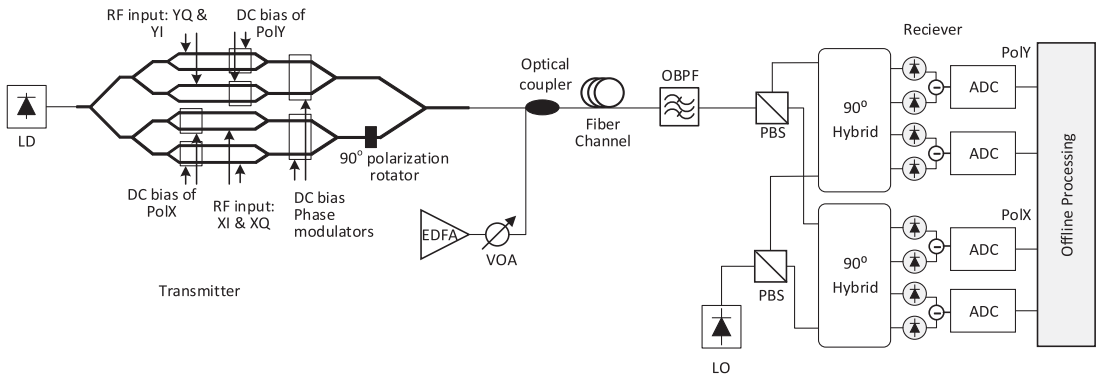



Fig. 3. Simulation setup illustrating the transmitter and receiver configurations that are used in this work. LD: Laser Diode, MZM: Mach-Zehnder modulator, EDFA: Erbium-doped fiber amplifier, VOA: variable optical attenuator, OBPF: optical band pass filter, PBS: polarization beam splitter, ADC: analog to digital convertor, LO: local oscillator.

TABLE 1
Impairments Values Used in Simulation

| | OSNR (dB) | CD (ps/nm) | DGD (ps) | PN (MHZ) |
|---|---|---|---|---|
| **Range** | 9-19 | 200-1600 | 10-70 | 0.01-10 |
| **Step size** | 2 | 200 | 10 | × 10 |

Figure 2 shows the effect of impairments on constellation and their corresponding AAH and IQH, for dual polarization-quadrature phase shift keying (DP-QPSK) 12.5 Gbaud coherent system. The data used in this work is taken after sampling and before demodulation stage. Figure 2, column 1, shows the back to back clean data without any impairments. Column 2 shows the effect when OSNR = 13 dB, we notice that the constellation becomes cloudier as the OSNR decreases, which is reflected in both AAH and IQH features by broadening their histogram profiles and reducing their peaks. Similarly, the CD effect (i.e., CD = 1200 ps/nm), in column 3, causes the constellation to become cloudier since CD results in broadening the optical pulse signal, which leads to inter-symbol interference (ISI) between consecutive data symbols. Thus, both AAH and IQH experience a stretch in their histograms and a reduction in their peak amplitudes. The effect of the first order PMD with a DGD value of 60 ps and principle states of polarization (PSP) angle of 45$^o$ is shown in column 4. Unlike the OSNR and CD effects, the effect of PMD on the constellation leads to the emergence of many peaks due to the overlap between the two polarizations. These peaks also appear in both AAH and IQH histograms. The effect of laser PN with linewidth (LW) of 1 MHz is shown in column 5. Since the PN affects only the signal phase, we get almost the same shape in case of AAH, in contrast to IQH where we get different shapes whenever the PN gets changed.

## 4. Simulation Setup

The data used in our simulation is generated using *VPI TransmissionMaker* 9.9 simulator, and according to the setup block diagram shown in Fig. 3. A DP-QPSK modulation format is generated at 12.5 Gbaud speed. At the transmitter side, a laser diode (LD) with 1550 nm wavelength is used. The linewidth of the laser is swept over different values to emulate the effect of PN. The amplified spontaneous emission (ASE) noise is added to the transmitted signal using an Erbium-doped fiber amplifier (EDFA). A variable optical attenuator (VOA) is used to vary the OSNR values. In addition, the optical fiber length is varied to provide different values of CD. Furthermore, we change the values of DGD and set the PSP angle at 45$^o$. All the impairments' values are summarized in Table 1. At the receiver side, an optical band pass filter (OBPF) with 50 GHz bandwidth is used. The received

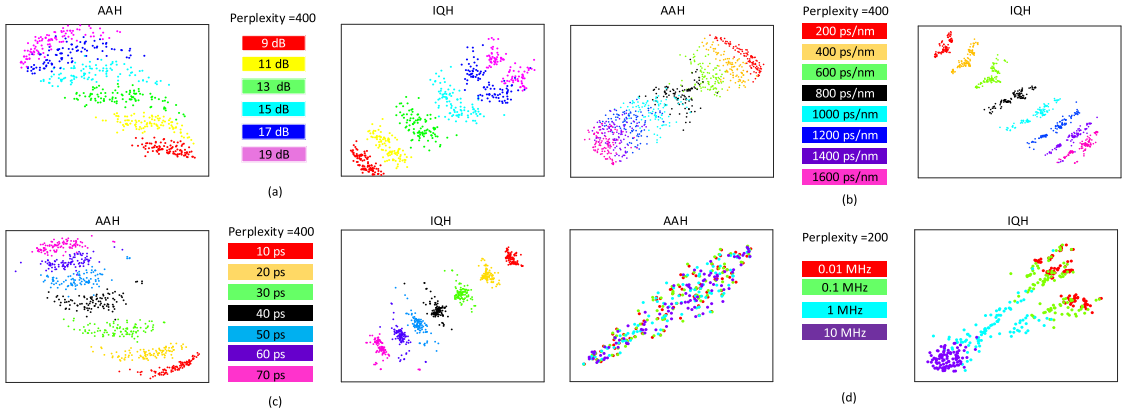



Fig. 4. Separability investigation of single impairment using AAH and IQH: (a) OSNR, (b) CD, (c) PMD, (c) and (d) PN.

optical signal is mixed with a local oscillator (LO) to demodulate the signal. Then, 90° hybrid and balanced detectors are used to convert the optical signal to an electrical waveform. The resultant electrical I and Q signals are sampled using two analog to digital converters (ADCs) to get 8192 amplitude samples for each I and Q signals. Detailed discussion regarding the effect of impairments is provided in the results section.

After obtaining the I and Q signals, we compute the AAH with 200 bins and IQH with ($80 \times 80$) bins. Both histograms are normalized by the standard deviation value. The IQH with ($80 \times 80$) bins is converted to an array with 6400 samples length. Therefore, the input to the t-SNE algorithm is a set of realizations, each of which is of 200 samples in case of AAH and 6400 samples in case of IQH. The t-SNE algorithm is used to visualize the 200 (6400) dimensional space of AAH (IQH) using a two-dimensional (2D) space. That is, the t-SNE algorithm represents each realization of length 200 (or 6400) samples by a single point in a 2D space. We studied 4 types of impairments. Each of these types is considered as a separate class, and we seek to find out the conditions under which these classes are separable. To do so, we use t-SNE MATLAB built-in function to visualize, on the same figure, the structure of the simulated data for all impairments levels. t-SNE is often used to investigate the separability of classes in a given dataset. This is done by first processing the data using t-SNE algorithm, and then visualizing the data in the reduced dimension (i.e., 2D). If different classes are plotted in contagious clusters with a clear separation between them, then the dataset is said to have separable classes. Note that t-SNE can only be used to confirm the separability of classes and cannot be used to extract features [23].

## 5. Results and Discussion

In this section, we discuss the obtained results for OPM investigation. We consider the effect of OSNR, CD, PMD and PN. We use the AAH bins and IQH histogram in conjunction with t-SNE. The OPM is investigated in case of single and multiple impairments. First, we use subjective measures (Good, Moderate, and Bad) based on visual inspection to evaluate the effectiveness of histogram-based features to separate impairments. Then, we verify our inspection results using a quantitative measure. In t-SNE simulation, the learning rate ($\eta$) and the momentum ($\alpha$) parameters are set to 0.01 and 0.5, respectively.

### *5.1 Single Impairment*

In this subsection, we discuss the results in case of data with one type of impairments. We generate 100 realizations for each impairment value shown in Table 1. The results show that OSNR, CD and PMD values are clustered as shown in Fig. 4(a)–(c). Referring to Fig. 4(a)–(c), it is observed that

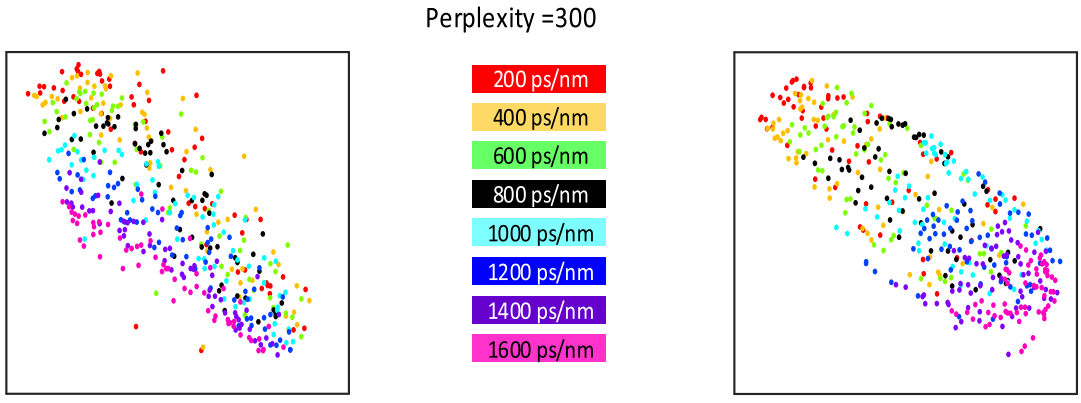



Fig. 5. Investigation of CD in case of joint OSNR-CD using AAH (left) and IQH histogram (right).

the separation of OSNR, CD and PMD for both AAH and IQH is good as each impairment value has a distinguishable contour. On the other hand, since the AAH only provides information about the amplitude, we cannot use it to monitor PN. This is not the case with the IQH since it provides information about the amplitude and phase, as shown in Fig. 4(d). It is worth to mention that the separation between small and close values of PN such as 0.01 and 0.1 MHz is difficult, as shown in Fig. 4(d). In general, the utilization of smaller step size for a given impairment will generate more clusters, which make the graph denser.

### *5.2 Multiple Impairments*

Next, we examine the effect of multiple impairments that occur simultaneously. Ten independent realizations have been generated for each joint impairment (i.e., in case of joint OSNR-CD, the total number of realizations is 6 (OSNR values) $\times$ 8 (CD values) $\times$ 10 $=$ 480 realizations).

*5.2.1 Joint OSNR-CD/PMD/PN:* First, we examine the possibility of separating two coexisting impairments. We started with the joint OSNR-CD and found that the separation possibility of CD is moderate, because there is some overlap between neighbor values but there is no overlap between divergent values, as shown in Fig. 5. On the other hand, the separation of OSNR is found to be bad, where all values are overlapped. Next, we consider the coexistence of the joint OSNR-PMD, where the separation quality of PMD is found to be moderate while the separation of OSNR is bad. Finally, we consider the Joint OSNR-PN, where the separation quality is found to be good (good) by IQH (AAH) for OSNR and moderate (bad) by IQH (AAH) for PN, since the AAH detects only the amplitude, which is not effective for PN monitoring.

*5.2.2 Joint OSNR-CD-PMD/OSNR-CD-PMD-PN:* In this subsection, we study the separability of the joint OSNR-CD-PMD and the Joint OSNR-CD-PMD-PN. For both cases, the results show that the separation quality of OSNR and CD are bad for both AAH and IQH techniques. This is due to the effect of other impairments and the similar effect of CD and OSNR, especially in case of high dispersion and low OSNR values, see Fig. 2. On the other hand, the monitoring quality of PMD is moderate for both AAH and IQH because the PMD has unique shapes for each PMD value. In case of Joint OSNR-CD-PMD-PN, the separation quality of PN is moderate in IQH while bad for AAH.

In Fig. 6, we summarize our simulation results for the single and multiple impairments by showing each impairment as a separate color contour (red: OSNR, blue: CD, green: PMD, and black: PN). Note that the intersection between contours shows the performance of joint impairments. The obtained results of the visual investigation can be summarized as follows:

- For the single impairment case, impairments levels are distinguishable for all types of impairments using AAH and IQH, excluding the PN when AAH is used.

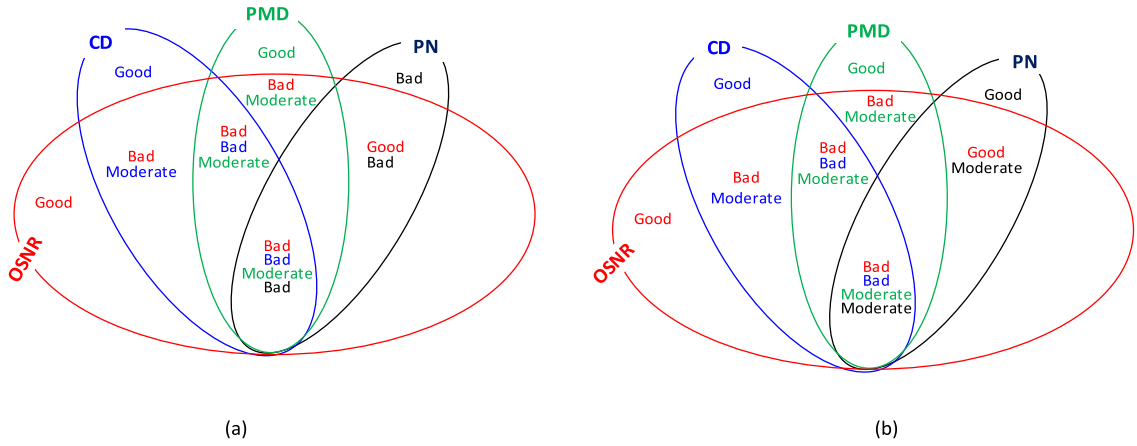



Fig. 6. Investigation results of single and multiple impairments using (a) AAH and (b) IQH.

- The PMD values can be separated to some extent (i.e., neighbor values cannot be separated while divergent values can be separated), while it is difficult to separate OSNR and CD values especially in case of presence of more than two impairments.

### 5.3 Results Validation

To validate the t-SNE results pertaining to the separability of AAH and IQH for optical performance monitoring, we have designed an estimator using linear support vector machine (SVM) regression to monitor the value of a given parameter. SVM has an advantage of being insensitive to the over-fitting problem. For a data set $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots \mathbf{x}_N\} \subset \mathrm{R}^{\mathrm{D}}$, and its target (labels) $\mathbf{t} = \{\mathrm{t}_1, \mathrm{t}_2, \ldots \mathrm{t}_N\} \subset \mathrm{R}$, the SVM finds the function $\mathrm{t}_n$ with at most $\varepsilon$ deviation from the labels and gives the best fit $\mathrm{R}^{\mathrm{D}} \rightarrow \mathrm{R}$.The fitting function is represented by the model's parameters $\omega$ and b as [31]

$$\mathrm{t}_n = \langle \omega, \mathrm{x}_n \rangle + \mathrm{b} \leq \varepsilon, \tag{11}$$

where $\langle ., . \rangle$ is the dot product, $\omega \subset \mathrm{R}^{\mathrm{D}}$ and b is a real number. In our work, the input to the SVM is the same as that of t-SNE; either the AAH or IQH of raw data. The data has been split into training and testing sets, where 70% of the data is used for training and 30% is used for testing. We have considered discrete values of impairments to build the dataset with different step sizes. Note that a smaller step size leads to a larger dataset. It has been demonstrated in the literature that as the training dataset size increases, the performance of SVM algorithms improves till a certain level [31]. Therefore, the size of training dataset should be properly selected.

In our development, a mathematical model is built using the training dataset with 10 fold cross-validation. Sequential minimal optimization (SMO) is used for solving the regression optimization problem [32]. The accuracy of estimation is measured by using the correlation coefficient parameter, given by [33]

$$\rho = 1 - \frac{\sum_{\mathrm{i}=1}^{\mathrm{N}} (\mathrm{t}_i - \hat{\mathrm{t}}_i)^2}{\sum_{\mathrm{i}=1}^{\mathrm{N}} (\mathrm{t}_i - \bar{\mathrm{t}})^2}, \tag{12}$$

where $\mathrm{t}_i$ is the actual data, $\hat{\mathrm{t}}_i$ is the estimated data, N is the total number of test samples, and $\bar{t}$ is the sample mean. We have computed the value of $\rho$ for all investigated results of single and multiple impairments using AAH and IQH, and used it to define a quantitative measure for the terms Bad, Moderate, and Good utilized in the previous subsection for separability classification. Specifically, the following criterion has been employed.

- Good: means $\rho$ is in the range of 0.9 to 1
- Moderate: means $\rho$ is in the range 0.7 to 0.89

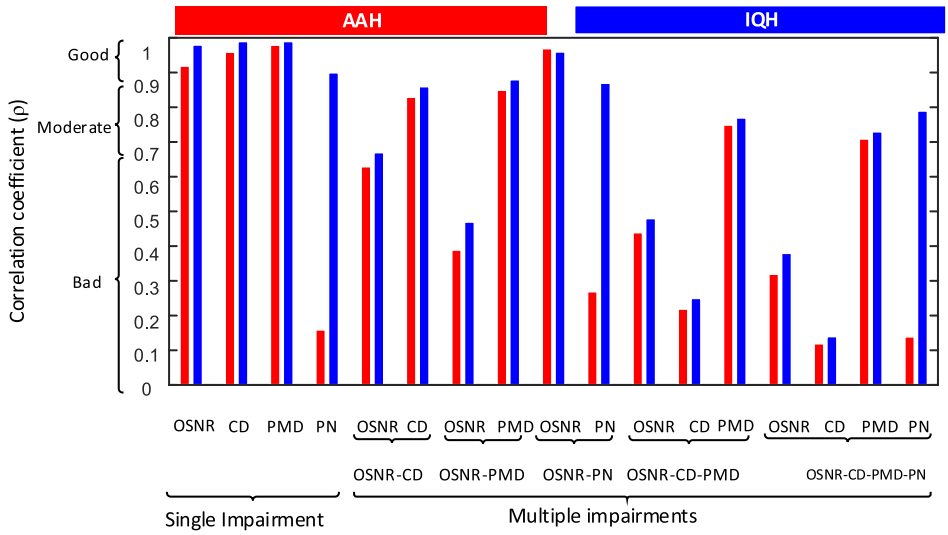



Fig. 7. Summary of accuracy results for both AAH and IQH.

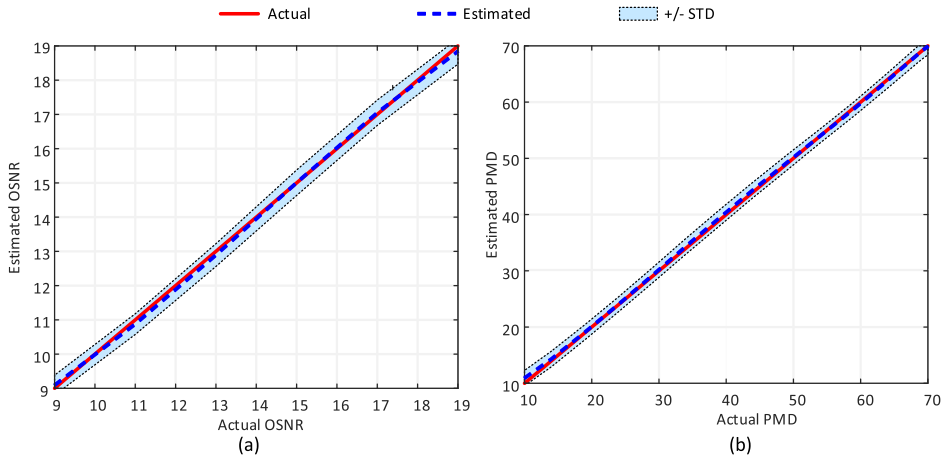



Fig. 8. Single impairment monitoring using IQH histogram (a) OSNR, and (b) PMD.

- Bad: means $\rho$ is in the range $<0.7$

The values of $\rho$ are displayed on the y-axis of Fig. 7, for all cases. Figure 8 and 9 show examples of monitoring performance results in case of single impairment (OSNR or PMD) and multiple impairments (OSNR, CD, and PMD), respectively, using IQH feature extraction. We plot the ground truth values versus the estimates. For the latter, we display the mean $+/-$ the standard deviation (STD). For the single impairment case, it is observed that the estimates are very close to the actual values.

Hence, the results confirm the possibility of separation, as shown earlier in Fig. 4(a) and (c). On the other hand, in case of joint OSNR-CD, the results accuracy reduces and we observe that the standard deviation becomes slightly higher, as shown in Fig. 9(a). Having more than two impairments, it becomes difficult to monitor the optical performance, and this is evident from the relatively large value of the standard deviation, as depicted in Fig. 9(b).

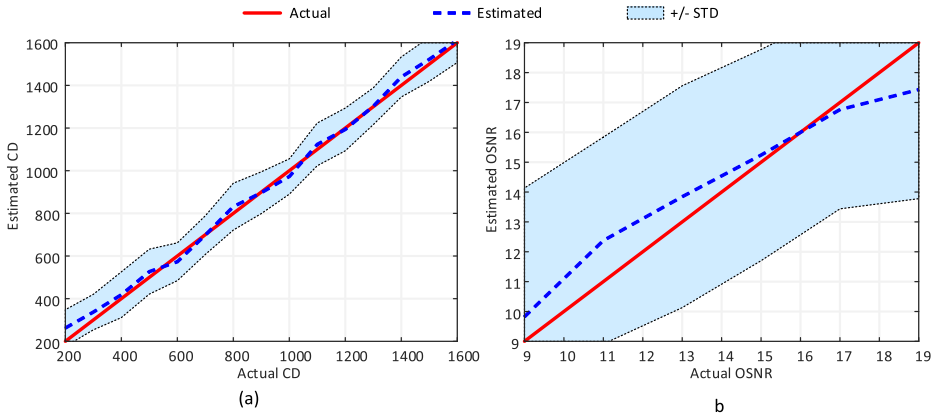



Fig. 9. (a) CD monitoring using IQH in case of OSNR-CD (b) OSNR monitoring using IQH in case of OSNR -CD-PMD.

## 6. Conclusions

In this paper, we investigated OPM monitoring using histogram-based features for 12.5 Gbaud DP-QPSK under realistic single and multiple optical channel impairments. Using simulated data generated from industry-standard VPI simulation package, we showed that histogram-based features, only under certain conditions, work well in distinguishing high and low impairment effects. Our study, systematically examined the separability of AAH and IQH features under various optical impairments. This is done by using t-SNE to reduce the dimensionality of the captured AAH and IQH features and visualize the approximate structure of the data's high dimension topology. Under some conditions, such as PN impairment, AAH is unable to distinguish between low and high levels of impairment indicating that such a feature does not contain useful information for OPM under PN. SVM is used to verify the separability results, where we show that when the number of joint impairments increases, the standard deviation between the actual results and the estimated results increases. It is relevant here to mention that even though the SVM algorithm was trained using discrete values, when deployed in a real system with continuously and dynamically varied environment, the trained SVM regression model will interpolate the results and provide accurate estimates. In a nutshell, our work provides important insights into the dynamics of various OPM features and utility of these features in estimation or classification of optical channel impairments, which is a necessary first step for researchers to consider before proceeding to the development of new OPM techniques.

## References

[1] M. Jinno, "Elastic optical networking: Roles and benefits in beyond 100-Gb/s era," *J. Lightw. Technol.*, vol. 35, no. 5, pp. 1116–1124, Mar. 2017.
[2] Y. Ou, M. Davis, A. Aguado, F. Meng, R. Nejabati, and D. Simeonidou, "Optical network virtualisation using multitechnology monitoring and SDN-enabled optical transceiver," *J. Lightw. Technol.*, vol. 36, no. 10, pp. 1890–1898, May 2018.
[3] Z. Dong, F. N. Khan, Q. Sui, K. Zhong, C. Lu, and A. P. T. Lau, "Optical performance monitoring: A review of current and future technologies," *J. Lightw. Technol.*, vol. 34, no. 2, pp. 525–543, Jan. 2016.
[4] A. E. Willner, Z. Pan, and C. Yu, "Optical performance monitoring," in *Optical Fiber Telecommunications VB*, 5th ed. Amsterdam, The Netherlands: Elsevier, 2008, pp. 233–292.
[5] I. Shake, W. Takara, S. Kawanishi, and Y. Yamabayashi, "Optical signal quality monitoring method based on optical sampling," *Electron. Lett.*, vol. 34, no. 22, pp. 2152–2154, 1998.
[6] N. Hanik, A. Gladisch, C. Caspar, and B. Strebel, "Application of amplitude histograms to monitor performance of optical channels," *Electron. Lett.*, vol. 35, no. 5, pp. 403–404, 1999.
[7] Z. Li, C. Lu, Y. Wang, and G. Li, "In-service signal quality monitoring and multi-impairment discrimination based on asynchronous amplitude histogram evaluation for NRZ-DPSK systems," *IEEE Photon. Technol. Lett.*, vol. 17, no. 9, pp. 1998–2000, Sep. 2005.

[8] B. Kozicki, O. Takuya, and T. Hidehiko, "Optical performance monitoring of phase-modulated signals using asynchronous amplitude histogram analysis," *J. Lightw. Technol.*, vol. 26, no. 10, pp. 1353–1361, 2008.
[9] T. S. R. Shen, A. P. T. Lau, and G. N. Liu, "OSNR monitoring for higher order modulation formats using asynchronous amplitude histogram," *IEEE Photon. Technol. Lett.*, vol. 22, no. 22, pp. 1632–1634, Nov. 2010.
[10] S. D. Dods and T. B. Anderson, "Optical performance monitoring technique using delay tap asynchronous waveform sampling," in *Proc. Opt. Fiber Commun. Conf.*, 2006, Paper OThP5.
[11] B. Kozicki, A. Maruta, and K.-I. Kitayama, "Asynchronous optical performance monitoring of RZ-DQPSK signals using delay-tap sampling," in *Proc. 33rd Eur. Conf. Exhib. Opt. Commun.*, 2007, pp. 1–2.
[12] B. Kozicki, A. Maruta, and K.-I. Kitayama, "Transparent performance monitoring of RZ-DQPSK systems employing delay-tap sampling," *J. Opt. Netw.*, vol. 6, no. 11, pp. 1257–1269, 2007.
[13] H. Y. Choi, Y. Takushima, and Y. C. Chung, "Multiple-impairment monitoring technique using optical field detection and asynchronous delay-tap sampling method," in *Proc. Opt. Fiber Commun. Conf.*, 2009, Paper OThJ5.
[14] Y. Yu and C. Yu, "Optical signal to noise ratio monitoring using variable phase difference phase portrait with software synchronization," *Opt. Exp.*, vol. 23, no. 9, pp. 11284–11289, 2015.
[15] V. Ribeiro, L. Costa, M. Lima, and A. L. Teixeira, "Optical performance monitoring using the novel parametric asynchronous eye diagram," *Opt. Exp.*, vol. 20, no. 9, pp. 9851–9861, 2012.
[16] 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, 2017.
[17] X. Lin, O. A. Dobre, T. M. 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. 2018.
[18] 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, 2010.
[19] 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. 2018.
[20] 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.*, 2018, pp. 1–3.
[21] J. Thrane, J. Wass, M. Piels, J. C. Diniz, R. Jones, and D. Zibar, "Machine learning techniques for optical performance monitoring from directly detected PDM-QAM signals," *J. Lightw. Technol.*, vol. 35, no. 4, pp. 868–875, Feb. 2017.
[22] F. Musumeci *et al.*, "A survey on application of machine learning techniques in optical networks," arXiv:1803.07976, 2018.
[23] L. V. D. Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, 2008.
[24] N. Pezzotti, A. Mordvintsev, T. Hollt, B. P. Lelieveldt, E. Eisemann, and A. Vilanova, "Linear tSNE optimization for the Web," arXiv:1805.10817, 2018.
[25] N. V. Acuff and J. Linden, "Using visualization of t-distributed stochastic neighbor embedding to identify immune cell subsets in mouse tumors," *J. Immunol.*, vol. 198, 2017, Art. no. 1602077.
[26] Y. Gordienko *et al.*, "Dimensionality reduction in deep learning for chest x-ray analysis of lung cancer," arXiv:1801.06495, 2018.
[27] J. A. Lee and M. Verleysen, *Nonlinear Dimensionality Reduction*. New York, NY, USA: Springer, 2007.
[28] Y. Song, J. Wang, L. Qi, W. Yuan, M. Yu, and J. Qu, "A new algorithm for preserving global and local structures in supervised dimensionality reduction," in *Proc. 7th IEEE Int. Conf. Electron. Inf. Emergency Commun.*, 2017, pp. 244–247.
[29] L. Van Der Maaten, E. Postma, and J. Van den Herik, "Dimensionality reduction: A comparative," *J. Mach. Learn. Res.*, vol. 10, pp. 66–71, 2009.
[30] M. Wattenberg, F. Viégas, and I. Johnson, "How to use t-sne effectively," *Distill*, vol. 1, no. 10, p. e2, 2016.
[31] V. Vapnik, *The Nature of Statistical Learning Theory*. New York, NY, USA: Springer, 2013.
[32] J. Platt, "Sequential minimal optimization: A fast algorithm for training support vector machines," 1998.
[33] C. Huber-Carol, N. Balakrishnan, M. Nikulin, and M. Mesbah, *Goodness-of-Fit Tests and Model Validity*. New York, NY, USA: Springer, 2012.