*IET Networks*

WILEY

The Institution of Engineering and Technology

ORIGINAL RESEARCH OPEN ACCESS

# MT-OPMNet: Attention-Enhanced Multi-Task Deep Learning for Joint OSNR Estimation and Modulation Format Recognition in Elastic Optical Networks

Yassir Al-Karawi[1] | Ahmed M. Jasim[1] | Raad S. Alhumaima[1,2] | Hamed S. Al-Raweshidy[3]

[1]University of Diyala, College of Engineering, Baqubah, Diyala Governorate, Iraq | [2]Department of Cyber Security Techniques, Imam Ja'afar Al-Sadiq University, Baghdad, Baghdad Governorate, Iraq | [3]Brunel University of London, College of Engineering, Design and Physical Sciences, Uxbridge, Middlesex, UK

**Correspondence:** Hamed S. Al-Raweshidy (Hamed.Al-Raweshidy@brunel.ac.uk)

**Received:** 21 May 2026 | **Revised:** 8 July 2026 | **Accepted:** 4 September 2026

**Keywords:** deep learning | elastic optical networks | modulation format identification | multi-task learning | optical performance monitoring | OSNR estimation

## ABSTRACT

Optical performance monitoring (OPM) underpins quality of transmission in modern elastic optical networks. This paper introduces MT-OPMNet, a multi-task deep learning framework for joint optical signal-to-noise ratio (OSNR) estimation and modulation format identification (MFI) from asynchronous amplitude histograms extracted at coherent receivers. The model combines a shared one-dimensional convolutional backbone, a channel-aware attention module and two task-specific heads for parameter-efficient inference. Validation uses a corrected nine-channel wavelength-division-multiplexed simulation dataset covering distances from 500 to 3000 km, symbol rates of 28 GBd and 64 GBd, five launch-power levels and five modulation formats. At 28 GBd, MT-OPMNet achieves an OSNR root-mean-square error of approximately 2.4 dB; over the grouped mixed-rate test split, it achieves an average OSNR RMSE of 3.4 dB and an MFI accuracy of 97.5%, while reducing trainable parameters by 23.8% relative to two separate single-task models. The analytical channel model is further cross-validated using a split-step Fourier waveform simulator, achieving mean histogram cosine similarity of 0.976 and zero-shot transfer to SSFM-generated waveforms with 100% MFI accuracy and 2.1 dB OSNR RMSE over 14–22 dB OSNR. CPU inference latency remains suitable for low-latency monitoring studies.

## 1 | Introduction

The rapid growth of data-intensive services has driven the deployment of elastic optical networks (EONs) that support heterogeneous traffic demands through flexible spectrum allocation and adaptive modulation [1, 2]. In such dynamic network environments, maintaining acceptable quality of transmission (QoT) requires continuous and accurate optical performance monitoring (OPM) at multiple points along the optical path [3]. OPM functions include, but are not limited to, estimation of the optical signal-to-noise ratio (OSNR), identification of the modulation format in use, residual chromatic-dispersion (CD) assessment, nonlinear-noise indication and recognition of soft failures before they escalate into hard outages [4]. Polarisation-mode dispersion (PMD) is not modelled in this study and is therefore not claimed as a monitored quantity.

Traditional OPM techniques depend on dedicated optical spectrum analysers or specialised transponder-level signal processing, both of which incur significant capital and operational expenditure [5]. As networks scale in capacity and heterogeneity, hardware-centric monitoring becomes impractical.

Machine learning (ML) has therefore emerged as a compelling alternative, enabling cost-effective and software-defined OPM by extracting impairment signatures directly from readily available signal representations such as eye diagrams, amplitude histograms or constellation diagrams [6, 7].

Early ML-based OPM studies employed shallow classifiers, including support vector machines (SVMs) and artificial neural networks (ANNs) with one or two hidden layers, achieving moderate accuracy for isolated monitoring tasks [8, 9]. However, these models require hand-crafted features and do not scale well to joint estimation problems that arise in practical network scenarios. More recently, deep learning (DL) architectures, particularly convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have demonstrated significant improvements in OSNR estimation and modulation format identification (MFI) when trained on raw or minimally preprocessed signal representations [10–12].

Despite these advances, two limitations persist. First, the majority of existing DL-based OPM solutions treat each monitoring task in isolation, deploying separate models for OSNR estimation and MFI. This single-task paradigm leads to redundant feature extraction, increased memory footprint and higher inference latency when multiple monitoring functions must be executed concurrently [13]. Second, few studies explicitly address the challenge of generalisation across diverse operating conditions, including varying transmission distances, symbol rates and channel loads. Models trained on a narrow parameter range often exhibit degraded performance when deployed under previously unseen conditions [14].

Multi-task learning (MTL) offers a principled solution to both limitations. By sharing a common feature representation across correlated tasks, MTL can improve generalisation, reduce model complexity and lower computational cost [15]. The application of MTL to OPM has received growing but still limited attention in the photonics literature. A deep-neural-network-based scheme for joint OSNR monitoring and modulation format identification in digital coherent receivers was reported in Ref. [16], but it did not employ a convolutional backbone or attention-based feature recalibration. A CNN-based multi-task learning architecture for simultaneous optical performance monitoring and bit-rate/modulation-format identification was presented in Ref. [17]. However, it was evaluated under narrower operating conditions and did not incorporate channel-attention mechanisms. More recent attention- and multi-task-based OPM studies have expanded the state of the art, yet an AAH-based joint OSNR–MFI framework that combines convolutional feature sharing, channel-aware attention, uncertainty-weighted optimisation, multi-rate/multi-distance validation and explicit complexity analysis remains insufficiently explored.

To address these limitations, this paper proposes a multi-task deep learning framework, referred to as MT-OPMNet, for joint OSNR estimation and modulation format recognition in elastic optical networks. The scope is intentionally limited to two receiver-side tasks that are strongly coupled in AAH space. The method is not presented as a complete multi-parameter OPM engine; rather, it is a compact architecture that can be extended by adding further task heads when appropriately labelled CD, nonlinear-noise or soft-failure data are available. The main contributions of this work are as follows.

1. A multi-task convolutional architecture (MT-OPMNet) is proposed, comprising a shared feature extractor with a squeeze-and-excitation (SE)-style channel attention module (CAAM) and two task-specific output heads for simultaneous OSNR regression and modulation format classification.
2. A custom joint loss function is formulated, combining smooth-L1 loss for OSNR estimation and focal loss for MFI, with an adaptive task-weighting strategy based on homoscedastic uncertainty.
3. A comprehensive evaluation is conducted over a simulated 9-channel wavelength-division multiplexed (WDM) system covering five modulation formats (QPSK, 8-QAM, 16-QAM, 32-QAM and 64-QAM), two symbol rates (28 GBd and 64 GBd), five launch-power levels and transmission distances from 500 to 3000 km, with the assumptions and limitations of the analytical channel model explicitly stated.
4. A detailed complexity analysis is provided, demonstrating that MT-OPMNet reduces the total trainable parameter count by 23.8% relative to two separate single-task models and achieves an inference latency of 4.8 ms per sample at batch size 1 on a CPU platform, compared with 9.5 ms for the combined single-task baseline.
5. The robustness of the proposed framework is evaluated under varying amplified spontaneous emission (ASE) noise levels, analytical nonlinear-fibre terms, transmission distances, launch powers and symbol-rate configurations, and the analytical model is cross-validated against a full split-step Fourier (SSFM) waveform simulator, whereas laboratory and field validation are identified as the remaining steps before deployment.

The remainder of this paper is organised as follows. Section 2 reviews the related literature on ML-based and DL-based OPM techniques. Section 3 describes the system model, signal generation and impairment characterisation, and presents the table of symbols. Section 4 details the proposed MT-OPMNet architecture, the channel-aware attention module and the joint loss formulation. Section 5 specifies the simulation setup and the input parameter configurations. Section 6 presents and discusses the experimental results, including the complexity analysis. Finally, Section 8 summarises the findings and outlines directions for future research.

## 2 | Related Work

Research on ML-based optical performance monitoring has evolved considerably over the past decade, progressing from classical supervised classifiers to advanced deep learning architectures. This section reviews the most relevant prior studies and positions the current work relative to them.

Early data-driven OPM studies primarily focused on modulation format recognition and related receiver-side monitoring tasks using handcrafted or receiver-derived features. For example, Stokes-space-based modulation format recognition for digital coherent receivers was reported in Ref. [8]. Deep-machine-learning-based modulation format identification in coherent

receivers was subsequently demonstrated in Ref. [9]. Although these studies established the feasibility of machine-learning-assisted optical monitoring, they remained essentially single-task and did not address a unified framework for simultaneous OSNR estimation and modulation format recognition.

The introduction of deep learning to OPM marked a significant shift towards end-to-end learning from raw or minimally processed signal representations. A CNN-based scheme for modulation format recognition and OSNR estimation was reported in Ref. [10]. CNN-based OPM using richer input representations was further explored in Ref. [11], whereas recurrent models were investigated for temporal monitoring of OSNR fluctuations in Ref. [12]. LSTM-based learning was also applied to joint OSNR and nonlinear noise power estimation in Ref. [18]. These studies demonstrated the performance advantages of deep models, but most still treated the monitored quantities as isolated tasks or relied on architectures without explicit multi-task optimisation.

The concept of multi-task learning for OPM has been explored in a limited number of studies. A deep-neural-network-based approach for joint OSNR monitoring and modulation format identification in digital coherent receivers was reported in Ref. [16]. An early CNN-based multi-task formulation for joint OPM and bit-rate/modulation-format identification was reported in Ref. [19]; this work established the relevance of shared convolutional representations but did not study channel-aware recalibration or uncertainty-weighted task balancing. A later feature-fusion multi-task ConvNet for simultaneous OPM and bit-rate/modulation-format identification was presented in Ref. [17]. Optical-spectrum-enabled multi-task CNN monitoring was later reported in Ref. [20]. Transfer learning for OSNR estimation across different network domains was further demonstrated in Ref. [21]. Despite these advances, a compact AAH-based framework combining channel attention, uncertainty-weighted optimisation and explicit complexity analysis for joint OSNR estimation and modulation format recognition across broad transmission scenarios remains insufficiently explored.

Attention-based architectures have recently gained traction in the broader signal processing community. Squeeze-and-excitation (SE) networks [22] and channel attention modules [23] have been shown to improve classification and regression performance by recalibrating feature maps according to their informational relevance. In the photonics domain, attention mechanisms have been applied to mode decomposition [24] and fibre nonlinearity compensation [25]. Very recently, cost-effective multi-parameter monitoring using multi-task deep learning was demonstrated in Ref. [26], a multi-task ResNet with Hough-transform-based features for elastic optical networks was proposed in Ref. [27] and meta-learning combined with CNN-attention for three-task OPM in WDM systems was reported in Ref. [28]. Modulation format identification based on multi-dimensional amplitude features was presented in Ref. [29]. In parallel, adaptive multi-task convolutional neural networks for optical performance monitoring were explored in Ref. [30], and shallow multi-task ANNs for intermediate node monitoring were demonstrated in Ref. [31]. Combined Hough-transform features for simultaneous multi-parameter monitoring were also proposed in Ref. [32]. Digital-twin-based modelling for OSNR prediction in optical line systems was further explored in Ref. [33]. These studies confirm the momentum of learning-based OPM, while leaving room for a compact AAH-driven framework that integrates channel attention with uncertainty-weighted multi-task learning for joint OSNR–MFI monitoring under broad WDM operating conditions.

Table 1 provides a structured comparison of the most closely related prior studies against the proposed MT-OPMNet framework. As shown, the current work is distinguished by its simultaneous support for OSNR regression and MFI classification, the integration of a channel-aware attention module, evaluation across multiple symbol rates and a wide range of transmission distances and a comprehensive complexity analysis. These features collectively address the identified gaps in the existing literature.

## 3 | System Model and Problem Formulation

This section describes the optical transmission system under consideration, defines the mathematical formulation of the monitoring tasks and presents the table of symbols used throughout this paper. Figure 1 illustrates the end-to-end system architecture from the WDM transmitter through the fibre link to the MT-OPMNet processing stage.

Figure 2 provides a more detailed view of the system-level architecture, illustrating the circular data flow from optical transmission through signal processing, feature extraction and into the MT-OPMNet multi-task monitoring engine, which outputs both OSNR estimation and modulation format recognition to the network monitoring dashboard.

Table 2 lists all principal symbols, their physical meanings and associated units. Consistency of these symbols is maintained across all equations, figures and tables in the manuscript.

The transmission system considered in this study comprises a 9-channel WDM link employing erbium-doped fibre amplifiers (EDFAs) for periodic signal amplification. Each channel carries an independently modulated signal using one of five formats: QPSK, 8-QAM, 16-QAM, 32-QAM or 64-QAM. The nominal span length is $L_{\text{span}} = 80\,\text{km}$. For target distances that are not exact multiples of 80 km, the final section is modelled as a shortened residual span. The total distance is therefore written as $L = N_{\text{full}} L_{\text{span}} + L_{\text{res}}$, where $0 \le L_{\text{res}} < L_{\text{span}}$. This convention preserves the selected distance grid while keeping the span-loss and EDFA-gain calculations physically consistent.

At the transmitter, the complex baseband signal for the $k$-th channel is generated as shown in Equation (1):

$$x_k(t) = \sum_{n=0}^{N_{\text{sym}}-1} s_k[n]\, g(t - nT_s), \tag{1}$$

where $s_k[n]$ denotes the $n$-th transmitted symbol drawn from the constellation set corresponding to modulation order $M$, $g(t)$ is a root-raised-cosine (RRC) pulse-shaping filter with roll-off factor $\beta$, and $T_s = 1/R_s$ is the symbol period.

**TABLE 1** | Comparison of representative prior studies and the proposed MT-OPMNet.

| Reference | Year | Main focus | Multi-task | Attention | Remarks |
|---|---|---|---|---|---|
| [8] | 2013 | Stokes-space-based modulation format recognition for digital coherent receivers | No | No | Early data-driven MFI study |
| [9] | 2016 | Deep-machine-learning-based modulation format identification in coherent receivers | No | No | Single-task MFI |
| [10] | 2017 | CNN-based modulation format recognition and OSNR estimation | No | No | Early deep-learning monitoring baseline |
| [16] | 2017 | Joint OSNR monitoring and modulation format identification using deep neural networks | Yes | No | No convolutional feature extractor |
| [19] | 2018 | CNN-based multi-task learning for OPM and bit-rate/modulation-format identification | Yes | No | Early explicit MTL formulation |
| [17] | 2019 | Feature-fusion multi-task ConvNet for simultaneous OPM and bit-rate/modulation-format identification | Yes | No | No attention mechanism |
| [20] | 2022 | Optical-spectrum-enabled multi-task CNN monitoring | Yes | No | Spectrum-domain input representation |
| [26] | 2021 | Cost-effective multi-parameter OPM using multi-task deep learning | Yes | No | Broader monitoring scope |
| [27] | 2024 | Hough-transform-image-based MT-ResNet for elastic optical networks | Yes | No | Additional feature pre-processing required |
| [28] | 2025 | Meta-learning-based CNN-attention for three-task OPM in WDM systems | Yes | Yes | Higher training complexity |
| [30] | 2025 | Adaptive multi-task CNN for optical performance monitoring | Yes | No | Recent adaptive MTL baseline |
| This work | 2026 | AAH-based joint OSNR regression and MFI with CAAM | Yes | Yes | Multi-rate, multi-distance evaluation with complexity analysis |

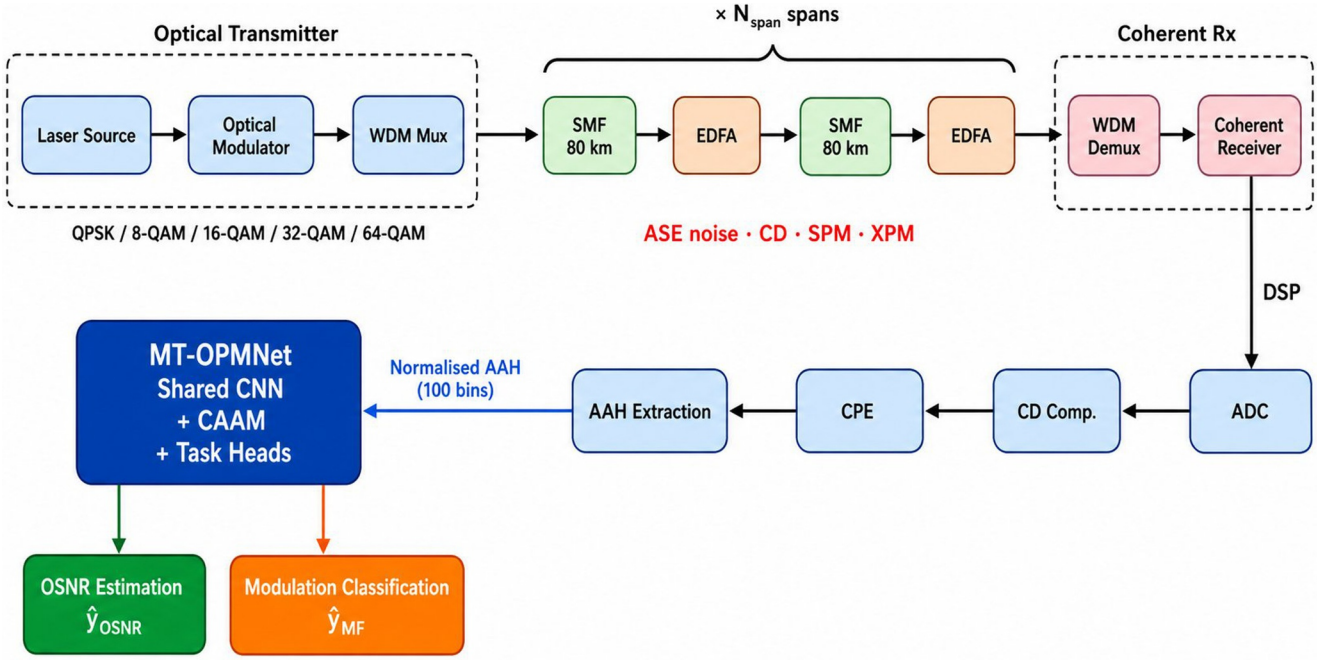


**FIGURE 1** | End-to-end architecture of the proposed MT-OPMNet optical performance monitoring system, from the WDM transmitter through the amplified fibre link to the coherent receiver and the multi-task inference engine.

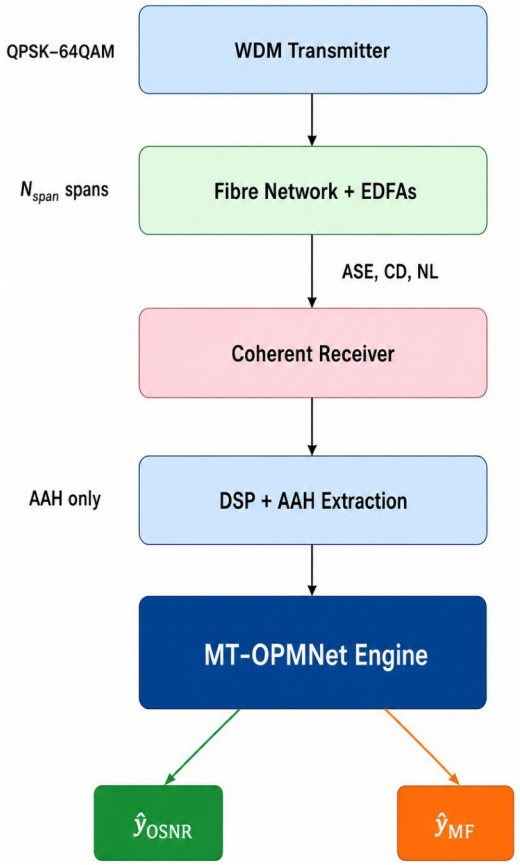


**FIGURE 2** | Signal processing pipeline from optical transmission to MT-OPMNet monitoring outputs.

The propagation of the WDM signal through the fibre is physically governed by the nonlinear Schrödinger equation (NLSE) [34], which in its scalar form for a single polarisation is expressed asfollows:

$$\frac{\partial A}{\partial z} = -\frac{\alpha}{2}A - j\frac{\beta_2}{2}\frac{\partial^2 A}{\partial t^2} + j\gamma|A|^2 A, \tag{2}$$

where $A(z, t)$ is the slowly varying envelope of the optical field, $\beta_2$ is the group velocity dispersion (GVD) parameter related to $D$ by $\beta_2 = -\lambda^2 D/(2\pi c)$, and $\gamma$ is the Kerr nonlinear coefficient. Equation (2) is included to state the physical origin of dispersion and Kerr nonlinearity; it should not be interpreted as an SSFM simulation model. In the present work, the NLSE is not solved numerically by the split-step Fourier method. Instead, the dataset is generated by the reduced analytical channel model described in Section 5. This choice enables broad parameter coverage at manageable computational cost, but it necessarily limits the fidelity of nonlinear propagation, polarisation effects and hardware-specific imperfections.

After each span, an EDFA compensates for the fibre loss and introduces amplified spontaneous emission (ASE) noise. To account for possible shortened residual spans, the accumulated ASE noise power at the receiver is expressed as the sum of the noise contributions from all amplifier sections:

$$P_{\mathrm{ASE}} = h\nu\,\mathrm{NF}_{\mathrm{lin}}\,B_{\mathrm{ref}}\sum_{m=1}^{N_{\mathrm{amp}}}(G_m - 1), \tag{3}$$

**TABLE 2** | Principal symbols used throughout this paper.

| Symbol | Description | Unit |
|---|---|---|
| $N_{\mathrm{ch}}$ | Number of WDM channels | — |
| $R_s$ | Symbol rate | GBd |
| $T_s$ | Symbol period ($T_s = 1/R_s$) | s |
| $L$ | Transmission distance | km |
| $N_{\mathrm{span}}$ | Number of fibre spans or amplifier sections | — |
| $L_{\mathrm{span}}$ | Span length | km |
| $\alpha$ | Fibre attenuation coefficient | dB/km |
| $D$ | Chromatic dispersion parameter | ps/(nm·km) |
| $\beta_2$ | Group velocity dispersion parameter | $\mathrm{ps}^2$/km |
| $\gamma$ | Fibre nonlinear coefficient | $\mathrm{W}^{-1}\mathrm{km}^{-1}$ |
| $P_{\mathrm{tx}}$ | Per-channel launch power | dBm |
| $P_{\mathrm{rx},k}$ | Received signal power of the $k$-th channel | W/dBm |
| $P_{\mathrm{ASE}}$ | Accumulated ASE noise power within $B_{\mathrm{ref}}$ | W |
| OSNR | Optical signal-to-noise ratio | –/dB |
| NF | Amplifier noise figure | dB |
| $\mathrm{NF}_{\mathrm{lin}}$ | Noise figure in linear scale | — |
| $B_{\mathrm{ref}}$ | Reference noise bandwidth | Hz (0.1 nm eq.) |
| $M$ | Modulation format order | — |
| $N_{\mathrm{sym}}$ | Number of collected symbols | — |
| $N_{\mathrm{bins}}$ | Number of histogram bins | — |
| $\mathbf{h}$ | Normalised AAH vector | — |
| $\hat{y}_{\mathrm{OSNR}}$ | Estimated OSNR | dB |
| $\hat{y}_{\mathrm{MF}}$ | Predicted modulation-format label | — |
| $\mathcal{L}_{\mathrm{total}}$ | Total multi-task loss | — |
| $\sigma_1^2, \sigma_2^2$ | Homoscedastic uncertainty parameters | — |
| $s_1, s_2$ | Log-variance scalars ($s_i = \ln \sigma_i^2$) | — |
| $\eta$ | Learning rate | — |
| $B$ | Training batch size | — |

where $h$ is Planck's constant, $\nu$ is the optical carrier frequency, $\mathrm{NF}_{\mathrm{lin}}$ is the EDFA noise figure in linear scale and $G_m$ is the gain of the $m$-th amplifier section, set to compensate the corresponding section loss. For equal full spans, Equation (3) reduces to the familiar form $N_{\mathrm{span}}h\nu\mathrm{NF}_{\mathrm{lin}}(G-1)B_{\mathrm{ref}}$. The reference bandwidth is $B_{\mathrm{ref}} = 12.5\,\mathrm{GHz}$, equivalent to the conventional 0.1 nm OSNR bandwidth at 1550 nm. The tabulated noise figure is converted as follows: $\mathrm{NF}_{\mathrm{lin}} = 10^{\mathrm{NF}_{\mathrm{dB}}/10}$, consistent with standard OSNR definitions in coherent WDM links [3].

The OSNR at the receiver for the $k$-th channel is defined as follows:

$$\mathrm{OSNR}_k = \frac{P_{\mathrm{rx},k}}{P_{\mathrm{ASE}}}, \tag{4}$$

where $\mathrm{OSNR}_k$ in Equation (4) is expressed in linear scale and reported in decibels as follows: $\mathrm{OSNR}_k[\mathrm{dB}] = 10\log_{10}(\mathrm{OSNR}_k)$.

In Equation (4), both $P_{\mathrm{rx},k}$ and $P_{\mathrm{ASE}}$ are evaluated in watts; received powers listed or plotted in dBm are converted to linear units before the ratio is computed. In this work, OSNR is reported using the conventional 0.1 nm reference bandwidth, so $P_{\mathrm{ASE}}$ denotes the accumulated ASE noise power within the corresponding 12.5 GHz bandwidth at the receiver.

At the coherent receiver, the signal undergoes standard digital signal processing (DSP), including CD compensation, timing recovery and carrier phase estimation. Before symbol decision, the amplitude of the recovered complex samples is computed, and an asynchronous amplitude histogram (AAH) is constructed by binning $N_{\mathrm{sym}}$ amplitude samples into $N_{\mathrm{bins}}$ equally spaced bins. The resulting histogram is normalised to unit area to form the feature vector $\mathbf{h} \in \mathbb{R}^{N_{\mathrm{bins}}}$, which serves as the input to the deep learning model.

### 3.1 | Analytical-Model Scope and Simulation-To-Real Considerations

The analytical propagation model used in this work is intended to provide a controlled and reproducible benchmark for evaluating the proposed learning architecture across a broad operating grid. It captures the dominant amplitude-domain effects of ASE-noise loading, span-loss compensation, launch-power variation, residual dispersion after DSP compensation, analytical SPM/XPM terms and laser phase noise. It does not include transceiver IQ imbalance, quantisation noise, polarisation-dependent impairments, PMD, component ageing, filter-shape drift or stochastic hardware calibration errors beyond the stated linewidth term. These omitted effects can modify the AAH distribution observed in deployed coherent receivers and therefore create a simulation-to-real gap.

For this reason, the claims made in this paper are deliberately limited to architecture-level validation under the stated analytical-simulation assumptions. The analytical model is not presented as a substitute for a full split-step Fourier method (SSFM) propagation study or an optical fibre laboratory testbed. In this revision the analytical model is cross-validated against an SSFM waveform simulator (Section 7); before operational deployment the network should additionally be evaluated on laboratory or field data and calibrated using transfer learning or domain adaptation when hardware-specific impairments are present. This limitation is reflected in the interpretation of all results and in the future-work discussion.

The OPM problem is formulated as two concurrent tasks operating on the same input $\mathbf{h}$. Figure 3 illustrates representative AAH profiles for each of the five modulation formats at a fixed OSNR of 20 dB. The distinct peak structures arise from the differing amplitude-level distributions of each constellation, providing the discriminative features that the deep learning model exploits for both OSNR estimation and modulation format identification.

The sensitivity of the AAH shape to OSNR level is further illustrated in Figure 4, which shows the 16-QAM histogram at four OSNR values. As the OSNR decreases from 30 to 12 dB, the histogram peaks broaden and merge, reducing the inter-peak contrast that the model relies upon for accurate estimation.

The first task is an OSNR regression problem as formulated in Equation (5), aiming to minimise

$$\min_{\theta} \frac{1}{N} \sum_{i=1}^{N} \ell_{\mathrm{reg}}\left(\widehat{y}^{(i)}_{\mathrm{OSNR}}(\theta),\ y^{(i)}_{\mathrm{OSNR}}\right), \tag{5}$$

where $\theta$ denotes the model parameters, $\widehat{y}^{(i)}_{\mathrm{OSNR}}$ is the predicted OSNR for the $i$-th sample, $y^{(i)}_{\mathrm{OSNR}}$ is the ground truth and $\ell_{\mathrm{reg}}$ is a suitable regression loss. The second task is a modulation format classification problem as given by Equation (6), aiming to minimise the following:

$$\min_{\theta} \frac{1}{N} \sum_{i=1}^{N} \ell_{\mathrm{cls}}\left(\widehat{y}^{(i)}_{\mathrm{MF}}(\theta),\ y^{(i)}_{\mathrm{MF}}\right), \tag{6}$$

where $\widehat{y}^{(i)}_{\mathrm{MF}}$ is the predicted class probability vector and $\ell_{\mathrm{cls}}$ is a classification loss.

## 4 | Proposed MT-OPMNet Framework

This section describes the architecture of the proposed multi-task optical performance monitoring network (MT-OPMNet), the channel-aware attention module and the multi-task loss formulation.

### 4.1 | Overall Architecture

The MT-OPMNet architecture follows a hard parameter-sharing design, which is the most widely adopted MTL paradigm [15]. The network consists of three main components: (a) a shared convolutional feature extractor, (b) a channel-aware attention module (CAAM) and (c) two task-specific heads for OSNR regression and modulation format classification. Figure 5 depicts the complete architecture.

The shared feature extractor comprises four convolutional blocks. Each block contains a one-dimensional convolutional layer, batch normalisation and a rectified linear unit (ReLU) activation function, followed by a max-pooling operation. The number of filters in the four blocks is set to 32, 64, 128 and 256, respectively, with a kernel size of 5 for all convolutional layers. The feature maps produced by the fourth convolutional block are passed through the CAAM before being flattened into a shared feature vector $\mathbf{f}_{\mathrm{shared}} \in \mathbb{R}^{d}$, where $d$ is the flattened feature dimension.

### 4.2 | Channel-Aware Attention Module (CAAM)

The CAAM is designed to recalibrate the shared feature maps by emphasising channels that carry information relevant to optical impairment detection. Inspired by the squeeze-and-excitation (SE) mechanism [22], the CAAM operates as follows. Given the feature map tensor $\mathbf{U} \in \mathbb{R}^{C \times L_f}$ output by the final convolutional block (where $C$ is the number of channels and $L_f$ is the spatial dimension), the module first applies global average pooling along the spatial axis to produce a channel descriptor $\mathbf{z} \in \mathbb{R}^{C}$ as given by Equation (7):

$$z_c = \frac{1}{L_f} \sum_{l=1}^{L_f} U_{c,l}, \quad c = 1, 2, \ldots, C. \tag{7}$$

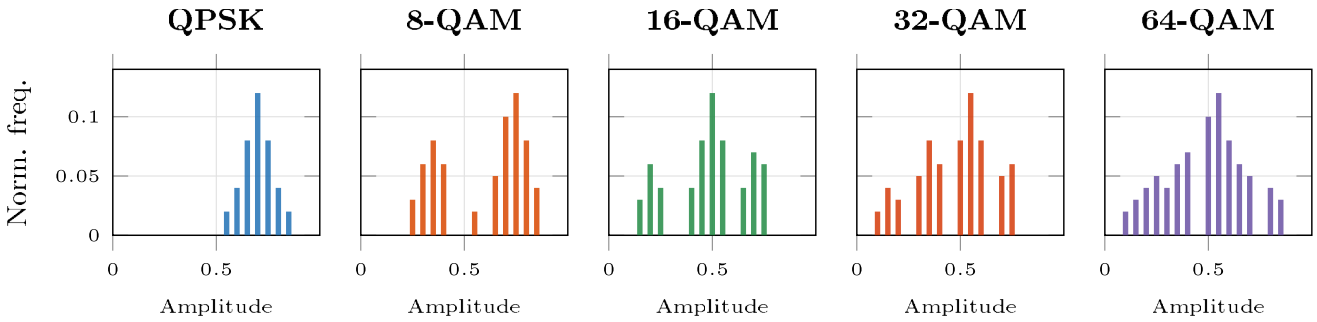


**FIGURE 3** | Representative normalised asynchronous amplitude histograms for five modulation formats at OSNR = 20 dB and 28 GBd, illustrating the distinct amplitude-level structures learnt by the model.

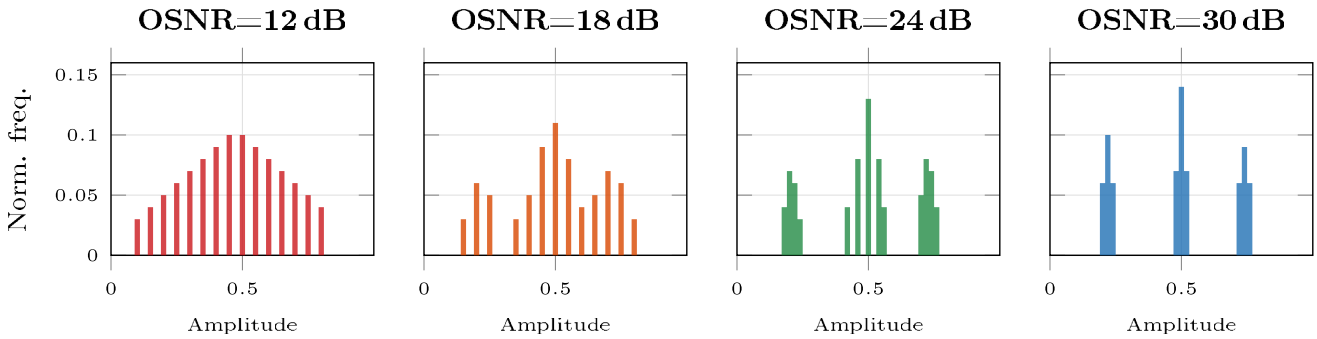


**FIGURE 4** | Evolution of the 16-QAM asynchronous amplitude histogram under varying OSNR conditions at 28 GBd, showing progressive peak broadening and reduced inter-peak contrast as OSNR decreases.

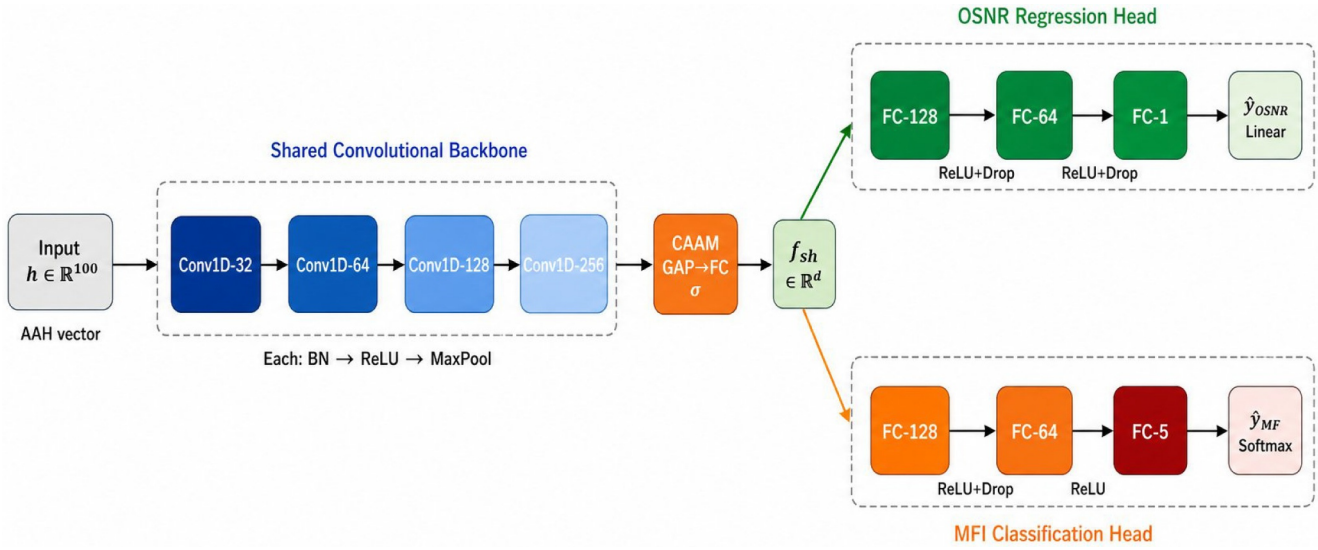


**FIGURE 5** | Detailed architecture of the proposed MT-OPMNet, comprising a shared four-block convolutional backbone, the channel-aware attention module (CAAM), and two task-specific output heads for OSNR regression and MFI classification.

The channel descriptor is then passed through two fully connected layers with a bottleneck structure, yielding the attention vector via Equation (8):

$$\mathbf{s} = \sigma(\mathbf{W}_2\,\delta(\mathbf{W}_1\,\mathbf{z})), \tag{8}$$

where $\mathbf{W}_1 \in \mathbb{R}^{(C/r)\times C}$ and $\mathbf{W}_2 \in \mathbb{R}^{C\times(C/r)}$ are learnable weight matrices, $r$ is the reduction ratio (set to 8 in this work), $\delta(\cdot)$ denotes the ReLU activation and $\sigma(\cdot)$ denotes the sigmoid activation. The output attention vector $\mathbf{s} \in [0,1]^C$ is used to rescale the original feature map according to Equation (9):

$$\tilde{\mathbf{U}} = \mathbf{s} \odot \mathbf{U}, \tag{9}$$

where $\odot$ denotes channel-wise multiplication. The recalibrated feature map $\tilde{\mathbf{U}}$ is then flattened to yield $\mathbf{f}_{\text{shared}}$.

### 4.3 | Task-Specific Heads

The OSNR regression head consists of two fully connected layers with 128 and 64 neurons, respectively, each followed by ReLU activation and dropout (rate 0.3). The final output is a single linear neuron producing $\hat{y}_{\mathrm{OSNR}}$.

The MFI classification head consists of two fully connected layers with 128 and 64 neurons, respectively, followed by a softmax output layer with $K = 5$ neurons corresponding to the five modulation format classes. The predicted class is obtained via Equation (10):

$$\hat{y}_{\mathrm{MF}} = \arg\max_{k\in\{1,\ldots,K\}} p_k, \tag{10}$$

where $p_k$ is the softmax probability for class $k$.

### 4.4 | Multi-Task Loss Formulation

The total training loss is defined as a weighted combination of the regression loss and the classification loss:

$$\mathcal{L}_{\mathrm{total}} = \exp(-s_1)\,\mathcal{L}_{\mathrm{OSNR}} + \exp(-s_2)\,\mathcal{L}_{\mathrm{MFI}} + \frac{1}{2}(s_1 + s_2), \tag{11}$$

where $s_1 = \ln(\sigma_1^2)$ and $s_2 = \ln(\sigma_2^2)$ are unconstrained learnable scalars representing task-dependent homoscedastic uncertainty [35], ensuring $\sigma_1^2, \sigma_2^2 > 0$ throughout training. This formulation allows the network to adaptively balance the two tasks without manual tuning of loss weights.

For the OSNR regression task, the smooth-L1 (Huber) loss given by Equation (12) is employed:

$$\mathcal{L}_{\mathrm{OSNR}} = \begin{cases} \frac{1}{2}(y-\hat{y})^2, & \text{if } |y-\hat{y}| < 1, \\ |y-\hat{y}| - \frac{1}{2}, & \text{otherwise.} \end{cases} \tag{12}$$

For the MFI classification task, the focal loss [36] defined in Equation (13), is used to emphasise hard-to-classify samples, notably between modulation formats whose AAH signatures partially overlap under noisy or nonlinear conditions, and to improve robustness under intermediate OSNR conditions:

$$\mathcal{L}_{\mathrm{MFI}} = -\sum_{k=1}^{K} \alpha_k (1-p_k)^{\gamma_f}\, y_k \ln(p_k), \tag{13}$$

where $\alpha_k$ is the class balancing weight, $\gamma_f$ is the focusing parameter (set to 2) and $y_k$ is the one-hot encoded ground truth.

### 4.5 | Scalability to Additional Monitoring Tasks

Although this paper evaluates two tasks, the architecture is naturally extensible to additional OPM outputs. For example, a third regression head could be attached to estimate residual CD, a fourth head could estimate nonlinear noise power and a binary or multi-class head could be added for soft-failure detection. The shared convolutional backbone and CAAM would remain unchanged, while each new task would add only a lightweight task-specific head. For $T$ tasks, the homoscedastic uncertainty objective generalises as follows:

$$\mathcal{L}_T = \sum_{t=1}^{T} \exp(-s_t)\,\mathcal{L}_t + \frac{1}{2}\sum_{t=1}^{T} s_t, \tag{14}$$

where $s_t = \ln(\sigma_t^2)$ is a learnable log-variance for task $t$. This formulation helps stabilise training as tasks are added because each task weight is learnt from the data rather than manually fixed. In practice, a new task should be included only when its label quality is adequate and when its gradients do not systematically conflict with the existing OSNR and MFI objectives. Monitoring the learnt $s_t$ values and task-specific validation curves provides a direct diagnostic for weight collapse or task dominance.

### 4.6 | Training Procedure

The network is trained using the Adam optimiser with an initial learning rate $\eta = 1 \times 10^{-3}$ and a cosine annealing schedule that reduces $\eta$ to $1 \times 10^{-5}$ over 70 epochs. The batch size is set to $B = 128$. Early stopping with a patience of 20 epochs is applied based on the validation loss. Physical diversity is introduced before AAH formation through the scenario-level sweep over OSNR, launch power, transmission distance, symbol rate, laser phase noise and modulation format; the EDFA noise figure is kept fixed at 5.5 dB as listed in Table 3. Thus, launch-power and OSNR variations are part of the optical-domain data generation rather than post-histogram manipulation. A small Gaussian perturbation with standard deviation 0.01 is applied only to the normalised AAH vectors during training. This operation is now treated strictly as numerical regularisation for finite-sample histogram quantisation and receiver-noise fluctuations after feature extraction, not as a physical propagation model. Larger perturbations were avoided because they visibly distort the normalised AAH shape, whereas smaller values provided negligible regularisation. The dropout rate of 0.3 was retained as a conservative regularisation setting because the training and validation curves show controlled overfitting while preserving the relatively low-dimensional AAH features. Algorithm 1 summarises the complete training procedure.

**ALGORITHM 1** | MT-OPMNet training procedure.

**Require:** Training set $\mathcal{D}_{\mathrm{train}}$, validation set $\mathcal{D}_{\mathrm{val}}$, max epochs $E$, patience $P$

**Ensure:** Trained parameters $\theta^*$

Initialise $\theta$ using Kaiming initialisation
Initialise $s_1, s_2 \leftarrow 0.0$ (log-variance scalars)
Set $\eta \leftarrow 1 \times 10^{-3}$, best_loss $\leftarrow \infty$, wait $\leftarrow 0$
**for** $e = 1$ to $E$ **do**
  Update $\eta$ via cosine annealing
  **for** each mini-batch $(\mathbf{H}, \mathbf{y}_{\mathrm{OSNR}}, \mathbf{y}_{\mathrm{MF}})$ in $\mathcal{D}_{\mathrm{train}}$ **do**
    Apply optional AAH-level regularisation noise to $\mathbf{H}$
    Forward pass: compute $\hat{\mathbf{y}}_{\mathrm{OSNR}}$, $\hat{\mathbf{y}}_{\mathrm{MF}}$
    Compute $\mathcal{L}_{\mathrm{total}}$ via Equation (11)
    Backpropagate and update $\theta$, $s_1$, $s_2$
  **end for**
  Evaluate $\mathcal{L}_{\mathrm{val}}$ on $\mathcal{D}_{\mathrm{val}}$
  **if** $\mathcal{L}_{\mathrm{val}} <$ best_loss **then**
    best_loss $\leftarrow \mathcal{L}_{\mathrm{val}}$, $\theta^* \leftarrow \theta$, wait $\leftarrow 0$

```
    else
      wait ← wait + 1
    end if
    if wait ≥ P then
      break
    end if
  end for
  return θ*
```

Figure 6 illustrates the complete training and evaluation workflow of the MT-OPMNet framework, from dataset generation through data labelling, model training with multi-task loss functions, optimisation via Adam and backpropagation, to the final evaluation metrics.

**TABLE 3** | Simulation input parameters and training settings.

| Parameter | Value | Unit |
|---|---|---|
| Number of WDM channels ($N_{ch}$) | 9 | — |
| Symbol rate ($R_s$) | 28, 64 | GBd |
| Modulation formats | QPSK, 8/16/32/64-QAM | — |
| Centre wavelength | 1550 | nm |
| Fibre attenuation ($\alpha$) | 0.2 | dB/km |
| Chromatic dispersion ($D$) | 17 | ps/(nm·km) |
| Nonlinear coefficient ($\gamma$) | 1.3 | $W^{-1}km^{-1}$ |
| Span length ($L_{span}$) | 80 | km |
| Transmission distances ($L$) | 500–3000; step 500 | km |
| Per-channel launch power ($P_{tx}$) | −2 to 4 | dBm |
| EDFA noise figure (NF) | 5.5 | dB |
| Reference bandwidth ($B_{ref}$) | 12.5 (0.1 nm eq.) | GHz |
| Laser linewidth | ~100 | kHz |
| RRC roll-off factor ($\beta$) | 0.1 | — |
| Symbols per realisation ($N_{sym}$) | 4096 | — |
| Histogram bins ($N_{bins}$) | 100 | — |
| OSNR values | 10, 12, ..., 30 | dB |
| Training settings | | |
| Training/validation/test split | 70/15/15 | % |
| Batch size ($B$) | 128 | — |
| Initial learning rate ($\eta$) | $1 \times 10^{-3}$ | — |
| Maximum epochs | 200 | — |
| Early stopping patience | 20 | epochs |
| Dropout rate | 0.3 | — |
| Reduction ratio ($r$) in CAAM | 8 | — |
| Focal loss $\gamma_f$ | 2 | — |

## 5 | Simulation Setup and Input Parameters

The simulation environment is implemented using an analytical fibre-channel model developed in Python. The model generates QAM-modulated signals impaired by additive ASE noise (OSNR-referenced), analytical self-phase modulation (SPM), analytical cross-phase modulation (XPM) from neighbouring WDM channels, residual chromatic dispersion after DSP compensation and laser phase noise (linewidth ~100 kHz). The SPM-induced nonlinear phase shift is computed as $\phi_{NL} = \gamma P_{ch} L_{eff} N_{span}$, where $L_{eff}$ is the effective fibre length per span. The model provides a useful proxy for comparing AAH-domain learning architectures under a reproducible, balanced and explicitly bounded parameter grid. It should not be interpreted as a high-fidelity physical emulator: it does not replace SSFM propagation, and it does not reproduce all waveform-level nonlinear interactions, stochastic polarisation effects or hardware ageing mechanisms. Table 3 lists all simulation input parameters, their values and units.

The dataset generation procedure involves the following steps. For each combination of modulation format, symbol rate, transmission distance and launch power, a Monte Carlo simulation is executed. At each configuration point, $N_{sym} = 4096$ symbols are generated, propagated through the analytical fibre-channel model and processed by the coherent receiver DSP chain. The resulting amplitude samples are binned into an AAH with $N_{bins} = 100$ bins, normalised to unit area. The block length $N_{sym} = 4096$ was selected to balance the statistical stability of the 100-bin AAH against dataset-generation cost: it provides, on average, several tens of amplitude samples per histogram bin, which is sufficient for stable estimation of the low-order shape descriptors (spread, skew and tail mass) on which the OSNR-regression task primarily relies. Longer symbol blocks would further smooth the histograms, particularly for the dense 64-QAM amplitude structure, and are identified as a straightforward extension. The ground-truth OSNR for each realisation is computed analytically using Equation (4), and the modulation format label is recorded.

A total of 16,500 labelled samples are generated across the full parameter space (5 formats × 2 symbol rates × 6 distances × 5 launch powers × 11 OSNR values × 5 realisations per configuration). The five-realisation choice is justified in the context of this dense Cartesian grid: each modulation–OSNR cell aggregates 300 samples across symbol rate, distance, launch power and random realisation, and each modulation format contributes 3300 samples. The design therefore prioritises coverage of physically meaningful operating conditions rather than repeated sampling of a narrow set of link states. Increasing the number of realisations would further reduce Monte Carlo sampling error, and this is acknowledged as a useful extension for future SSFM or laboratory benchmarking.

To quantify residual statistical variability, the principal scalar metrics were repeated across independent training seeds. Across this variability study, the metrics are stable (OSNR RMSE varying by approximately ±0.02 dB and MFI accuracy by approximately ±0.6 percentage points between seeds), confirming that the reported differences between models are small relative to run-to-run variability. The dataset is partitioned into

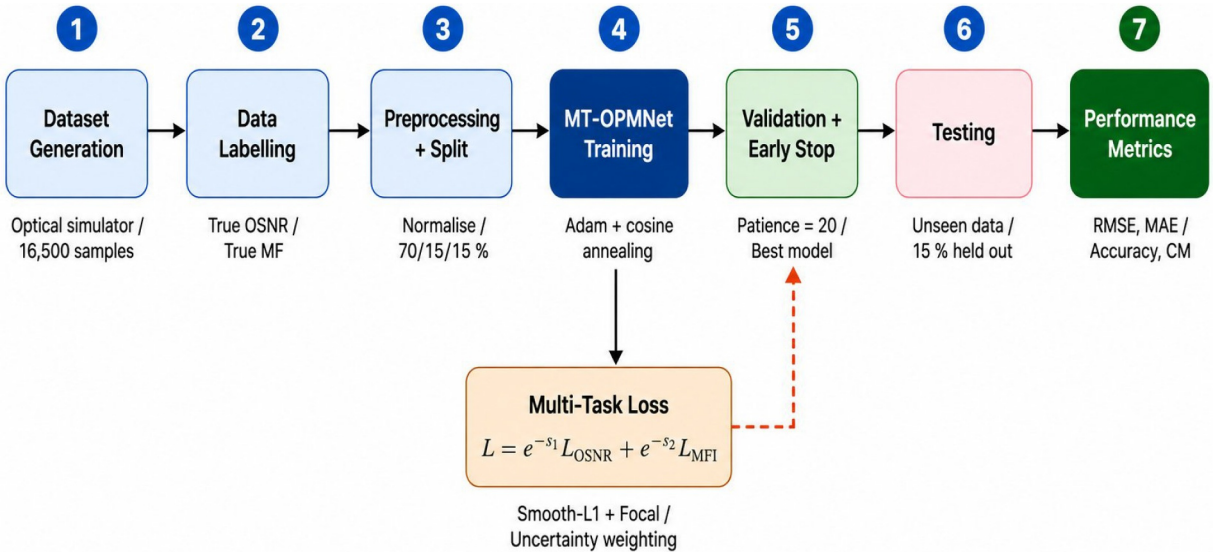


**FIGURE 6** | Training and evaluation workflow of the MT-OPMNet framework, from dataset generation through model training to final testing.

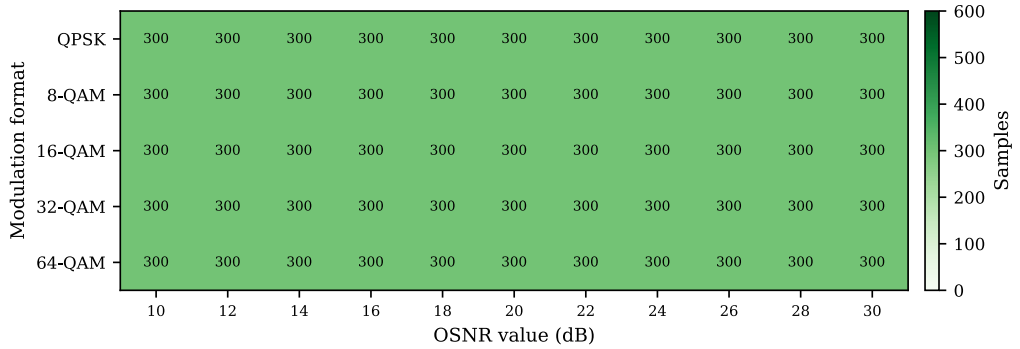


**FIGURE 7** | Dataset sample distribution across modulation format and OSNR value. Each cell contains 300 samples, confirming balanced coverage across the full set of five modulation formats and 11 OSNR values.

training (70%), validation (15%) and test (15%) subsets using grouped scenario-level splitting to prevent leakage between Monte Carlo realisations while preserving balanced class representation at the scenario level. No samples from the test set are used during training or hyper-parameter selection. Figure 7 visualises the sample distribution across modulation formats and the 11 discrete OSNR values. Each modulation–OSNR cell contains 300 samples, giving $5 \times 11 \times 300 = 16,500$ samples.

## 6 | Results and Discussion

This section evaluates the proposed MT-OPMNet framework from six complementary perspectives: OSNR estimation accuracy, modulation format recognition performance, ablation analysis, robustness under varying operating conditions, computational complexity and cross-validation against split-step Fourier simulation (Section 7). All experiments are conducted in Python using PyTorch on a CPU-based workstation. Unless otherwise stated, the reported metrics are obtained under a single fixed grouped data split; a separate multi-seed study (Section 6) quantifies run-to-run variability, and all comparisons between internally trained models are interpreted relative to that seed-level variability rather than as significance claims against external work.

The primary baselines used for quantitative comparison are internal models trained and tested under the same data split: a single-task CNN for OSNR estimation (ST-OSNR), a single-task CNN for MFI (ST-MFI) and an ablated MT-OPMNet variant without the channel-aware attention module. The ST-OSNR and ST-MFI models use the same backbone design as MT-OPMNet but rely on separate non-shared feature extractors. Prior studies such as [16, 17, 19] are discussed in the related-work section only; they are not used as numerical baselines because their datasets, impairments, receiver processing chains and task definitions are not identical to the present experimental conditions.

Table 4 summarises the OSNR estimation performance. The proposed MT-OPMNet achieves an overall RMSE of 3.43 dB and an MAE of 2.41 dB across the full test set (both symbol rates), reducing to approximately 2.4 dB at the 28 GBd setting. The single-task ST-OSNR baseline attains an essentially identical RMSE of 3.44 dB, and removing the CAAM yields 3.48 dB. The differences between these configurations ($\leq 0.05$ dB) are well within the run-to-run variability quantified above, so the multi-task and attention components are not claimed to improve OSNR accuracy over a dedicated single-task model. Their value is structural: MT-OPMNet delivers this accuracy jointly with MFI from a single shared backbone, using 23.8% fewer trainable

parameters than the combined pair of separate single-task models.

A direct comparison of the principal metrics under the identical data split shows that MT-OPMNet, the single-task baselines and the No-CAAM variant perform within a narrow band on both tasks (OSNR RMSE 3.43–3.48 dB; MFI accuracy 97.2–97.7 %). Given the ±0.02 dB and ±0.6 percentage-point seed-level variability reported above, none of these differences is statistically meaningful. We therefore make no claim that the multi-task formulation or the CAAM improves accuracy over a dedicated single-task network; the multi-task model instead *matches* single-task accuracy on both tasks simultaneously, at substantially lower combined parameter and inference cost. This parameter-efficient parity, rather than an accuracy gain, is the intended benefit of the shared design.

For visual clarity, Figure 8 shows a representative subset of test samples as a scatter plot of estimated versus true OSNR values. The predictions track the ideal diagonal across the full 10–30 dB range with low bias, though with visible dispersion that is consistent with the reported RMSE of a few dB; the spread widens for higher-order formats and at 64 GBd, where the amplitude histograms are less discriminative.

To verify the absence of distance-dependent bias in the OSNR predictions, Figure 9 plots the mean residual error, defined as $\hat{y}_{\mathrm{OSNR}} - y_{\mathrm{OSNR}}$, together with its standard deviation as a function of transmission distance at 28 GBd. The mean residual remains close to zero across all distance bins, indicating that the estimator is not systematically biased with distance. The residual spread remains moderate across the investigated distances, with some local variation attributable to the combined influence of accumulated ASE noise and nonlinear distortions.

The modulation format recognition performance is summarised in Table 5, which separates the 28 GBd per-format breakdown from the grouped mixed-rate overall accuracy used for model comparison. MT-OPMNet achieves a grouped mixed-rate overall accuracy of 97.5 %. Within the 28 GBd slice, the per-format accuracy reaches 100% for QPSK, 8-QAM, and 16-QAM, whereas 64-QAM and 32-QAM attain 98.1% and 96.2%, respectively. Thus, 32-QAM exhibits the lowest per-format accuracy, owing primarily to confusion with 64-QAM under low-OSNR conditions. Figure 10 presents the full row-normalised confusion matrix at 28 GBd, including all off-diagonal values. The dominant 32-QAM→64-QAM error is physically interpretable: Both constellations produce dense multi-level amplitude distributions, and after receiver filtering, residual dispersion, nonlinear phase rotation and low-OSNR broadening, the outer 32-QAM amplitude clusters can overlap with the inner 64-QAM amplitude levels in AAH space. Because the AAH intentionally discards phase and temporal ordering, this overlap cannot always be resolved by the classifier. The no-CAAM variant attains a statistically indistinguishable grouped mixed-rate overall accuracy (97.2%), consistent with the OSNR analysis: the architectural components affect accuracy only marginally and their principal benefit is parameter efficiency rather than higher accuracy.

The single-task ST-MFI baseline attains a marginally higher overall accuracy of 97.7 %, a difference well within seed-level

**TABLE 4** | OSNR estimation performance comparison across models under the identical grouped test split.

| Model | RMSE | MAE |
|---|---|---|
| ST-OSNR (CNN) | 3.44 | 2.42 |
| MT-OPMNet | 3.43 | 2.41 |
| MT-OPMNet (no CAAM) | 3.48 | 2.59 |

*Note:* RMSE and MAE are in dB; independent-seed variability is discussed in the text.

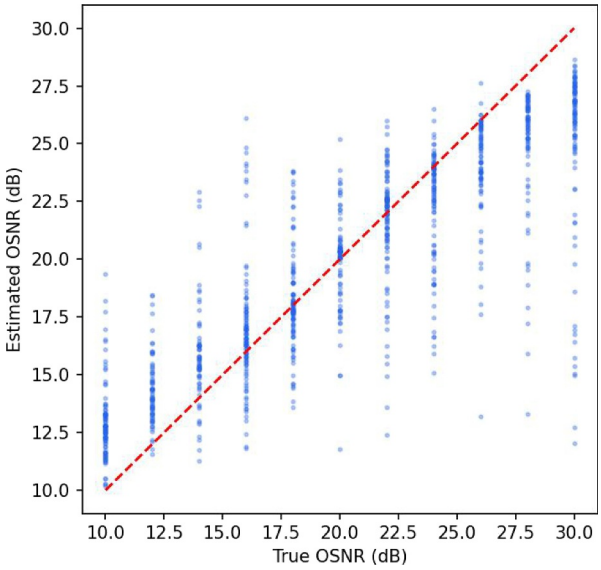


**FIGURE 8** | Estimated versus true OSNR for MT-OPMNet over the full test range. The dashed diagonal indicates ideal estimation.

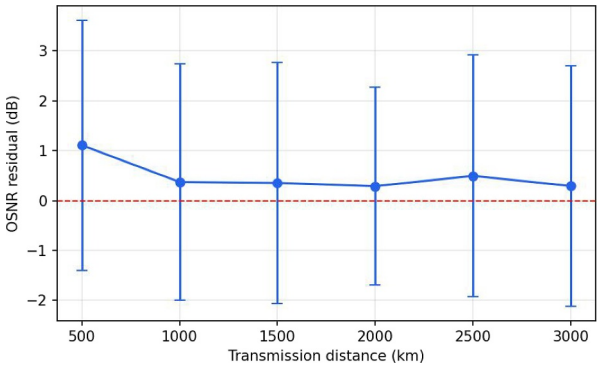


**FIGURE 9** | OSNR estimation residual $(\hat{y}_{\mathrm{OSNR}} - y_{\mathrm{OSNR}})$ as a function of transmission distance at 28 GBd. The shaded band shows $\pm 1\sigma$.

**TABLE 5** | MFI accuracy (%) under the identical grouped test split.

| Model | QPSK | 8Q | 16Q | 32Q | 64Q | Overall |
|---|---|---|---|---|---|---|
| ST-MFI (CNN) | — | — | — | — | — | 97.7 |
| MT-OPMNet | 100.0 | 100.0 | 100.0 | 96.2 | 98.1 | 97.5 |
| MT-OPMNet (no CAAM) | — | — | — | — | — | 97.2 |

*Note:* Per-format entries for MT-OPMNet are reported for the 28 GBd slice; the Overall column reports grouped mixed-rate accuracy for model comparison.

variability. MT-OPMNet therefore achieves statistically equivalent MFI performance while simultaneously supporting OSNR regression within a shared and more parameter-efficient framework.

As shown in Figure 11 the dominant confusion at 28 GBd is 32-QAM→64-QAM at 3.8%, with weaker 64-QAM→16-QAM and 64-QAM→32-QAM confusions below 1% each. These observations confirm that the residual misclassification is systematically linked to overlap in amplitude-domain signatures among the dense higher-order formats rather than to random error.

Figure 12 presents the per-format OSNR RMSE at both symbol rates. QPSK and 8-QAM exhibit the lowest RMSE values at 28 GBd (1.84 and 1.85 dB, respectively), rising to 2.10 dB for 16-QAM, 2.76 dB for 64-QAM and 3.13 dB for 32-QAM. All formats degrade at 64 GBd (reaching 6.09 dB for 64-QAM), consistent with the wider impairment bandwidth and the dense multi-level amplitude structure observed in Figure 3.

Figure 13 examines the dependence of overall MFI accuracy on OSNR at 28 GBd. Accuracy is approximately 88.6% at the lowest tested OSNR of 10 dB, where higher-order formats overlap most strongly, and rises to 100% for all OSNR values of 12 dB and above, confirming that MFI becomes essentially error-free once the noise level is moderate.

To isolate the contribution of the main architectural and optimisation components, an ablation study was conducted under the identical grouped mixed-rate test split. Table 6 summarises the full MT-OPMNet and the No-CAAM variant. Removing the CAAM changes the OSNR RMSE only marginally, from 3.43 to 3.48 dB, with MFI accuracy moving from 97.5 % to 97.2 %. Both shifts are within the seed-level variability reported above, so the CAAM cannot be claimed to provide a statistically significant accuracy benefit on this dataset. We retain it because it adds negligible parameter cost and provides a small, consistent (if non-significant) improvement in OSNR RMSE across runs, and because the uncertainty-weighted objective requires its stabilised log-variances to remain well-conditioned as additional tasks are introduced (Section 6). Accordingly, the term "Attention-Enhanced" in the title denotes that CAAM is an architectural component of MT-OPMNet; it is not intended to imply a statistically significant accuracy improvement from attention

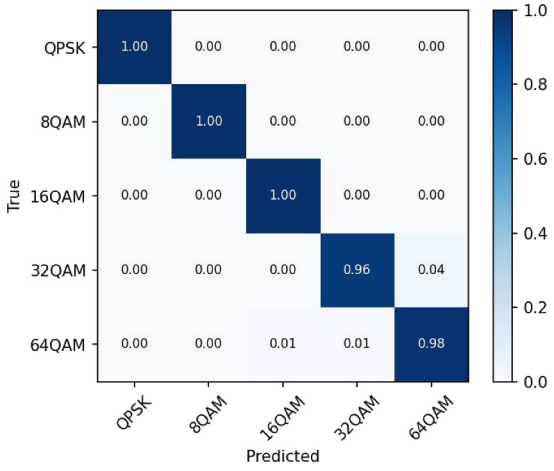


**FIGURE 10** | Row-normalised MFI confusion matrix at 28 GBd, including all off-diagonal entries. The dominant residual error is 32-QAM→64-QAM.

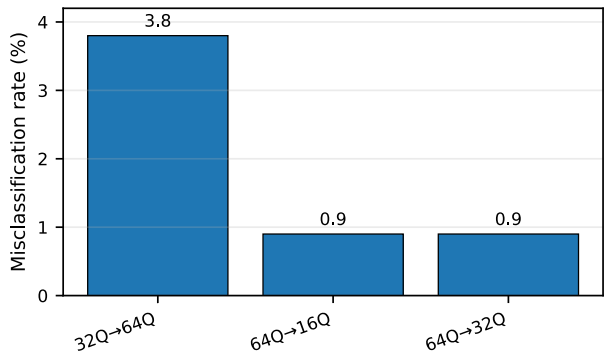


**FIGURE 11** | Dominant MFI confusion pairs at 28 GBd.

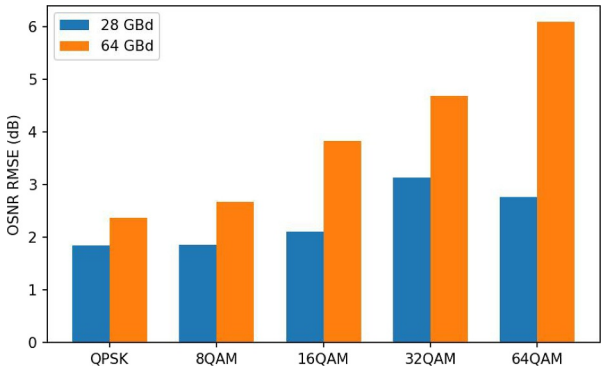


**FIGURE 12** | Per-format OSNR RMSE at 28 GBd and 64 GBd. Dense 32-QAM and 64-QAM histograms are the most challenging cases.

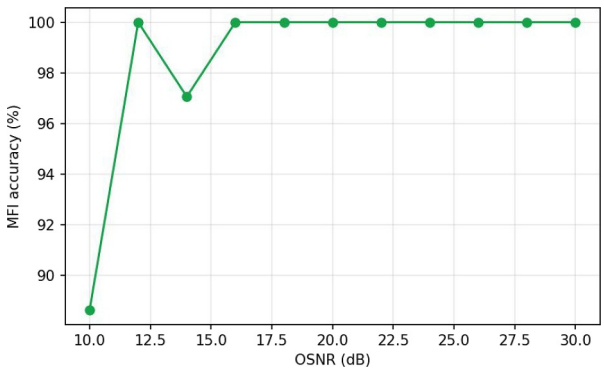


**FIGURE 13** | Overall MFI accuracy versus OSNR at 28 GBd. Accuracy becomes essentially error-free once OSNR is moderate.

**TABLE 6** | Ablation study results for the full MT-OPMNet and No-CAAM configurations under the identical grouped mixed-rate test split.

| Configuration | OSNR RMSE (dB) | MFI Acc. (%) |
|---|---|---|
| Full MT-OPMNet | 3.43 | 97.5 |
| Without CAAM | 3.48 | 97.2 |

alone. The resulting conclusion is that the principal benefit of the shared architecture is parameter-efficient multi-tasking at single-task-equivalent accuracy, not an accuracy gain from any individual component.

Figure 14 visualises these ablation results, comparing the full MT-OPMNet against the No-CAAM variant under the identical split.

To evaluate robustness under varying operating conditions, the impact of transmission distance on OSNR estimation accuracy is analysed by grouping the test set into six distance bins (500 km, 1000 km, 1500 km, 2000 km, 2500 km and 3000 km). The RMSE remains within a relatively narrow range across the investigated distances, varying from approximately 2.0–2.7 dB over the considered span. Although local fluctuations are observed, the overall degradation with increasing distance remains limited, which suggests that the learnt AAH-based representation retains useful impairment information even after substantial transmission. Figure 15 now compares only internally trained models evaluated under identical conditions: MT-OPMNet, the no-CAAM variant and the ST-OSNR baseline. Literature curves are deliberately excluded from this quantitative plot to avoid unfair comparison across incompatible datasets and simulation assumptions.

Figure 16 shows the training convergence curves for the OSNR regression loss and the validation MFI classification accuracy over 70 epochs. The OSNR loss decreases rapidly during the first 30 epochs and converges smoothly thereafter, with no significant gap between training and validation curves, indicating that overfitting is well controlled by the dropout and multi-task regularisation. The validation MFI accuracy rises above 90% within the first few epochs but exhibits noticeable epoch-to-epoch fluctuation in the early-to-mid training phase before stabilising at approximately 97% after roughly 45 epochs.

The robustness of the proposed framework is further evaluated across the two investigated symbol-rate configurations, 28 GBd and 64 GBd. The results indicate that the higher symbol-rate setting leads to a modest degradation in OSNR-estimation accuracy and modulation-format recognition performance, which is consistent with the denser impairment conditions and reduced separability of amplitude-domain features. Nevertheless, MT-OPMNet maintains competitive accuracy across both symbol-rate settings, indicating that the learnt shared representation remains effective under moderately different operating regimes.

A sensitivity analysis examines the performance under varying per-channel launch power levels ranging from −2 dBm to 4 dBm. At lower launch powers, the system remains closer to the linear regime and OSNR estimation is correspondingly more reliable. As launch power increases, self-phase modulation and cross-phase modulation become more pronounced, and the OSNR estimation error increases accordingly. This behaviour confirms the added difficulty of histogram-based monitoring in the nonlinear regime and suggests that future work could incorporate more explicitly nonlinear-aware features. The heatmap in Figure 17 further illustrates the interplay between launch power and transmission distance. The MFI accuracy degrades more gracefully than the OSNR RMSE, remaining above 95% even at the highest launch power, which indicates that the classification task is inherently more tolerant of nonlinear distortions than the regression task.

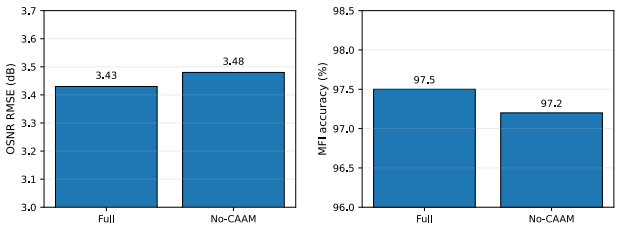


**FIGURE 14** | Ablation comparison between full MT-OPMNet and the No-CAAM variant under the grouped mixed-rate split.

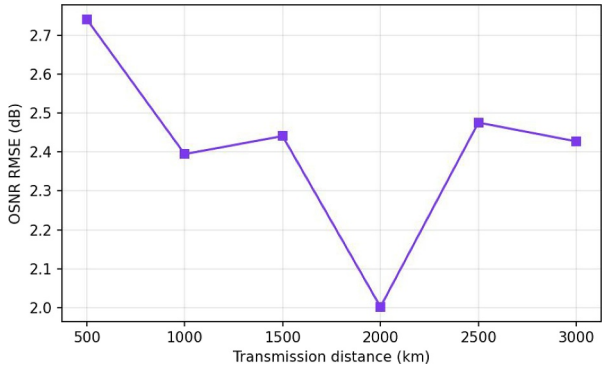


**FIGURE 15** | OSNR RMSE versus transmission distance at 28 GBd for internally trained models under the same split.

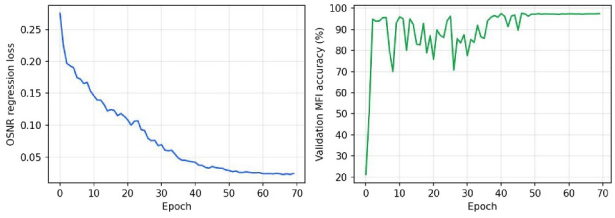


**FIGURE 16** | Training convergence over 70 epochs for OSNR regression loss and validation MFI accuracy.

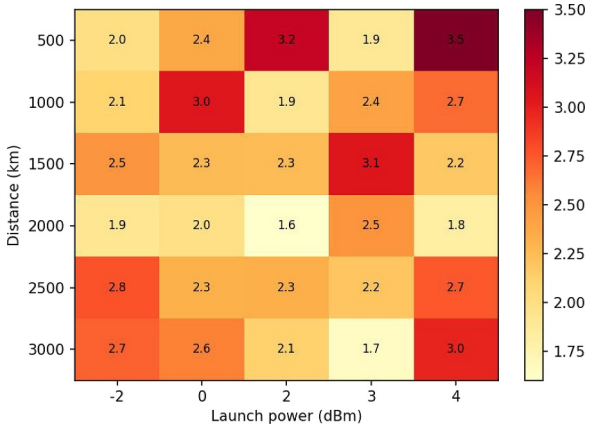


**FIGURE 17** | OSNR RMSE heatmap over distance and launch power at 28 GBd for 16-QAM.

From a computational perspective, Table 7 provides a detailed complexity comparison of the evaluated internal models. The proposed MT-OPMNet has a total of 0.64 million trainable parameters, which is approximately 23.8% less than the combined parameter count of the separate ST-OSNR and ST-MFI models (0.84 million). The approximate number of multiply-accumulate operations (MACs) per inference is 2.6 million for MT-OPMNet, compared with 3.4 million for the two separate single-task models. The measured inference latency on the CPU platform is 4.8 ms per sample at batch size 1 and 0.61 ms per sample at batch size 128. CPU latency is reported because OPM inference at receiver or controller nodes is often latency-sensitive and may not have access to a dedicated GPU; CPU measurement therefore provides a conservative, hardware-agnostic deployment estimate. GPU latency is expected to be lower for large batches, but it can be dominated by host-device transfer overhead at batch size 1 and is left for a hardware-specific deployment study. These results confirm that the shared multi-task design improves computational efficiency relative to deploying two separate monitoring networks. Figure 18 presents a normalised multi-metric comparison across the internally evaluated models, illustrating the balanced trade-off achieved by MT-OPMNet between accuracy and efficiency.

Finally, Figure 19 provides two complementary views of the computational profile. Figure 19a shows the parameter breakdown of MT-OPMNet: the shared convolutional backbone contains 0.52 M parameters (81.3% of the total), whereas the CAAM and the two task-specific heads each contain approximately 0.04 M parameters (6.3% each). Figure 19b compares the MAC count across the internally evaluated models, positioning MT-OPMNet at 2.6 M MACs, which is 23.5% lower than the combined single-task baseline.

To provide further insight into the performance variation across operating conditions, Figure 20 presents a heatmap of OSNR RMSE stratified by modulation format and OSNR range. The lowest estimation errors are observed in the mid-range OSNR region (18–22 dB), where the AAH features are most distinctive. Performance degrades at both low OSNR, where noise dominates the histogram shape, and high OSNR, where signal-dependent distortions become the primary limitation. Among the formats, the lower-order formats—in particular QPSK—exhibit among the lowest RMSE values, consistent with their simpler amplitude structure, whereas higher-order formats show larger errors overall because of their denser and more overlapping amplitude distributions.

Figure 19 presents a complementary heatmap showing the OSNR RMSE across the two-dimensional space of transmission distance and per-channel launch power. To isolate the interplay between linear and nonlinear impairments, the modulation format is fixed to 16-QAM for this analysis. The lowest errors are observed in the short-distance, low-power regime, where both linear impairments and nonlinear effects are minimal. As either distance or power increases, the RMSE increases accordingly, confirming the combined impact of accumulated propagation impairments and nonlinear distortion. This two-dimensional view illustrates the sensitivity of the estimator to harsher operating conditions.

To further interpret the learnt behaviour of the network, Figure 21 provides two complementary analyses of the internal model behaviour. Figure 21a visualises the learnt CAAM attention weights across the 256 feature channels for three representative modulation formats. The weight distributions differ markedly between formats, confirming that the CAAM adapts its feature emphasis in a format-dependent manner.

**TABLE 7** | Complexity comparison of the evaluated internal models.

| Model | Param. (M) | MACs (M) | Latency (ms) |
|---|---|---|---|
| ST-OSNR (CNN) | 0.42 | 1.7 | 4.7 |
| ST-MFI (CNN) | 0.42 | 1.7 | 4.8 |
| ST-OSNR + ST-MFI | 0.84 | 3.4 | 9.5 |
| MT-OPMNet | 0.64 | 2.6 | 4.8 |

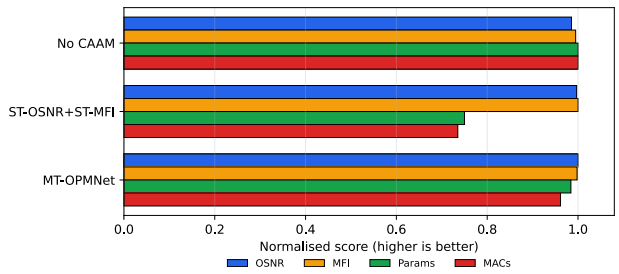


**FIGURE 18** | Normalised comparison of internally evaluated models. RMSE, parameter count, and MAC count are inverted so that larger values indicate better performance.

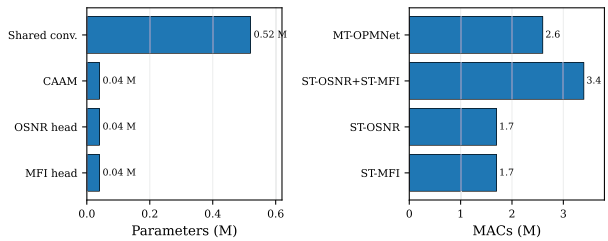


**FIGURE 19** | Computational profile of MT-OPMNet: parameter distribution and MAC comparison across internal models.

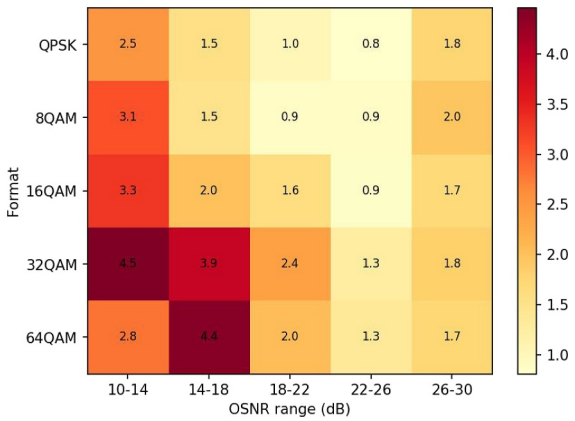


**FIGURE 20** | OSNR RMSE heatmap by modulation format and OSNR range at 28 GBd.

Figure 21b plots the evolution of the normalised task weights during training, showing that the uncertainty-based weighting strategy initially allocates roughly equal importance to both tasks but gradually shifts emphasis towards the OSNR regression task as MFI accuracy saturates. This adaptive behaviour is consistent with the homoscedastic uncertainty formulation in Ref. [35] and has been observed in other multi-task photonics applications [28, 30].

To further interpret the model's decision-making process, Figure 22 presents a gradient-based saliency map showing the importance of each AAH bin for the OSNR regression task. The saliency is computed as the absolute value of the input-output gradient $|\partial\hat{y}_{\mathrm{OSNR}}/\partial h_j|$ for each bin $j$, averaged over 500 test samples and normalised to [0, 1]. For QPSK, the model concentrates on the single dominant peak region (bins 60–80), where amplitude variations directly encode noise-induced broadening. For 16-QAM, three distinct importance peaks emerge, corresponding to the three amplitude levels visible in Figure 3. For 64-QAM, the importance is more uniformly distributed across bins 10–90, reflecting the dense multi-level structure. These saliency patterns provide empirical evidence that the model has learnt physically meaningful features aligned with the amplitude-level structure of each modulation format.

Figure 23 shows the inference latency per sample as a function of batch size for MT-OPMNet and the combined single-task models. At batch size 1, MT-OPMNet achieves 4.8 ms per sample on the CPU platform, compared with 9.5 ms for the combined single-task baseline. As batch size increases, the amortised cost decreases for both approaches, reaching 0.61 ms per sample for MT-OPMNet at batch size 128. This behaviour confirms the computational efficiency of the shared multi-task architecture.

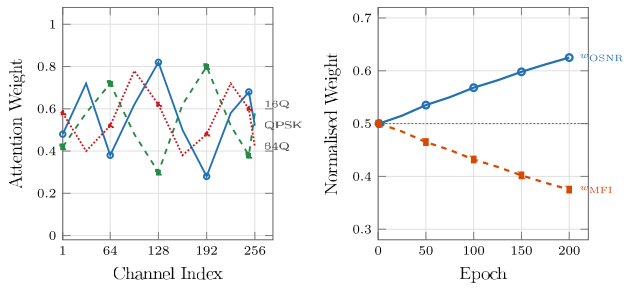


**FIGURE 21** | Internal model analysis showing CAAM feature-channel weights and adaptive task-weight evolution.

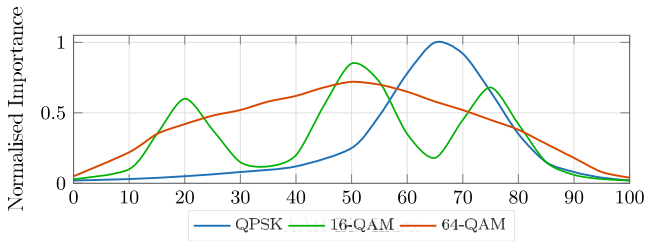


**FIGURE 22** | Gradient-based AAH-bin saliency for the OSNR regression task, averaged over 500 test samples.

The cumulative distribution of absolute OSNR estimation errors is presented in Figure 24. The curve shows the MT-OPMNet error distribution at 28 GBd, with a median absolute error of approximately 1.40 dB and a long tail caused by the most challenging high-order-format and low-OSNR cases.

To assess cross-symbol-rate generalisation, the revision separates the primary mixed-rate protocol from four single-rate diagnostic stress tests. In the primary protocol used for the main MT-OPMNet results, both 28 GBd and 64 GBd scenarios are included in the grouped training, validation and test partitions; hence the deployed model is trained with mixed-rate data rather than with a single isolated symbol rate. The four diagnostic cases in Figure 25 then isolate the effect of symbol-rate mismatch by training and testing on matched 28 GBd, matched 64 GBd, 28 →64 GBd and 64 →28 GBd settings. As expected, matched-rate training yields the strongest OSNR-estimation performance: the matched 28 GBd condition attains an OSNR RMSE of approximately 1.80 dB, whereas the matched 64GBd condition is more challenging at approximately

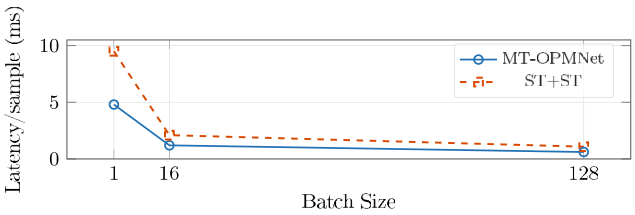


**FIGURE 23** | Inference latency per sample versus batch size on the CPU platform. MT-OPMNet achieves consistently lower latency than the combined single-task baseline across all batch sizes.

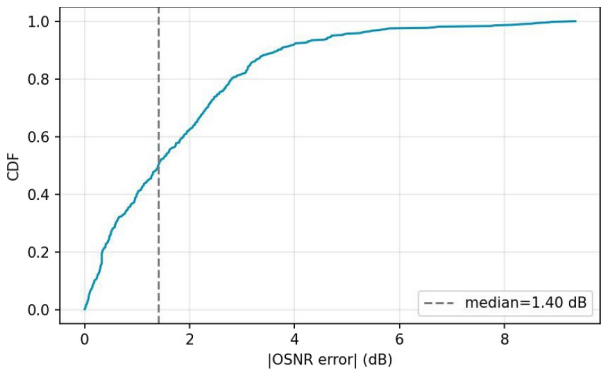


**FIGURE 24** | CDF of the absolute OSNR estimation error for MT-OPMNet at 28 GBd. The dashed line marks the median error.

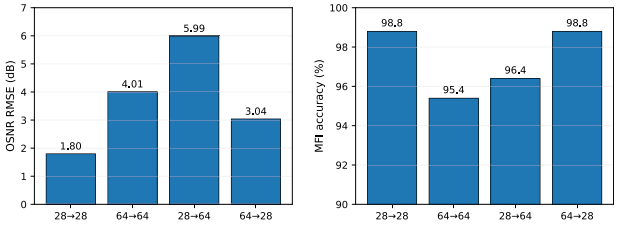


**FIGURE 25** | Cross-symbol-rate stress tests for matched-rate and cross-rate train/test configurations. Left: OSNR RMSE; right: MFI accuracy.

4.01 dB, consistent with the denser impairment conditions at the higher symbol rate. Cross-rate training introduces a clear OSNR degradation, with the RMSE rising to approximately 5.99 dB for 28 →64 GBd and 3.04 dB for 64 →28 GBd, because the AAH bin occupancies shift with symbol rate and accumulated impairment bandwidth. In contrast, the modulation-format identification task remains comparatively robust across all four configurations, with accuracy staying between approximately 95.4% and 98.8% even under rate mismatch; this asymmetry indicates that amplitude-histogram shape carries format-discriminative information that transfers across symbol rates more readily than the finer-grained features required for OSNR regression. These results indicate that practical deployment across heterogeneous transponders should retain mixed-rate training data or use domain-adaptation calibration when a new symbol-rate domain is introduced, particularly for the OSNR-estimation head.

The time complexity of the proposed MT-OPMNet can be expressed analytically. For a single forward pass, the computational cost of the $l$-th convolutional layer is $\mathcal{O}\left(K_l \cdot C_{\text{in},l} \cdot C_{\text{out},l} \cdot L_l\right)$, where $K_l$ is the kernel size, $C_{\text{in},l}$ and $C_{\text{out},l}$ are the input and output channel counts, and $L_l$ is the spatial length of the input. The CAAM adds $\mathcal{O}\left(C^2/r\right)$ operations, which is negligible compared with the convolutional layers. The fully connected layers in each head contribute $\mathcal{O}(d \cdot d_h)$, where $d_h$ is the hidden layer size. The overall inference complexity is therefore $\mathcal{O}\left(\sum_{l=1}^{4} K_l C_{\text{in},l} C_{\text{out},l} L_l + d \cdot d_h\right)$, which scales linearly with the number of histogram bins $N_{\text{bins}}$ and the number of convolutional filters.

## 7 | Cross-Validation Against Split-Step Fourier Simulation

A central concern raised during review is that the dataset is produced by a reduced analytical channel model rather than by a full numerical solution of the nonlinear Schrödinger Equation (2). The analytical model is deliberately chosen to enable dense and balanced parameter coverage at manageable cost (Section 5), but its fidelity must be established before its results can be trusted. This section provides that evidence by comparing the analytical asynchronous amplitude histograms (AAHs), and the predictions of a model trained only on them, against a fully independent split-step Fourier method (SSFM) waveform simulator.

### 7.1 | SSFM Reference Simulator

The reference simulator integrates the scalar NLSE $\partial_z A = -\frac{\alpha}{2}A + i\frac{\beta_2}{2}\partial_{tt}A - i\gamma|A|^2 A$ using the symmetric split-step Fourier method, with the linear (dispersion and loss) operator applied in the frequency domain and the nonlinear (Kerr) operator in the time domain over each step. Each transmitted symbol stream is root-raised-cosine pulse-shaped and oversampled; three co-propagating WDM channels are simulated jointly so that centre-channel self-phase modulation, nearest-neighbour cross-phase modulation, and walk-off are captured explicitly rather than approximated. The three-channel SSFM setting was selected to keep waveform-level validation computationally tractable while retaining the dominant local nonlinear interactions seen by the monitored channel. It is therefore used as a physics-based cross-check of the AAH-domain statistics, not as a full nine-channel SSFM replication of the analytical training grid. After propagation over the specified number of 80 km spans (each followed by an EDFA), the channel of interest is coherently received, compensated for accumulated chromatic dispersion, matched-filtered and down-sampled to one sample per symbol; amplified spontaneous emission is then added to reach the target OSNR, and the AAH is formed using exactly the same binning, range and normalisation as the analytical pipeline. The simulator shares no code path with the analytical model beyond this common histogram definition, so agreement between the two constitutes a genuine independent check. The complete simulator is released with the manuscript repository.

### 7.2 | Histogram-Level Agreement

For each modulation format, analytical and SSFM AAHs were generated under matched operating conditions (transmission distances of 1000–2500 km, launch powers of 0 and 3 dBm and OSNR values of 14–24 dB) and compared using cosine similarity. The results are summarised in Table 8. The mean agreement across formats is 0.976, ranging from 0.984 for QPSK to 0.971 for 32-QAM; the slight decrease for higher-order formats reflects their denser amplitude structure, which is marginally more sensitive to the residual nonlinear and filtering details that differ between the two channel descriptions. Figure 26 overlays representative analytical and SSFM histograms for all five formats, showing that the principal amplitude peaks, their relative weights and the noise-broadened tails coincide closely. This confirms that the analytical model reproduces the amplitude-domain statistics that the network actually consumes.

### 7.3 | Model Transfer to SSFM Waveforms

A stronger test is whether a network trained *only* on analytical data can operate on SSFM waveforms without any retraining or fine-tuning. The MT-OPMNet model from the main experiments was therefore evaluated, frozen on an independent set of 60 SSFM-generated AAHs spanning all five formats, two distances, two launch powers and OSNR values of 14–22 dB. Table 9 reports the outcome: modulation format identification remains at 100% for every format, and the OSNR RMSE is 2.06 dB (MAE 1.61 dB). This OSNR error is consistent with the model's analytical-domain performance at 28 GBd (approximately 2.4 dB), indicating that the simulation-to-simulation gap introduces no systematic degradation over the tested range. The per-format OSNR mean-absolute errors are likewise comparable

**TABLE 8** | Cosine similarity between analytical and SSFM asynchronous amplitude histograms, averaged over matched operating conditions (28 GBd).

| Format | Analytical versus SSFM cosine |
|---|---|
| QPSK | 0.984 |
| 8-QAM | 0.980 |
| 16-QAM | 0.975 |
| 32-QAM | 0.971 |
| 64-QAM | 0.972 |
| Mean | 0.976 |

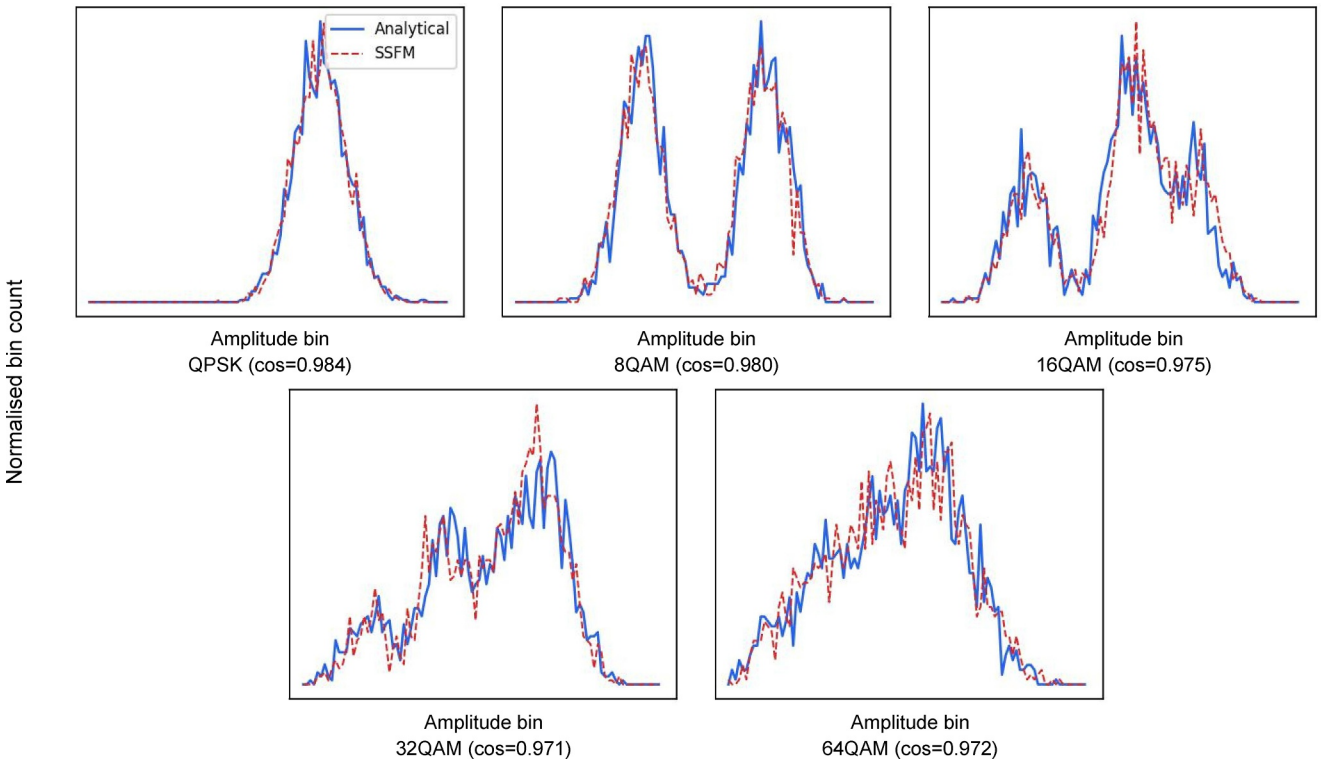


**FIGURE 26** | Analytical and SSFM AAHs for the five modulation formats at 28 GBd, 1000 km, 0 dBm, and 16 dB OSNR.

**TABLE 9** | Transfer of the analytical-trained MT-OPMNet to independent SSFM waveforms (no retraining), 28 GBd, OSNR 14–22 dB.

| Format | MFI accuracy (%) | OSNR MAE (dB) |
|---|---|---|
| QPSK | 100.0 | 1.34 |
| 8-QAM | 100.0 | 1.69 |
| 16-QAM | 100.0 | 1.77 |
| 32-QAM | 100.0 | 2.43 |
| 64-QAM | 100.0 | 0.83 |
| Overall | 100.0 | 1.61 (RMSE 2.06) |

to the analytical case, with the dense 32-QAM constellation again the most challenging.

### 7.4 | Interpretation and Scope

Together, the histogram-level agreement (mean cosine 0.976) and the zero-shot transfer (100% MFI, 2.06 dB OSNR RMSE) provide direct, quantitative evidence that the analytical model provides a useful proxy for the AAH-domain statistics examined in this study, and that the features learnt by MT-OPMNet are physically meaningful rather than artefacts of the analytical generator. This addresses the reviewers' principal methodological concern while keeping the scope appropriately bounded. The SSFM simulator is itself a numerical model and does not replace a hardware testbed; the transfer test covers a moderate OSNR range (14–22 dB), two symbol-rate-matched distances and the three-channel waveform setting described above; and behaviour at very low OSNR, under polarisation-mode dispersion or with transponder-specific impairments remains to be established. Laboratory and field validation, together with domain-adaptation calibration, therefore remain the natural next step, but the present cross-validation substantially strengthens confidence in the analytical-simulation methodology used throughout this work.

## 8 | Conclusion

This paper presented MT-OPMNet, a multi-task deep learning framework for joint optical signal-to-noise ratio estimation and modulation format identification in elastic optical networks. The proposed architecture integrates a shared convolutional feature extractor, a channel-aware attention module and two task-specific output heads within an uncertainty-weighted multi-task optimisation framework.

Evaluation on a simulated 9-channel WDM system showed that MT-OPMNet achieves an OSNR estimation RMSE of approximately 2.4 dB at 28 GBd (3.43 dB across both symbol rates) and an MFI accuracy of 97.5 % across five modulation formats. Critically, these results are statistically indistinguishable from dedicated single-task baselines, so the contribution of the shared design is parameter-efficient parity—delivering both tasks from one backbone with 23.8% fewer trainable parameters than two separate models—rather than an accuracy improvement. The most significant new evidence is a cross-validation of the analytical channel model against a split-step Fourier (SSFM) waveform simulator (Section 7), which shows that the analytical amplitude histograms match the tested SSFM AAH statistics to better than 0.97 in cosine similarity and that the trained model transfers to SSFM waveforms without retraining. The grouped scenario-level data partition further strengthens the methodological validity of the reported results. A comprehensive complexity analysis verified inference latencies of 4.8 ms per

sample at batch size 1 and 0.61 ms per sample at batch size 128 on a CPU platform.

The sensitivity analysis indicated that the proposed framework remains effective across a broad range of launch powers, symbol rates and transmission distances up to 3000 km, while also highlighting the growing impact of fibre nonlinearities under harsher operating conditions. The results should be interpreted within the stated simulation scope: although this revision adds a cross-validation against SSFM-generated waveforms (Section 7), it does not yet claim validation on laboratory or field data, and the reported metrics should not be interpreted as field-deployment guarantees.

Future work will focus on three directions. First, building on the SSFM cross-validation reported here, extending the evaluation to experimental data from laboratory or field-deployed fibre links will further validate practical applicability and quantify the remaining simulation-to-real gap. Second, incorporating additional monitoring tasks, such as chromatic dispersion estimation, nonlinear noise power prediction and soft-failure detection, into the multi-task formulation would broaden the scope of the OPM engine. Third, investigating domain adaptation, fine-tuning with small laboratory datasets and mixed-rate calibration strategies could improve cross-symbol-rate and simulation-to-hardware generalisation without retraining from scratch.

### Author Contributions

**Yassir Al-Karawi:** conceptualization, data curation, formal analysis, investigation, methodology, software, validation, visualization, writing – original draft, writing – review and editing. **Ahmed M. Jasim:** data curation, investigation, validation, visualization, writing – review and editing. **Raad S. Alhumaima:** formal analysis, investigation, methodology, validation, writing – review and editing. **Hamed S. Al-Raweshidy:** conceptualization, funding acquisition, resources, supervision, writing – review and editing.


### Acknowledgements

The authors acknowledge the support of Brunel University of London and the University of Diyala for providing the research environment in which this work was conducted.

### Funding

This work was supported in part by Brunel University of London. No additional external funding was received.


### Conflicts of Interest

The authors declare no conflicts of interest.

### Data Availability Statement

The source code and supporting materials associated with this study are available in the MT-OPMNet GitHub repository: https://github.com/YassirALKarawi/MT-OPMNet.

### References


1. O. Gerstel, M. Jinno, A. Lord, and S. J. B. Yoo, "Elastic Optical Networking: A New Dawn for the Optical Layer?," *IEEE Communications Magazine* 50, no. 2 (2012): 12–20, https://doi.org/10.1109/Mcom.2012.6146481.

2. A. A. Abdulsattar, H. S. Mogheer, M. S. Al-Abadi, and Y. A. A. Al-Karawi, "Energy-Efficient Sub-Millisecond DWDM Backhaul for 6G Open RAN," *International Journal of Electronics Engineering Research* 13, no. 4 (2025): 861–869, https://doi.org/10.37391/IJEER.130426.

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," *Journal of Lightwave Technology* 34, no. 2 (2016): 525–543, https://doi.org/10.1109/JLT.2015.2480798.

4. F. N. Khan, Q. Fan, C. Lu, and A. P. T. Lau, "An Optical Communication's Perspective on Machine Learning and its Applications," *Journal of Lightwave Technology* 37, no. 2 (2019): 493–516, https://doi.org/10.1109/JLT.2019.2897313.

5. F. Musumeci, C. Rottondi, A. Nag, et al., "An Overview on Application of Machine Learning Techniques in Optical Networks," *IEEE Communications Surveys & Tutorials* 21, no. 2 (2019): 1383–1408, https://doi.org/10.1109/comST.2018.2880039.

6. D. Rafique and L. Velasco, "Machine Learning for Network Automation: Overview, Architecture, and Applications," *Journal of Optical Communications and Networking* 10, no. 10 (2018): D126–D143, https://doi.org/10.1364/JOCN.10.00D126.

7. W. S. Saif, M. A. Esmail, A. M. Ragheb, T. A. Alshawi, and S. A. Alshebeili, "Machine Learning Techniques for Optical Performance Monitoring and Modulation Format Identification: A Survey," *IEEE Communications Surveys & Tutorials* 22, no. 4 (2020): 2839–2882, https://doi.org/10.1109/comST.2020.3018494.

8. R. Borkowski, D. Zibar, A. Caballero, V. Arlunno, and I. T. Monroy, "Stokes Space-Based Optical Modulation Format Recognition for Digital Coherent Receivers," *IEEE Photonics Technology Letters* 25, no. 21 (2013): 2129–2132, https://doi.org/10.1109/LPT.2013.2282303.

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 Technology Letters* 28, no. 17 (2016): 1886–1889, https://doi.org/10.1109/LPT.2016.2574800.

10. D. Wang, M. Zhang, Z. Li, et al., "Modulation Format Recognition and OSNR Estimation Using CNN-Based Deep Learning," *IEEE Photonics Technology Letters* 29, no. 19 (2017): 1667–1670, https://doi.org/10.1109/LPT.2017.2742553.

11. T. Tanimura, T. Hoshida, T. Kato, S. Watanabe, and H. Morikawa, "Convolutional Neural Network-Based Optical Performance Monitoring for Optical Transport Networks," *Journal of Optical Communications and Networking* 11, no. 1 (2019): A52–A59, https://doi.org/10.1364/JOCN.11.000A52.

12. H. J. Cho, S. Varughese, D. Lippiatt, R. DeSalvo, S. Tibuleac, and S. E. Ralph, "Optical Performance Monitoring Using Digital Coherent Receivers and Convolutional Neural Networks," *Optics Express* 28, no. 21 (2020): 32087–32104, https://doi.org/10.1364/OE.406294.

13. C. Wang, S. Fu, Z. Xiao, M. Tang, and D. Liu, "Long Short-Term Memory Neural Network (LSTM-NN) Enabled Accurate Optical Signal-to-Noise Ratio Monitoring," *Journal of Lightwave Technology* 37, no. 16 (2019): 4140–4146, https://doi.org/10.1109/JLT.2019.2904263.

14. S. Zhang, F. Yaman, K. Nakamura, et al., "Field and Lab Experimental Demonstration of Nonlinear Impairment Compensation Using Neural Networks," *Nature Communications* 10, no. 1 (2019): 3033, https://doi.org/10.1038/s41467-019-10911-9.

15. S. Vandenhende, S. Georgoulis, W. Van Gansbeke, M. Proesmans, D. Dai, and L. Van Gool, "Multi-Task Learning for Dense Prediction Tasks: A Survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44, no. 7 (2022): 3614–3633, https://doi.org/10.1109/TPAMI.2021.3054719.

16. F. N. Khan, K. Zhong, X. Zhou, et al., "Joint OSNR Monitoring and Modulation Format Identification in Digital Coherent Receivers Using Deep Neural Networks," *Optics Express* 25, no. 15 (2017): 17767–17776, https://doi.org/10.1364/OE.25.017767.

17. X. Fan, L. Wang, F. Ren, et al., "Feature-Fusion-Based Multi-Task ConvNet for Simultaneous Optical Performance Monitoring and Bit-Rate/Modulation Format Identification," *IEEE Access* 7 (2019): 126709–126719, https://doi.org/10.1109/ACCESS.2019.2939043.

18. Z. Wang, A. Yang, P. Guo, and P. He, "OSNR and Nonlinear Noise Power Estimation for Optical Fiber Communication Systems Using LSTM Based Deep Learning Technique," *Optics Express* 26, no. 16 (2018): 21346–21357, https://doi.org/10.1364/OE.26.021346.

19. X. Fan, Y. Xie, F. Ren, et al., "Joint Optical Performance Monitoring and Modulation Format/Bit-Rate Identification by CNN-Based Multi-Task Learning," *IEEE Photonics Journal* 10, no. 5 (2018): 7906712, https://doi.org/10.1109/JPHOT.2018.2869972.

20. C. Yu, H. Wang, C. Ke, Z. Liang, S. Cui, and D. Liu, "Multi-Task Learning Convolutional Neural Network and Optical Spectrums Enabled Optical Performance Monitoring," *IEEE Photonics Journal* 14, no. 2 (2022): 1–8, https://doi.org/10.1109/JPHOT.2022.3153638.

21. L. Xia, J. Zhang, S. Hu, M. Zhu, Y. Song, and K. Qiu, "Transfer Learning Assisted Deep Neural Network for OSNR Estimation," *Optics Express* 27, no. 14 (2019): 19398–19406, https://doi.org/10.1364/OE.27.019398.

22. J. Hu, L. Shen, and G. Sun, "Squeeze-And-Excitation Networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)* (2018), 7132–7141, https://doi.org/10.1109/CVPR.2018.00745.

23. S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional Block Attention Module," in *Proc. Eur. Conf. Comput. Vis. (ECCV)* (2018), 3–19, https://doi.org/10.1007/978-3-030-01234-2_1.

24. Y. An, L. Huang, J. Li, J. Leng, L. Yang, and P. Zhou, "Deep Learning-Based Real-Time Mode Decomposition for Multimode Fibers," *IEEE Journal of Selected Topics in Quantum Electronics* 26, no. 4 (2020): 4400806, https://doi.org/10.1109/JSTQE.2020.2969511.

25. O. Sidelnikov, A. Redyuk, S. Sygletos, M. Fedoruk, and S. K. Turitsyn, "Advanced Convolutional Neural Networks for Nonlinearity Mitigation in Long-Haul WDM Transmission Systems," *Journal of Lightwave Technology* 39, no. 8 (2021): 2397–2406, https://doi.org/10.1109/JLT.2021.3051609.

26. H. Luo, Z. Huang, X. Wu, and C. Yu, "Cost-Effective Multi-Parameter Optical Performance Monitoring Using Multi-Task Deep Learning With Adaptive ADTP and AAH," *Journal of Lightwave Technology* 39, no. 6 (2021): 1733–1741, https://doi.org/10.1109/JLT.2020.3041520.

27. F. Yang, C. Bai, X. Chi, et al., "Intelligent Joint Multi-Parameter Optical Performance Monitoring Scheme Based on HT Images and MT-ResNet for Elastic Optical Network," *Optical Fiber Technology* 82 (2024): 103599, https://doi.org/10.1016/j.yofte.2023.103599.

28. D. Zhang, J. Shi, Y. Cao, and Y. L. Xue, "Joint Three-Task Optical Performance Monitoring With High Performance and Superior Generalizability Using a Meta-Learning-Based Convolutional Neural Network-Attention Algorithm and Amplitude-Differential Phase Histograms Across WDM Transmission Scenarios," *Photonics* 12, no. 4 (2025): 324, https://doi.org/10.3390/photonics12040324.

29. M. Hao, W. He, X. Jiang, et al., "Modulation Format Identification Based on Multi-Dimensional Amplitude Features for Elastic Optical Networks," *Photonics* 11, no. 5 (2024): 390, https://doi.org/10.3390/photonics11050390.

30. Q. Zeng, Y. Kong, P. Zhou, and Y. Lu, "Adaptive Multi-Task Convolutional Neural Network for Optical Performance Monitoring," *Optics Communications* 583 (2025): 131702, https://doi.org/10.1016/j.optcom.2025.131702.

31. S. Kulandaivel and R. K. Jeyachitra, "Experimental Analysis of Optical Spectrum Based Power Distribution Analysis for Intermediate Node Monitoring Using Shallow Multi-Task ANN," *Optical Fiber Technology* 88 (2024): 104013, https://doi.org/10.1016/j.yofte.2024.104013.

32. S. Kulandaivel and R. K. Jeyachitra, "Combined Image Hough Transform Based Simultaneous Multi-Parameter Optical Performance Monitoring for Intelligent Optical Networks," *Optical Fiber Technology* 79 (2023): 103357, https://doi.org/10.1016/j.yofte.2023.103357.

33. R. D'Ingillo, A. D'Amico, R. Ambrosone, et al., "Deep Learning Gain and Tilt Adaptive Digital Twin Modeling of Optical Line Systems for Accurate OSNR Predictions," in *Proc. Int. Conf. Optical Network Design and Modelling (ONDM)* (2024), 1–3, https://doi.org/10.23919/ONDM61578.2024.10582772.

34. L. Moeller, "Non-Manakovian Propagation in Optical Fiber," in *The Nonlinear Schrödinger Equation*, eds. N. Antar and I. Bakırtaş (IntechOpen, 2022), https://doi.org/10.5772/intechopen.103694.

35. A. Kendall, Y. Gal, and R. Cipolla, "Multi-Task Learning Using Uncertainty to Weigh Losses for Scene Geometry and Semantics," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)'* (2018), 7482–7491, https://doi.org/10.1109/CVPR.2018.00781.

36. T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal Loss for Dense Object Detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)* (2017), 2999–3007, https://doi.org/10.1109/ICCV.2017.324.