



# Deep intelligent spectral labelling and receiver signal distribution for optical links

TONGYANG XU,<sup>1</sup>  TIANHUA XU,<sup>1,2,3,\*</sup>  AND IZZAT DARWAZEH<sup>1</sup>

<sup>1</sup>*Department of Electronic and Electrical Engineering, University College London (UCL), London, WC1E 7JE, United Kingdom*

<sup>2</sup>*School of Engineering, University of Warwick, Coventry CV4 7AL, United Kingdom*

<sup>3</sup>*School of Precision Instruments & Opto-Electronics Engineering, Tianjin University, Tianjin 300072, China*

\*[tianhua.xu@ieee.org](mailto:tianhua.xu@ieee.org)

**Abstract:** A unique automatic receiver signal distribution strategy is proposed for private optical networks based on the concept of non-orthogonality. A non-orthogonal signal waveform can compress the spectral bandwidth, which not only fits a signal in a bandwidth limited scenario, but also enables the compression ratio information for labelling. Depending on a unique value of spectral compression, an end user destination can be correlated. A network edge node will rely on deep learning to intelligently identify each raw signal and forward it to corresponding end users with no sophisticated digital signal pre-processing. In this case, signal identification and distribution are faster while computationally intensive signal compensation and detection will be shifted to each end user since the receiver is highly dynamic and user-defined in private optical networks. An intelligent signal classifier will be trained considering various fiber transmission factors such as transmission distance, training dataset size and launch power. At the end, a universal classifier is obtained, which can be used to identify signals in a system for any fiber transmission distance and launch power.

Published by The Optical Society under the terms of the [Creative Commons Attribution 4.0 License](https://creativecommons.org/licenses/by/4.0/). Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

## 1. Introduction

With the evolution to future networks [1–4], the diversity of application services requires an optical network to be flexible and automatically reconfigurable. The advancement of software-defined networks (SDN) and network function virtualization (NFV) has made dynamic networks a reality for on-demand application services. With the demand of wider bandwidth, lower latency, enhanced privacy and more flexibility, end users are becoming interested in private optical networks. A private optical network is not only for research such as UK National DarkFiber Facility (NDFF), but also for financial services, data center, healthcare, school campus, etc. The dedicated resources of a private network can provide tailored services to unique customers. Private networks often belong to legacy connection facilities, which initially aimed to deploy more fiber links than demand required for future needs. Such networks have been existing for many years underground but are not being efficiently used. Therefore, both software protocols and hardware architectures have not been standardized. In addition, customers, who will use a private network, can have a high degree of control of the network infrastructure. Therefore, this gives users freedom in designing, deploying and optimizing a tailored network for specialist applications. Since a private network is not connected to the public Internet, its private features will provide security and privacy to sensitive communications. Moreover, communication latency is reduced since the data in a private network avoids the transmission over contention networks.

The development of artificial intelligence (AI) has made significant achievements in computer vision (CV), natural language processing (NLP), pattern recognition, autonomous driving,

healthcare, etc. In wireless communications, AI [5,6] has solved many traditionally remaining challenges in signal detection, modulation identification, channel estimation, channel coding, sensing and localization. Unfortunately, due to time-variant frequency selective channel conditions, practical applications of AI in wireless communications remain limited. In optical communications, however, channel condition is relevantly better since optical signals are transmitted via fibers with more stable channel parameters. Therefore, the concept of AI is being widely studied and implemented for optical fiber communications. There are many legacy optical transmission challenges that can be efficiently solved via AI. The review work [7] proposed AI friendly optical applications in fiber nonlinear transmission with non-linearity mitigation specifically addressed in [8] and performance monitoring and cross-layer optimization in [9]. Furthermore, a network level overview of AI potentials is comprehensively investigated in [10].

Typical resource allocation [1,11] has to consider real-world conditions for both channels and end users. For a bandwidth limited application scenario, the traditional way is to pack a smaller number of sub-carriers or wavelengths, so as to reduce occupied bandwidth within a bandwidth limited channel, at the expense of reduced information rate. In addition, conventional optical communication systems would rely on cooperative signal processing. Signals would be forwarded to target users based on pre-known signalling configurations. With the increasing size of optical networks and the growing demand from consumers [12], the methodology of optical transmission has to be optimized. An intelligent receiver would avoid signalling overhead and directly extract signal information from received signals. In this case, the signalling control overhead would be simplified and signal distribution would be more efficient.

Network with features such as flexibility and intelligence will pave the way into the next era of network evolution; upper layer network algorithms have played important roles for such an evolution. This work will deal with these matters from a physical layer signal waveform perspective. Appropriately designed signals can bring additional gains in spectral efficiency. The widely used waveform is orthogonal frequency division multiplexing (OFDM), which has been standardized in 4G [13] and 5G [14]. Non-orthogonal signal waveforms can break the orthogonality in OFDM and therefore bring unique advantages such as spectral bandwidth saving in spectrally efficient frequency division multiplexing (SEFDM) [15], timing stream compression in faster than Nyquist (FTN) [16] and out-of-band power leakage suppression in filterbank based multi-carrier (FBMC) [17] and generalized frequency division multiplexing (GFDM) [18]. Considering system compatibility, this work will focus on the bandwidth compressed SEFDM signal waveform, which has been proved to have better performance than traditional orthogonal waveforms in optical fiber systems [19,20], millimeter wave systems [21,22] and visible light communication (VLC) systems [23].

This work will investigate an intelligent receiver signal distribution scheme, which ensures a signal band fitting in a bandwidth limited scenario and more importantly the spectral bandwidth compression ratio of the non-orthogonal signal will be used as a user address label to an end user. At the end, an optimal way to train a convolutional neural network (CNN) signal classifier is designed and a universal classifier is obtained that can be used for long-haul fiber communication systems at a random distance.

The main discoveries of this work are as the following.

- A spectral compression labelling automatic signal distribution scheme is proposed to enhance flexibility and intelligence in private optical networks. The spectral compression strategy, with a compression ratio  $\alpha$ , can dynamically squeeze signals to fit in a given spectral bandwidth, which may be highly spectrally efficient in bandwidth-limited scenarios. In addition, the  $\alpha$ -labelling strategy can avoid end user address overhead and corresponding digital signal processing for signal compensation and data recovery. Thus, a faster signal distribution is realized to raw signals at the level of near-optical domain, which happens immediately after optical-to-electrical signal conversion. In this case, bandwidth-limited

spectrally efficient communications and faster end user addressing are simultaneously achieved.

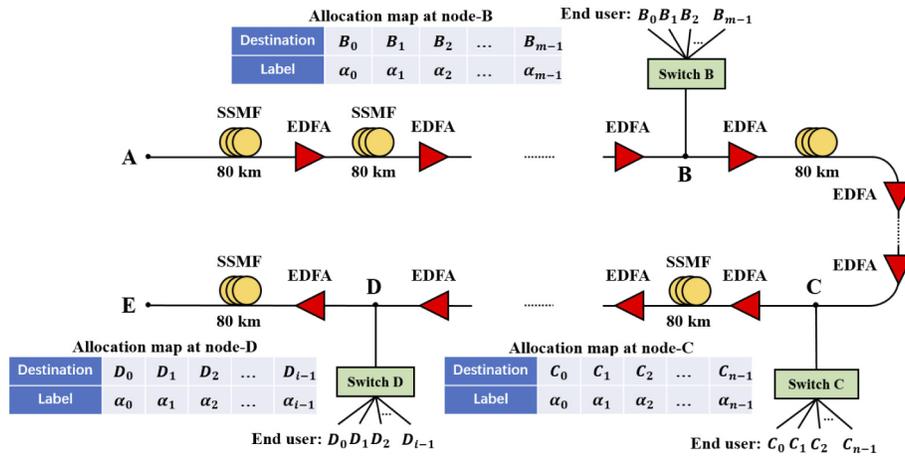
- Intelligent deep learning classifier is designed to avoid prior-knowledge of signal format and automatically identify each  $\alpha$ -labelling signal and forward it to corresponding end users. In this case, each network edge node will manage signal forwarding at near-optical domain at a faster speed and all the computationally intensive digital signal processing will be shifted to each end user. Therefore, parallel processing is achieved with increased network capacity.
- Comprehensive performance investigations of intelligent signal classifiers under various optical environment factors are considered such as fiber transmission distance, training dataset size and launch power.
- A universal signal classifier is obtained by using an optimal training methodology. It is possible to train a neural network model on one fiber communication system and reuse it efficiently on any other long-haul optical fiber systems.

## 2. Application scenario

The commonly used technique behind a private optical network, such as a dark fiber system, is dense wavelength division multiplexing (DWDM), in which a single or multiple wavelengths are used to deliver data to users. In such cases users are allocated one or more wavelengths, according to demand and availability of network resources. However, it is always the case that the whole resource of a wavelength is used, in other words no fractional wavelength resource is allowed. We propose to use a multi-carrier non-orthogonal waveform to compress the original signal, thereby increasing spectral efficiency. The spectral bandwidth is compressed based on a spectral compression ratio termed  $\alpha$ , which ranges from 0 to 1. When  $\alpha=1$ , the signal is OFDM while when  $\alpha<1$ , spectrally compressed signals will be applied. Beneficially, the spectral compression ratio will be allocated uniquely to end users. Therefore, an intelligent signal classifier will be possible to identify a target end user based on the spectral compression features. It is noted that traditional single-carrier based modulation patterns are also potential for intelligent signal classification. However, the modulation classification is limited by the number of available modulation patterns. A more challenging issue is that one modulation pattern could be misclassified into another modulation pattern due to constellation similarity. Therefore, this work will focus on the multi-carrier signal classification rather than single-carrier modulation classification.

Since a private network, such as the NDFF in [24], has no standardized system requirement, therefore this work can flexibly define a system aiming to present a concept of a new communication framework. As shown in Fig. 1, a long-haul fiber connection link is designed with five network edge nodes, labelled as A, B, C, D and E. Multiple fiber spans are serially connected with erbium doped fiber amplifier (EDFA) for signal power amplification. Node-A is the signal transmitter and it will generate data based on channel/hardware conditions from different end users. Signals will be compressed when one wavelength is not sufficient to an end user. Since end users have unique data bandwidth requirements, the compression ratio  $\alpha$  is thus various and can be used for end user addressing. The final signal will be stamped with the address of network edge node such as B, C, D, but the address for each end user is not included. The spectral compression ratio  $\alpha$  will be used as a new domain information for end user addressing. At node-B, an optical signal is received and converted to its electrical equivalent. An intelligent signal classifier will automatically identify the signal format and its associated  $\alpha$  and forward the signal to its corresponding end user in  $(B_0, B_1, B_2, \dots, B_{m-1})$ . It should be noted that each end user will handle the subsequent digital signal processing for imperfect timing/frequency/phase compensation

and signal detection. Therefore, the digital signal processing pressure is shifted from network edge node to each end user. This intelligent forward strategy is helpful to cut round-trip latency. A traditional receiver will compensate signal timing/frequency/phase impairments, channel estimation, channel equalization, multi-carrier demultiplexing, demapping, channel decoding and then cyclic redundancy check (CRC) check. Basically, the CRC check will also assist to find the correct receiver identity. If the signal passes CRC check, the signal reception is successful. Otherwise, a re-transmission is required, which will cause extra round-trip latency. Our proposed strategy can avoid all of the digital signal processing and make a decision immediately after receiving a signal. This will save the signal processing time and the intelligent classifier will locate the intended end-user quickly and leave the digital signal processing to each end-user. In this case, parallel processing is possible for network capacity increase. In addition, end users will have more flexibility to structure a user-defined receiver in a private optical network. An end user may be an individual with limited data requirements or an enterprise with high requirement; in both cases the needs may be accommodated by adjusting the receiver architecture and destination spectral labeling characteristics according to technical needs, energy and cost restrictions. Similarly, the signal identification and forward process will be applied to node-C and node-D as well. However, the optical channel environment is different for node-B, C, D since different fiber transmission distances will include unequal amplified spontaneous emission (ASE) noise. With higher launch power for longer fiber transmission especially for node-D, fiber non-linearities become more detrimental. Therefore, one motivation of this work is to train a universal intelligent classifier that can work for any network edge node within an optical network.



**Fig. 1.** Application scenario for the spectral compression labelling signal distribution in private optical networks. The parameter  $\alpha$ , ranging from 0 to 1, indicates the spectral compression ratio for the multi-carrier signals.

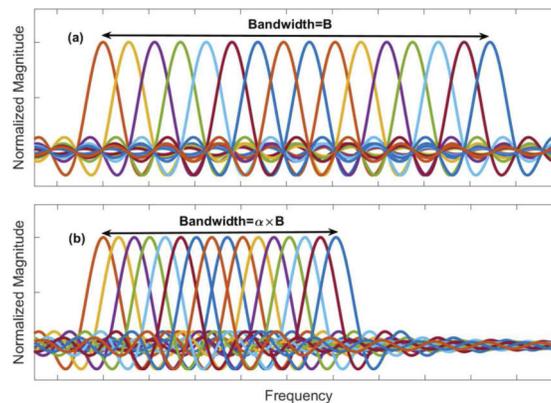
The proposed scenario can be implemented in DWDM similar systems and each network edge node will occupy a unique wavelength. The intelligent signal distribution will happen in a time division multiplexing (TDM) mode at a wavelength. In a time slot  $t_0$ , a unique  $\alpha_0$  labelled signal will be transmitted and then go through the signal classifier for an end user signal distribution. In a second time slot  $t_1$ , a signal labelled by  $\alpha_1$  will go through the same process and find its target end user. Theoretically, all the nodes and their end users can share the same wavelength assuming that no two identical  $\alpha$  labels are being used in a network. The number of end user in this work is determined by both the wavelength and the number of designed signal patterns. The more wavelength options and signal patterns have been used, the more end users can be supported. When the number of end users increases, it indicates the network requires more wavelengths

and signal patterns. As we have mentioned above, an end user normally indicates a group. Therefore, the received data would be further allocated to specific end persons. The allocation strategy is decided by each end user and this also shows the advantage of using dedicated receiver signal processing since it can define the end person allocation strategy by itself. It is noted that more signal patterns would challenge accurate signal identification and therefore lead to further research on intelligent signal identifiers.

This intelligent signal distribution not only saves signalling overhead but also simplifies and speeds up signal processing at the edge. Traditionally, the end user address is included within the overhead of a signal. The accurate address formatted in digital bits can only be extracted after sophisticated signal compensation algorithms. After a series of compensation algorithms, signals will be forwarded to specific end users by each node. This will cause pressure at edge nodes such as node-B, C, D as they will do all the receiver side signal processing for their end users. The proposed intelligent signal distribution scheme will merely assist each network edge node to forward signals to each unique end user. In this case, the subsequent signal processing pressure is distributed to each end user and the processing speed at each network edge node will be improved.

### 3. Spectral compression signal fundamentals

The key principle of the spectral compression waveform is to compress sub-carriers spacing and break the orthogonality. The sub-carrier packing principle is illustrated in Fig. 2 where 16 sub-carriers are used for both the OFDM waveform and the SEFDM waveform. Both signals have the same sub-carrier bandwidth while the sub-carrier spacing in SEFDM is narrower leading to a compressed spectral bandwidth.



**Fig. 2.** Sub-carrier packing illustration. (a) OFDM. (b) SEFDM. The parameter  $\alpha$  indicates bandwidth compression factor.

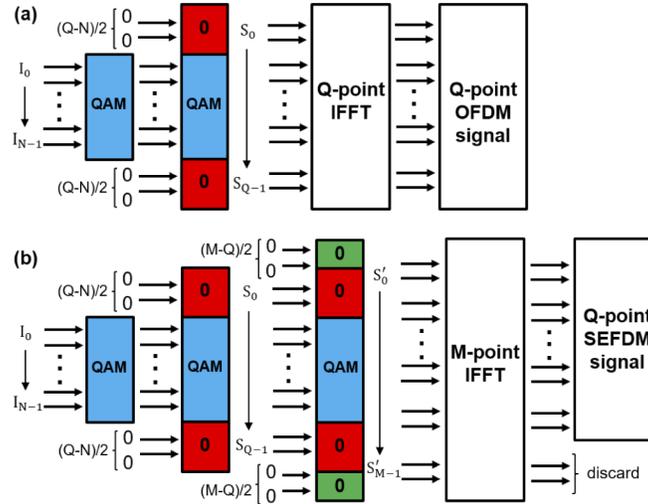
The non-orthogonal SEFDM waveform is a multi-carrier signal and has common features as OFDM. Therefore, the expression of an SEFDM signal is straightforward by adding  $\alpha$  in a traditional OFDM signal as

$$X_k = \frac{1}{\sqrt{Q}} \sum_{n=0}^{Q-1} S_n \exp\left(\frac{j2\pi nk\alpha}{Q}\right), \quad (1)$$

where  $Q = \rho N$  indicates the number of samples with the oversampling factor  $\rho$  and the number of sub-carriers  $N$ .  $\alpha$  is the bandwidth compression factor (BCF).  $X_k$  is the  $k^{\text{th}}$  time sample with the index  $k = 0, 1, \dots, Q - 1$ ,  $S_n$  is the  $n^{\text{th}}$  single-carrier symbol with the index  $n = 0, 1, \dots, Q - 1$ .

Signal generation for OFDM is straightforward via using inverse fast Fourier transform (IFFT) as shown in Fig. 3(a). A protection guard band is commonly required and zeros are padded on both sides of an input QAM symbol vector. Therefore, in Fig. 3(a), the original input symbol vector  $[I_0, I_1, \dots, I_{N-1}]$  is expanded to a  $Q$ -dimensional vector as

$$[S_0, S_1, \dots, S_{Q-1}] = \underbrace{[0, \dots, 0]}_{(Q-N)/2}, [I_0, I_1, \dots, I_{N-1}], \underbrace{[0, \dots, 0]}_{(Q-N)/2}, \quad (2)$$



**Fig. 3.** Signal generation block diagram. (a) OFDM. (b) SEFDM.

Then a  $Q$ -point IFFT is applied and operated on the  $Q$ -dimensional zero padded signal vector  $S$  leading to an oversampled  $Q$ -dimensional OFDM signal.

The direct operation for SEFDM signal generation in Eq. (1) will cause high computational complexity due to the fractional value of  $\alpha$ . The parameter  $\alpha$  can be removed via defining a new parameter  $M = Q/\alpha$  where  $M$  should be rounded to its closest integer. Correspondingly, the original vector  $S$  will be further expanded to a longer vector  $S'$  with the following stage-II zero padding operation

$$S'_n = \begin{cases} 0 & 0 \leq n < (M - Q)/2 \\ S_n & (M - Q)/2 \leq n < (M + Q)/2 \\ 0 & (M + Q)/2 \leq n < M \end{cases} \quad (3)$$

The stage-II zero padding is demonstrated in Fig. 3(b) where a vector of  $(M - Q)/2$  zeros are padded on both sides of  $S$  as

$$[S'_0, S'_1, \dots, S'_{M-1}] = \underbrace{[0, \dots, 0]}_{(M-Q)/2}, [S_0, S_1, \dots, S_{Q-1}], \underbrace{[0, \dots, 0]}_{(M-Q)/2}. \quad (4)$$

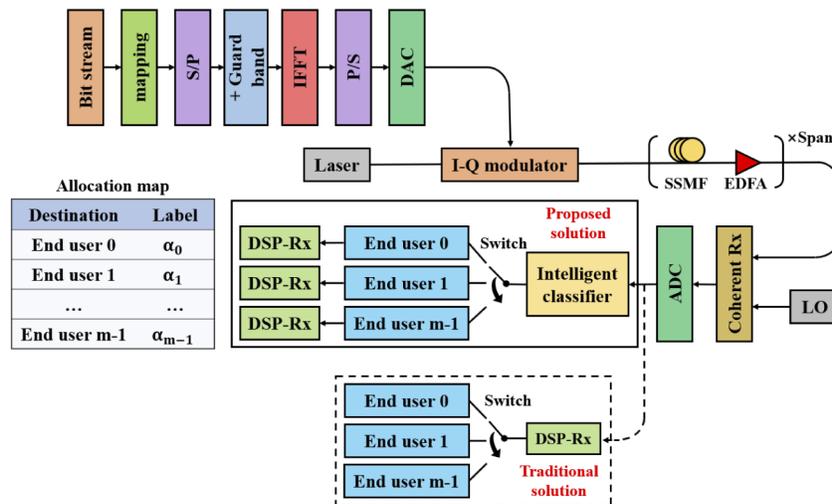
Therefore, the zero padding strategy in 4 will simplify the direct signal generation in 1 into an  $M$ -point IFFT operation as

$$X'_k = \frac{1}{\sqrt{M}} \sum_{n=0}^{M-1} S'_n \exp\left(\frac{j2\pi nk}{M}\right), \quad (5)$$

where  $n, k = [0, 1, \dots, M - 1]$ . The output will be truncated with only  $Q$  samples reserved while the rest of the samples are discarded leading to a  $Q$ -point SEFDM signal. Thus, the signal generation of SEFDM waveform is realistic using IFFT and will work well in low-cost hardware.

#### 4. Transmission setup

A non-cooperative optical transmission system is designed in Fig. 4. A random bit stream is generated by the bit generator. Serial complex symbols are obtained after the QPSK mapping. To efficiently use IFFT for multi-carrier signal generation, the serial symbols are converted into multiple parallel symbol vectors. Prior to the IFFT operation, each symbol vector will be oversampled, which is equivalent to guard band packing on both sides of each symbol vector. The final serial multi-carrier OFDM/SEFDM symbol stream will be obtained after the parallel-to-serial convertor. The digital signals will be fed to a digital-to-analogue converter (DAC) for analog signal conversion. An I-Q modulator, including two Mach-Zehnder modulators (MZMs), is applied to up-convert the base band electrical signal to an optical signal at a laser central wavelength of 1550 nm.



**Fig. 4.** Simulation model of a long-haul optical fiber transmission system. Proposed edge node intelligent receiver architecture and traditional edge node receiver architecture. The intelligent classifier will be explained in detail in the following sections.

Multiple fiber spans are serially connected in the long-haul transmission link and each span is 80 km. Optical fiber channel is simulated based on the nonlinear Schrödinger equation [25] using split-step Fourier method (SSFM) with a step size of 0.05 km. The fiber model is accurate to simulate power attenuation, chromatic dispersion (CD) effect and Kerr fiber non-linearities such as self-phase modulation (SPM), cross-phase modulation (XPM) and four-wave mixing (FWM). The EDFA is applied in the fiber transmission system after each span to amplify attenuated optical signals. The detailed simulation parameters are listed in Table 1.

After the optical fiber transmission, the multi-factor impaired optical signals will be received and down-converted to an electrical signal in the coherent receiver. The digital signals will be obtained after the analogue-to-digital converter (ADC) module. It is noted that the intelligent signal classifier is directly applied after the ADC block without any complex digital signal processing (DSP) for fast Fourier transform (FFT), timing, CD or fiber non-linearity compensation. Therefore, the classifier has to consider comprehensively the imperfect timing synchronization, CD effect and fiber non-linearity. The principle of the intelligent signal distribution is based

**Table 1. Simulation model parameters**

Parameter	Value
Central wavelength	1550 nm
Fiber attenuation constant	0.2 dB/km
Chromatic dispersion parameter	17 ps/(nm·km)
Nonlinear fiber parameter	1.2 (W·km)
Fiber length per span	80 km
EDFA noise figure	4.5 dB

on spectral feature variations. Therefore, any other factors that will not affect overall occupied spectral resources will not be taken into account. In this case, the phase noise from the transmitter and local oscillator (LO) lasers, the frequency offset between the transmitter and LO lasers, as well as the polarization mode dispersion (PMD) in the fibre are neglected.

The signal classifier in this work will automatically identify signal format, which is determined by the BCF  $\alpha$ . An allocation map will be obtained and saved initially and will be used for the signal distribution at edge node. Once a signal format is identified with its associated  $\alpha$ , the signal will be forwarded by the classifier (i.e. network edge node) to its corresponding end user destination by adaptively switching to the target node. Each end user will have its own DSP module to do further processing such as timing recovery, CD compensation and fiber non-linearity compensation. A traditional distribution scheme is demonstrated as well in Fig. 4, in which a centralized DSP module is used to switch and forward signals. Compared with the proposed intelligent signal distribution scheme in this work, traditional strategies can only do serial processing while the proposed solution can do processing in parallel. Therefore, the proposed intelligent solution can speed up signal processing and therefore effective throughput of the network. In addition, the DSP module at each end user is highly user defined and the cost can be controlled by each end user.

## 5. Intelligent classifier design

### 5.1. Principle of intelligent signal classifier

The traditional solution for modulation classification is the model-driven based maximum likelihood (ML) function [26], which can derive an accurate mathematical model for the classification task. The likelihood function in an additive white Gaussian noise (AWGN) channel with perfect knowledge of all parameters except the modulation format is expressed as

$$L(r|\mathfrak{M}, \sigma) = \frac{1}{P} \prod_{n=0}^{N-1} \sum_{p=0}^{P-1} \frac{1}{2\pi\sigma^2} \exp\left(-\frac{|r(n) - \mathfrak{M}(i,p)|^2}{2\sigma^2}\right), \quad (6)$$

where  $\mathfrak{M}$  indicates modulation candidates,  $\mathfrak{M}(i,p)$  represents the  $p^{\text{th}}$  constellation symbol in the  $i^{\text{th}}$  modulation scheme. Each modulation scheme has up to  $P$  constellation symbols.  $N$  is the number of symbols for each observation, which indicates the number of sub-carriers in multi-carrier signals.  $\sigma^2$  is noise variance and  $r(n)$  is the  $n^{\text{th}}$  single-carrier complex symbol.

The maximum likelihood classification is to maximize the likelihood function among all the modulation candidates. Assuming the entire potential solution set is  $\Theta$ , the maximum likelihood based solution  $\hat{\mathfrak{M}}$  is give by

$$\hat{\mathfrak{M}} = \arg \max_{\mathfrak{M}(i) \in \Theta} L(r|\mathfrak{M}, \sigma). \quad (7)$$

It is clearly seen that to accurately classify modulation formats, the maximum likelihood method requires perfect pre-known parameters, which are sometimes unachievable in practice.

For the signal classification in this work, the continuous variations of the SEFDM signal waveform parameter  $\alpha$  make maximum likelihood estimate impossible. Therefore, intelligent receivers, relying on artificial intelligence, is more desirable. For modulation classification, intelligent solutions have been proposed and implemented in wireless-domain [27] and later applied in optical-domain [28–32].

Signal classification in this work is more challenging than modulation format classification. Traditional modulation classification aims for single-carrier signals, which will more easily show distinguishable signal features when different modulation formats are applied. However, the signal classification in this work targets on multi-carrier signals and all single-carrier distinguishable features will disappear when multiple sub-carriers overlap. It is noted that perfect signal classification will accurately demultiplex multi-carrier signals to multiple single-carrier signals, after which the traditional modulation classification can be applied. Therefore, signal classification works regardless of modulation format but determines the subsequent modulation classification.

Rather than manually extracting features via traditional machine learning algorithms for the signal classification, this work applies deep learning since it can automatically extract sophisticated signal hidden features, in which unique features are applied to identify each signal. One representative deep learning (DL) method is CNN, which employs multiple convolutional layers for automatic feature extraction and its feature extraction principle is mathematically explained below

$$\begin{aligned} L_0 &= f(Y; \phi_0, \phi_1, \dots, \phi_{K-1}) \\ L_1 &= f(L_0; \omega_0, \omega_1, \dots, \omega_{K-1}) \\ &\vdots \\ L_{\Omega-1} &= f(L_{\Omega-2}; \sigma_0, \sigma_1, \dots, \sigma_{K-1}), \end{aligned} \quad (8)$$

where  $L_0$  indicates the feature maps after the 1st convolutional layer operations,  $f(\cdot)$  indicates a convolution operation,  $Y$  is the received signal immediately after the ADC module,  $(\phi_0, \phi_1, \dots, \phi_{K-1})$  represents  $K$  feature filters at the 1st convolutional layer. There will be  $K$  convolution operations between  $Y$  and  $\phi$ . Therefore,  $L_0$  will include  $K$  feature maps. The similar operations will be repeated for the following convolutional layers until all the  $\Omega$  convolutional layers are went through. The final feature maps are noted as  $L_{\Omega-1}$ , which will be used by a classification function as below

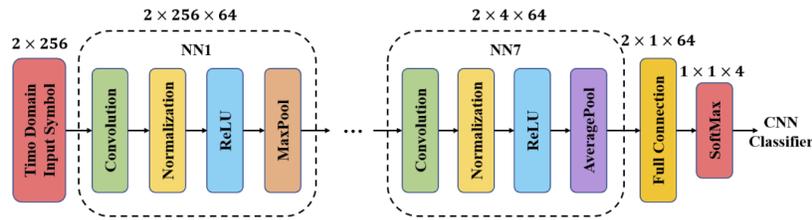
$$C_{cnn} = \xi(L_{\Omega-1}; W), \quad (9)$$

where  $\xi(\cdot)$  represents classification function and  $W$  indicates full connection functions.

It is noted that the input to the CNN feature learning process Eq. (8) is the original signal  $Y$ , which is not pre-processed by any compensation algorithms. Unlike the maximum likelihood method Eq. (6) requiring perfect signal conditions, the CNN will automatically learn hidden features that are robust to channel/hardware impairments such as imperfect timing, CD effect and fiber non-linearity.

The specific CNN architecture demonstrated in Fig. 5 follows the previous work in [33] where multiple neural network (NN) sub-blocks are packed to realize a deep neural network structure. This work will configure multi-carrier signals with 512 sub-carriers. Each input time-domain training symbol is randomly truncated to emulate imperfect timing conditions. This work considers a 50% truncation scheme and therefore the size of the effective input symbol will be 256. Taking into account QPSK modulated OFDM/SEFDM symbols, real and imaginary part of a symbol will lead to a  $2 \times 256$  input symbol matrix.

For the first six NN sub-blocks, four sub-layers are included, namely convolutional layer, normalization layer, ReLU layer and MaxPool layer. Each layer has its unique function. The convolutional layer works as a filter. This work designs  $K=64$  feature filters and after the



**Fig. 5.** CNN signal classifier architecture.

convolutional layer there will be 64 independent feature maps. Therefore, the dimension of the first NN sub-block becomes  $2 \times 256 \times 64$ . Each feature map will contain useful hidden feature information. The convolutional layer output will be normalized and then fed to the ReLU activation function. The MaxPool layer is used to down sample the ReLU layer output and reserve extreme features. The dimension of each NN sub-block will be decreased and it will be simplified into  $2 \times 4 \times 64$  at the last NN sub-block. The same sub-layer structure will be followed for other NN sub-blocks except the last NN sub-block, in which an AveragePool layer is applied instead of the MaxPool layer. The reason is to get smooth features here instead of extreme features. Once the hidden feature is extracted, a full connection layer and a SoftMax layer will work together for signal classification. The cross-entropy loss between predicted signal classes and true signal classes will be minimized by the stochastic gradient descent with momentum (SGDM) optimizer. After pre-defined training iterations with backpropagation operations, the optimal CNN classifier will be obtained.

## 5.2. Methodology of training

The CNN classifier training is offline since the non-linear convolutional feature learning process is time-consuming. Once the model is learnt, it will be reused for testing without any further re-training. This work will evaluate various training methodologies in order to find the most efficient scheme when ASE, fiber non-linearity and signal non-orthogonality are jointly taken into account.

At the training stage, training dataset size would affect training accuracy and the number of training symbols per signal class (TSPSC) will be considered. We start from TSPSC=200 OFDM/SEFDM symbols for each signal class at one specific launch power for a specific fiber length. Since we consider a wide range of launch power from  $-20$  dBm to  $24$  dBm with an increment step of  $4$  dB, therefore the total number of training symbols for each signal class is  $2,400$ .

**Table 2.** CNN classifier training configurations

Parameter	Configurations
Signal class	OFDM, SEFDM ( $\alpha = 0.9, 0.8, 0.7$ )
Signal time sample	512
Samples per symbol ( $\rho$ )	2
Launch power (dBm)	$-20:24, 4$ dB step
Training dataset size (TSPSC)	200, 1000, 2000
Number of Fiber Span	1, 10, 20, 30, 40, 50
Number of training epochs	30
Mini-Batch size	128
Learning rate	0.02
Optimizer	SGDM

For simplicity, this work considers four signal classes and the total number of training symbols for each training is 9,600. For other training dataset size of TSPSC=1,000 OFDM/SEFDM symbols and TSPSC=2,000 OFDM/SEFDM symbols for each signal class, the total number of training symbols will be 48,000 and 96,000, respectively. The testing stage follows the same data generation methodology as the training stage. The length of optical fiber is another factor that would affect the training quality. This work will consider a fiber span of 80 km and up to 50 spans will be simulated. Therefore, the simulation fiber distance will be between 80 km and 4000 km. The detailed impact of fiber length on training model accuracy will be explained in the next section. The choice of training dataset launch power is also an impact factor. A low launch power indicates strong effect from ASE and a high launch power will cause fiber non-linearity issue. The optimal way of using launch power for the training will be discussed in the next section. The detailed configurations for the CNN classifier training is presented in Table 2.

## 6. Investigations on classification accuracy

The performance of CNN classifier is related to neural network hyperparameters, which are determined by multiple training factors. This section will evaluate the impact of different training factors on the signal classification accuracy.

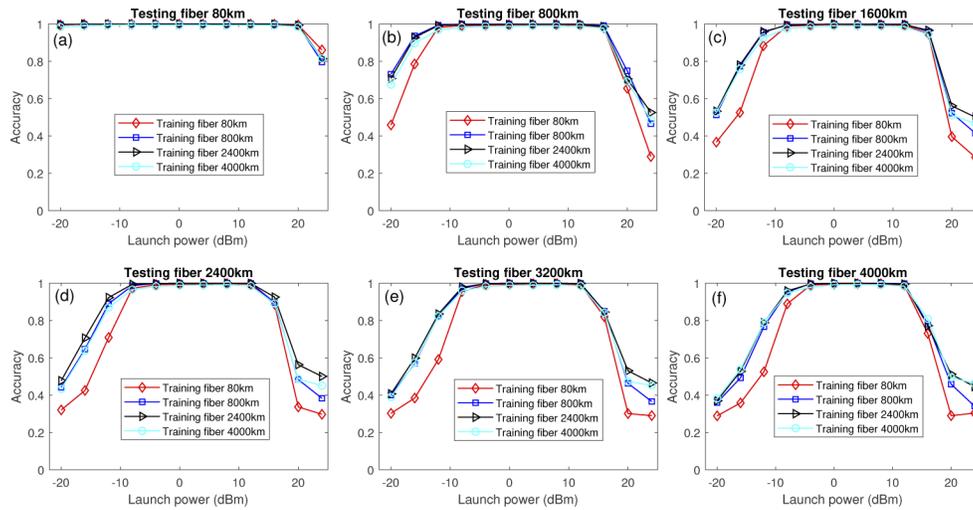
### 6.1. Impact of fiber length

Fiber length variations in long-haul optical fiber communications will lead to different performance due to accumulated ASE introduced by EDFA. In practice, it would be challenging to frequently re-train a classifier for an optical communication with a specific fiber length. This indicates the necessity of training a robust classifier that can be flexibly used by any distance fiber transmission systems.

The fiber length of an optical system is defined by the number of spans and each span is 80 km distance in this work. At the beginning, we collect four datasets after 80 km (1 span), 800 km (10 spans), 2400 km (30 spans) and 4000 km (50 spans) fiber transmissions. All the training datasets are appropriately distorted by our modelled fiber channels with fiber non-linearities, CD and ASE noise. For the purpose of universal classifier training, the datasets will cover launch power from  $-20$  dBm to 24 dBm, which is also the testing launch power range. The channel impaired training data will be applied to train four separate CNN classifiers, which are later used to test optical signals after 1-span, 10-span, 20-span, 30-span, 40-span and 50-span distance fiber transmissions.

Accuracy results are presented in Fig. 6. It is apparent that the 800 km (10 spans), 2400 km (30 spans) and 4000 km (50 spans) trained classifiers can be universally applied to classify testing signals after different fiber transmission distances. A special case is the 80 km (1 span) trained classifier, which shows degraded accuracy at both low and high launch power ranges relative to other fiber span trained classifiers. The reason could be that the 80 km (1 span) data trained classifier is overfitting to the specific 80 km case. The 80 km fiber is too short to effectively include the ASE noise. In addition, the inter carrier interference (ICI) within SEFDM signals follows the Gaussian distribution [34], which makes SEFDM similar to an OFDM signal with high ASE noise. In this case, a signal would be misclassified into another signal format. Therefore, a proper amount of ASE should be emulated in the training data for universal classifier training.

This section reveals that training data of 2400 km fiber transmission is sufficient for a robust classifier that can be universally used to classify testing signals after a random fiber distance transmission. With further increase of training fiber length, no accuracy improvement is observed.



**Fig. 6.** Classifier accuracy under different training and testing fiber transmission distances. Classifiers are trained with fiber distances at 80 km, 800 km, 2400km and 4000 km. TSPSC is fixed at 1000. Training data covers launch power ranging from  $-20$  dBm to  $24$  dBm. (a) Testing fiber distance at 80 km. (b) Testing fiber distance at 800 km. (c) Testing fiber distance at 1600 km. (d) Testing fiber distance at 2400 km. (e) Testing fiber distance at 3200 km. (f) Testing fiber distance at 4000 km.

## 6.2. Impact of dataset size

As observed from previous results, the ASE dominated accuracy area (i.e. low launch power range) and non-linearity dominated accuracy area (i.e. high launch power range), classification accuracy is greatly decreased. One possible reason might be insufficient training data. Therefore, this section will investigate the impact of training dataset size on classification accuracy.

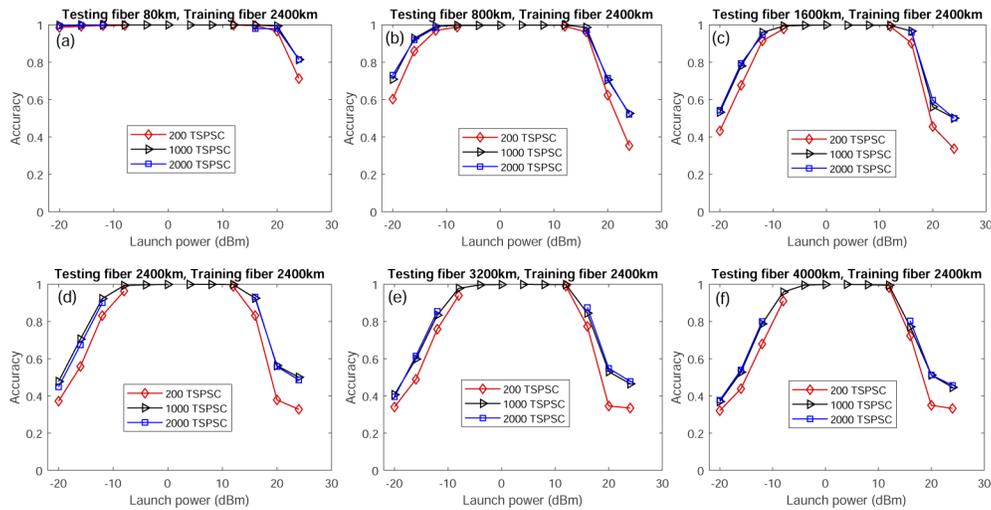
Since the optimal training fiber length is 2400 km, which is concluded from the previous observation, this section will merely focus on the 2400 km fiber data trained classifiers. For the results in Section 6.1, the training dataset size is 1,000 TSPSC. In this section, the TSPSC is extended with two new options, TSPSC=200 and TSPSC=2000.

Results in Fig. 7 reveal that TSPSC=200 can reach reasonable accuracy but better performance is achievable with a larger number of TSPSC. It shows that TSPSC=1000 can slightly improve the accuracy for all the testing fiber transmission distances at ASE dominated area and fiber non-linearity dominated area. With further increase of the training dataset size, no classification accuracy improvement is observed. Therefore, based on the results, TSPSC=1000 is the optimal training dataset size and its trained classifier is robust for different testing fiber length scenarios.

## 6.3. Impact of launch power

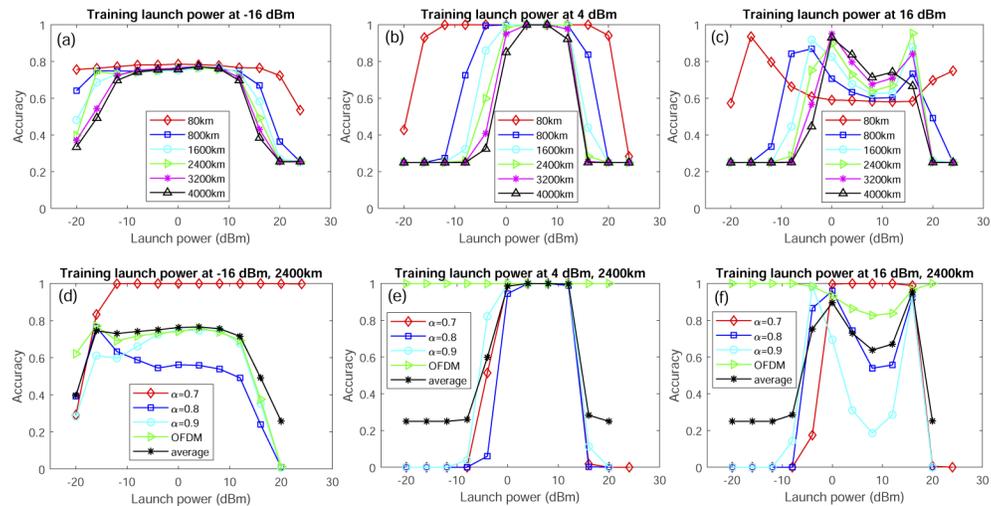
Launch power is also an important factor affecting classification accuracy since a signal needs sufficient transmission power to reach a location in a long-haul optical fiber transmission system. Previous studies focus on a wide range of training launch power covering from  $-20$  dBm to  $24$  dBm. This section will study the training impact of a single launch power on the classification accuracy. We choose three launch power candidates to test the classifier sensitivity.

Firstly,  $-16$  dBm powered data will be used to train a classifier, which will be used to classify signals ranging from  $-20$  dBm to  $24$  dBm. Since the training power is very low, the ASE effect will be significant. As shown in Fig. 8(a), The overall accuracy is reduced to below 80%. The detailed accuracy rate for each BCF  $\alpha$  is also included in Fig. 8(d), it shows that the peak accuracy



**Fig. 7.** Classifier accuracy under different training dataset sizes. TSPSC=200, 1000, 2000 are tested. Training fiber distance is fixed to 2400 km. Training data covers launch power ranging from  $-20$  dBm to  $24$  dBm. (a) Testing fiber distance at 80 km. (b) Testing fiber distance at 800 km. (c) Testing fiber distance at 1600 km. (d) Testing fiber distance at 2400 km. (e) Testing fiber distance at 3200 km. (f) Testing fiber distance at 4000 km.

rate for each  $\alpha$  is achieved at  $-16$  dBm, which is the power used for training. It is inferred that a single launch power trained classifier might only work for testing signals at that specific launch power.



**Fig. 8.** Classifier accuracy under different training launch power effects. Three launch power candidates are  $-16$  dBm,  $4$  dBm and  $16$  dBm. Training fiber distance is fixed to 2400 km. TSPSC is fixed at 1000. (a) Training launch power at  $-16$  dBm for different testing fiber lengths. (b) Training launch power at  $4$  dBm for different testing fiber lengths. (c) Training launch power at  $16$  dBm for different testing fiber lengths. (d) Training launch power at  $-16$  dBm for specific signals. (e) Training launch power at  $4$  dBm for specific signals. (f) Training launch power at  $16$  dBm for specific signals.

The second evaluated launch power is 4 dBm, which is the optimal launch power leading to the peak accuracy rate as previously shown in Fig. 6 and Fig. 7. In Fig. 8(b), the peak accuracy is obtained around the training power 4 dBm and the accuracy rate is greatly reduced towards the ASE dominated zone and the fiber non-linearity dominated zone. Its performance on each  $\alpha$  is shown in Fig. 8(e) with a similar observation that a single training launch power would merely work well in a limited range around the trained launch power.

Further moving to the third launch power 16 dBm, which is a high launch power and will cause non-linearity distortion to signals. Training on this launch power, the average accuracy performance and independent accuracy rates are demonstrated in Fig. 8(c) and Fig. 8(f), respectively. The similar observation is that the accuracy peak appears at 16 dBm, which makes sense since the classifier is trained at 16 dBm. One additional observation is that a second peak exists at low launch power. The two accuracy peaks are potentially symmetric.

It seems that the CNN classifier has learnt unique features from fiber non-linearity dominated signals. The extracted features by CNN can help to identify either ASE dominated signals or fiber non-linearity dominated signals. The results also reveal that an ASE dominated signal and a fiber non-linearity dominated signal have common features. But the fiber non-linearity dominated signal has special features that are not included by an ASE dominated signal. This can be observed from Fig. 8(d), in which only one accuracy peak is obtained and the ASE dominated signal trained classifier cannot identify signals at fiber non-linearity range.

#### 6.4. *Optimal training solution*

Three key impact factors of classifier training have been investigated for long-haul optical fiber transmissions. The main conclusion is the following:

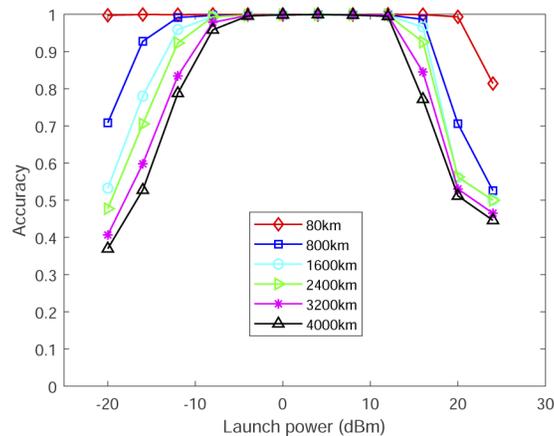
- In practice, a fiber transmission distance is flexible and training a classifier for each fiber length is not realistic. The study in this work reveals that training a classifier with a proper fiber length of 2400 km (30 spans) will be sufficient and can be universally employed to classify signals with random fiber communication distances.
- Commonly, a large number of training samples will help a black-box based deep learning classifier automatically learn and extract signal features. However, the training complexity will be increased as well. This work investigates the impact of training dataset size and reveals that a proper selection of TSPSC=1000 will be sufficient for robust classifier training. Smaller than this value will get slightly performance loss and further increase of this value will have no accuracy improvement but training complexity.
- Launch power determines a signal experiencing a strong ASE effect or a strong fiber non-linearity effect. Signals trained at a single launch power will merely work for that specific launch power in a testing stage. One exception is the classifier training by signals at fiber non-linearity dominated range. The trained classifier not only works for testing symbols at fiber non-linearity range but also achieves high accuracy for ASE dominated signals. This is due to the automatic feature learning mechanism by CNN, in which it can extract mathematically unachievable hidden features. This infers that a fiber non-linearity dominated signal has common features with an ASE dominated signal. Meanwhile, the non-linearity dominated signal might have unique features that the ASE signal doesn't have. Therefore, a robust training strategy should cover a wide range of training launch power, which ranges from -20 dBm to 24 dBm with a 4 dB increment step in this work.

The optimal and universal training methodology is summarized as

- Length of fiber: 2400 km (30 spans).
- Training dataset size: TSPSC=1000.

- Launch power range:  $-20$  dBm to  $24$  dBm with a  $4$  dB increment step.

The optimal and universal signal classifiers are trained and tested with accuracy results showing in Fig. 9, in which six fiber transmission scenarios are tested. At the low launch power range, the classification accuracy drops with the increase of fiber length. This makes sense since the effect of ASE noise is more significant for longer fiber communications. With the increase of launch power, fiber non-linearity effect starts to appear showing a similar trend as the ASE dominated situation where the accuracy drops with the increase of fiber length. In addition, the accuracy dropping rate is faster due to the joint effect of ASE noise and fiber non-linearity. The result also reveals that fiber non-linearity has more effects on accuracy than ASE noise. This is observed from the  $80$  km transmission result, which shows a high and flat accuracy rate at low launch power while the accuracy drops sharply at high launch power range. The middle area between the ASE dominated degradation zone and the non-linearity dominated degradation zone is flat and nearly  $100\%$  accuracy. This reveals that the imperfect signal classifier, due to the mismatch of training data and testing data, is robust to different fiber communication distances. This also verifies the hypothesis that ASE and fiber non-linearity dominate the classification accuracy rather than the imperfect signal classifier.



**Fig. 9.** Classification accuracy under different fiber transmission distances using the optimally and universally trained signal classifier.

Figure 10 shows the accuracy for each individual signal class. It is observed that each signal class has its own accuracy performance. The signals labelled with  $\alpha=0.8$  and  $\alpha=0.9$  show the worst accuracy rates because they could be misclassified into adjacent signal classes. For example, the one with  $\alpha=0.8$  could be misclassified into either  $\alpha=0.7$  or  $\alpha=0.9$ . Similarly, the one with  $\alpha=0.9$  could be mistakenly identified into either  $\alpha=0.8$  or OFDM. However, for the signals with  $\alpha=0.7$  and OFDM, they can achieve better accuracy since only one side misclassification is possible. The reduced misclassification possibility will lead to increased accuracy. The detailed classification accuracy can be visualized using confusion matrix. Three confusion matrices, evaluated launch power at  $-16$  dBm,  $4$  dBm and  $16$  dBm are presented in Fig. 11, Fig. 12 and Fig. 13, respectively.

As shown in Fig. 11, four true signal classes are compared with their predicted classes. The diagonal cells indicate correct classification without any errors and the non-diagonal cells indicate incorrect classification. The bottom row indicates the identification ratios of all the predicted signal classes belonging to a given true signal class that are correctly (upper ratio) or incorrectly (lower ratio) classified. The far right column indicates the identification ratios of all the predicted signal classes belonging to a given predicted signal class that are correctly (upper ratio) or

incorrectly (lower ratio) classified. The first column indicates the identification accuracy rates for OFDM signal class when  $\alpha=1$ . It is shown that 833 out of 1000 OFDM symbols are correctly classified leading to the classification accuracy rate 83.3%. The rest of symbols are incorrectly classified into other classes. The second, third and fourth columns are for SEFDM signals when  $\alpha=0.7, 0.8$  and  $0.9$ , respectively. Their correct classification accuracy rates are 90%, 57.9% and 48.5%.

Viewing the confusion matrix in Fig. 11 horizontally, 833 correctly predicted OFDM signals out of all the predicted OFDM signals leads to a correct classification rate of 67.9%, which is another way to judge the quality of the classifier accuracy. In addition, the percentage in each cell is the ratio among all the testing symbols. In this work, the testing symbols per signal class is 1000, therefore there will be overall 4000 testing symbols. Those 833 correctly classified OFDM symbols out of 4000 will correspond to 20.8%. It is noted that the cell on the bottom right shows the overall correct and incorrect classification ratios to be 69.9% and 30.1%, respectively.

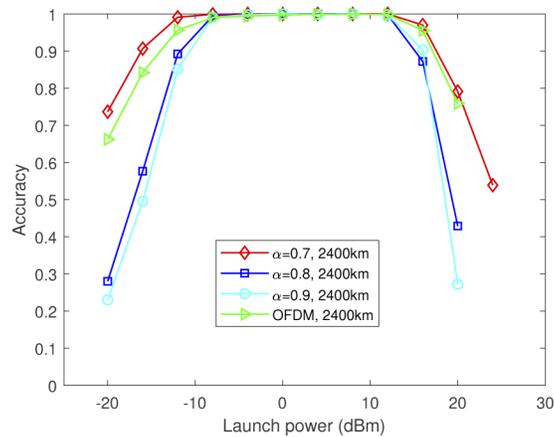


Fig. 10. Classification accuracy for each signal class ( $\alpha$ ) testing at 2400 km fiber transmission distance.

**Testing data at -16 dBm launch power**

Predicted Class	1	833 20.8%	30 0.8%	85 2.1%	278 7.0%	67.9% 32.1%
	0.7	47 1.2%	900 22.5%	256 6.4%	110 2.8%	68.5% 31.5%
	0.8	25 0.6%	58 1.5%	579 14.5%	127 3.2%	73.4% 26.6%
	0.9	95 2.4%	12 0.3%	80 2.0%	485 12.1%	72.2% 27.8%
		83.3% 16.7%	90.0% 10.0%	57.9% 42.1%	48.5% 51.5%	69.9% 30.1%
	$\alpha$	0.7	0.8	0.9		
		True Class				

Fig. 11. Confusion matrix for testing data in Fig. 10 at -16 dBm launch power.

**Testing data at 4 dBm launch power**

Predicted Class	1	1000 25.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
	0.7	0 0.0%	1000 25.0%	0 0.0%	0 0.0%	100% 0.0%
	0.8	0 0.0%	0 0.0%	1000 25.0%	0 0.0%	100% 0.0%
	0.9	0 0.0%	0 0.0%	0 0.0%	1000 25.0%	100% 0.0%
		100% 0.0%	100% 0.0%	100% 0.0%	100% 0.0%	100% 0.0%
	^	0.1	0.8	0.9		True Class

**Fig. 12.** Confusion matrix for testing data in Fig. 10 at 4 dBm launch power.

**Testing data at 16 dBm launch power**

Predicted Class	1	962 24.1%	0 0.0%	2 0.1%	42 1.1%	95.6% 4.4%
	0.7	0 0.0%	963 24.1%	91 2.3%	7 0.2%	90.8% 9.2%
	0.8	0 0.0%	37 0.9%	873 21.8%	41 1.0%	91.8% 8.2%
	0.9	38 0.9%	0 0.0%	34 0.9%	910 22.8%	92.7% 7.3%
		96.2% 3.8%	96.3% 3.7%	87.3% 12.7%	91.0% 9.0%	92.7% 7.3%
	^	0.1	0.8	0.9		True Class

**Fig. 13.** Confusion matrix for testing data in Fig. 10 at 16 dBm launch power.

It is apparent from Fig. 11 that OFDM and SEFDM of  $\alpha=0.7$  have higher correct accuracy rates than the other two signal classes since these two signals can only be misclassified into their one side signal classes. In the first column, OFDM signal is easily misclassified into SEFDM signals with  $\alpha=0.9$ . This is also the case for  $\alpha=0.7$  since the highest misclassification rate occurs at  $\alpha=0.8$ . However, for  $\alpha=0.9$  and  $\alpha=0.8$ , since they can be incorrectly classified into adjacent signal classes on both sides. Therefore, these two signals have doubled chance for misclassification and therefore lower accuracy. The confusion matrix gives detailed explanation on the results in Fig. 10, in which OFDM and SEFDM with  $\alpha=0.7$  outperform the other two signal classes at  $-16$  dBm launch power within the ASE dominated range.

Increasing the launch power to 4 dBm, perfect classification is shown in Fig. 12. Both individual accuracy rate and overall accuracy rate are 100%. This also reflects the results obtained in Fig. 10. Further increasing launch power to 16 dBm, a similar trend is observed in Fig. 13 compared with the results obtained in Fig. 11. The details of correct and incorrect classification accuracy rates in each cell explains the results at 16 dBm in Fig. 10.

In summary, the results in Fig. 9 give us a general idea on the signal classification accuracy under various optical fiber transmission scenarios. The performance study on each signal class in Fig. 10 gives us a detailed instruction on how to design the signal distribution more efficiently since each signal class has different sensitivity to ASE noise and fiber non-linearities. At the reasonable launch power range between  $-8$  dBm and 12 dBm, four signal classes could be used since they all achieve nearly 100% accuracy. Therefore, a convincing design should be within this regime. In an ASE dominated optical transmission system, the signals of  $\alpha=0.7$  and OFDM could be used since they have much higher accuracy than the others. The same strategy is also for fiber non-linearity dominated systems at beyond 12 dBm launch power, in which  $\alpha=0.7$  and OFDM could have better accuracy rates.

### 6.5. Explanations on feature patterns

To illustrate unique features learnt by the CNN in this work, new results are obtained by outputting the patterns from the last convolutional layer, which is the NN7 block in Fig. 5. In principle, all seven convolutional layers can output feature patterns. As shown in Fig. 5, the output dimension of the last convolutional layer is  $2 \times 4 \times 64$ . It is noted that the  $2 \times 4$  feature matrix is a downsampled version from the initial  $2 \times 256$  input matrix (considering both real and imaginary signal parts). Due to the pooling operation after each convolution operation, a small size (high-level) feature matrix will be obtained. The third dimension, 64, indicates the number of feature filters. Here, due to space limitations, we will only extract and show patterns from the first four feature filters (i.e. not showing the other 60), each with feature patterns of size  $2 \times 4$ . For an OFDM signal input, the feature pattern is shown in the first row of Fig. 14. To have a fair and intuitive comparison, we limit the heatmap range from  $-1$  to  $+1$ . In this case, any values beyond  $+1$  will be rounded to  $+1$  and any values smaller than  $-1$  will be rounded to  $-1$ . This will thereby avoid minor differences and limit extreme values. The same operations are repeated for SEFDM (with  $\alpha=0.9, \alpha=0.8$ ,

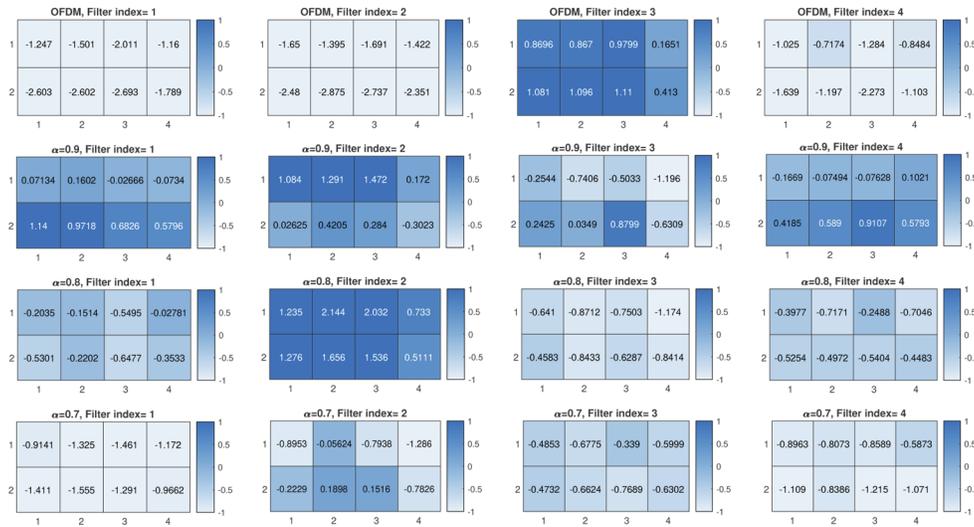
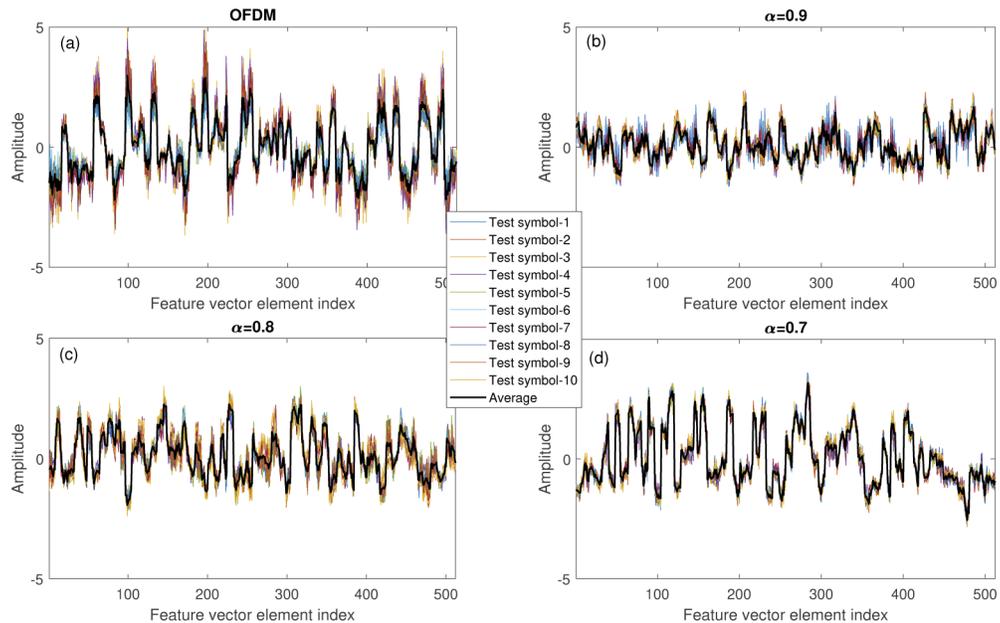


Fig. 14. Feature maps illustration after the convolutional layer in NN7.

$\alpha=0.7$ ). Comparing feature maps of OFDM and SEFDM signals, it is apparent that feature maps have different patterns, indicating that specific signals have distinct features and can accordingly be identified. The difference in features is key to the signal identification.

Figure 14 presents feature maps for a single input symbol per signal class. We have also tested more symbols per signal class, to show the robustness of learnt features. Results of these tests are shown in Fig. 15. To simplify the feature illustrations, we flatten the  $2 \times 4 \times 64$  three-dimensional object to a  $512 \times 1$  one-dimensional vector. In this case, all 64 feature maps will be preserved in the  $512 \times 1$  one-dimensional vector. For each signal class, we have tested 10 random symbol inputs. In this case, there will be 10 output feature vectors. We superimpose these vectors in one figure, together with a plot of their average feature vector. It is clearly seen that for a given signal class, output feature vectors are very similar between 10 random input symbols, and they all overlap with the average feature vector. It is inferred that the CNN tries to learn a neural network architecture that can output a static feature vector independent of the variance of input symbols, once a signal class is determined. Based on the static feature output, a classifier can easily decide the original input signal class. Similarly, unique feature vectors are obtained for other signal classes in Fig. 15.



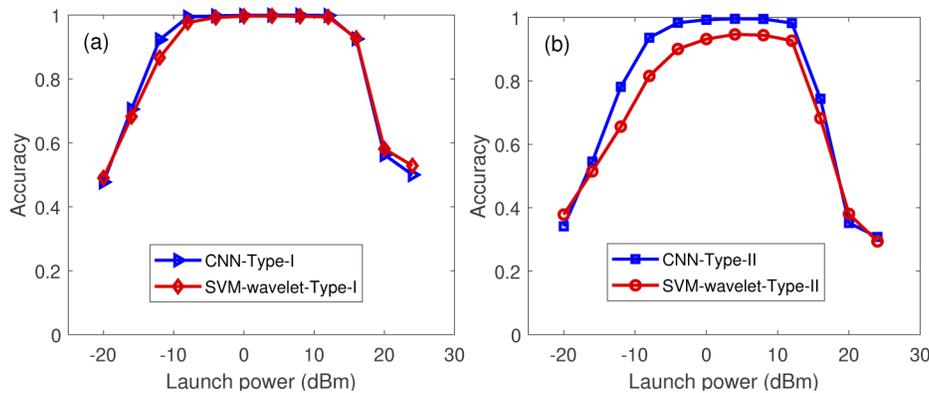
**Fig. 15.** Flatten feature vectors illustration after the convolutional layer in NN7.

Therefore, unlike the image classification by CNN, the signal features learnt by CNN are abstract, in a numerical sense. It is difficult to intuitively figure out the meaning of extracted signal features. However, the CNN can learn a neural network architecture that can output deterministic features for a specific signal class. This indicates that once the value of  $\alpha$  is determined, any input symbols to the CNN classifier will output a fixed and exclusive feature pattern. This allows the CNN classifier to decide the input signal class and therefore facilitates the correct signal detection.

#### 6.6. Comparison with support vector machine (SVM) classification

In addition to the CNN based signal classification, other machine learning based signal classification is also possible for SEFDM signals. One representative machine learning method

is support vector machine (SVM). Therefore, here we will compare CNN with SVM in the performance of signal classification. The initial purpose of using CNN in this work is due to its 'full' intelligence where signal features will be automatically learnt. However, SVM relies on manual feature engineering and its performance would be variable according to expert knowledge. To assist the SVM classification, we manually extract two-dimensional time-frequency features using wavelet transform. The accuracy comparisons are shown in Fig. 16. For the scenario in Fig. 16(a), Type-I signal patterns from Table 1 are considered with signal classes of OFDM and SEFDM (with  $\alpha=0.9, 0.8, 0.7$ ). It is found that the SVM-wavelet classifier can achieve similar performance as the CNN classifier but with minor performance loss at low launch power regime. This indicates that SVM may work well for the SEFDM signal classification. However, when challenging Type-II signal patterns are considered in Fig. 16(b) where more similar signal classes of OFDM and SEFDM (with  $\alpha=0.95, 0.9, 0.85, 0.8, 0.75, 0.7$ ) are designed, SVM starts to show its weakness. It is shown in Fig. 16(b) that SVM has lower accuracy at both low launch power regime and high launch power regime. More importantly, in the middle launch power regime, SVM can not reach 100% accuracy any more. The CNN classifier is still robust to achieve 100% accuracy. The performance difference is due to the inefficient manual feature extractions in the SVM classification while CNN can automatically adjust the neural network architecture and still achieves the optimal performance. It is therefore inferred that CNN is more flexible and robust in terms of feature extractions.



**Fig. 16.** Comparisons between deep learning CNN classification and machine learning SVM-wavelet classification. Type-I signal patterns: OFDM, SEFDM ( $\alpha=0.9, 0.8, 0.7$ ). Type-II signal patterns: OFDM, SEFDM ( $\alpha=0.95, 0.9, 0.85, 0.8, 0.75, 0.7$ ).

### 6.7. Further discussion

The proposed intelligent receiver signal distribution is based on the spectral compression label. The label, which is defined by the value of  $\alpha$ , will determine the performance of the distribution. This work gives two example solutions, in which four signal classes and seven signal classes are designed with an equal label gap,  $\Delta\alpha$ . Due to the continuous variations of  $\alpha$ , the number of signal classes could be infinite. In addition, the signal class label gap  $\Delta\alpha$  could be unequal. Therefore, it indicates that the number of supported users could be large as well. However, with the reducing gap of  $\Delta\alpha$ , signal identification will be more challenging, which has been briefly studied in the previous section. The optimal configurations of  $\alpha$  are beyond the scope of this work and this work will merely demonstrate the concept and extended proposals of integrating more signal classes could be investigated in the future.

Meanwhile, our simulations were carried out based on point-to-point transmission links. This is used to study the feasibility of the signal distribution in private optical networks. Therefore,

we designed the simulation setup and investigated the signal recognition ability at the receiver side for different transmission scenarios. In practical optical networks, add-drop functions can be implemented using reconfigurable optical add-drop multiplexers (ROADMs) based on the switching of wavelength channels. In this case, the linear and nonlinear cross-talks, including Kerr nonlinearities, between wavelength channels and the loss in ROADMs will be major additional impairments besides the transmission distortions considered in the discussed simulations. The impact of these additional impairments will be investigated in our future work.

## 7. Conclusion

This work proposed an intelligent signal distribution scheme for private optical networks based on the non-orthogonal waveform principle. Spectral efficiency is improved via compressing signals to fit in a bandwidth-limited scenario. The spectral bandwidth compression ratio is also used as end user labelling for signal distribution. The scheme will help to forward signals automatically to unique end users. This work investigated the ability of artificial intelligence for identifying different signals with no signalling control overheads. In this case, signalling of the entire network will be simplified leading to a non-cooperative communication network. To train a universal signal classifier that can work for a randomly received signal over the network, multiple long-haul optical environmental factors are evaluated. Results show that a training dataset, in which each signal class has 1000 training symbols, impaired by transmission over a 2400 km fiber and covering launch power from  $-24$  dBm to 20 dBm, will be robust to classify all received signals within a private optical network.

**Funding.** Engineering and Physical Sciences Research Council (EP/S028455/1); Horizon 2020 Framework Programme (101008280).

**Disclosures.** The authors declare that there are no conflicts of interest related to this article.

**Data availability.** Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

## References

1. G. Zhang, M. De Leenheer, A. Morea, and B. Mukherjee, "A survey on OFDM-based elastic core optical networking," *IEEE Commun. Surv. Tutorials* **15**(1), 65–87 (2013).
2. B. Kantarci and H. T. Mouftah, "Bandwidth distribution solutions for performance enhancement in long-reach passive optical networks," *IEEE Commun. Surv. Tutorials* **14**(3), 714–733 (2011).
3. A. Napoli, M. Bohn, D. Rafique, A. Stavdas, N. Sambo, L. Poti, M. Nölle, J. K. Fischer, E. Riccardi, A. Pagano, A. Di Giglio, M. S. Moreolo, J. M. Fabrega, E. Hugues-Salas, G. Zervas, D. Simeonidou, P. Layec, A. D'Errico, T. Rahman, and J. P. F. Giménez, "Next generation elastic optical networks: The vision of the european research project IDEALIST," *IEEE Commun. Mag.* **53**(2), 152–162 (2015).
4. B. C. Chatterjee, N. Sarma, and E. Oki, "Routing and spectrum allocation in elastic optical networks: A tutorial," *IEEE Commun. Surv. Tutorials* **17**(3), 1776–1800 (2015).
5. C. Jiang, H. Zhang, Y. Ren, Z. Han, K. C. Chen, and L. Hanzo, "Machine learning paradigms for next-generation wireless networks," *IEEE Wireless Commun.* **24**(2), 98–105 (2017).
6. T. O'Shea and J. Hoydis, "An introduction to deep learning for the physical layer," *IEEE Trans. Cogn. Commun. Netw.* **3**(4), 563–575 (2017).
7. F. N. Khan, Q. Fan, C. Lu, and A. P. T. Lau, "An optical communication's perspective on machine learning and its applications," *J. Lightwave Technol.* **37**(2), 493–516 (2019).
8. T. Nguyen, S. Mhatli, E. Giacomidis, L. Van Compernelle, M. Wuilpart, and P. Mégret, "Fiber nonlinearity equalizer based on support vector classification for coherent optical OFDM," *IEEE Photonics J.* **8**(2), 1–9 (2016).
9. 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 Photonics Technol. Lett.* **28**(17), 1886–1889 (2016).
10. F. Musumeci, C. Rottondi, A. Nag, I. Macaluso, D. Zibar, M. Ruffini, and M. Tornatore, "An overview on application of machine learning techniques in optical networks," *IEEE Commun. Surv. Tutorials* **21**(2), 1383–1408 (2019).
11. R. Munoz, R. Martínez, and R. Casellas, "Challenges for GMPLS lightpath provisioning in transparent optical networks: Wavelength constraints in routing and signaling," *IEEE Commun. Mag.* **47**(8), 26–34 (2009).
12. P. J. Winzer, D. T. Neilson, and A. R. Chraplyvy, "Fiber-optic transmission and networking: the previous 20 and the next 20 years (invited)," *Opt. Express* **26**(18), 24190–24239 (2018).
13. 3GPP TS 36.213 v.14.2.0, "LTE: Evolved universal terrestrial radio access (e-utra), physical layer procedures, Tech. rep. (2017).,"

14. 3GPP TS 38.212 version 15.2.0, "5G NR: Multiplexing and channel coding, Tech. rep. (2018).,"
15. T. Xu and I. Darwazeh, "Transmission experiment of bandwidth compressed carrier aggregation in a realistic fading channel," *IEEE Trans. Veh. Technol.* **66**(5), 4087–4097 (2017).
16. J. Anderson, F. Rusek, and V. Öwall, "Faster-than-Nyquist signaling," *Proc. IEEE* **101**(8), 1817–1830 (2013).
17. B. Farhang-Boroujeny, "OFDM versus filter bank multicarrier," *IEEE Signal Process. Mag.* **28**(3), 92–112 (2011).
18. N. Michailow, M. Matthe, I. Gaspar, A. Caldevilla, L. Mendes, A. Festag, and G. Fettweis, "Generalized frequency division multiplexing for 5th generation cellular networks," *IEEE Trans. Commun.* **62**(9), 3045–3061 (2014).
19. I. Darwazeh, T. Xu, T. Gui, Y. Bao, and Z. Li, "Optical SEFDM system; bandwidth saving using non-orthogonal sub-carriers," *IEEE Photonics Technol. Lett.* **26**(4), 352–355 (2014).
20. J. Zhou, Y. Qiao, Z. Yang, Q. Cheng, Q. Wang, M. Guo, and X. Tang, "Capacity limit for faster-than-Nyquist non-orthogonal frequency-division multiplexing signaling," *Sci. Rep.* **7**(1), 3380 (2017).
21. T. Xu, S. Mikroulis, J. E. Mitchell, and I. Darwazeh, "Bandwidth compressed waveform for 60-GHz millimeter-wave radio over fiber experiment," *J. Lightwave Technol.* **34**(14), 3458–3465 (2016).
22. H. Ghannam, D. Nopchinda, M. Gavell, H. Zirath, and I. Darwazeh, "Experimental demonstration of spectrally efficient frequency division multiplexing transmissions at E-Band," *IEEE Trans. Microwave Theory Tech.* **67**(5), 1911–1923 (2019).
23. Y. Wang, Y. Zhou, T. Gui, K. Zhong, X. Zhou, L. Wang, A. P. T. Lau, C. Lu, and N. Chi, "Efficient MMSE-SQRD-based MIMO decoder for SEFDM-based 2.4-Gb/s-spectrum-compressed WDM VLC system," *IEEE Photonics J.* **8**(4), 1–9 (2016).
24. Z. Gao, S. Yan, J. Zhang, M. Mascarenhas, R. Nejabati, Y. Ji, and D. Simeonidou, "ANN-based multi-channel QoT-prediction over a 563.4-km field-trial testbed," *J. Lightwave Technol.* **38**(9), 2646–2655 (2020).
25. G. Agrawal, *Nonlinear Fiber Optics* (Academic, 2013), 5th ed.
26. F. Hameed, O. A. Dobre, and D. C. Popescu, "On the likelihood-based approach to modulation classification," *IEEE Trans. Wireless Commun.* **8**(12), 5884–5892 (2009).
27. T. J. O'Shea, T. Roy, and T. C. Clancy, "Over-the-air deep learning based radio signal classification," *IEEE J. Sel. Top. Signal Process.* **12**(1), 168–179 (2018).
28. F. N. Khan, Y. Zhou, A. P. T. Lau, and C. Lu, "Modulation format identification in heterogeneous fiber-optic networks using artificial neural networks," *Opt. Express* **20**(11), 12422–12431 (2012).
29. X. Lin, Y. A. Eldemerdash, O. A. Dobre, S. Zhang, and C. Li, "Modulation classification using received signals amplitude distribution for coherent receivers," *IEEE Photonics Technol. Lett.* **29**(21), 1872–1875 (2017).
30. D. Wang, M. Zhang, Z. Li, J. Li, M. Fu, Y. Cui, and X. Chen, "Modulation format recognition and OSNR estimation using CNN-based deep learning," *IEEE Photonics Technol. Lett.* **29**(19), 1667–1670 (2017).
31. D. Wang, M. Zhang, J. Li, Z. Li, J. Li, C. Song, and X. Chen, "Intelligent constellation diagram analyzer using convolutional neural network-based deep learning," *Opt. Express* **25**(15), 17150–17166 (2017).
32. L. Jiang, L. Yan, A. Yi, Y. Pan, M. Hao, W. Pan, and B. Luo, "An effective modulation format identification based on intensity profile features for digital coherent receivers," *J. Lightwave Technol.* **37**(19), 5067–5075 (2019).
33. T. Xu and I. Darwazeh, "Deep learning for over-the-air non-orthogonal signal classification," in *Proceedings of IEEE Vehicular Technology Conference*, (IEEE, 2020), pp. 1–5.
34. H. Ghannam and I. Darwazeh, "SEFDM: Spectral efficiency upper bound and interference distribution," in *Proceedings of International Symposium on Communication Systems, Networks & Digital Signal Processing*, (IEEE, 2018), pp. 1–6.