APPLICATIONS OF
MODELLING AND SIMULATION
http://arqiipubl.com/ams eISSN 2600-8084 VOL 8, 2024, 26-39

Lightweight Multi-Channel Gated Recurrent Deep Neural Network for Automatic Modulation Recognition in Spatial Cognitive Radio

Avani Vithalani[1*] and Chintan Shah[2]

[1]Department of Electronics and Communication Engineering, Government Engineering College, Sector-28, Gandhinagar, Gujarat, India
[2]Department of Biomedical Engineering, Government Engineering College, Gandhinagar, Gujarat, India

[*]Corresponding author: avanivithalani3008@gmail.com

Submitted 27 September 2023, Revised 30 December 2023, Accepted 19 January 2024, Available online 22 January 2024.

Abstract: Automatic modulation recognition (AMR) is a promising technology for intelligent communication receivers to detect signal modulation schemes. Recently, the emerging deep learning (DL) research has facilitated high-performance DL-AMR approaches. This research presents a novel and versatile Multi-Channel Gated Recurrent Deep Neural Network framework (MCGDNN) designed to tackle the intricate challenges of automatic modulation recognition. MCGDNN integrates two dedicated Deep Learning Networks (DLNs) to address specific signal types: one DLN specializes in classifying In-phase Quadrature (IQ) signals, overcoming limited training data with data augmentation and model optimization through pruning by Differentiable Annealing Indicator Search, resulting in a streamlined, lightweight model. The other DLN focuses on Frequency-Domain Amplitude-Phase signals, leveraging a modified Fast Fourier Transform (FFT) with data normalization which avoids the numerical distance between different features for enhancing feature extraction. Additionally, it introduces the Adaptive Moment Estimation Maximum (Adamax) Bi-directional Gated Recurrent Unit (Optimized BiGRU3) network that accurately extracts amplitude and phase spectrum features within the frequency domain. Furthermore, the research presents an innovative approach to signal classification by introducing a modified FFT technique for the extraction of amplitude and phase feature information from Amplitude Modulated-Double Sideband and Wideband Frequency Modulation signals in the frequency domain. This development culminates in the creation of a two-class dataset named DW, based on these amplitude and phase characteristics. In summary, this research signifies a significant stride in the field of AMR, offering a comprehensive framework (MCGDNN) capable of handling diverse signal types, an optimized feature extraction network (BiGRU3), and a novel dataset (DW) with enhanced classification accuracy. These advancements hold immense promise for applications in modern communication systems and signal processing.

Keywords: Adamax-BiGRU3; Automatic modulation recognition; Deep learning; Light-MCGDNN.

1. INTRODUCTION

Automatic modulation recognition (AMR) plays a crucial part in a variety of scenarios, including signal surveillance, cognitive radio, interference identification, spectrum sensing, etc. Moreover, the AMR technique gives the crucial modulation data of the received radio signals, particularly non-cooperative radio transmissions. Recent years have seen a substantial increase in research interest in this technique, which tries to automatically identify the modulation system of wireless communications signals without previous knowledge. During transmission, the transmitter's signals are typically changed by opposing forces in the radio frequency chain, including shadow fading, noise, sample rate offset, multi-path fading, and centre frequency offset. It may be challenging to distinguish between various modulation schemes since the structural properties of these signals may be altered because of the crystal oscillator drifting or short hardware scheme. AMR delivers the significant function of identifying modulation systems as a crucial stage between signal detection and demodulation. The modulation schemes used in signals are expected to become increasingly complex and diverse to meet the demands of evolving wireless communication technologies in rapidly changing communication scenarios. Therefore, effective AMR models must be created quickly that are reliable in challenging radio situations.

The early AMR approaches are separated into two types: feature-based AMR (FB-AMR) and likelihood theory-based AMR (LB-AMR). Based on their features like high-order spectral features, repeated features, and wavelet features, the FB-AMR makes a distinction between various types of modulation signals. Then the LB-AMR approaches frequently provide the best identification accuracy in terms of Bayesian approximation, however, they have arrow applicability and great

computational difficulty [1]. Deep learning has recently received extensive research in the areas of speech recognition, object detection, and image categorization. Moreover, it has also received a lot of interest and is used in modulation recognition [2]. In order to take phase offset (PO) into account in the orthogonal frequency division multiplexing (OFDM) system, [3] suggested an AMR approach in terms of Convolutional Neural Network (CNN) network. To attain excellent classification accuracy, the suggested approach is employed to remove the PO. In [4], they proposed a deep learning (DL)-based AMR system that can automatically extract useful and representative characteristics. Then, over the Rayleigh channel and additive white Gaussian noise channel, the proposed DL-based AMR algorithm performs significantly better than the conventional technique. However, this approach takes a lot of computational time.

To detect radio signals in a communication system, [2] describe the AMR framework. This is the mixture of deep CNN and an Long-Short Term Memory (LSTM) network. Moreover, in order to represent modulated signals in data, they suggested the combination of IQ and fourth-order statistics (IQ-FOC) methods. The proposed CNN and LSTM models can perform 8% better on their test dataset. However, this approach leads the high computational complexity. [5] suggest an effective DL-AMR approach based on the estimate and the phase parameters transformation, with CNN and Gated Recurrent Unit (GRU) serving as the feature extraction layers. This approach decreases the volume of their parameters while still achieving great recognition accuracy comparable to current state-of-the-art approaches. For AMR, [6] suggested the spatiotemporal multi-channel learning framework (MCLDNN). When viewed from a temporal and spatial perspective, this network might extract characteristics more efficiently. At high signal-to-noise ratios (SNRs), the network has made a good distinction between QAM16 and QAM64, however, half of the Wideband Frequency Modulation (WBFM) signals continue to be incorrect for AM-Double Sideband (DSB). The joint AMR approach presented in [7] is the combination of expert characteristics and deep learning to accurately distinguish QAM16 from QAM64. Moreover, in this research, the performance of the network's overall recognition in spatial cognitive communication is improved. However, the signals of WBFM and AM-DSB are still confused because of the high computational complexity. To overcome this problem, we proposed the channel pruning approach. Moreover, in order to attain higher accuracy AMR, this research presents a time-frequency domain combined recognition approach that integrates the light-MCGDNN and Adamax-BiGRU3 deep learning networks. The primary contributions of this paper are:

a) Propose a multi-domain fusion-based deep learning framework combining two neural networks MCGDNN.
   - One DLN is used to classify IQ signals, in that for dealing with the lack of training data the data augmentation techniques are applied. Then the proposed DLN network is pruned by Differentiable Annealing Indicator Search (DAIS) to reduce the model redundancy while preserving the network. Thus, the network becomes a lightweight model.
   - Other DLN is used to classify frequency domain amplitude and phase (FDAP) signals. The signals are obtained by modified FFT of IQ signals that are difficult to distinguish in the former DLN. The FFT in this network is modified by data normalization which avoids the numerical distance between different features.

b) Built a novel deep learning network, optimized BiGRU3, which can accurately extract amplitude spectrum and phase spectrum features in the frequency domain.

c) Introduce modified FFT to obtain amplitude and phase feature information of AM-DSB and WBFM in the frequency domain. And form a new two-class dataset called DW based on the amplitude and phase characteristics. At the same time, optimized BiGRU3 is used to classify DW, and the accuracy is improved on the two publicly available datasets, respectively, compared to the IQ sequence classification of the two types of signals using MCGDNN.

The following section explains the remaining part of this article. The related work is detailed in Section 2. Section 3 explains the system model, which includes the methods of proposed approaches. The experimental outcomes are explained in Section 4, that includes the datasets and results. Then Section 5 details the conclusion.

2. RELATED WORK

Current automatic modulation classification techniques have a tough time balancing classification accuracy and model complexity because of the impact of noise in the established signal in non-cooperative communication. So, the Convolutional Adaptive Noise Reduction (CANR) presented by [8] is the combination of a Convolutional Feature Extraction (CFE) module and an Adaptive Noise Reduction (ANR) module. These two modules remove the noise from the joint input while recording the spatiotemporal characteristics of the time-series. In order to improve the information interaction among the input data's I/Q channels for AMR, [9] suggest an auto-encoder based technique. To associate the characteristics of I/Q information and retrieve the communication feature from the middle layer, the suggested technique makes use of an auto-encoder constructed through the fully connected layers. Moreover, they are combined as model inputs with the unique I/Q data. DL-AMR techniques are becoming more famous because of the development of deep learning. But, the majority of the DL-based existing approaches are incapable of handling complicated data formats, and learning the direct mappings from the time-series or its modified demonstration to the factual modulation kind is challenging. Therefore, the complicated-valued convolution and frequency global filter unit (CGFU) and the hybrid neural network (CGF-HNN) proposed in [10] prevent these issues. Moreover, this proposed approach can effectively utilize the features from many domains. In the field of radio monitoring, the automatic modulation categorization (AMC) and specific emitter identification (SEI) are usually two different tasks. However, these two tasks are integrated to enhance the classification accuracy for each task and decrease the total computational complexity. So, [11] presented the AMSCN approach, a dual-task NN that concurrently categorises the modulation and the transmitter of the established signal. To extract distinguishing characteristics for the AMSCN, they first combine DenseNet and Transformer as the backbone network. Next, they construct a mask-based dual-head classifier (MDHC) to strengthen the combined learning of the 2 tasks.

The CNN-based modulation recognition framework is presented by [12] for radio signal detection in communication networks. Because frequency fluctuation with time is the primary way that radio signals with different modulation patterns can be distinguished from one another in their work, 1-D radio signals are transformed into spectrogram images utilizing the short-time discrete Fourier transform. Moreover, they evaluate the statistical characteristics of the radio signals and decrease the noise by using the Gaussian filter. In [13], suggest a multi-task learning (MTL) method for identifying the modulation scheme being utilised among a particular collection of analog and digital modulations. In this method, the required features are extracted utilizing a deep CNN in order to categorise the various modulation schemes. Moreover, the modulation classes that typically cause a lot of confusion are separately trained using the MTL scheme, so the overall classification accuracy is increased. Cognitive radio (CR) is one of the growing fields where Automated Modulation Classification (AMC) has been used. DL is the most effective approach for classification so, [14] focus on DL to address communication issues. Moreover, to improve the categorization accuracy of communication signal modulation, they suggest a new data conversion algorithm. [15] have proposed a CNN model based on seven optimization approaches Nadam, AdaGrad, Adamax, SGD, AdaDelta, RMSprop, and Adam with eight different learning rates for effectively forecasting hourly typhoon rainfall. The learning rates are proved to a very sensitive to forecast sequence based hourly typhoon rainfall.

A deep learning framework based on multi-domain fusion is proposed by [16] as a combination of two neural networks. The first phase is made by the MCLDNN, which has been trained with sample features made of IQ signals to discriminate modulation modes with ease, when the receiver gets the unidentified signals. Then the second phase of FFT is utilized for removing the FDAP data. Jialang Xu has presented a unique three-stream DL framework based MCLDNN to excerpt features more efficiently from the perspective of space and time for AMR. [17] suggested Graph convolution based adaptive regularized accompaniment of the loss of cross entropy based learning of deep network through adding Laplacian basis matrix with the regularization period to acknowledge the spatial data of removed features for achieving the most accurate classification performance, a complement entropy approach is integrated based on the loss of cross-entropy the misclassification issues of samples are overwhelmed. For constructing a Bi-GRU learning model, [18] proposed an adaptive moment estimation maximum optimization approach (Adamax) based on Adamax-BiGRU with the MSE as the loss function. This proposed two-way GRU learning model is used to explore the GRU model parameters of the iterations, hidden layer, and hidden nodes for predicting the gas concentration with time series sequences. The batch size and optimizer, activation function are utilized as hyper-parameters to explore the best settings of the 1DCNN-GRU approach through grid search. This CNN and Stacked GRU based model has been proposed by [19] for the prediction of cryptocurrency price with four optimizers namely RMSProp, Adam, Adamax, and SGD. The five activation functions such as softmax, tanh, linear, sigmoid and ReLU to perform and learn more difficult tasks. The error gradient computation of this model is defined by the batch size for updating the weights of each layer neuron.

One of the most popular methods for deep model compression is network pruning. Moreover, the space and time complexity is decreased by pruning the big neural network while preserving their performance. The current pruning techniques are mostly attention to the significance of filters in the overall process. The matrix similarity measure (MSM) is presented in [12], this approach is utilized to calculate the similarity between feature maps or filters for the same layer. The automatic channel pruning (ACP) approach is proposed by [20] for accelerating and compressing the CNNs. For the channel clustering and pruning, they compute the cosine distance between feature maps. Moreover, based on the candidate population initiated through the clustering pruned substructure, the PSO technique is presented to repeatedly search for the most compact network. In [21] proposed the Recursive Bayesian Pruning method (RBP) approach for solving the NP-hard problem. The redundant channels are successfully discovered and immediately pruned layer by layer using the proposed approach RBP. [22] have proposed an annealing-based DAIS For attaining appropriate pruned models by reducing the system redundancy while protecting the network structure with sparsity requirements.

3. SYSTEM MODEL

3.1 SCR Uses an Intelligent Receiver System Based on Zero IF Design

The most popular receiver architecture is Quadrature sampling, which samples the IQ signal to acquire the data stream. The network design in this paper will make high use of this architecture. Figure 1 shows the structure of the quadrature sampling zero-IF intelligent receiver. I and Q signals of dimension N×2 were acquired for each sample, where the time length of the signal is corresponded to N and the component of I/Q is corresponded to 2. The workflow procedure for the high satellite communication receiver is as follows: In order to choose a frequency and amplification, at first the signal of RF passes over the middle-pass filter and the little-noise amplifier. After that, the mixer receives the signal, and the components of I/Q are effectively produced by the mixing at the local oscillator frequency. In order to generate a digital IQ baseband signal, I and Q signals are first amplified and clearly filtered, sampled and then deeply extracted. In order to fully complete the signal modulation kind identification, the AMR receives the obtained IQ baseband signal.

3.2 The Proposed Model

Based on the structure of deep learning, this article suggested a multi-domain fusion that combines two neural networks. In order to recognize multiple modulated signals that are extensively utilized in contemporary wireless communication systems by using the joint model is shown in Figure 2. The first stage of signal identification will be made by light-MCGDNN when the unknown signals are obtained by the receiver. Furthermore, in this stage, AM-DSB and WBFM are referred as DW which is regarded as the same class. During the second stage, the sampling features of the frequency domain are created using FFT. A brand-new dataset was established, then for the recognition of the binary classification of DW, it will be fed into the Adamax-BIGRU3 network.

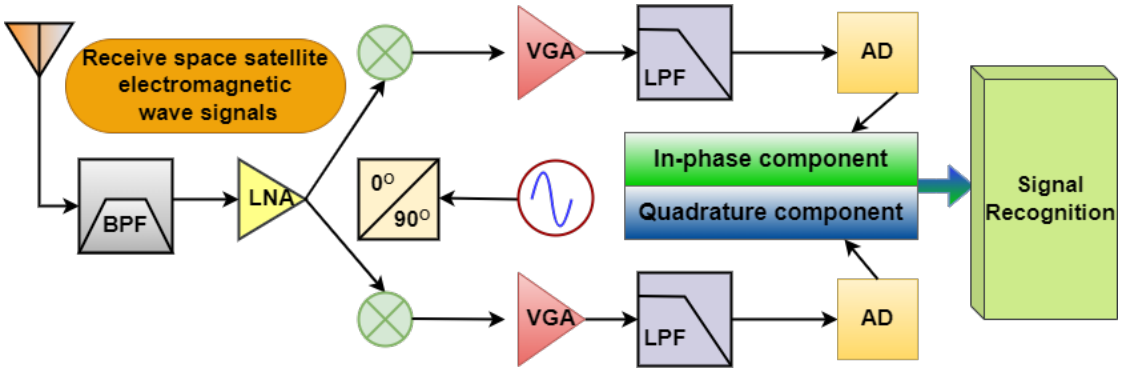


Figure 1. Architecture of Quadrature sampling zero-IF intelligent receiver.

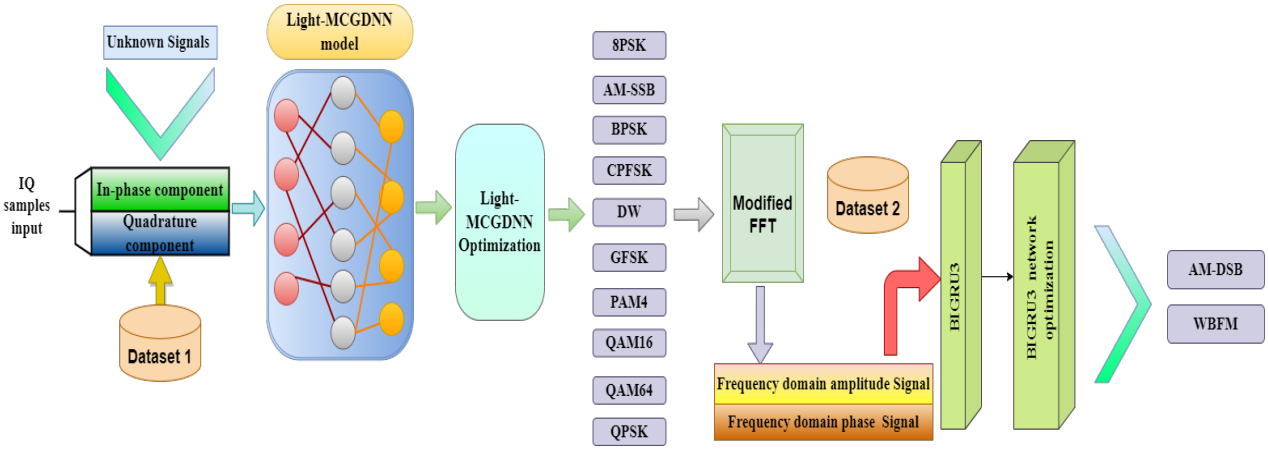


Figure 2. Architecture of AMR model.

3.3 Methods of Proposing Model

3.3.1 Light-MCGDNN

A unique three-stream deep learning framework called light-MCGDNN is used to excerpt the features from the separate and joint I/Q symbols of modulated information. The suggested framework combines GRU layers, two-dimensional (2D) convolutional layers, and one-dimensional (1D) convolutional layers to excerpt information more efficiently from a space and time perspective. Figure 3 depicts the structure of light-MCGDNN approach.

The network separates the input IQ signals into 3 channels as shown in Figure 3. First, the input signals go to the augmentation process for improving the mapping of biological signals and then excerpts the time sequence properties of the routes I, Q, and IQ, in that order. Moreover, data augmentation is an approach that is frequently utilized to address the shortage of training data and it also prevents the overfitting produced by the shortage of training data. [128, Batch-size, channels, 2] is the four-dimensional data input form used for IQ during model training. When the channels merge, light-MCGDNN models are joined in the channel dimension. For more feature extraction, use the convolution kernel of (2,5). Then the pruning process is applied in the extracting features of the CNN network. In this article, we present the DAIS approach for channel pruning in convolutional networks. Moreover, this DAIS technique is utilized to decrease the memory footprint, model parameters, and FLOPs. After the CNN pruning, the two-layer GRU is used with the extracted features to further extract the time sequence feature, which is then connected to a two-layer DNN for classification.

The following describes the pruning techniques considered in this work.

Differentiable Indicator Search: The channel pruning problem is rarely addressed using the differentiable search process, despite being frequently utilised in Neural Architecture Search (NAS). Each binarized channel indicator $E_i^d, d \in [1, c_i]$ is relaxed into a relaxed channel indicator $\tilde{E}_i^d$ parameterized through an auxiliary parameter α_i^d in order to create a continuous search space. The aim of the indicator search after the relaxation is to simultaneously learn auxiliary parameters α model parameters $\mathcal{M}$. A logical alternative for optimisation is to simultaneously update $\mathcal{M}$ and α on the training set, however this α will make over-fitting on the training set, which could lead to a pruned outcome with low generalizability. As a result, we use the bi-level optimisation with upper-level variable (α) and lower-level variable ($\mathcal{M}$) in the differentiable indicator search technique. It specifically looks for α^* for reduces the sum of regularizers $r(\alpha)$and validation loss $Loss_{val}(\mathcal{M}^*(\alpha), \alpha)$, where $\mathcal{M}^*$ stands for the model parameters that are acquired by reducing training loss $Loss_{train}(\mathcal{M}, \alpha)$ that is expressed in Equations (1) and (2).

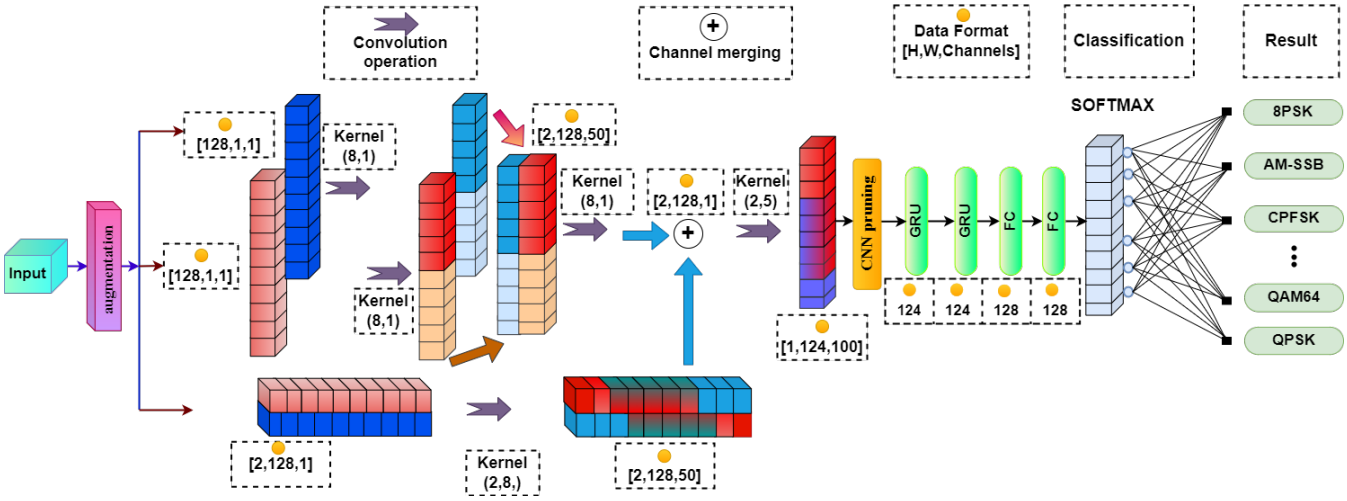


Figure 3. The architecture of light-MCGDNN model

$$\min_{\alpha} Loss_{val}(\mathcal{M}^*(\alpha), \alpha) + \lambda r(\alpha) \tag{1}$$

$$\mathcal{M}^*(\alpha) = arg \min_{\mathcal{M}} Loss_{train}(\mathcal{M}, \alpha) \tag{2}$$

Gradient descent on training and validation sets is used to update $\mathcal{M}$ and α in a multi-step process until they achieve the local minima, which is the solution to the bi-level optimisation issue.

Relaxed channel indicator for annealing: The value range for each input $\tilde{E}_i^d$ should be restricted to the range of 0 to 1, making the relaxed-channel indicator a more accurate representation of the binarized channel indicator. Therefore, using a sigmoid function over sub-parameters is a simple fix is shown in Equation (3).

$$\tilde{E}_i^d = \frac{1}{1+e^{-\alpha_i^d}} \tag{3}$$

But there are two inescapable issues with the above estimation. The first one is sparsity based on the continuous sub-parameters is not guaranteed, therefore the resulting $\tilde{E}_i^d$ might not converge to values near 0 or 1. The second one is the continuous search result and the binarized models differ significantly as a result of discretization. When the combined parameters of α and $\mathcal{M}$ are trapped in severe local minima where a minor perturbation might result in significant performance loss, this could damage the pruned model's accuracy. Then the annealing relaxed channel is proposed to avoid these issues. A temperature variable Tem is added to the sigmoid function, at the start of training that is set to be high-level and slowly anneals to 0. In order to allow the gradient update of the auxiliary parameters, the indicator is initially continuous. However, it eventually joins the binarized state, which results in the pruned model as shown in Equation (4).

$$\tilde{E}_i^d = F_{Tem}(\alpha_i^d) = \frac{1}{1+e^{-\alpha_i^d/Tem}}, E_i^d = \lim_{Tem \to 0} F_{Tem}(\alpha_i^d) \tag{4}$$

where the auxiliary parameters are denoted by α_i^d, $F_{Tem}(\cdot)$ is represented as the annealing relaxed function that is started with the maximum temperature $Tem = Tem_0$. The temperature of F_{Tem} anneals to $Tem_0/\psi(\mathrm{n})$ as the training enters the n-th epoch. Then the temperature annealing is represented by $\psi(\cdot)$. After the search, it is possible to approach the discrete E_i^d with $Tem \to 0$.

Structure Restrictions through Regularisation: Inducing a priori structural limits, such as the quantity of floating point operations (FLOPs), which are essential to pruning, with the vanilla cross-entropy loss alone is impractical. Hence, when upgrading the auxiliary parameters, we add three regularizers to the search process.

(i) Lasso regularizer: The lasso regularise is a commonly utilized sparsity regularizer, which is formulated in Equation (5).

$$r_{lasso} = \sum_{i=1}^{I} \sum_{d=1}^{c_i} \left\| F_{Tem}(\alpha_i^d) \right\|_1 \tag{5}$$

Some channel indications can be efficiently zeroed out by r_{lasso} using the lasso/L1 regularisation, however, the pruning rate it produces from the r_{lasso} is uncontrollable and highly reliant on the weight it places on the entire loss function.

(ii) Regularizer for the continuous FLOPs estimator: The possibility of protection of the associated channel in the last pruned model can be thought of as the value of the annealing-relaxed channel indicator $F_{Tem}(\alpha_i^d)$. We may calculate the expected total FLOPs of the pruned model by adding the FLOPs of every output channel using the annealing-relaxed indicator of overall output channels. The continuous FLOPs estimator for the i[th] convolutional layer might be modelled as Equation (6).

$$Est_{FLOPs}(\alpha) = \sum_{i=1}^{I} \left(p_i \left(\sum_{d=1}^{c_{i-1}} F_{Tem}(\alpha_{i=1}^d) \right) \cdot \left(\sum_{d=1}^{c_i} F_{Tem}(\alpha_i^d) \right) \right) \tag{6}$$

where $p_l = x_l \times y_l \times z_l^2$, the kernel size is represented by z_l, then the output feature maps spatial size is denoted by x_l and y_l. To construct the regularizer given the continuous FLOPs estimator in the manner suggested in Equation (7) [23].

$$r_{FLOPs}(\alpha) = \begin{cases} \log\left(Est_{FLOps}(\alpha)\right), \frac{Est_{FLOPs}(\alpha)}{Fs} > 1 \\ -\log\left(Est_{FLOPs}(\alpha)\right), \frac{Est_{FLOPs}(\alpha)}{Fs} < 1 - \epsilon \\ 0, \ otherwise \end{cases} \tag{7}$$

where expected FLOPs denoted by Fs. In order to find the best pruned model with FLOPs in the range of $[(1-\epsilon) * Fs, Fs]$, DAIS is guided by the continuous FLOPs estimator regularizer.

(iii) Symmetry regularizer: Recent network topologies frequently use residual blocks, considerably enhancing the potential of gradient propagation across multiple layers. The recognition function is replaced with a 1 x 1 convolutional layer for dimension matching because the current channel pruning techniques can't guarantee that the number of output and input channels of a remaining block is identical. The issues of vanishing and exploding gradients may result from these substitutions, which could make gradient propagation across numerous layers more challenging. So, the symmetry regularizer is presented for channel pruning on networks with remaining connections that are defined as following in Equation (8).

$$r_{sym} = \sum_{i,i'} \left| \left(\sum_{d=1}^{c_i} F_{Tem}(\alpha_i^d) \right) - \left(\sum_{d=1}^{c_{i'}} F_{Tem}(\alpha_i^d) \right) \right| \tag{8}$$

where residual blocks are represented by (i, i'), input and output channels are denoted by c_i and $c_{i'}$.

3.3.2 Adamax-BIGRU3 Network Optimization

GRU is the most widely used neural network because it is very effective and better performance compared to the LSTM network. In addition to being able to resolve the issue of long-term reliance, GRU provides outstanding feature extraction capabilities for sequence data. The GRU2 network was proposed by [24], they also showed that GRU contains high feature extraction capabilities for phase spectrum and time-domain amplitude spectrum characteristics, with 91.3% classification accuracy for several signals. So, we suggest an Adamax-BiGRU3 model that excerpts phase and amplitude characteristics from the frequency domain. The architecture of Adamax-BiGRU3 is shown in Figure 4. In this Figure 4, there are three BiGRU layers to provide more fully extract the features from the data. Then to integrate the features, we add a fully-connected layer and a Guassian Dropout layer. Moreover, the Adamax optimization is applied in BiGRU3 network to implement classification accuracy.

Optimization framework: In the DNN, the parameter that needs to be optimised is called ω, and the goal function is called $f(\omega)$. The deep learning optimisation problem can be characterised as the search for an appropriate collection of parameters ω to reduce the loss function. The DNNs' optimisation framework can be summarised as follows: Iterative optimisation starts after the first learning rate is set at ω. The following steps are used to calculate the update parameters for the *t*-th training cycle (epoch):

1) Determine the objective function's gradient g_t with regard to the present parameters that is shown in Equation (9)

$$g_t = \nabla f(\omega) \tag{9}$$

2) Based on the historical gradient, determine the first order momentum m_t and second order momentum V_t using the following formulas:
   (i) Stochastic Gradient Descent (SGD): the characteristic makes its gradient update m_t directly equal to g_t because it does not employ momentum, converges gradually, and falls into local extremes quickly, which is expressed in Equations (10) and (11).

$$m_t = g_t \tag{10}$$
$$\eta_t = \alpha \cdot m_t \tag{11}$$

   (ii) SGD with momentum: utilizes the gradient's momentum and converges more quickly than SGD that is shown in Equations (12) and (13).

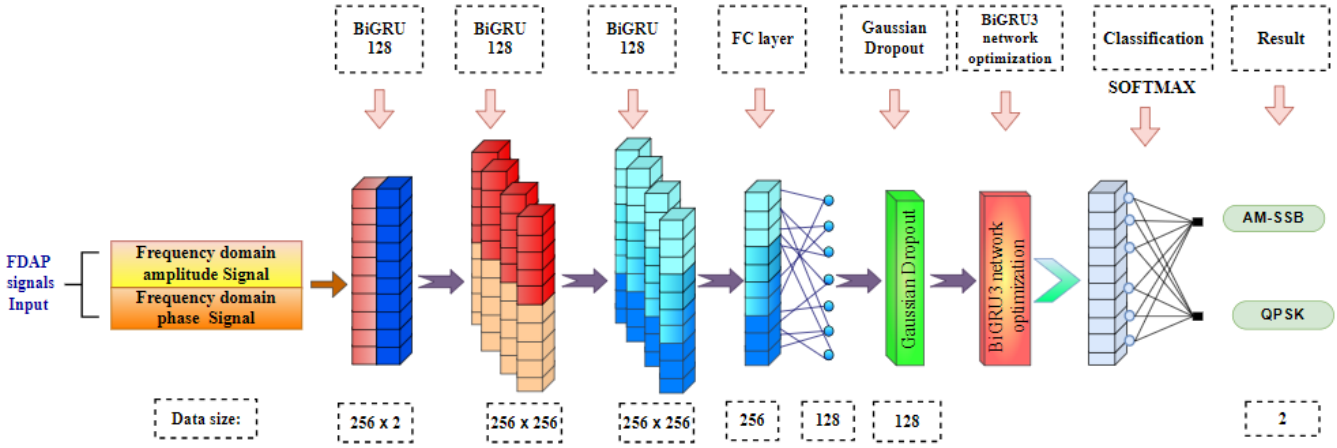


Figure 4. The architecture of Adamax-BiGRU3 optimization model.

$$m_t = \beta \cdot m_t - 1 + (1 - \beta) \cdot g_t \quad (12)$$
$$\eta_t = \alpha \cdot m_t \quad (13)$$

(iii) Adagrad: Equation (14) determines the phase size that can be adaptively altered when second order momentum is utilised, but because second-order momentum gathers all of the history, learning may be halted early.

$$v_t = \sum_{i=1}^{t} g_i^2 \quad (14)$$

(iv) AdaDelta/RMSProp: The update mechanism for second-order momentum is distinct from AdaGrad's. All historical second-order momentum is accumulated by AdaGrad, and some of it is accumulated by AdaDelta is shown in Equation (15). It can prevent an early study from ending.

$$v_t = \beta \cdot v_{t-1} + (1 - \beta) \cdot g_t^2 \quad (15)$$

(v) Adam: It mixes the 1^{st} and 2^{nd} order momentums, which significantly speeds up the convergence rate. Moreover, it uses the same 1^{st} order momentum update mechanism as SGD with momentum and the same 2^{nd} order momentum update approach as AdaGrad that is shown in Equations (16) and (17).

$$m_t = \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot g_t \quad (16)$$
$$v_t = \beta_2 \cdot v_{t-1} + (1 - \beta_2) \cdot g_t^2 \quad (17)$$

(vi) Adamax: A variation of Adam is Adamax. With this approach, the top limit of the learning rate can be changed in a simplified range formula is shown in Equation (18).

$$\eta_t = max(v \times \eta_{t-1}, |g_t|) \quad (18)$$

3) Determine the current decreasing gradient η_t by using Equation (19).

$$\eta_t = \alpha \cdot m_t / \sqrt{v_t} \quad (19)$$

4) According to the decreasing gradient, the ω value is updated by using Equation (20).

$$\omega_{t+1} = \omega_t - \eta_t \quad (20)$$

The various ways that g_t, m_t, and $v_t$are calculated in steps (1) and (2) represent the differences in the various optimisation algorithms.

Adamax Optimization Algorithm: The improved version of Adam is known as Adamax. It establishes a more similar range for the maximum learning rate.max algorithm is utilized to optimize the BiGRU approach. Therefore, In Adamax, Equation (9) is used to calculate the gradient of formula in Equation (21).

$$g_t = \nabla_\omega f_t(\omega_{t-1}) \quad (21)$$

Equation (22) is used to evaluate the first-order momentum.

$$m_t = \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot g_t \tag{22}$$

Equation (23) is utilized to calculate the second-order momentum.

$$v_t = max(\beta_2 \cdot v_{t-1}, |g_t|) \tag{23}$$

The first order and second-order momentum are each controlled by hyper-parameters β_1 and β_2, respectively, in this set of variables. The optimization algorithm is used to calculate its values.

3.3.3 Fast Fourier Transforms (FFT)

FFT is a discrete Fourier transform (DFT) optimisation technique that is highly effective than DFT. However, their outcomes are practically identical. The primary objective of this phase is to gather frequency domain data from power quality disturbances (PQDs) signals. Time series signals are inappropriate for certain sorts of recognition, particularly when there are harmonic aberrations, so FFT is used to extract frequency domain data from PQDs signals, which break the signal down into several sinusoid frequencies. The FFT procedure can be described as following in Equation (24).

$$X(s) = \sum_{p=0}^{N/2-1} x(2p) W_N^{s(2p)} + \sum_{p=0}^{N/2-1} x(2p+1) W_N^{s(2p+1)} \tag{24}$$

where $W_n = e^{-j2\pi/N}$ and the sampling point between 0 and $N - 1$ is represented by s. It produced one frequency output after applying FFT, designated as $X_d = \{x_{1d}, x_{2d}, \ldots, x_{Nd}\}$ where $x_{id} = x_{id1}, x_{id2}, \ldots, x_{idp}$ and the frequency characteristics from FFT are represented by the variable p.

The typical PQD waveform in a noisy environment is shown in Figures 5(a) – 5(d) from top to bottom, along with frequency component charts. The fundamental frequency, which can be seen from the plots of each frequency component, is 50 Hz. The third, fifth, and seventh harmonic components of the waveform are shown in Figure 5(b) for the frequency component. The precise harmonic components are difficult to separate from the original PQDs waveform, though.

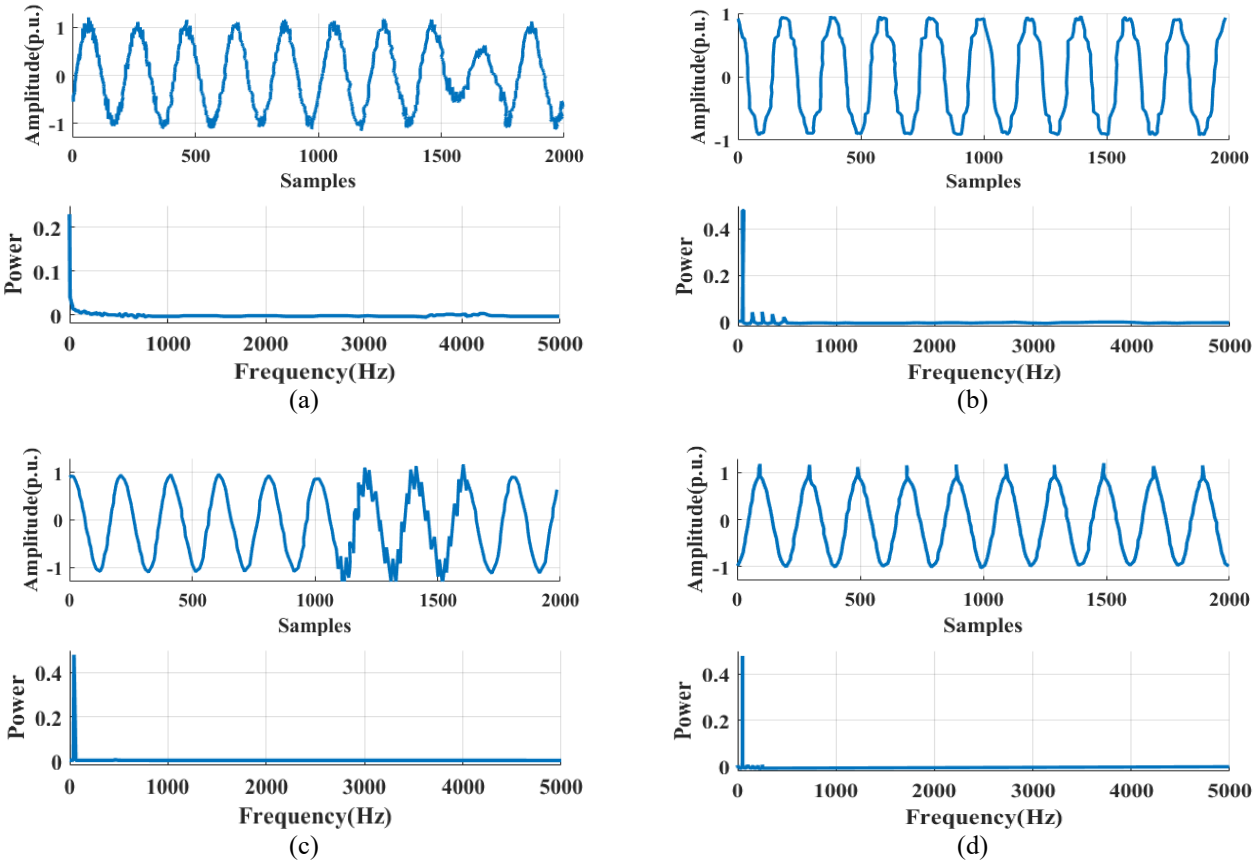


Figure 5. PQDs frequency component and its waveform in the noisy surroundings (a) sag, (b) harmonics, (c) oscillatory transient and (d) spike.

Data Normalization: Prior to training the classification model, data normalisation is required, which rescales each feature in the fixed range. Examples of this normalisation include decimal scaling, Z-score (ZSN), minimum-maximum, and others. The ZSN is used for the frequency characteristics X_d and original signal X_r. $x'_{ir} = \langle x'_{ir1}, x'_{ir2}, \dots, x'_{irp} \rangle$ and $x'_{id} = \langle x'_{id1}, x'_{id2}, \dots, x'_{idp} \rangle$ are represents the appropriate normalised time and frequency features. The formula for each feature x'_{irn} and x'_{idp} values are shown in Equations (25) and (26).

$$x'_{irn} = \frac{x_{irn} - \mu_{rn}}{\sigma_{rn}} \tag{25}$$

$$x'_{idp} = \frac{x_{idp} - \mu_{dp}}{\sigma_{dp}} \tag{26}$$

where the standard deviation and mean of the *i*-th characteristics in the frequency and time domains are denoted by μ_{rn}, μ_{dp} and σ_{rn}, σ_{dp}. Each variable that was scaled for time and frequency has a mean and standard deviation of 0 and 1. Data normalisation prevents the numerical separation between various characteristics; however, it cannot assure the absence of unnecessary features. These undesirable characteristics might reduce classification accuracy and raise computing complexity. Therefore, removing unwanted features is essential for raising classification accuracy. $X(s)$ in Equation (24) is set as $A(s)$ in shown is Equation (27).

$$A(s) = \sum_{p=0}^{\mathrm{N}/2-1} x(2p) W_{\mathrm{N}}^{s(2p)} + \sum_{p=0}^{\mathrm{N}/2-1} x(2p+1) W_{\mathrm{N}}^{s(2p+1)} \tag{27}$$

Equation (14) illustrates the phase of $X(s)$ in the frequency domain that is shown in Equation (28).

$$\theta(s) = arctan\left(\frac{Im[A(s)]}{Re[A(s)]}\right) \tag{28}$$

In the frequency domain, the outcome *X*(*k*) is the sample point value, and that the value is expressed as a complex number, according to the above discussed DFT principle. Equation (29) shows the format:

$$X(s) = [p_1 + jq_1, p_2 + jq_2, \dots, p_s + jq_s], s = 1, \dots, \mathrm{N} \tag{29}$$

Therefore, determine the signal's amplitude $K_s$and phase Φ_s in the frequency domain according to Equations (30) and (31).

$$K_s = |X(s)| = \sqrt{p_s^2 + q_s^2},\ s = 1,2, \dots, \mathrm{N} \tag{30}$$

$$\Phi_s = \theta(s) = arctan\left(\frac{q_s}{p_s}\right),\ s = 1,2, \dots, \mathrm{N} \tag{31}$$

The FFT transform can translate the discrete data in the time domain into separate magnitude and phase values in the frequency domain. The time domain formulation of the amplitude modulation signal (AM) for AM-DSB and WBFM signals is displayed in Equation (32).

$$K_{AM}(t) = [A_0 + m(t)] \cos \omega_c t \tag{32}$$

where the deterministic signal is denoted by $m(t)$, carrier frequency is represented by ω_c and A_0 is the component of applied direct current. Equation (33) demonstrates the time domain format of the AM-DSB signal.

$$K_{\mathrm{DSB}}(t) = m(t) \cos \omega_c\, t \tag{33}$$

Then, Equation (34) displays the time domain format of the WBFM signal.

$$K_{\mathrm{WBFM}}(t) = A \cos(t)(\omega_c t + S_{FM} \int f(t) dt) \tag{34}$$

where the coefficient of the FM is represented by $S_{FM} \int f(t)dt$, the modulated signal is denoted by $f(t)$.

4. EXPERIMENTAL RESULTS

4.1 Dataset Description

The proposed model uses 2 datasets RadioML2016.10a and RadioML2016.10b. These two datasets are produced in [25,26] utilizing the GNU-Radio software platform. Eight different kinds of analog modulation signals and three different digital modulation signal types are included in the first dataset. Then the second dataset contains seven different kinds of analogue modulation signals and three different digital modulation signal types. This has one fewer class than the first dataset, but the second signal count is higher. The SNR of different modulated signals varies from -20 dB to 18 dB with a 2 dB gap. SNRs come in 20 different varieties. Each modulated signal has a size of (2, 128), where 128 is the number of sample points per signal and 2 is the number of quadrature IQ two-way signals used as the input signal. An entire of 220,000 signals are present in the RML2016.10a dataset while 1,200,000 signals are present in the RML2016.10b dataset. The parameters of the

MCGDNN are shown in Table 1. The Batch-size parameter is set to 128, 0.001 is the starting learning rate and the Epoch size is 100. Using stochastic gradient descent with the Adam-max optimizer. The RadioML2016.10a RadioML2016.10b Dataset detailed parameters are given in Table 2.

4.2 Result

Table 3 contains a variety of variables that are chosen for complexity analysis and performance comparison, such as the number of parameters, training and test times, and the average and greatest accuracy of all models. Then the number of parameters varies between the models, as indicated in this table, since each model is evaluated using the two datasets *A* and *B*. Moreover, when compared to benchmark models, it is evident that the suggested model has the fewest parameters, so it takes less time for training and testing process, and it has higher accuracy.

In a confusion matrix, each square's colour depth represents the precision of the recognition result. The classification of diverse modulated signals via network classification is represented by the abscissa of the confusion matrix. The actual modulation techniques are shown in the ordinate. The accuracy of the model's predictions is higher if the matrix's diagonal is darker. The recognition effect is less than optimal if the colour blocks are dispersed across the matrix and not focused along the diagonal. Figure 6 displays the confusion matrices for the first dataset when signal modulation is recognised using MCLDNN networks with SNRs of 0 dB, 4 dB, and 12 dB. Still, the AM-DSB and WBFM transmissions are not in the correct order. As revealed in light colour in the figure, even with strong SNRs, half of WBFM transmissions is still not identified as AM-DSB. When utilising MCLDNN networks for signal modulation identification and selecting SNRs of 0, 4, and 12 dB, Figure 7 displays the confusion matrices in the second dataset. The AM-DSB and WBFM signals in the second dataset is not in order, as seen in Figure 7, which has an impact on the overall recognition rate.

Table 1. Parameters for training.

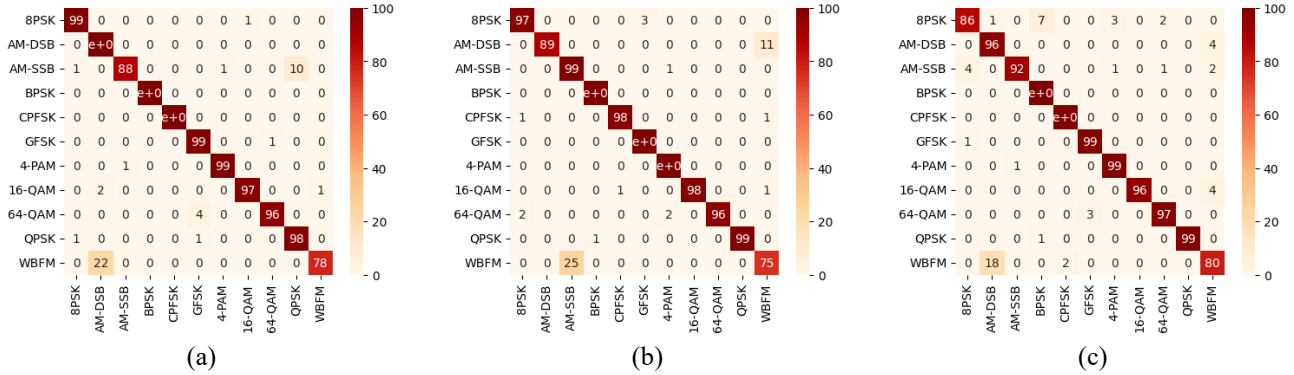
Parameters for training	MCGDNN
Epoch	100
Optimizer	Adam-max
Dataset partitioning ratio	7:2:1
starting learning rate	0.001
Batch-size	128

Table 2. RadioML2016.10a RadioML2016.10b Dataset detailed parameters.

Details	RadioML2016.10a	RadioML2016.10b
Modulations	QPSK, 16QAM, 32QAM, 64QAM, 64-OFDM, 128-OFDM, 256-OFDM, 512-OFDM	Digital Modulations: BPSK, QPSK, 8PSK, 16QAM, 64QAM, CPFSK, PAM4 and GFSK Analog Modulations: AM-DSB and WBFM
Samples per symbol	8	2*128
SNR range	-10dB to +10dB (2dB step), [20/dB to 30/dB] (10 dB step)	[-20dB, -18dB, ..., 16dB, 18dB]
Number of training samples	51200 frames	840000 vectors
Number of testing samples	51200 frames	360000 vectors

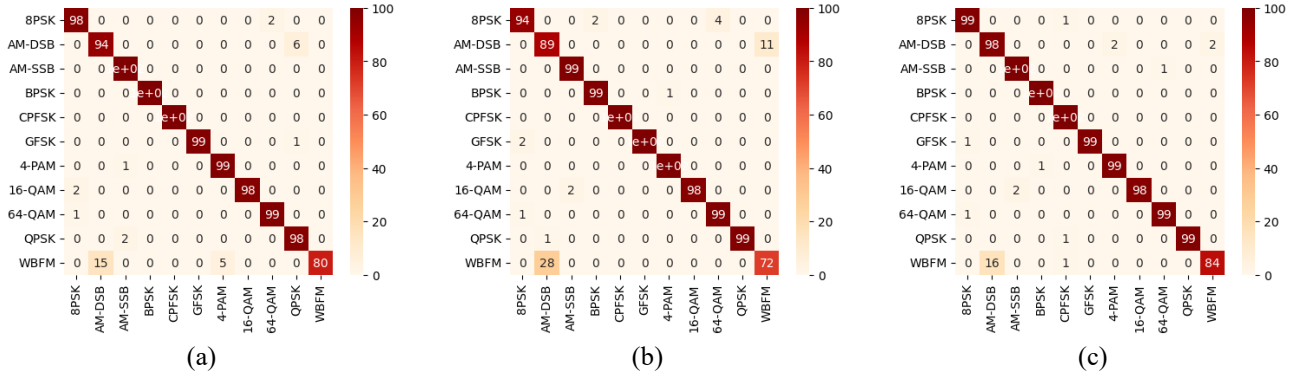
Table 3. Two dataset model comparisons (A: RadioML2016.10a, B: RadioML2016.10b).

Models	Datasets	Parameters	Training Time	Testing Time	Highest accuracy (%)	Average accuracy (%)
CNN	A	132,504	63 m	54m	85.3	64.2
	B	132,491	3h 51m	3h 02m	83.9	63.5
MCGDNN	A	10,204	58m	55m	91.7	58.6
	B	10,176	3h 18m	2h 27m	93.2	60.7
Bi-GRU3	A	94,959	50m	44m	92.6	60.9
	B	94,820	2h 47m	1h 58m	90.9	57.2
MCLDNN	A	80,011	42m	38m	89.8	62.7
	B	79,729	1h 29m	1h 02m	95.4	63.6
Suggested technique	A	64,532	36m	33m	97.2	59.7
	B	64,209	1h 12m	58m	96.1	60.1



(a) (b) (c)

Figure 6. Confusion matrices of first dataset at various SNRs (a) SNR=0 dB, (b) SNR=4 dB, (c) SNR=12 dB.



(a) (b) (c)

Figure 7. Confusion matrices of second dataset at various SNRs (a) SNR=0 dB, (b) SNR=4 dB, (c) SNR=12 dB.

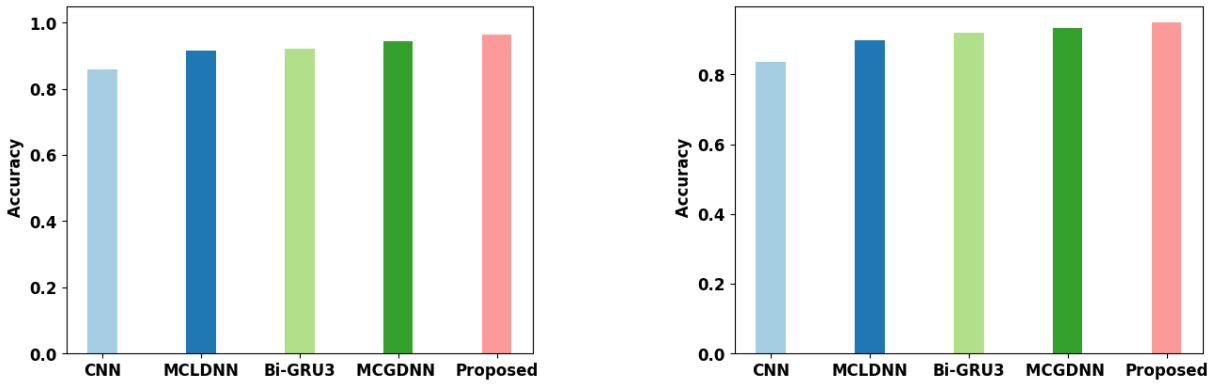
In a confusion matrix, each square's colour depth represents the precision of the recognition result. The classification of diverse modulated signals via network classification is represented by the abscissa of the confusion matrix. The actual modulation techniques are shown in the ordinate. The accuracy of the model's predictions is higher if the matrix's diagonal is darker. The recognition effect is less than optimal if the colour blocks are dispersed across the matrix and not focused along the diagonal. Figure 6 displays the confusion matrices for the first dataset when signal modulation is recognised using MCLDNN networks with SNRs of 0 dB, 4 dB, and 12 dB. Still, the AM-DSB and WBFM transmissions are not in the correct order. As revealed in light colour in the figure, even with strong SNRs, half of WBFM transmissions are still not identified as AM-DSB. When utilising MCLDNN networks for signal modulation identification and selecting SNRs of 0, 4, and 12 dB, Figure 7 displays the confusion matrices in the second dataset. The AM-DSB and WBFM signals in the second dataset is not in order, as seen in Figure 7, which has an impact on the overall recognition rate.

In the context of signal classification, Table 4 presents an evaluation of various methods on two distinct datasets, RadioML2016.10a and RadioML2016.10b, with a focus on key performance metrics. In the case of the RadioML2016.10a dataset, the "Proposed Model" stands out as it achieved the highest levels of accuracy (0.9632), F1 score (0.9814), and recall (0.9616) among the methods, all while maintaining a relatively efficient training time of 36 minutes. This underscores its superiority in discerning signal patterns within this dataset. For the RadioML2016.10b dataset, the proposed model once again demonstrates remarkable performance with an accuracy of 0.9482, an F1 score of 0.9654, and a recall of 0.9482, representing notable improvements over the other methods. These findings emphasize the efficacy and efficiency of the proposed Model" in signal classification, making it a compelling choice for both datasets, RadioML2016.10a and RadioML2016.10b. Moreover, the suggested and existing techniques accuracy performance is shown in Figure 8.

Table 4. Performance of the proposed and existing models in two datasets.

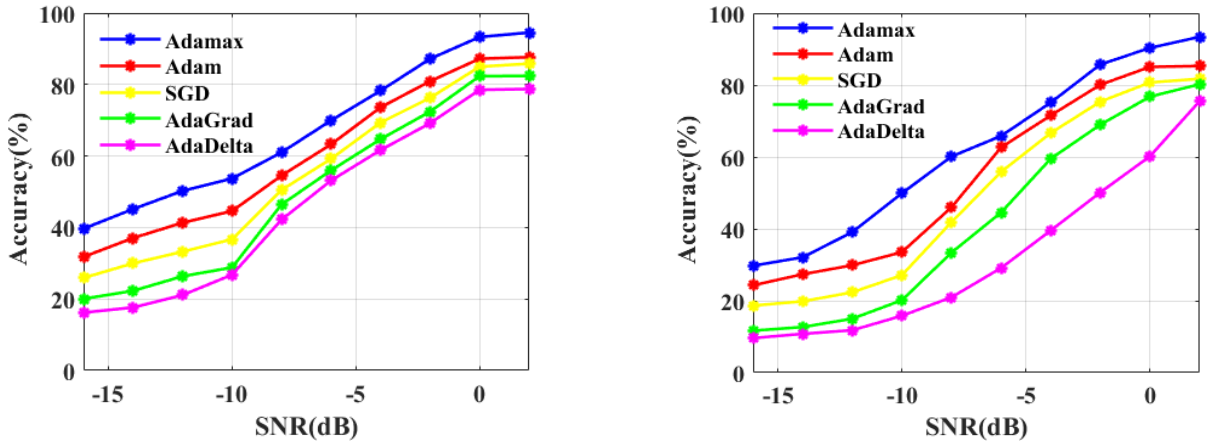
Dataset	Method	Accuracy	F1 score	Recall
RadioML2016.10a	CNN	0.8574	0.8874	0.8560
	MCGDNN	0.9166	0.9285	0.9134
	Bi-GRU3	0.9211	0.9413	0.9222
	MCGDNN	0.9427	0.9466	0.9404
	Proposed Model	0.9632	0.9814	0.9616
RadioML2016.10b	CNN	0.8366	0.8532	0.8345
	MCGDNN	0.8971	09175	0.9011
	Bi-GRU3	0.9187	0.9318	0.9172
	MCLDNN	0.9324	0.9482	0.9343
	Proposed Model	0.9482	0.9654	0.9482

In the training period, the optimizer plays a significant role. Figures 9(a) and 9(b) show the comparison of the existing and proposed optimizer's accuracy values. The proposed Adamax optimizer gives a better performance of accuracy compared to the existing optimizers such as Adam, SGD, AdaGrad and AdaDelta. The performance of the model recognition after pruning is shown in Figures 10(a) and 10(b) compared to the existing pruning techniques like RBP, SBP and the proposed pruning DAIS produces the high identification accuracy value.



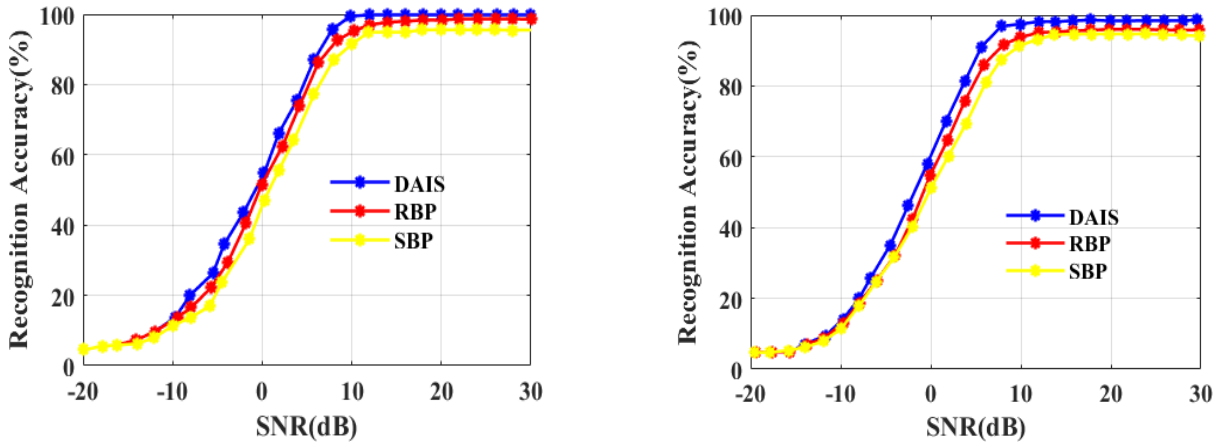
(a) RadioML2016.10a (b) RadioML2016.10b

Figure 8. Performance of the accuracy of proposed and existing approaches (a) RadioML2016.10a, (b) RadioML2016.10b.



(a) RadioML2016.10a (b) RadioML2016.10b

Figure 9. Performance of the proposed and existing optimizer's accuracy (a) RadioML2016.10a, (b) RadioML2016.10b.



(a) RadioML2016.10a (b) RadioML2016.10b

Figure 10. Performance of the model recognition accuracy after pruning (a) RadioML2016.10a, (b) RadioML2016.10b.

5. CONCLUSION

This work has made significant advancements in the field of automatic modulation recognition (AMR) within the context of wireless communication technologies. The Multi-Channel Gated Recurrent Deep Neural Network framework, MCGDNN, proved to be a pivotal contribution, enabling more effective signal classification across diverse datasets. By incorporating two specialized Deep Learning Networks (DLNs), we addressed the complexity of signal recognition, particularly in distinguishing In-phase Quadrature (IQ) and Frequency-Domain Amplitude-Phase (FDAP) signals. The optimization of the Bi-directional Gated Recurrent Unit (BiGRU3) network further enhanced feature extraction, leading to improved classification accuracy. The results presented in this study demonstrated the superiority of the proposed model across both RadioML2016.10a and RadioML2016.10b datasets. This model consistently outperformed alternative methods in terms of accuracy, F1 score, and recall. Moreover, this suggested model is mainly reduces the model's complexity and also correctly identify the signals. However, this model has a narrow applicability and also it does not give the better performance in real-world communication system. As future work, further research can explore the adaptability of the proposed model performs the more robust model to real-world communication systems and datasets.

ACKNOWLEDGEMENT AND FUNDING

The authors receive no financial support for the research, authorship, and publication of this article.

DECLARATION OF CONFLICTING INTERESTS

The authors declare no potential conflicts of interest with respect to the research and publication of this article.

REFERENCES

[1] S. Hou, Y. Dong, Y. Li, Q. Yan, M. Wang and S. Fang, Multi-domain-fusion deep learning for automatic modulation recognition in spatial cognitive radio, *Scientific Reports*, 13, 2023, 10736.

[2] M. Zhang, Y. Zeng, Z. Han and Y. Gong, Automatic modulation recognition using deep learning architectures, *Proceedings of the 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC),* Kalamata, Greece, 2018, 1-5.

[3] J. Shi, S. Hong, C. Cai, Y. Wang, H. Huang and G. Gui, Deep learning-based automatic modulation recognition method in the presence of phase offset, *IEEE Access*, 8, 2020, 42841-42847.

[4] C. Yang, Z. He, Y. Peng, Y. Wang and J. Yang, Deep learning aided method for automatic modulation recognition, *IEEE Access*, 7, 2019, 109063-109068.

[5] F. Zhang, C. Luo, J. Xu and Y. Luo, An efficient deep learning model for automatic modulation recognition based on parameter estimation and transformation, *IEEE Communications Letters*, 25(10), 2021, 3287-3290.

[6] J. Xu, C. Luo, G. Parr and Y. Luo, A Spatiotemporal multi-channel learning framework for automatic modulation recognition, *IEEE Wireless Communications Letters*, 9(10), 2020, 1629–1632.

[7] S. Cen, D. O. Kim and C. G. Lim, A fused CNN-LSTM model using FFT with application to real-time power quality disturbances recognition, *Energy Science & Engineering*, 11(7), 2023, 2267-2280.

[8] H. Bai, M. Huang and J. Yang, An efficient automatic modulation classification method based on the convolution adaptive noise reduction network, *ICT Express*, 9, 2023, 834-840.

[9] F. Zhang, C. Luo, J. Xu and Y. Luo, An Autoencoder-based I/Q channel interaction enhancement method for automatic modulation recognition, *IEEE Transactions on Vehicular Technology*, 72(7), 2023, 9620-9625.

[10] Q. Luo, M. M. Zhao, Z. Chen, Z. Su and M. J. Zhao, Complex-valued convolution and frequency global filter for automatic modulation recognition, *IEEE Communications Letters,* 27(7), 2023, 1779-14783.

[11] S. Ying, S. Huang, S. Chang, J. He and Z. Feng, AMSCN: A novel dual-task model for automatic modulation classification and specific emitter Identification, *Sensors*, 23(5), 2023, 2476.

[12] Y. Zeng, M. Zhang, F. Han, Y. Gong and J. Zhang, Spectrum analysis and convolutional neural network for automatic modulation recognition, *IEEE Wireless Communications Letters*, 8(3), 2019, 929-932.

[13] O. S. Mossad, M. ElNainay and M. Torki, Deep convolutional neural network with multi-task learning scheme for modulations recognition, *Proceedings of the 15th International Wireless Communications & Mobile Computing Conference (IWCMC)*, Morocco, 2019, 1644-1649.

[14] Y. Lin, Y. Tu, Z. Dou and Z. Wu, The application of deep learning in communication signal modulation recognition, *Proceedings of the IEEE/CIC International Conference on Communications in China (ICCC)*, Qingdao, China, 2017, 1-5.

[15] Z. Chen, H. Cui, J. Xiang, K. Qiu, L. Huang, S. Zheng, S. Chen, Q. Xuan and X. Yang, SigNet: A novel deep learning framework for radio signal classification, *IEEE Transactions on Cognitive Communications and Networking*, 8(2), 2021, 529-541.

[16] S. Hou, Y. Fan, B. Han, Y. Li and S. Fang, Signal modulation recognition algorithm based on improved spatiotemporal multi-channel network, *Electronics*, 12(2), 2023, 422.

[17] M. Wang, Y. Fan, S. Fang, T. Cui and D. Cheng, A joint automatic modulation classification scheme in spatial cognitive communication, *Sensors*, 22(17), 2022, 6500.

[18] R. Liang, X. Chang, P. Jia and C. Xu, Mine gas concentration forecasting model based on an optimized BiGRU network, *ACS Omega*, 5(44), 2020, 28579-28586.

[19] C. Y. Kang, C. P. Lee and K. M. Lim, Cryptocurrency price prediction with convolutional neural network and stacked gated recurrent unit, *Data*, 7(11), 2022, 149.
[20] J. Chang, Y. Lu, P. Xue, Y. Xu and Z. Wei, Automatic channel pruning via clustering and swarm intelligence optimization for CNN, *Applied Intelligence*, 52(15), 2022, 17751-17771.
[21] Y. Zhou, Y. Zhang, Y. Wang and Q. Tian, Accelerate CNN via recursive bayesian pruning, *Proceedings of the IEEE/CVF International Conference on Computer Vision*, Seoul, South Korea, 2019, 3306-3315.
[22] Y. Guan, N. Liu, P. Zhao, Z. Che, K. Bian, Y. Wang and J. Tang, Dais: Automatic channel pruning via differentiable annealing indicator search, *IEEE Transactions on Neural Networks and Learning Systems*, 34(12), 2022, 9847-9858.
[23] X. Dong and Y. Yang, Network pruning via transformable architecture search, *Advances in Neural Information Processing Systems*, 2019, 760–771.
[24] D. Hong, Z. Zhang and X. Xu, Automatic modulation classification using recurrent neural networks, *Proceedings of IEEE International Conference on Computer and Communications (ICCC)*, Chengdu, China, 2017.
[25] https://www.kaggle.com/datasets/gustavopolicarpo/rml201610a-dict, 2023 (accessed 27.09.2023).
[26] https://www.kaggle.com/datasets/marwanabudeeb/rml201610b, 2023 (accessed 27.09.2023).