International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 11
Article Received: *25 October 2022* ***Revised:*** *12 November 2022* ***Accepted:*** *30 November 2022*

Machine Learning-Based Modulation Format Identification Under Varying OSNR

Paulomi Nilesh Patel

Lecturer, Department Of Electronics And Communication Engg., Government Polytechnic for Girls, Surat

gpgsecpgp@gmail.com

Abstract— A receiver must frequently determine the format of an incoming signal without previous knowledge because elastic optical networks alter a channel's modulation format to fit the quality of the link. In this work, four forms (QPSK, 8-QAM, 16-QAM, and 64-QAM) at optical signal-to-noise ratios (OSNR) ranging from 4 to 26 dB are identified using machine learning. Three classifiers—a support vector machine (SVM), a random forest (RF), and a tiny convolutional neural network (CNN) using 32 x 32 constellation images—are evaluated using synthetic 32 GBd signal frames with residual phase error. At OSNR ≥ 20 dB, all three models achieve 100% accuracy, whereas at 4 dB, they approach roughly 71–74%. At 12 dB, the CNN achieves 95% accuracy before the SVM and RF, which require 14 dB. The accuracy is 91.0% (SVM), 91.0% (RF), and 92.0% (CNN) on average across all OSNR values. Confusion between 16-QAM and 64-QAM is the primary cause of error below 12 dB. The CNN requires roughly 32 seconds to train, compared to less than a second for the SVM and RF. This means that the CNN's increased accuracy comes at a significantly higher training cost. The findings clearly illustrate how OSNR restricts format identification.

Index Terms— modulation format identification, OSNR, machine learning, convolutional neural network, support vector machine, random forest, elastic optical networks.

I. Introduction

In optical networks, fixed-rate channels are replaced with flexible or elastic networks. In an elastic network, the transmitter can select the modulation format, symbol rate and spectrum width of each channel depending on the length and quality of the path [1], [2]. A long and noisy path requires a robust format like QPSK, but a short and clean path can use a dense format like 64-QAM to carry more bits per symbol. This flexibility allows the network to cope with more, but it also introduces a new challenge for the receiver: it has to first figure out what format it is receiving before it can retrieve the data.

Many of the algorithms used by a coherent receiver with digital signal processing [3] can be adapted to the format but first each algorithm has to know the format. The format can be sent over a control channel, but this increases the signalling cost and the possibility of a mismatch between transmitter and receiver. Thus, blind identification based on the received samples is attractive. It provides automatic reconfiguration and decouples the receiver from the transmitter [4].

The noise plays an important role in determining the difficulty of the task. The constellations of the four formats at two OSNR levels are shown in Fig. 1. Each format shows discrete, separate clusters at 26 dB. At 12dB, the clusters of 16-QAM and 64-QAM converge and have similar shapes. The quantity which is easy to understand is the signal amplitude distribution, which is not affected by the phases. Fig. 2 shows the average amplitude histogram for each format, which are different and have a distinct pattern at 26 dB, which is smeared at 12 dB.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 11
Article Received: *25 October 2022* ***Revised:*** *12 November 2022* ***Accepted:*** *30 November 2022*

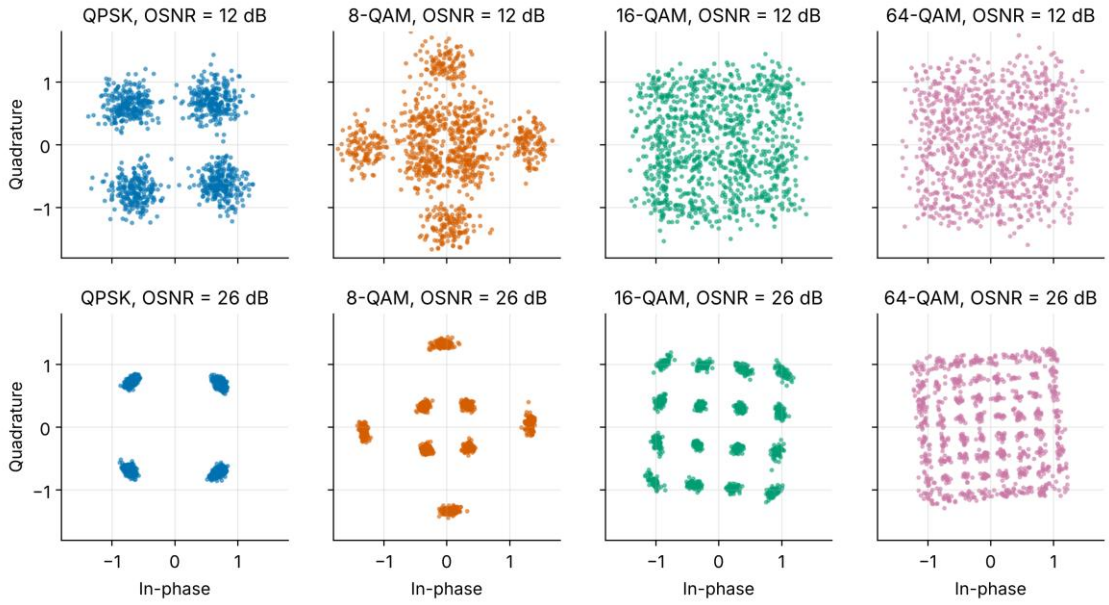


Fig. 1. Constellations at OSNR of 12 dB (top) and 26 dB (bottom).

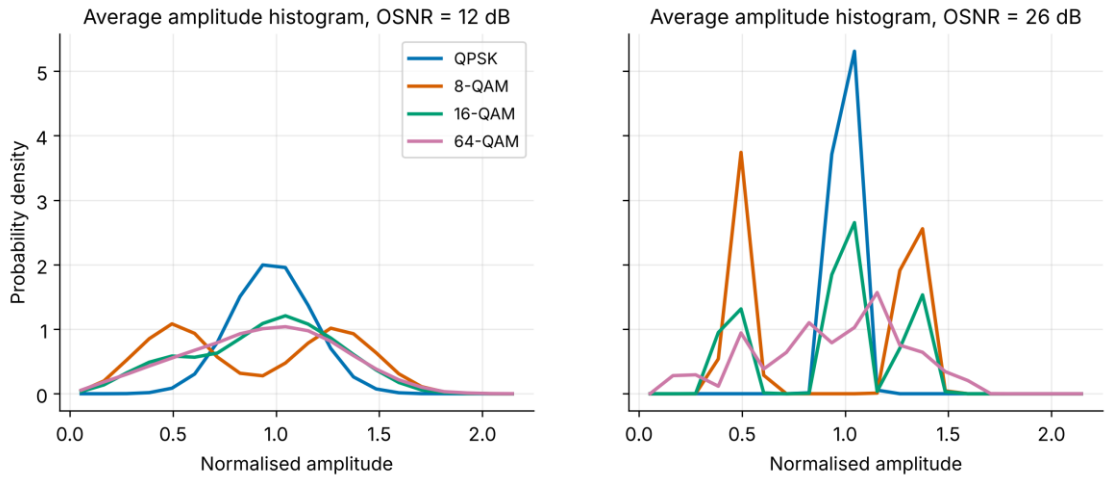


Fig. 2. Average amplitude histograms at OSNR of 12 dB and 26 dB.

Traditional modulation classification methods rely on higher order cumulants and other statistical features [5], [6]. Recent research on optical communication has employed machine learning techniques, including neural networks, clustering, and deep learning, to learn the decision rule from data [4], [7]–[10]. 2022 work generalises this to super-channel networks and coupled format monitoring with other signal properties [11]–[14]. Most of these studies show the accuracy for a single model or a few OSNR values. It is less common to compare different types of models on the same data, noise conditions, and training and usage costs.

This paper makes the following contributions:

(1) a simulation model of coherent optical signal frames for four modulation formats with OSNR from 4 to 26 dB and residual phase error;

(2) a fair comparison of an SVM, a random forest and a CNN on identical train, validation and test sets;

(3) an analysis of which formats are confused at low OSNR, with a simple explanation based on the distance between constellation points; and

(4) a comparison of the training time and the time per frame of the three models.

The remainder of the paper is organised as follows. Section II reviews related work. Section III describes the

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 11
Article Received: *25 October 2022* ***Revised:*** *12 November 2022* ***Accepted:*** *30 November 2022*

signal model, the features and the classifiers. Section IV gives the simulation setup. Section V presents the results, Section VI discusses practical points and limitations, and Section VII concludes.

II. Related Work

Elastic optical networking was proposed to increase the spectrum utilisation over fixed-grid systems, by varying the rate and format of each connection [1]. Gerstel et al. talk about how this flexibility impacts the design of the optical layer and why the network needs to be able to automatically tune the transponder settings [2]. A receiver that can identify the format of a signal is beneficial to such networks and helps to simplify the control plane.

Automatic modulation classification has been exploited long ago outside optics. Higher order cumulants of the received samples were used by Swami and Sadler to classify digital modulations in a hierarchical manner [6]. Dobre et al. review more recent advances in this subject besides the basic traditional methods based on characteristics and likelihood [5]. Cumulant characteristics are attractive due to their simplicity and robustness to the carrier phase.

The techniques and applications of machine learning in optical communication are extensively discussed by Khan et al. [4]. Format identification was carried out in heterogeneous networks with signal amplitude statistics using artificial neural networks [7]. Boada et al. used clustering techniques in the Stokes space to identify the format [8]. Later, Wang et al. proposed the combination of the format recognition and OSNR estimation based on convolutional neural network [10]. Khan et al. also applied deep learning in coherent receivers for the same purpose [9]. This trend continues with more recent research conducted in 2022. Saif et al. [11] studied the effectiveness of format identification in super-channel optical networks. Li et al. applied a multi-task residual network for optical signal monitoring [13]. Yu et al. applied a multi-task convolutional neural network with optical spectra for optical performance monitoring [12]. Jha and Mishra used transfer learning to simultaneously monitor data rate and modulation format [14]. These works demonstrate that learning based methods can achieve high accuracy. However, their results are not directly comparable as they use different signals, features, noise levels and test settings.

This paper does not propose a new classifier. The same settings are used to compare three popular models: a convolutional neural network [17], a random forest [16] and a support vector machine [15]. The comparison includes confidence intervals, confusion matrices and computing costs through the entire OSNR range, from highly noisy to clean signals, so that the advantages of each model are evident.

III. Background

A. Signal Model

A frame contains $N = 1024$ symbols s_k drawn at random, with equal probability, from the constellation of the format. All constellations are scaled to unit average power. The received symbol after carrier recovery is modelled as

$$r_k = s_k \cdot \exp(j(\theta + \psi_k)) + n_k \quad (1)$$

where θ is a constant residual phase error, uniform in ±5°, ψ_k is residual phase noise with standard deviation 3° per symbol, and n_k is complex Gaussian noise from the optical amplifiers. The noise is described by the optical signal-to-noise ratio, measured in the standard reference bandwidth $B_{ref} = 12.5$ GHz (0.1 nm). For a symbol rate R_s, the SNR per polarisation is

$$\mathrm{SNR} = \mathrm{OSNR} \cdot 2\, B_{ref} / R_s \quad (2)$$

For $R_s = 32$ GBd, this gives SNR = OSNR − 1.07 dB, so OSNR values of 4 to 26 dB correspond to SNR values of 2.9 to 24.9 dB. The noise has variance 1/SNR per complex sample. The receiver scales each frame to unit average power, because the signal level is not known in advance.

B. Features

The SVM and the random forest use 28 features computed from one frame. Let $a_k = |r_k|$ be the amplitude. The first 20 features are a histogram of a_k in 20 equal bins from 0 to 2.2, with the area normalised to 1. Four more features are the mean, standard deviation, skewness and kurtosis of a_k. Two more are the fourth and sixth moments, $E[a^4]$ and $E[a^6]$. The last two are the cumulants

$$C_{40} = E[r^4] - 3\,(E[r^2])^2, \quad C_{42} = E[|r|^4] - |E[r^2]|^2 - 2\,(E[|r|^2])^2 \quad (3)$$

of which $|C_{40}|$ and C_{42} are used [6]. All 28 features depend only on the amplitude, on $|C_{40}|$ or on C_{42}, so a constant phase rotation does not change them, and they are robust to the residual phase error. Table I lists the features.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 11
Article Received: 25 October 2022 ***Revised:*** 12 November 2022 ***Accepted:*** 30 November 2022

TABLE I. Features used by the SVM and random forest.

Feature group	Number	Purpose
Amplitude histogram (20 bins, 0–2.2)	20	Shape of the amplitude distribution
Mean, std, skewness, kurtosis of amplitude	4	Summary of the distribution
Moments $E[a^4]$, $E[a^6]$	2	Weight of large amplitudes
$\|C_{40}\|$, C_{42}	2	Higher-order cumulants

C. Classifiers

Support vector machine: The SVM finds a decision boundary that maximises margin between classes [15]. It learns curved boundaries using RBF kernel. Before training, the features are standardised and C = 10 is the punishment parameter.

A random forest is an ensemble of decision tree models that vote and are aggregated, each trained on a different subset of the data [16]. This forest has 200 trees. It trains quickly and doesn't require feature scaling.

Convolutional neural network: The 32 x 32 constellation image is sent to the CNN. Each pixel counts the symbols that fall inside a small square of the I/Q plane (range 1.8 in both axes). The image is divided by its largest count. Local forms are learned by two convolution layers and a fully connected portion makes the choice [17]. In theory, as the network views the entire I/Q plane it can utilise phase information that the amplitude features ignore.

D. Evaluation Measures

The main measure is the identification accuracy, the fraction of test frames with the correct format. Because the number of test frames is limited, a 95 % binomial confidence interval is given as

$$\Delta = 1.96 \cdot \sqrt{(p (1 - p) / n)} \quad (4)$$

where p is the accuracy and n is the number of test frames. With n = 800 frames at one OSNR value and p = 0.75, Δ is about 3 percentage points. With the n = 9600 frames of the whole test set and p = 0.91, Δ is about 0.6 points. Two accuracies whose intervals overlap should not be called different. In addition, confusion matrices, training time and the time per frame are reported.

IV. Simulation Setup

Fig. 3 shows the processing chain. Random symbols are mapped to one of four formats, passed through the noisy channel of (1), and converted either to the 28 features or to the constellation image. The classifiers then predict the format. The whole study is implemented in Python with NumPy, SciPy, scikit-learn and PyTorch.

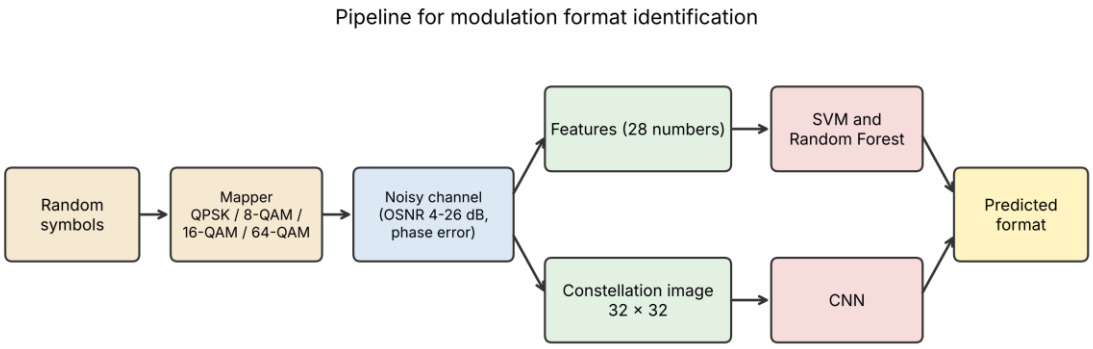


Fig. 3. Processing chain for format identification.

A. Formats and Channel

QPSK has 4 points. In the 16-QAM and 64-QAM formats, we use square grids of 4 × 4 and 8 × 8 points respectively. The star type of 8-QAM format has the outer ring of four points with radius 1+√3 and inner ring of four points with radius 1. In the discussion in Section V we use the minimum distance between points which is 1.41 for QPSK, 0.63 for 16-QAM and 0.31 for 64-QAM after scaling to unit average power.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 11
Article Received: 25 October 2022 *Revised:* 12 November 2022 *Accepted:* 30 November 2022

B. Data Sets

The data are generated at 12 OSNR values from 4 dB to 26 dB in 2 dB steps. 100 frames are used for training, 20 frames for validation, and 200 frames for testing for each format and OSNR value. The three sets have no common frames and were generated with different random seeds. There are 800 test frames for each OSNR value. This results in 4800 training, 960 validation and 9600 test frames. The receiver does not need any prior knowledge of the OSNR since the models are trained on the combined frames of all OSNR values.

C. Model Settings

The CNN consists of two convolution layers (16 and 32 filters of size 3 x 3 followed by ReLU and 2 x 2 max pooling) and a fully connected layer with 64 units, ReLU, dropout of 0.3 and an output layer with 4 units. It is trained for 25 epochs using the Adam optimiser [18] with learning rate 0.001, batch size 128 and cross-entropy loss. The settings are given in Table II.

TABLE II. Simulation and model settings.

Item	Setting
Formats	QPSK, 8-QAM (star), 16-QAM, 64-QAM
Symbol rate, frame length	32 GBd, 1024 symbols
OSNR range	4 to 26 dB, step 2 dB (reference 12.5 GHz)
Residual phase error	±5° constant, 3° per symbol (std)
Frames per format per OSNR	100 train, 20 validation, 200 test
SVM	RBF kernel, C = 10, standardised features
Random forest	200 trees
CNN	2 conv layers (16, 32), FC 64, dropout 0.3, Adam, 25 epochs
Hardware	CPU only

V. Results and Discussion

A. Accuracy versus OSNR

Fig. 4 shows the accuracy of the three models for each OSNR value, and Table III lists selected values. All models reach 100 % accuracy at OSNR ≥ 20 dB, and 99.5 % or better from 16 dB. Below 12 dB the accuracy falls to a plateau of about 71–76 % between 4 and 8 dB. Because there are four equally likely formats, the accuracy of random guessing is 25 %, so the models still carry much information at low OSNR.

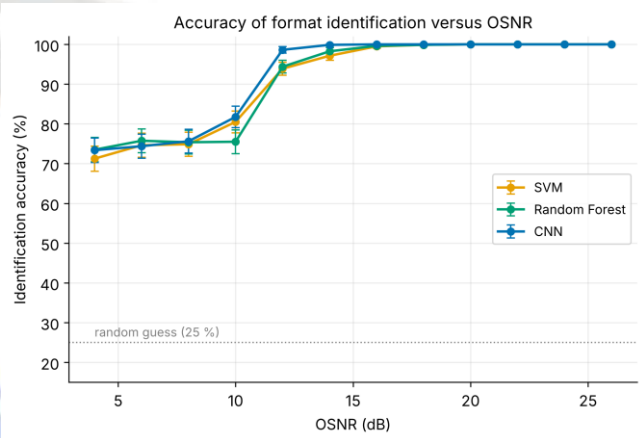


Fig. 4. Accuracy versus OSNR (bars: 95 % confidence interval).

TABLE III. Accuracy at each OSNR value (800 test frames each).

OSNR (dB)	SNR (dB)	SVM (%)	RF (%)	CNN (%)
4	2.9	71.2	73.5	73.4
6	4.9	74.6	75.8	74.4
8	6.9	74.9	75.4	75.6
10	8.9	80.5	75.5	81.8
12	10.9	93.9	94.4	98.6
14	12.9	97.1	98.2	99.9
16	14.9	99.5	99.6	100.0
18	16.9	99.9	99.9	100.0
20	18.9	100.0	100.0	100.0
22	20.9	100.0	100.0	100.0

OSNR (dB)	SNR (dB)	SVM (%)	RF (%)	CNN (%)
24	22.9	100.0	100.0	100.0
26	24.9	100.0	100.0	100.0

The results of the three models are similar at 4, 6 and 8 dB and their confidence intervals (around ±3 points) overlap. The transition region is between 10 and 14 dB where the models differ from one another. At 12 dB, the CNN reaches 98.6%, while the SVM and random forest reach 93.9% and 94.4% respectively; these discrepancies are outside the confidence intervals (around ±1 to 2 points). CNN reports 99.9% at 14 dB, the other two report 97.1% and 98.2%. Thus, the minimum OSNR with an accuracy of at least 95 % is 12 dB for the CNN and 14 dB for the SVM and random forest for the 2 dB grid (a gain of 2 dB for the CNN).

The average accuracy over the entire range is 91.0% for SVM, 91.0% for random forest and 92.0% for CNN. Thus, the CNN is superior to the other two by 1.0 points and the 95% confidence interval of each average is about ±0.6 points. Not a great deal of difference. The majority of the CNN gain comes from the transition zone between 10 and 14 dB, rather than an overall improvement across all OSNR values.

The advantage of CNN in this area may be because it looks at the whole I/Q plane, including the ring structure of the constellation, while the SVM and forest only look at the amplitude statistics. The ring structure can still be seen in the image when the points overlap, even after the amplitude histogram has become smooth. The results are in line with this theory, although not tested independently here (e.g. by training the CNN on amplitude-only images).

B. Confusion Between Formats

Fig. 5 shows confusion matrices at OSNR = 10 dB, where the models differ and the task is still hard. QPSK and 8-QAM are identified without error by all three models. All errors come from the two square formats: many 16-QAM frames are predicted as 64-QAM, and some 64-QAM frames are predicted as 16-QAM. Table IV gives the accuracy for each format.

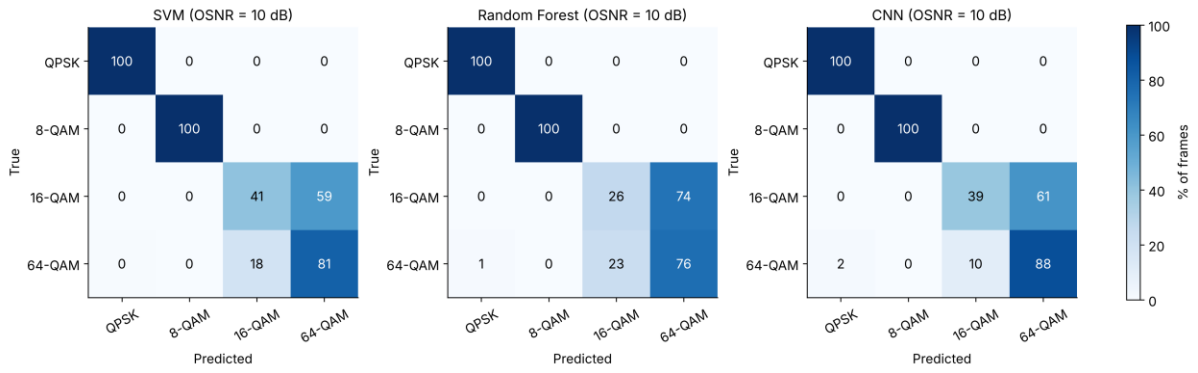


Fig. 5. Confusion matrices at OSNR of 10 dB.

TABLE IV. Accuracy per format at OSNR of 10 dB.

Format	SVM (%)	RF (%)	CNN (%)	Frames
QPSK	100	100	100	200
8-QAM	100	100	100	200
16-QAM	41	26	39	200
64-QAM	81	76	88	200
Average	80.5	75.5	81.8	800

That's why you see this pattern when you look at the distances between points. The SNR is 8.9 dB and the noise standard deviation is about 0.25 per real component at OSNR=10 dB. For 16- QAM the neighbouring points overlap significantly as half the minimum distance is 0.32, which is only 1.25 times the noise standard deviation. With 64-QAM, the 64 points coalesce into one smooth cloud at half the minimum distance of 0.15, or 0.6 times the noise standard deviation. The two formats then possess similar amplitude statistics and the two clouds look similar. The QPSK and 8-QAM are distinguished from the square formats due to their much larger separation between points and their amplitude pattern.

For 64-QAM, the CNN is better than the SVM and the forest (88% vs. 81% and 76%), but not for 16-QAM, when the models are all below 50%. At 10 dB the forest picks 64-QAM for both square formats (74% of 16-QAM frames) and lowers its average accuracy. This shows the importance of the confusion matrix and how average accuracy can hide the different errors the models are making.

C. Computing Cost

The comparison of accuracy, training time and one frame classification time is shown in Fig. 6 and Table V. The CNN takes 31.7 seconds to train, which is about 240 times longer than the SVM, which takes 0.13 seconds, or the random forest, which takes 0.68 seconds. All three are fast to use after training: 0.045 ms for SVM, 0.008 ms for RF and 0.13 ms for CNN per CPU frame. Even 0.13 ms is not much, but the CNN is approximately three times slower in use than the SVM and sixteen times slower than the random forest. These classifiers would act on certain frames rather than on all frames of a live signal since a frame of 1024 symbols at 32 GBd has a duration of 32 ns.

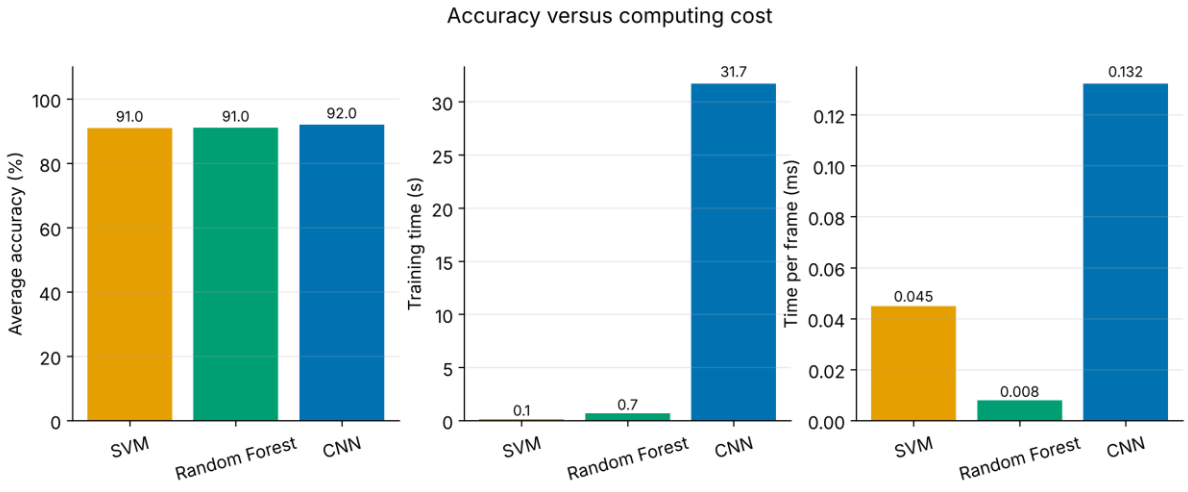


Fig. 6. Accuracy and computing cost of the three models.

TABLE V. Accuracy and cost.

Model	Average accuracy (%)	Training time (s)	Time per frame (ms)
SVM	91.0	0.13	0.045
Random forest	91.0	0.68	0.008
CNN	92.0	31.7	0.132

The training time of the CNN will be much shorter on a graphics processor, and the time for the SVM and forest would rise if the training set were larger. The numbers are therefore given for the CPU system used here and should be read as relative values.

D. CNN Training Behaviour

Fig. 7 shows the loss and the accuracy of the CNN during training. The training loss falls from 0.44 to 0.11 and the validation loss from 0.44 to about 0.15. The validation

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 11
Article Received: *25 October 2022* ***Revised:*** *12 November 2022* ***Accepted:*** *30 November 2022*

accuracy stays at about 91–92 % after the tenth epoch, while the training accuracy rises to about 94 %. A gap of about two to three points between training and validation indicates mild over-fitting, but the validation loss does not rise, so longer training does not make the network worse. Dropout helps to control this gap.



Fig. 7. CNN loss and accuracy during training.

E. Summary of the Comparison

Table VI summarises the main findings. For OSNR above 16 dB, any of the three models identifies the format almost perfectly, and the simple SVM or forest is enough. Between 10 and 14 dB, the CNN gives the best accuracy. Below 10 dB, no model separates 16-QAM from 64-QAM reliably, which suggests that more information than a single frame of 1024 symbols, or a different feature, is needed in that region.

TABLE VI. Summary of the comparison.

Measure	SVM	RF	CNN
Lowest OSNR with accuracy ≥ 95 %	14 dB	14 dB	12 dB
Accuracy at 4 dB (%)	71.2	73.5	73.4
Lowest OSNR with 100.0 % (rounded)	20 dB	20 dB	16 dB
Training time	0.13 s	0.68 s	31.7 s

VI. Practical Considerations and Limitations

A. Practical Considerations

The model selection is driven by the running OSNR. In a metro or regional network, the received signal OSNR is often much above 15 dB, and a basic SVM or random forest with 28 features can identify the formats. These models are compact, train in a split second, and are simple to describe as the features have a physical meaning. The CNN does provide some added accuracy for a long link close to the format's OSNR limit. Alternatively, a hybrid architecture could be used, where the basic model operates initially and the CNN is only activated when the model lacks confidence.

An OSNR estimate is not required since the models were trained on data from all OSNR values combined. However, the training data should span the noise, phase error and symbol speeds that an actual receiver will see. A model trained under one situation may be less accurate under another situation before being applied.

B. Limitations

We employ a simple phase error model and synthetic data with additive Gaussian noise. The absence of

residual chromatic dispersion, polarisation effects, frequency offset, pulse shaping and fibre nonlinearity modify the shape of the constellation. The symbols are independent and equiprobable and only four formats are investigated. The accuracy is measured on 9600 test frames to obtain wide confidence intervals at single OSNR values. The CNN was trained only once, with a single random seed, so there is no way to quantify the variation across training runs. Thus the conclusions are valid for the conditions of Table II and the measured signals will have different absolute accuracies.

VII. Conclusion

This study compares an SVM, a random forest and a CNN to detect QPSK, 8-QAM, 16-QAM and 64-QAM signals in the OSNR range of 4 to 26 dB. All models are 100% accurate at 20 dB and above. The CNN achieves the highest mean accuracy (92.0% vs. 91.0%) and 95% accuracy at 12 dB, which is 2 dB lower than the SVM and the forest. The most significant errors are located in the region below 12 dB for 16-QAM and 64-QAM. The distance between constellation sites explains this behaviour. For high OSNR and low computational resources, the SVM and random forest are an excellent choice, training 50–240 times faster than the CNN.

Future work may include measured signals, other modulations like 32-QAM, probabilistically structured signals, and adding residual dispersion, frequency offset and nonlinear noise. Combining the amplitude features with the constellation image or combining multiple frames may help to identify 16-QAM and 64-QAM at low OSNR. Another useful extension is the joint estimation of OSNR and format.

References

[1] M. Jinno, H. Takara, B. Kozicki, Y. Tsukishima, Y. Sone, and S. Matsuoka, “Spectrum-efficient and scalable elastic optical path network: Architecture, benefits, and enabling technologies,” *IEEE Commun. Mag.*, vol. 47, no. 11, pp. 66–73, Nov. 2009.

[2] O. Gerstel, M. Jinno, A. Lord, and S. J. B. Yoo, “Elastic optical networking: A new dawn for the optical layer?” *IEEE Commun. Mag.*, vol. 50, no. 2, pp. s12–s20, Feb. 2012.

[3] S. J. Savory, “Digital coherent optical receivers: Algorithms and subsystems,” *IEEE J. Sel. Topics Quantum Electron.*, vol. 16, no. 5, pp. 1164–1179, Sep./Oct. 2010.

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

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

[6] A. Swami and B. M. Sadler, “Hierarchical digital modulation classification using cumulants,” *IEEE Trans. Commun.*, vol. 48, no. 3, pp. 416–429, Mar. 2000.

[7] F. N. Khan, Y. Zhou, A. P. T. Lau, and C. Lu, “Modulation format identification in heterogeneous fiber-optic networks using artificial neural networks,” *Opt. Express*, vol. 20, no. 11, pp. 12422–12431, 2012.

[8] R. Boada, R. Borkowski, and I. T. Monroy, “Clustering algorithms for Stokes space modulation format recognition,” *Opt. Express*, vol. 23, no. 12, pp. 15521–15531, 2015.

[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 Photon. Technol. Lett.*, vol. 28, no. 17, pp. 1886–1889, Sep. 2016.

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

[11] W. S. Saif, A. M. Ragheb, B. Nebendahl, T. Alshawi, M. Marey, and S. A. Alshebeili, “Performance investigation of modulation format identification in super-channel optical networks,” *IEEE Photon. J.*, vol. 14, no. 2, pp. 1–10, 2022.

[12] 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 Photon. J.*, vol. 14, no. 2, pp. 1–8, 2022.

[13] J. Li *et al.*, “Optical signal monitoring using multi-task residual network,” *Optik*, vol. 259, Art. no. 169003, 2022.

[14] D. K. Jha and J. K. Mishra, “Transfer learning approach toward joint monitoring of bit rate and

modulation format," *Appl. Opt.*, vol. 61, no. 13, pp. 3695–3701, 2022.

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

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

[17] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.

[18] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, San Diego, CA, USA, 2015.