CISTER

Journal Paper

3D convolutional neural networks based automatic modulation classification in the presence of channel noise

Rahim Khan

Qiang Yang

Inam Ullah

Ateeq Ur Rehman

Ahsan Bin Tufail

Alam Noor*

Abdul Rehman

Korhan Cengiz

*CISTER Research Centre

3D convolutional neural networks based automatic modulation classification in the presence of channel noise

Rahim Khan, Qiang Yang, Inam Ullah, Ateeq Ur Rehman, Ahsan Bin Tufail, Alam Noor*, Abdul Rehman, Korhan Cengiz

*CISTER Research Centre

Polytechnic Institute of Porto (ISEP P.Porto)

Rua Dr. António Bernardino de Almeida, 431

4200-072 Porto

Portugal

Tel.: +351.22.8340509, Fax: +351.22.8321159

E-mail: yq@hit.edu.cn, ahsantufail1985@gmail.com, alamn@isep.ipp.pt

https://www.cister-labs.pt

Abstract

Automatic modulation classification is a task that is essentially required in many intelligent communication systems such as fibre-optic, next-generation 5G or 6G systems, cognitive radio as well as multimedia internet-of-things networks etc. Deep learning (DL) is a representation learning method that takes raw data and finds representations for different tasks such as classification and detection. DL techniques like Convolutional Neural Networks (CNNs) have a strong potential to process and analyse large chunks of data. In this work, we considered the problem of multiclass (eight classes) classification of modulated signals, which are, Binary Phase Shift Keying, Quadrature Phase Shift Keying, 16 and 64 Quadrature Amplitude Modulation corrupted by Additive White Gaussian Noise, Rician and Rayleigh fading channels using 3D-CNN architectures in both frequency and spatial domains while deploying three approaches for data augmentation, which are, random zoomed in/out, random shift and random weak Gaussian blurring augmentation techniques with a cross-validation (CV) based hyperparameter selection statistical approach. Simulation results testify the performance of 10-fold CV without augmentation in the spatial domain to be the best while the worst performing method happens to be 10-fold CV without augmentation in the frequency domain and we found learning in the spatial domain to be better than learning in the frequency domain.

IET Communications

IET The Institution of Engineering and Technology | WILEY

ORIGINAL RESEARCH PAPER

3D convolutional neural networks based automatic modulation classification in the presence of channel noise

Rahim Khan[1] | **Qiang Yang[1]** | **Inam Ullah[2]** | **Ateeq Ur Rehman[3]** | **Ahsan Bin Tufail[1,5]** | **Alam Noor[4]** | **Abdul Rehman[6,7]** | **Korhan Cengiz[8]**

[1] School of Electronics and Information Engineering, Harbin Institute of Technology, Harbin, China

[2] College of Internet of Things (IoT) Engineering, Hohai University (HHU), Changzhou Campus, Changzhou, China

[3] Department of Electrical Engineering, Government College University, Lahore, Pakistan

[4] CISTER Research Centre, ISEP, Politécnico do Porto, Portugal

[5] Department of Electrical and Computer Engineering, COMSATS University Islamabad, Sahiwal Campus, Sahiwal, Pakistan

[6] Department of Computer Science and Engineering, Kyungpook National University, Daegu, South Korea

[7] Faculty of Media and Communication Studies, University of Central Punjab, Lahore, Pakistan

[8] Department of Electrical - Electronics Engineering, Trakya University, Edirne, Turkey

Correspondence

Qiang Yang, School of Electronics and Information Engineering, Harbin Institute of Technology, Harbin 150001, China.
Email: yq@hit.edu.cn

Funding information

National Natural Science Foundation of China, Grant/Award Number: 62031014; Key Research and Development Program of Hainan Province (China), Grant/Award Number: ZDYF2019195

Abstract

Automatic modulation classification is a task that is essentially required in many intelligent communication systems such as fibre-optic, next-generation 5G or 6G systems, cognitive radio as well as multimedia internet-of-things networks etc. Deep learning (DL) is a representation learning method that takes raw data and finds representations for different tasks such as classification and detection. DL techniques like Convolutional Neural Networks (CNNs) have a strong potential to process and analyse large chunks of data. In this work, we considered the problem of multiclass (eight classes) classification of modulated signals, which are, Binary Phase Shift Keying, Quadrature Phase Shift Keying, 16 and 64 Quadrature Amplitude Modulation corrupted by Additive White Gaussian Noise, Rician and Rayleigh fading channels using 3D-CNN architectures in both frequency and spatial domains while deploying three approaches for data augmentation, which are, random zoomed in/out, random shift and random weak Gaussian blurring augmentation techniques with a cross-validation (CV) based hyperparameter selection statistical approach. Simulation results testify the performance of 10-fold CV without augmentation in the spatial domain to be the best while the worst performing method happens to be 10-fold CV without augmentation in the frequency domain and we found learning in the spatial domain to be better than learning in the frequency domain.

1 | INTRODUCTION

Automatic modulation classification (AMC) is a task that is essentially required in many intelligent communication systems. Software defined radio (SDR) is a new age technology that has driven communication systems to become more flexible to take maximum advantage of the available limited bandwidth. In a cooperative system, AMC enables transceiver components to switch modulation schemes in the presence of noise. AMC supports monitoring, intelligence and recognition tasks by identifying the modulation scheme of an undefined signal in a non-cooperative system [1, 2]. In addition to SDR technology, AMC is also deployed in cognitive radio (CR) to sense the spectrum band [3] and utilize the best available band

for services to avoid user interference and congestion [4–6]. Other applications of AMC are in fibre optic channels for the transport of radio and millimetre wave signals at long distances for the next generation 5G and sub terahertz communications [7], the Internet of Things (IoT) technology [8, 9], automated modulation of Orthogonal Frequency Division Multiplexing (OFDM) systems [10–13], beyond 5G low-latency communication [14], multimedia IoT [15], and spectrum monitoring and signal interception [16] etc.

Deep learning (DL) technology has revolutionized the landscape of modern technologies. Its applications such as neural networks have been instrumental in bringing changes to the communication technologies worldwide. Convolutional Neural Networks (CNNs) are a powerful family of neural networks for learning from data and have wide applications in image recognition, object detection [17, 18], and semantic segmentation tasks etc. Typical desirable properties of the features learned by CNNs are spatial invariance, translation invariance, and locality while typical components of these networks are convolutional layer, padding and stride operations, maximum and average pooling layers, fully connected (FC) layer, dropout layer, and batch normalization layer etc. For object recognition purposes, deformations like pose, affine transformations like scaling, translation, optical flow as well as rotation or shear are widely used as augmentation methods to synthetically increase the size of the datasets [19]. Colour information instead of a grayscale image may also improve prediction performance [20]. Spatial transformation methods such as per-pixel flow, mean blur and differentiable bilinear interpolation can also be used to deform the input images benefiting many visual recognition tasks [21]. CNNs performance in a classification process is based on several aspects, including fulfilling the requirements of the Nyquist sampling principle. Small shifts in the input can drastically change the output of a CNN. Classic anti-aliasing may improve shift equivariance of deep networks leading towards better generalization of the network. Corruptions like salt and pepper noise, masking noise and additive isotropic Gaussian noise makes it difficult to learn a useful representation. In general, CNNs performance is robust to corruptions such as rotation, scaling, blurring, and noise variants. Literal shift equivariance cannot hold when subsampling and recovered only when features can be extracted densely. Shift equivariance is lost in modern deep networks as commonly used down sampling layers such as strided convolution, max pooling and average pooling ignores Nyquist sampling theorem [22].

Modulation is a technique by which a message-carrying signal is superimposed on a carrier signal for transmission. Commonly used modulation schemes are Binary Phase Shift Keying (BPSK), Quadrature Phase Shift Keying (QPSK) and Quadrature Amplitude Modulation (QAM). The term PSK is widely used in a radio communication system. This method is largely compatible with data transmissions. It allows information to be transmitted through a radio communications signal in a more efficient manner as compared to other modulation techniques. QAM combines amplitude and phase information at different

levels and has wide applications in internet services and digital cable television.

Blurring is often used as a first step before operations such as thresholding, edge detection, or before finding the contours of an image. Applying a low-pass blurring filter smooths the edges and removes noise from an image. A Gaussian blur is a low-pass frequency filter and thus blurring is tolerant of changes in the high frequency range. The down sampling of an image allows sharper blurred areas to exist. A complicated, but largely image-independent relationship exists between corresponding blur levels in images at different resolutions, which can be clarified by a blur magnitude model studied as a function corresponding to spatial frequency.

The human visual system has an uneven response to various frequency components. It is vulnerable to many frequency elements and treats them unequally. CNN models are more vulnerable to low frequency components than higher frequency components and are thus close to the human visual system. Spectral bias of the CNN models can't allow the CNN models to keep unimportant frequency channels during inference without the loss of accuracy during inference. Discrete cosine transform (DCT) represents a finite data sequence in terms of a sum of cosine functions that oscillates at various frequencies. DCT is used in a wide range of applications such as digital audio, speech coding, digital image, digital video and digital radio. DCT is a Fourier related transform but uses only real numbers [23].

The Additive White Gaussian Noise (AWGN) is invariant with respect to its signal space origin rotation. It is circularly symmetrical in any direction of the signal space, applied to any noise inherent to the information system with a constant magnitude around the frequency band, and it has a regular distribution with an average time domain value of zero. Rayleigh fading is however a reasonable pattern when several objects disperse the incoming signal in the atmosphere before it reaches the receiver. If the scattering is high enough, the central limit theorem states that the responses of the channel impulses are modelled as a Gaussian process irrespective of the distribution of each component. If the scatter has no dominant element, then the mean would be zero and the phase will be evenly dispersed between zero and 2π radians. The Rician fading is also used to characterize fading in environments where the transmitter and the receiver have a line of sight (LOS) or clear specular direction. The factor Rice is measured as the ratio of LOS or specular power to distributed power and follows closely a non-central chi-square distribution with two degrees of freedom.

A large number of AMC approaches suggested in the literature can be divided into two distinct categories: feature based (FB) [24]–[27] and likelihood based (LB) methods [28]–[32]. The LB processes are known to obtain optimal outcomes by considering AMC as a problem of hypothesis testing to have an optimum solution in the Bayesian sense, thus decreasing probability of misclassification, however under restricted computing resources the time per decision is not feasible. On the other hand, FB classifier is computationally proficient and can achieve nearly optimum performance when properly designed.

FB algorithms including cyclic statistics, wavelet transforms and cumulant-based methods extract features in order to identify modulation schemes. FB methods are favoured as suboptimal classifiers in practice. The feature extraction is performed in FB approaches during the pre-processing stage preceded by the classification stage. Traditional FB methods focus largely on professional expertise, enabling them to do well in some contexts, but suffering from higher computational complexity and limited generalization issues.

Several studies in the literature have been proposed that are aimed at designing specialized features for the recognition of breast cancer [39]–[41], future robust networks for 6G [42], classification of digital modulated images [43], in-band spectral variation and deviation from unit circle while utilizing Nesterov accelerated adaptive moment estimation technique and the classifier based on the artificial neural network, which carries out AMC across a wide range of signal to noise ratio (SNR), multi-gene genetic programming (MGP) based on features that transforms cumulative sample estimates into highly discriminatory, iterative features before maximal MGP features are achieved and to determine the final classification performance of the MGP features, while taking advantage of the structural risk minimization principle [4]. Other works include the integrating of a new Nelder-Mead channel estimator into the radio frequency distinct features fingerprinting technique, as well as utilizing a multipath system with degraded SNR [8], a blind modulation classification algorithm using discrete Fourier transform to check the existence of a synchronization defect, that is a timing-phase offset and frequency without previous knowledge on the signal and channel parameters for the QPSK, BPSK, Minimal Shift Keying (MSK), 16-QAM and Offset-Quadrature Phase Shift Keying (OQPSK) schemes [10], utilizing higher order cumulants and signal spectral features to train K-Nearest Neighbour (KNN) classifiers and Support Vector Machine (SVM) [15], and a block coordinate descent dictionary learning algorithm for multiclass classification between QPSK, 8-PSK, 8-QAM, 16-QAM, Quadrature Amplitude Shift Keying (QASK), and 8-ASK modulation schemes [16].

In addition to these traditional methods for AMC tasks, DL has risen as an emerging field for AMC tasks. In the literature, different methods have been employed for the use of auto-encoding neural networks for the extraction of features and classify them using millimetre waves over fibre optic communication systems [7], an auto encoder focused on DL to extract spectrum representative features to accurately classify waveforms as idle, jammer, or Wireless Fidelity [5], a deep neural network made of FC layers for multiple input and multiple output (MIMO) OFDM system for QPSK, BPSK, 64-QAM and 16-QAM modulation schemes [10], an improved CNN based AMC network to classify among 8-PSK, Double Side Band Amplitude Modulation (AM-DSB), BPSK, Wideband Frequency Modulation (WBFM), Gaussian FSK, 16-QAM, 64-QAM, Continuous Phase Frequency Shift Keying (CPFSK) and 4-Pulse Amplitude Modulation (4-PAM) schemes in beyond fifth generation communication systems [14]. In the OFDM system, CNN based AMC approach is used to consider phase

offset for the classification of 8-PSK, 16-QAM, QPSK and BPSK modulation schemes [12], an ensemble deep neural network employing Euclidean distance based rectified linear unit (ReLU) activation functions for the classification of 16-QAM, 64-QAM, QPSK and BPSK modulation schemes [13], a DL based radio frequency signal classifier for the classification of BPSK, QPSK, Continuous Phase Modulation (CPM), Gaussian FSK, 16-QAM and Gaussian MSK modulation schemes [14], a combination of two CNNs for the classification of CR based signals representing BPSK, QPSK, 8-PSK, Gaussian FSK, CPFSK, 4-PAM, 16-QAM and 64-QAM modulation schemes [6], as well as a feed-forward deep neural network based multi-class classifier, which is made using FC layers, for adaptive spatial modulation MIMO systems [33].

Different from other works, in this paper, we have considered the problem of multiclass (eight classes) classification of modulated signals, which are, BPSK, 64-QAM, 16-QAM and QPSK signals affected by Rician, AWGN and Rayleigh fading channels employing a 3D-CNN architecture in both frequency and spatial domains with three data augmentation techniques such as random zoomed in/out, random weak Gaussian blurring and random shift augmentation techniques. We employed augmentation only in the spatial domain during training of the DL architecture. The remaining paper is set accordingly.

Mathematical background is given in Section 2, datasets are explained in Section 3, 3D-CNN architectures are explained in Section 4, Section 5 explains the experiments, and results and discussion is given in Section 6. Finally, conclusions are given in Section 7.

2 | MATHEMATICAL BACKGROUND

In this part, we will give a mathematical background for fading channels which are used for simulations in this study such as AWGN, Rayleigh and Rician fading channels. Furthermore, the convolution operation will be described followed by modulation schemes which are used in the experiments such as BPSK, QPSK, 64-QAM and 16-QAM modulation schemes as well as frequency domain (DCT) operation.

The AWGN channel is defined at discrete time event index i by a series of outputs Y_i. Y_i is the sum of the IN_i and noise NO_i inputs, where NO_i is distributed independently and identically and taken from a normal zero-mean distribution of variance N. Furthermore, it is assumed that the NO_i is not correlated with the IN_i. Mathematically,

$$NO_i \sim \mathcal{N}(0, N), \tag{1}$$

$$Y_i = IN_i + NO_i. \tag{2}$$

The response to the channel impulses is better modelled as a Gaussian mechanism no matter how each component is distributed. If the scatter does not have a dominant component, the scatter will be spread uniformly between 0 and 2π

radians with zero mean and phase components. The channel response envelope is thus distributed as a Rayleigh function. Mathematically,

$$\mathcal{P}_R \ (r) = \frac{2r}{\Omega} \ e^{-r^2/\Omega}, \ r \geq 0, \qquad (3)$$

where $\Omega \ = \ E(R^2)$.

There are two parameters to characterize a Rician fading channel. The first one, K is the ratio of the power in the direct direction to the power in the other dispersed directions.

$$K \ = \ \frac{v^2}{2\sigma^2}. \qquad (4)$$

The second one, W is the total power from both paths and acts as a scaling factor to the distribution

$$\Omega \ = \ v^2 + 2\sigma^2. \qquad (5)$$

The 3D kernel is convolved into the cube to accomplish 3D convolution, made up of several contiguous frames that are stacked together. By this design, the feature maps in the convolutional layer are connected to the previous layer by several contiguous frames together. The value of the jth feature map in the ith layer at point (x, y, z) is formally defined with:

$$v_{ij}^{xyz} \ = \ f\left(b_{ij} + \sum_m \sum_{p=0}^{P_i-1} \sum_{q=0}^{Q_i-1} \sum_{r=0}^{R_i-1} \mathfrak{w}_{ijm}^{pqr} v_{m(i-1)}^{(x+p)(y+q)(z+r)}\right), \qquad (6)$$

where R_i is a 3D kernel size in the temporal dimension, $\mathfrak{w}_{ijm}^{pqr}$ is the (p, q, r)th kernel value that is connected to the previous layer of the mth feature map. Since multiple kernels are convolved with the input layer, the output contains a stack of activation maps when several kernels are convolved with the input layer. BPSK, functionally equivalent to 2-QAM modulation, uses two phases separated by $180^°$ and $0°$.

Mathematically,

$$s_n \ (t) = \sqrt{\frac{2E_b}{T_b}} \ cos\left(2\pi ft + \pi \ (1-n)\right), n \ = \ 0, 1. \qquad (7)$$

QPSK uses four points on the constellation map equally distributed around a circle functionally equivalent to 4-QAM modulation, which are separated by $7\pi/4, 5\pi/4, 3\pi/4$ and $\pi/4$. Mathematically,

$$s_n \ (t) = \sqrt{\frac{2E_s}{T_s}} \ cos\left(2\pi f_c t + (2n-1)\pi/4\right), n \ = \ 0, 1, 2, 3. \qquad (8)$$

One carrier lags the other by $90^°$ in a QAM signal and is generally defined as the in-phase component I(t) for its amplitude information. The quadrature component Q (t) is the other modulating function for its phase information. Mathematically,

$$s_n \ (t) \stackrel{\Delta}{=} \ \sin\left(2\pi f_c t\right) I \ (t) + \cos\left(2\pi f_c t + \frac{\pi}{2}\right) Q \ (t). \qquad (9)$$

The finite sequence of data points is expressed by DCT in terms of the number of cosine functions that oscillate at various frequencies. Mathematically,

$$X_{k_1, k_2, k_3} = \sum_{n1=0}^{N_1-1} \sum_{n2=0}^{N_2-1} \sum_{n3=0}^{N_3-1} x_{n1}, x_{n2}, x_{n3} \ cos\left[\frac{\pi}{N_1}\left(\left(n1 + \frac{1}{2}\right)\right) k_1\right]$$
$$cos\left[\frac{\pi}{N_2}\left(\left(n2 + \frac{1}{2}\right)\right) k_2\right] cos\left[\frac{\pi}{N_3}\left(\left(n3 + \frac{1}{2}\right)\right) k_3\right].$$
$$\forall \ k_i = 0, 1, 2, \ldots, N_{i-1} \qquad (10)$$

3 | DATASET DESCRIPTION

We used 16-QAM, 64-QAM, QPSK and BPSK modulation schemes to modulate random 2D lung X-ray images of COVID-19 and healthy subjects downloaded from the internet, added AWGN, Rayleigh and Rician noise to them; and finally formed 3D volumes by stacking multiple contiguous 2D frames together. There are eight classes present in the dataset which are BPSK-AWGN-COVID-19, BPSK-RICIAN-HEALTHY, 16-QAM-AWGN-COVID-19, 16-QAM-RAYLEIGH-HEALTHY, 64-QAM-RAYLEIGH-COVID-19, 64-QAM-RICIAN-HEALTHY, QPSK-AWGN-COVID-19 and QPSK-RICIAN-HEALTHY. A sample digital modulated image of these classes is shown in Figure 1.

4 | DESCRIPTION OF THE 3D-CNN ARCHITECTURES

We used two 3D-CNN architectures for the experiments which are shown in Figures 2 and 3. The only difference between these architectures is the number of filters. We used more filters for the experiments with more data in the training set such as those involving augmentation methods. As we can see in Figures 2 and 3, there is an input layer with a size of 297 × 167 × 10 with zero-centre normalization to obtain data dimensions (channels) of approximately the same scale through division of each dimension (channel) by its standard deviation once it has been zero-centred. This is done by subtracting the mean from each of these dimensions (channels) so that the data cloud is centred at the origin. After that, feature maps are created using a 3D convolutional layer by moving a filter of size 3×3×3 with number of feature maps set to 12 or 11 depending on the number of samples in the training set. Here, we set the values of bias and weight L2 factors to 0.00005 as smaller but non-zero weights generate simpler model that is able to learn complex data patterns and thus helps avoid overfitting by mitigating noise in the samples. After that, a batch normalization layer [34] dynamically normalizes the inputs on a per mini-batch basis which has shown to
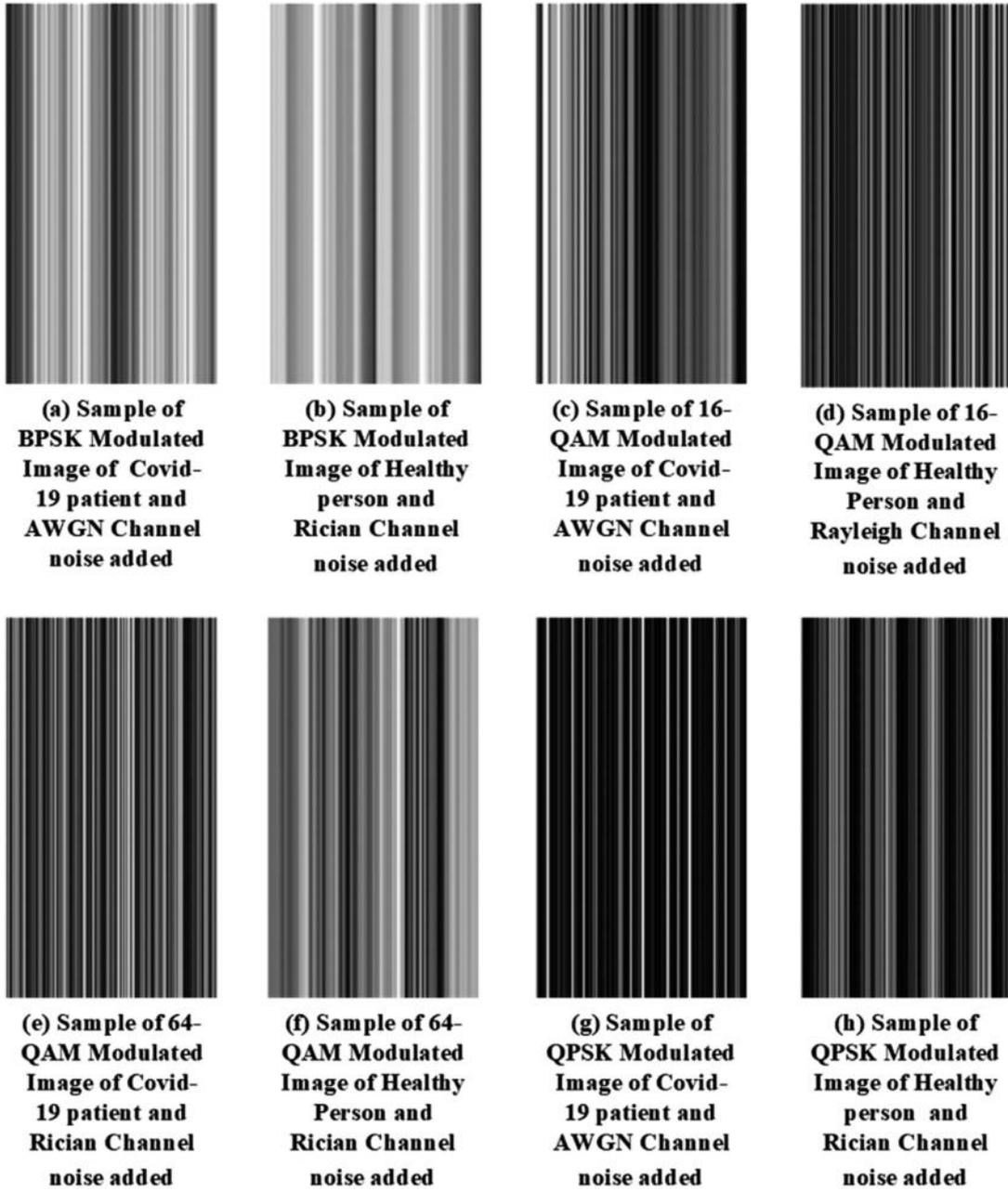
(a) Sample of BPSK Modulated Image of Covid-19 patient and AWGN Channel noise added

(b) Sample of BPSK Modulated Image of Healthy person and Rician Channel noise added

(c) Sample of 16–QAM Modulated Image of Covid-19 patient and AWGN Channel noise added

(d) Sample of 16–QAM Modulated Image of Healthy Person and Rayleigh Channel noise added

(e) Sample of 64–QAM Modulated Image of Covid-19 patient and Rician Channel noise added

(f) Sample of 64–QAM Modulated Image of Healthy Person and Rician Channel noise added

(g) Sample of QPSK Modulated Image of Covid-19 patient and AWGN Channel noise added

(h) Sample of QPSK Modulated Image of Healthy person and Rician Channel noise added

FIGURE 1   A sample digital modulated image of all classes used for the multiclass classification task

improve the training time while avoiding overfitting. After that, an exponential linear unit (ELU) layer [35] is added to speed up learning by pushing mean activations closer to zero. Mathematically, it could be described as:

$$ELU: \begin{cases} x, \ x \geq 0 \\ \alpha \left(e^x - 1 \right), \ x < 0 \end{cases} \qquad (11)$$

The pooling process progressively decreases the spatial resolution of hidden representations by aggregating information such that greater receptive fields in the higher layers (in the input) are sensitive to each hidden node. The pooling operator consists of a fixed window that slides over all input regions in

compliance with its strides to measure single output at each position crossed by a fixed-shaped window. Max pooling operators compute the highest factor value in the pooling window. FC layer (also known as dense layer or the inner product layer) has full links to all activation maps of the previous layer. The input is simply multiplied by a weight matrix and a bias offset is added. The FC layer is somehow similar to the convolutional layer. Although the convolutional layer is connected to a local region of the input, all inputs are connected to the FC layer. One can be converted into the other easily. The dropout operation [36] helps in mitigating the overfitting phenomenon. It works by injecting noise into each layer of the network during training zeroing out some fraction of the nodes in the individual
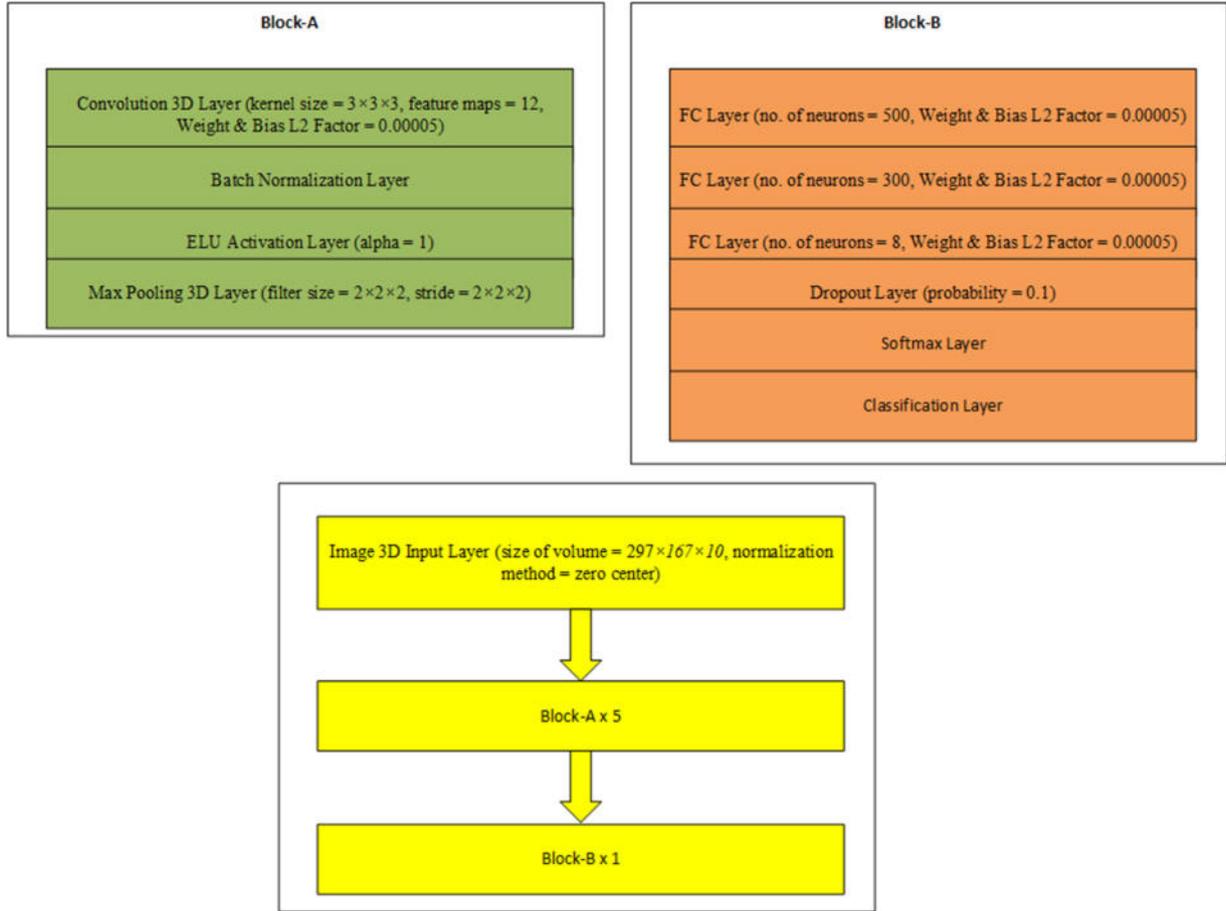
FIGURE 2 3D-CNN architecture with more number of samples in the training set for the experiments

layer before computing the next layer in forward propagation. Finally, the softmax function helps in interpreting the outputs of the 3D-CNN architectures as probabilities by optimizing the model parameters to produce probabilities that maximize the likelihood of the observed data. The classification layer places the outputs into one of the eight classes.

In our proposed architecture, we have added three FC or dense layers with the final dense layer having 8 neurons to place the input into one of the 8 categories. The first two dense layers have 500 and 300 neurons each to capture the feature activations as they are getting passed from the convolutional layers. After that, we added a dropout layer right before the softmax layer with a probability of 0.1 to mitigate the disharmony between batch normalization and dropout techniques caused by the variance shift phenomenon [38]. We chose the architecture to keep the number of parameters to a minimum without sacrificing the performance. Figure 4 displays another view of the proposed 3D-CNN architecture

5 | EXPERIMENTS

We conducted a number of experiments using 10-fold and 5-fold cross-validation (CV) approach to choose the optimum set of hyperparameters. The experiments were carried out in the spatial domain with and without augmentation techniques as well as without augmentation in the frequency (DCT) domain. We performed the following experiments:

1. Multiclass (8-classes) classification of modulated signals augmented by the combined random shift, random weak Gaussian blurred and random zoomed in/out augmentation techniques in the spatial domain using 5-fold CV approach.
2. Multiclass (8-classes) classification of modulated signals augmented by the combined random weak Gaussian blurred, random shift and random zoomed in/out augmentation schemes in the spatial domain using 10-fold CV approach.
3. Multiclass (8-classes) classification of modulated signals augmented only by the random weak Gaussian blurred augmentation scheme in the spatial domain using 5-fold CV approach.
4. Multiclass (8-classes) classification of modulated signals augmented only by the random weak Gaussian blurred augmentation scheme in the spatial domain using 10-fold CV approach.
5. Multiclass (8-classes) classification of modulated signals augmented only by the random shifted augmentation scheme in the spatial domain using 5-fold CV approach.
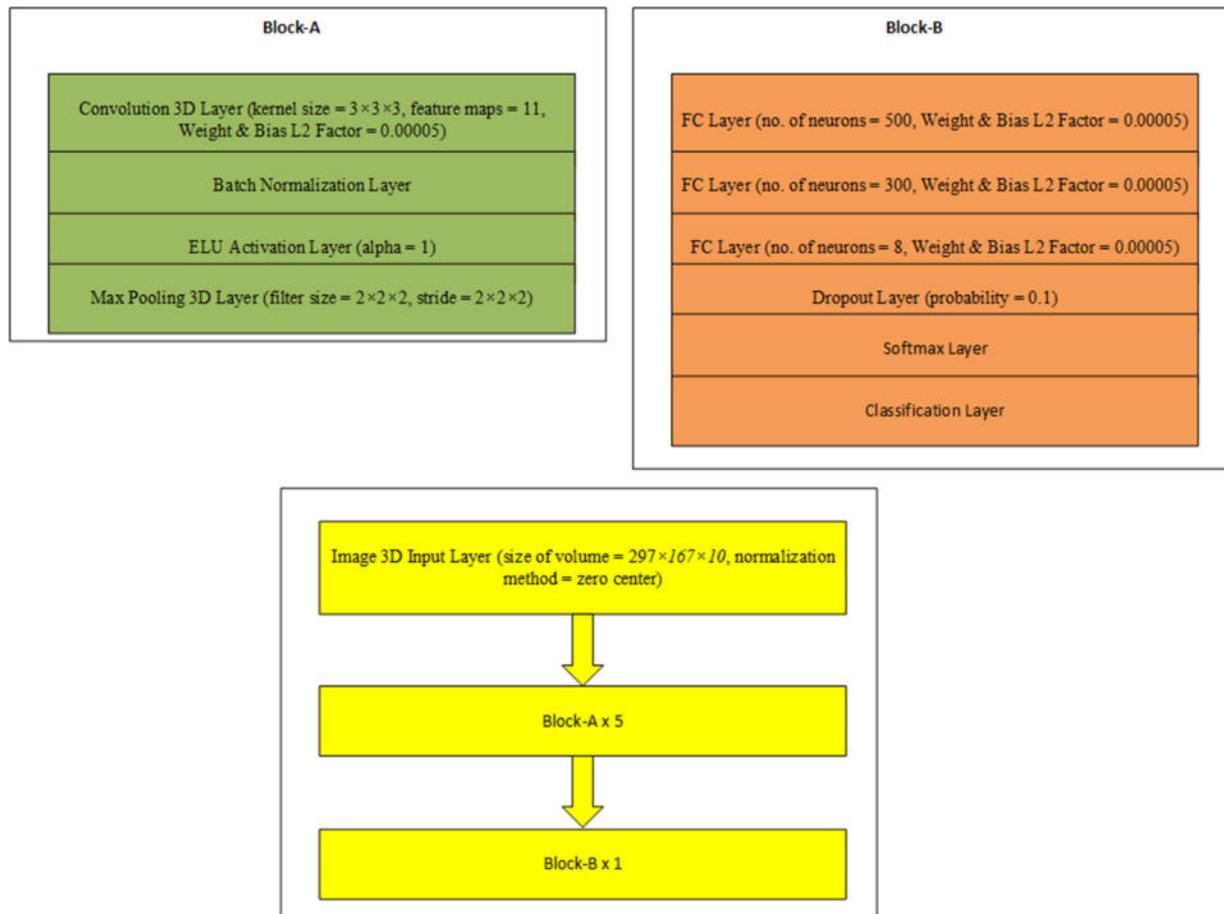
FIGURE 3 3D-CNN architecture with less number of samples in the training set for the experiments
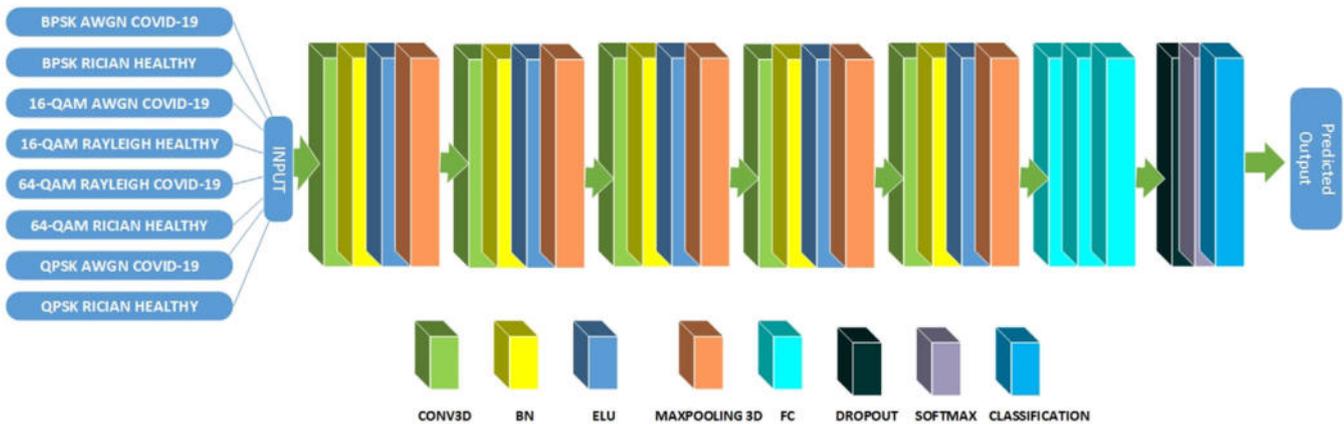

FIGURE 4 Proposed 3D-CNN architecture with a different view

6. Multiclass (8-classes) classification of modulated signals augmented only by the random shifted augmentation scheme in the spatial domain using 10-fold CV approach.

7. Multiclass (8-classes) classification of modulated signals augmented only by the random zoomed in/out augmentation scheme in the spatial domain using 5-fold CV approach.

8. Multiclass (8-classes) classification of modulated signals augmented only by the random zoomed in/out augmentation scheme in the spatial domain using 10-fold CV approach.

9. Multiclass (8-classes) classification of modulated signals without augmentation in the spatial domain using 5-fold CV approach.

10. Multiclass (8-classes) classification of modulated signals without augmentation in the spatial domain using 10-fold CV approach.
11. Multiclass (8-classes) classification of modulated signals without augmentation in the frequency domain using 5-fold CV approach.
12. Multiclass (8-classes) classification of modulated signals without augmentation in the frequency domain using 10-fold CV approach.

For all the experiments that involve 10-fold CV approach as well as 5-fold CV with augmentation(s), we choose a mini-batch of size 2, an initial learning rate of 0.001, maximum number of epochs were set to 30, training set was shuffled after every epoch to mitigate overfitting, piecewise learning rate was selected that lowers the learning rate after every 5 epochs by multiplying with a factor of 0.1, while Adam [37] was used as the optimizer and categorical cross-entropy as the loss function. Feature maps in the convolutional layer of the 3D-CNN architecture were set to 12. The training was conducted on a single NVIDIA Titan RTX graphical processing unit.

For all the experiments that involve 5-fold CV approach without augmentation(s) in the frequency and spatial domains, we choose a mini-batch of size 2, an initial learning rate of 0.001, maximum number of epochs were set to 30, training set was shuffled after every epoch to mitigate overfitting, piecewise learning rate was selected that reduces the learning rate after every 5 epochs by multiplying with a factor of 0.1, while Adam was used as the optimizer and categorical cross-entropy as the loss function. Feature maps in the convolutional layer of the 3D-CNN Architecture were set to 11.

For the experiments involving 5-fold CV approach without augmentation in the frequency and spatial domains, the validation set has 96 samples per class while the training set has 384 samples per class. For the experiments involving 5-fold CV approach with single augmentation technique in the spatial domain, the validation set has 96 samples per class while the training set has 768 samples per class. For the experiments involving 5-fold CV approach with combined augmentation techniques in the spatial domain, the validation set has 96 samples per class while the training set has 1536 samples per class.

For the experiments involving 10-fold CV approach without augmentation in the frequency and spatial domains, the validation set has 48 samples per class while the training set has 432 samples per class. For the experiments involving 10-fold CV approach with single augmentation technique in the spatial domain, the validation set has 48 samples per class while there are 864 samples per class in the training set. For the experiments involving 10-fold CV approach with combined augmentation techniques in the spatial domain, the validation set has 48 samples per class while the training set has 1728 samples of each class.

Total time for the experiments was approximately 28,418 min, 474 h, or 20 days while time for a single experiment was approximately 316 min or 5.27 h. Approximate time for the experiments that used 5-fold CV approach was around

TABLE 1 Average CEN values for all methods

Serial #	Methods	Average CEN
1	10-fold CV without augmentation in the spatial domain	0.0605
2	5-fold CV without augmentation in the spatial domain	0.0732
3	10-fold CV with random weak Gaussian blurred augmentation in the spatial domain	0.082
4	5-fold CV with random weak Gaussian blurred augmentation in the spatial domain	0.0891
5	10-fold CV with random zoomed in/out augmentation in the spatial domain	0.092
6	5-fold CV with combined augmentations in the spatial domain	0.1002
7	5-fold CV with random zoomed in/out augmentation in the spatial domain	0.1044
8	5-fold CV without augmentation in the frequency domain	0.1135
9	10-fold CV with random shifted augmentation in the spatial domain	0.1144
10	10-fold CV with combined augmentations in the spatial domain	0.115
11	5-fold CV with random shift augmentation in the spatial domain	0.1161
12	10-fold CV without augmentation in the frequency domain	0.1296

8942 min or 149 h while approximate time for the experiments that used 10-fold CV approach was around 19476 min or 325 h.

6 | RESULTS AND DISCUSSION

Tables 1–6 represent the results of the experiments performed for the multiclass classification of signals in the presence of fading. We used RCI, CEN, IBA, GM and MCC as our performance metrics.

The RCI metric is an entropy-based measure that quantifies how much the uncertainty of the decision problem is reduced by the classifier, relative to classifying by simply using the prior probabilities of each class. It corrects for differences in prior probabilities of the diagnostic categories, as well as the number of categories. Values of this measure lie in the interval between 0 and 1, where values close to 1 represent better classification.

CEN is an information theoretic measure based upon the idea of entropy for measuring classifier performances. It evaluates the confusion level of the class distribution of misclassified samples. CEN measures generated entropy from misclassified cases considering not only how the cases of each class have been misclassified into other classes, but also how the cases of the other classes have been misclassified as belonging to this class, as well as entropy inside well-classified cases. Small values of CEN represent less information loss and better classification.

TABLE 2 Average IBA values for all methods

Serial #	Methods	Average IBA
1	10-fold CV without augmentation in the spatial domain	0.9406
2	5-fold CV without augmentation in the spatial domain	0.9295
3	10-fold CV with random weak Gaussian blurred augmentation in the spatial domain	0.9215
4	5-fold CV with random weak Gaussian blurred augmentation in the spatial domain	0.9141
5	10-fold CV with random zoomed in/out augmentation in the spatial domain	0.9106
6	5-fold CV with combined augmentations in the spatial domain	0.9018
7	5-fold CV with random zoomed in/out augmentation in the spatial domain	0.8993
8	5-fold CV without augmentation in the frequency domain	0.892
9	10-fold CV with random shifted augmentation in the spatial domain	0.891
10	10-fold CV with combined augmentations in the spatial domain	0.889
11	5-fold CV with random shift augmentation in the spatial domain	0.8881
12	10-fold CV without augmentation in the frequency domain	0.8716

TABLE 4 Average MCC values for all methods

Serial #	Methods	Average MCC
1	10-fold CV without augmentation in the spatial domain	0.9655
2	5-fold CV without augmentation in the spatial domain	0.9589
3	10-fold CV with random weak Gaussian blurred augmentation in the spatial domain	0.9542
4	5-fold CV with random weak Gaussian blurred augmentation in the spatial domain	0.9498
5	10-fold CV with random zoomed in/out augmentation in the spatial domain	0.9478
6	5-fold CV with combined augmentations in the spatial domain	0.9424
7	5-fold CV with random zoomed in/out augmentation in the spatial domain	0.9409
8	5-fold CV without augmentation in the frequency domain	0.9364
9	10-fold CV with random shifted augmentation in the spatial domain	0.9357
10	10-fold CV with combined augmentations in the spatial domain	0.9346
11	5-fold CV with random shift augmentation in the spatial domain	0.9339
12	10-fold CV without augmentation in the frequency domain	0.924

TABLE 3 Average GM values for all methods

Serial #	Methods	Average GM
1	10-fold CV without augmentation in the spatial domain	0.9826
2	5-fold CV without augmentation in the spatial domain	0.9793
3	10-fold CV with random weak Gaussian blurred augmentation in the spatial domain	0.9769
4	5-fold CV with random weak Gaussian blurred augmentation in the spatial domain	0.9746
5	10-fold CV with random zoomed in/out augmentation in the spatial domain	0.9736
6	5-fold CV with combined augmentations in the spatial domain	0.9708
7	5-fold CV with random zoomed in/out augmentation in the spatial domain	0.9701
8	5-fold CV without augmentation in the frequency domain	0.9678
9	10-fold CV with random shifted augmentation in the spatial domain	0.9675
10	10-fold CV with combined augmentations in the spatial domain	0.9669
11	5-fold CV with random shift augmentation in the spatial domain	0.9666
12	10-fold CV without augmentation in the frequency domain	0.9614

TABLE 5 RCI values for all methods

Serial #	Methods	RCI
1	10-fold CV without augmentation in the spatial domain	0.9189
2	5-fold CV without augmentation in the spatial domain	0.902
3	10-fold CV with random weak Gaussian blurred augmentation in the spatial domain	0.8902
4	5-fold CV with random weak Gaussian blurred augmentation in the spatial domain	0.8805
5	10-fold CV with random zoomed in/out augmentation in the spatial domain	0.8765
6	5-fold CV with combined augmentations in the spatial domain	0.8656
7	5-fold CV with random zoomed in/out augmentation in the spatial domain	0.8602
8	5-fold CV without augmentation in the frequency domain	0.8483
9	10-fold CV with random shifted augmentation in the spatial domain	0.8471
10	10-fold CV with combined augmentations in the spatial domain	0.8459
11	5-fold CV with random shift augmentation in the spatial domain	0.8448
12	10-fold CV without augmentation in the frequency domain	0.8263

TABLE 6 Individual and overall ranking of the methods

Serial #	Methods	Individual and overall ranking
1	10-fold CV without augmentation in the spatial domain	1
2	5-fold CV without augmentation in the spatial domain	2
3	10-fold CV with random weak Gaussian blurred augmentation in the spatial domain	3
4	5-fold CV with random weak Gaussian blurred augmentation in the spatial domain	4
5	10-fold CV with random zoomed in/out augmentation in the spatial domain	5
6	5-fold CV with combined augmentations in the spatial domain	6
7	5-fold CV with random zoomed in/out augmentation in the spatial domain	7
8	5-fold CV without augmentation in the frequency domain	8
9	10-fold CV with random shifted augmentation in the spatial domain	9
10	10-fold CV with combined augmentations in the spatial domain	10
11	5-fold CV with random shift augmentation in the spatial domain	11
12	10-fold CV without augmentation in the frequency domain	12

IBA provides information on the dominancy of a class with the highest degree of individual accuracy. The method combines an unbiased index of its overall accuracy and a measure of how dominant the class with the highest individual accuracy rate is. Higher values indicate better classification.

GM focuses only on the recall of each class which is aggregated multiplicatively. It is defined as the product of sensitivity and specificity under a square root. Higher values of this measure indicate better classification performances.

MCC is a coefficient of correlation between the classifications that are observed and predicted. Its values lie in the interval between -1 and +1, where +1 indicates perfect classification.

As given in Tables 1–6, we considered average values of classwise statistics for CEN, IBA, GM and MCC metrics for the eight classes, the values of the RCI metric as well as individual and overall ranking of the methods. The average values are calculated by taking the sum of values in the eight classes and dividing that sum by 8.

As a visual aid, Figures 5–10 display the values of each of these performance metrics as well as the overall ranking of the methods based on these metrics.

The procedure for forming the ranking system will be explained next. To form the RCI based ranking of methods, we took the maximum values and sorted all the methods based on these values. To form the CEN based ranking of methods, we took the minimum of average values of this metric and sorted all of our methods based on these values with the best performing
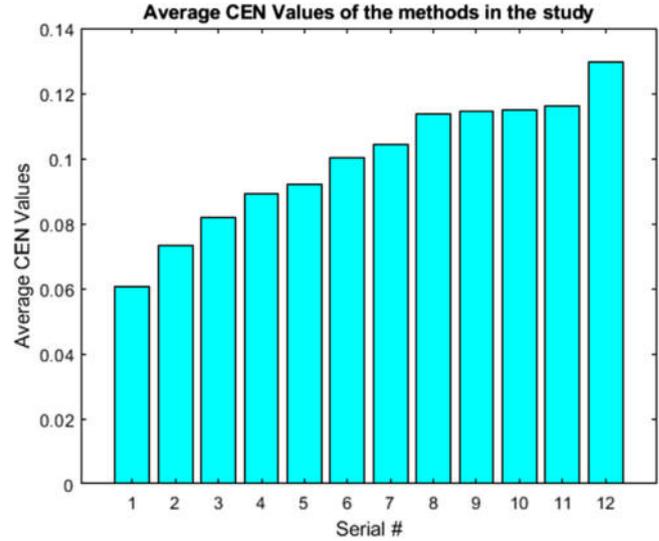

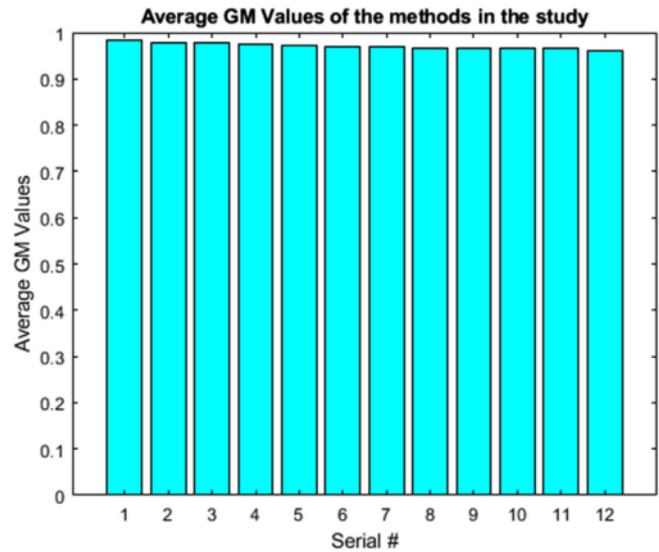**FIGURE 5** Average CEN values of the methods in the study

FIGURE 6 Average GM values of the methods in the study

method been given the minimum average CEN value. To form the IBA, GM and MCC based ranking of the methods, we took the maximum of average values of these metrics and sorted all of our methods based on these values with the best performing method been given the maximum average IBA, GM or MCC value. Finally, we formed the overall ranking of the methods based on the individual RCI, CEN, IBA, GM and MCC rankings. In this case, our overall and all individual metrics based rankings are exactly the same which shows strong correlation between the individual metrics as given in Table 6. As a visual aid, Figure 10 displays the ranking of all the methods considered in this study as given in Table 6.

As can be seen in Table 6, the best performing method turns out to be 10-fold CV without augmentation in the spatial domain while the worst performing method happens to be 10-fold CV without augmentation in the frequency domain. We

TABLE 7 Comparison with the state-of-the-art

Author(s)	Data	Method(s)	Accuracy	Classification task
K Bu et al. [44]	RadioML2016.10a	Adversarial transfer learning architecture	82.4%	Multiclass classification
Kim S.H et al. [45]	RadioML2018.01A	CNN architecture	91.48%	Multiclass classification
P. Ghasemzadeh et al. [46]	RadioML2018.01A	DBN based model	85%	Multiclass classification
Ma H et al. [47]	RML2016.10a	Cross model deep learning scheme	90%	Multiclass classification
Y Lin et al. [48]	RML2016.10a	Filter-level pruning technique	80%	Multiclass classification
Hu S et al. [49]	RML2016.10a	DNN based classifiers	80%	Multiclass classification
Lin R et al. [50]	RadioML2016.10b	CNN + bi-directional GRU	89.25%	Multiclass classification
Li L et al. [51]	RML2016.04C	Capsule network	90%	Multiclass classification
Zhao Y et al. [52]	Custom	DL based intelligent edge surveillance technique	89%	Multiclass classification
Our approach	**Custom**	**3D-CNN**	**96.97%**	**Multiclass classification**

DBN: Deep Belief Network, GRU: Gated Recurrent Unit.
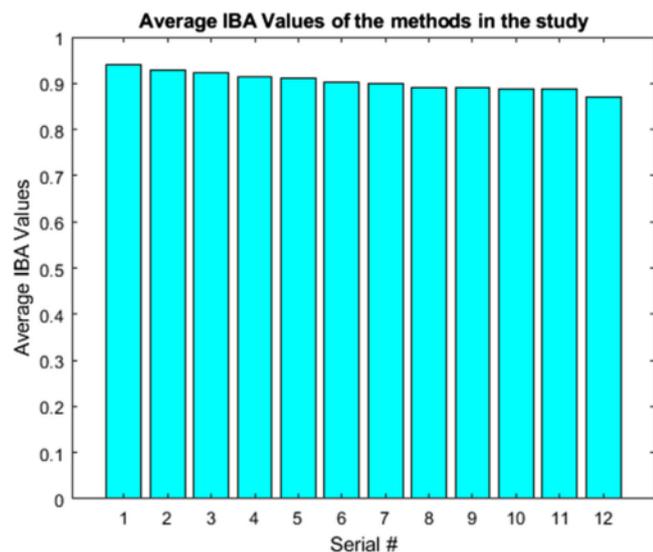
FIGURE 7 Average IBA values of the methods in the study
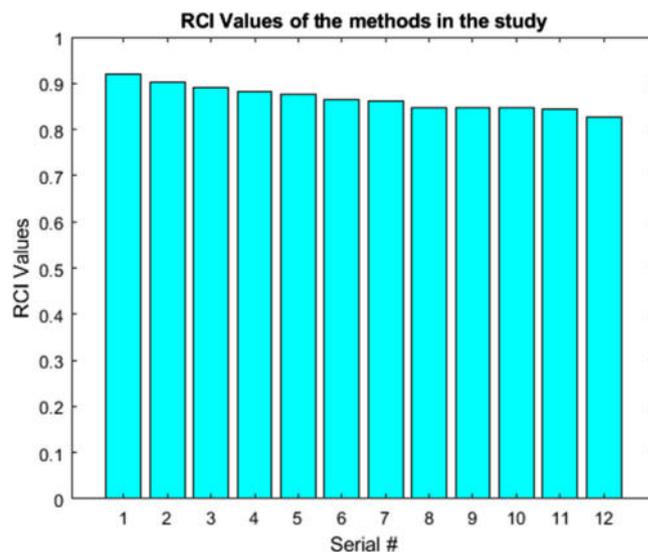
FIGURE 9 RCI values of the methods in the study
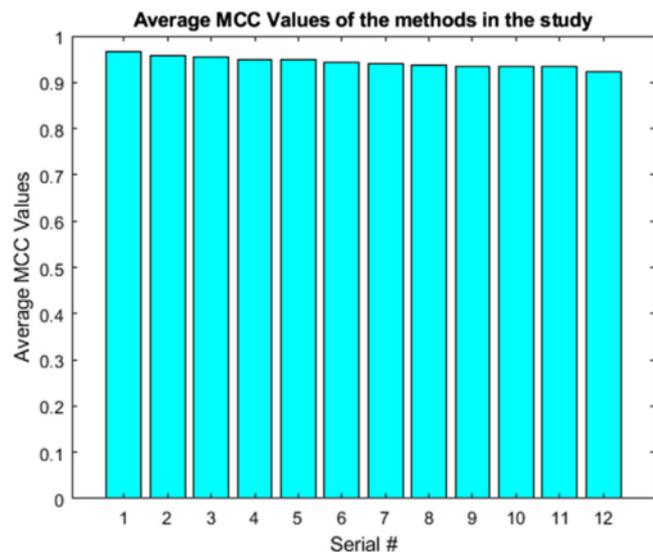
FIGURE 8 Average MCC values of the methods in the study

can also observe that spatial domain methods have an edge over frequency domain methods. One reason for the better performance of spatial domain methods could be that they work on data with larger variation in intensity values of inputs which allows them to capture intrinsic information of a volume better. Furthermore, combining the augmentation methods in the training set resulted in performance degradation in comparison with the methods that used single augmentation methods. In addition to that, we can observe that random weak Gaussian blurring augmentation method is the best performing augmentation method in comparison to random zoomed in/out augmentation and random shifted augmentation methods. Although the best performing method uses more data in the training set, methods that used more data during training may not necessarily be the best. The performance of methods that used random shifted augmentation during training was found to be inferior to the other methods. One reason for this phenomenon is that small translations or rescalings of the input
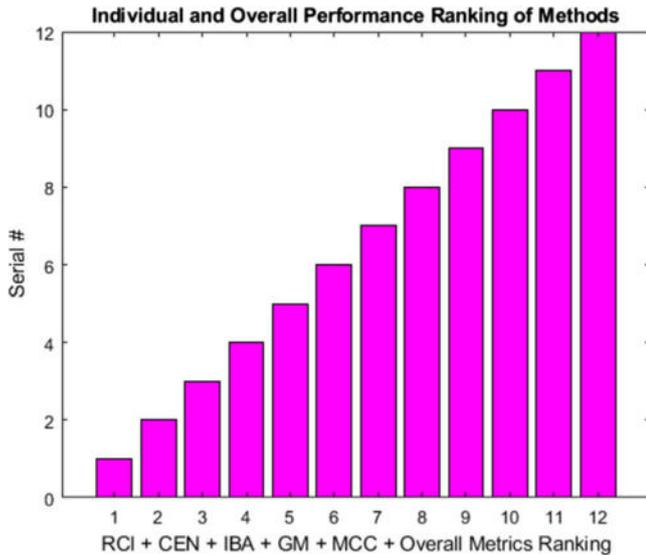
FIGURE 10 Individual and overall ranking of methods based on performance metrics

image can drastically change the prediction of a CNN model as CNNs are not invariant to such transformations due to the ignorance of classical sampling theorem. The better performance of random weak Gaussian blurring augmentation method could be explained by the blurring operation that smooths the output of the non-linearity which helps in preventing high frequency activations, and also helps in isolating the aliasing phenomenon.

AMC is an important task with wide range of civil and military applications and has a number of uses in different scenarios. Our work lies at the intersection of DL and AMC task. We deployed state-of-the-art 3D CNN architectures for the multiclass classification of modulation schemes in the presence of noise in both spatial and frequency domains with data augmentation procedures to carry out this task.

The proposed architecture is optimally designed to carry out multiclass classification of modulation schemes. The main gist behind the design is to attain maximum performance from the proposed design. Number of neurons in the FC layers, feature maps in the convolutional layers and other hyperparameters have been chosen to avoid overfitting, long training time and other problems. It can be seen that 3D CNN architectures are a useful tool for this task in achieving better performances.

Comparison of our work with the other studies reported in the literature is given in Table 7.

7 | CONCLUSION

In this work, we compared and contrasted the performance of several DL architectures for the multiclass (8-classes) classification of modulated signals in the presence of noise in both spatial and frequency (DCT) domains. The best performing model has been found to be 10-fold CV without using augmentation in the spatial domain while the worst performing model has been found to be 10-fold CV without augmentation in the frequency

domain. Furthermore, we note that spatial domain methods performed better than their frequency domain counterparts. This study can be extended further by considering other modulation schemes such as frequency modulation methods as well as noise models such as Nakagami model and other DL architectures such as graph convolutional networks.

ORCID
Inam Ullah https://orcid.org/0000-0002-5879-569X

REFERENCES
1. Jagannath, J., et al.: Artificial neural network based automatic modulation classification over a software defined radio testbed. In: 2018 IEEE International Conference on Communications (ICC), pp. 1–6. IEEE, Piscataway (2018)
2. Soltani, S., et al.: Real-time and embedded deep learning on FPGA for RF signal classification. In: 2019 IEEE Military Communications Conference (MILCOM), pp. 1–6. IEEE, Piscataway (2019)
3. Yu, S., et al.: Adaptive double-threshold cooperative spectrum sensing algorithm based on history energy detection. Wirel. Commun. Mobile Comput. 2020, 4794136 (2020)
4. Huang, S., et al.: Automatic modulation classification of overlapped sources using multi-gene genetic programming with structural risk minimization principle. IEEE Access. 6, 48827–48839 (2018)
5. Davaslioglu, K., et al.: DeepWiFi: Cognitive WiFi with deep learning. IEEE Trans. Mob. Comput. 20(2), 429–444 (2019)
6. Wang, Y., et al.: Data-driven deep learning for automatic modulation recognition in cognitive radios. IEEE Trans. Veh. Technol. 68(4), 4074–4077 (2019)
7. Alharbi, M., et al.: Automatic modulation classification: investigation for millimeter wave over fiber channels. IEEE Photonics Technol. Lett. 31(3), 1092–1095 (2019)
8. Fadul, M.K. et al.: RF-DNA fingerprint classification of ofdm signals using a rayleigh fading channel model. In: 2019 IEEE Wireless Communications and Networking Conference (WCNC), pp. 1–7. IEEE, Piscataway (2019)
9. Ullah, I., et al.: Extended Kalman filter-based localization algorithm by edge computing in wireless sensor networks. Digital Commun. Netw. 7(2), 187–195 (2020)
10. Gupta, R., et al.: Blind modulation classification for asynchronous ofdm systems over unknown signal parameters and channel statistics. IEEE Trans. Veh. Technol. 69(5), 5281–5292 (2020)
11. Ha, C.B., et al.: Machine learning model for adaptive modulation of multi-stream in MIMO-OFDM system. IEEE Access 7, 5141–5152 (2018)
12. Shi, J., et al.: Deep learning-based automatic modulation recognition method in the presence of phase offset. IEEE Access 8, 42841–42847 (2020)
13. Ha, C.B., Song, H.K.: Signal detection scheme based on adaptive ensemble deep learning model. IEEE Access 6, 21342–21349 (2018)
14. Hermawan, A.P., et al.: CNN-based automatic modulation classification for beyond 5G communications. IEEE Commun. Lett. 24(5), 1038–1041 (2020)
15. Wahla, A.H., et al.: Automatic wireless signal classification in multimedia Internet of Things: An adaptive boosting enabled approach. IEEE Access 7, 160334–160344 (2019)
16. Zhang, K., et al.: A dictionary learning based automatic modulation classification method. IEEE Access 6, 5607–5617 (2018)
17. Ahmad, I., et al.: Efficient algorithms for e-healthcare to solve multiobject fuse detection problem. J. Healthcare Eng. 2021 (2021)

18. Ullah, I., et al.: Localization and detection of targets in underwater wireless sensor using distance and angle based algorithms. IEEE Access 7, 45693–45704 (2019)

19. Azulay, A., Weiss, Y.: Why do deep convolutional networks generalize so poorly to small image transformations? J. Mach. Learning Res. 20(184), 1–25 (2019)

20. Wong, S.C., et al.: Understanding data augmentation for classification: When to warp?. In: 2016 International Conference on Digital Image Computing: Techniques and Applications (DICTA),pp. 1–6. IEEE, Piscataway (2016)

21. Lenc, K., Vedaldi, A.: Understanding image representations by measuring their equivariance and equivalence. Int. J. Comput. Vision 127(5), 456–479 (2019)

22. Sifre, L., Mallat, S.: Rotation, scaling and deformation invariant scattering for texture discrimination. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1233–1240. IEEE, Piscataway (2013)

23. Xu, K., et al.: Learning in the frequency domain. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1740–1749. IEEE, Piscataway (2020)

24. Alharbi, H., et al.: Automatic modulation classification of digital modulations in presence of HF noise. EURASIP J. Adv. Signal Process. 2012(1), 238 (2012)

25. Chang, D.C., Shih, P.K.: Cumulants-based modulation classification technique in multipath fading channels. IET Commun. 9(6), 828–835 (2015)

26. Swami, A., Sadler, B.M.: Hierarchical digital modulation classification using cumulants. IEEE Trans. Commun. 48(3), 416–429 (2000)

27. Han, L., et al.: Low complexity automatic modulation classification based on order-statistics. IEEE Trans. Wireless Commun. 16(1), 400–411 (2016)

28. Hameed, F., et al.: On the likelihood-based approach to modulation classification. IEEE Trans. Wireless Commun. 8(12), 5884–5892 (2009)

29. Wimalajeewa, T., et al.: Distributed asynchronous modulation classification based on hybrid maximum likelihood approach. In: 2015 IEEE Military Communications Conference (MILCOM), pp. 1519–1523. IEEE, Piscataway (2015)

30. Zhang, Y., et al.: Optimal decision fusion based automatic modulation classification by using wireless sensor networks in multipath fading channel. In: 2011 IEEE Global Telecommunications Conference-(GLOBECOM), pp. 1–5. IEEE, Piscataway (2011)

31. Dulek, B., et al.: Distributed maximum likelihood classification of linear modulations over nonidentical flat block-fading Gaussian channels. IEEE Trans. Wireless Commun. 14(2), 724–737 (2014)

32. Ozdemir, O., et al.: Asynchronous linear modulation classification with multiple sensors via generalized EM algorithm. IEEE Trans. Wireless Commun. 14(11), 6389–6400 (2015)

33. Yang, P., et al.: Adaptive spatial modulation MIMO based on machine learning. IEEE J. Sel. Areas Commun. 37(9), 2117–2131 (2019)

34. Ioffe, S., Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: International Conference of Machine Learning (ICML), pp. 448–456. Curran Associates, Inc., Red Hook (2015)

35. Clevert, D.A., et al.: Fast and accurate deep network learning by exponential linear units (elus). arXiv preprint, arXiv:1511.07289, 2015

36. Srivastava, N., et al.: Dropout: A simple way to prevent neural networks from overfitting. J. Mach. Learning Res. 15(1), 1929–1958 (2014)

37. Da, K., Adam, J.B.: A method for stochastic optimization. In: International Conference on Learning Representation (ICLR). Curran Associates, Inc., Red Hook (2015)

38. Xiang, L., et al.: Understanding the disharmony between dropout and batch normalization by variance shift. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2677–2685. IEEE, Piscataway (2019)

39. Yap, M.H., et al.: Breast ultrasound lesions recognition: End-to-end deep learning approaches. J. Med. Imaging 6(1), 1–7 (2019)

40. Hu, Q., et al.: A deep learning methodology for improved breast cancer diagnosis using multiparametric MRI. Sci. Rep. 10(1), 1–11 (2020)

41. Toğaçar, M., et al.: BreastNet: A novel convolutional neural network model through histopathological images for the diagnosis of breast cancer. Physica A 545, 123592 (2020)

42. Junejo, A., et al.: Future robust networks: current scenario and beyond for 6G. IMCC J. Sci. 1(1), 67–81 (2021)

43. Khan, R., et al.: Classification of digital modulated covid-19 images in the presence of channel noise using 2D convolutional neural networks. Wireless Communication and Mobile Computing 2021, 5539907 (2021). https://doi.org/10.1155/2021/5539907

44. Bu, K., et al.: Adversarial transfer learning for deep learning based automatic modulation classification. IEEE Signal Process Lett. 27, 880–884 (2020)

45. Kim, S.H., et al.: Lightweight deep learning model for automatic modulation classification in cognitive radio networks. IEEE Access. 8, 197532–197541 (2020)

46. Ghasemzadeh, P., et al.: A novel deep learning and polar transformation framework for an adaptive automatic modulation classification. IEEE Trans. Veh. Technol. 69(11), 13243–13258 (2020)

47. Ma, H., et al.: Cross model deep learning scheme for automatic modulation classification. IEEE Access. 8, 78923–78931 (2020)

48. Lin, Y., et al.: An improved neural network pruning technology for automatic modulation classification in edge devices. IEEE Trans. Veh. Technol. 69(5), 5703–5706 (2020)

49. Hu, S., et al.: Deep neural network for robust modulation classification under uncertain noise conditions. IEEE Trans. Veh. Technol. 69(1), 564–577 (2019)

50. Lin, R., et al.: A hybrid neural network for fast automatic modulation classification. IEEE Access. 8, 130314–130322 (2020)

51. Li, L., et al.: Automatic Modulation Recognition: A few-shot learning method based on the capsule network. IEEE Wireless Commun. Letters. 10(3), 474–477 (2020)

52. Zhao, Y., et al.: Lightweight deep learning based intelligent edge surveillance techniques. IEEE Trans. Cognit. Commun. Networking. 6(4), 1146–1154 (2020)