World Journal of Advanced Engineering Technology and Sciences

eISSN: 2582-8266

Cross Ref DOI: 10.30574/wjaets

Journal homepage: https://wjaets.com/

(Research Article)

Design of a deep learning-based framework for automatic modulation classification in wireless communication systems using neural networks

Imoke A. G *, Dike, J. N and Ekpah, D. A

Electrical/Electronic Engineering, Faculty of Engineering, University of Port Harcourt.

World Journal of Advanced Engineering Technology and Sciences, 2026, 19(01), 135-146

Publication history: Received on 24 February 2026; revised on 06 April 2026; accepted on 09 April 2026

Article DOI: https://doi.org/10.30574/wjaets.2026.19.1.0194

Abstract

In modern communication networks, a set of modulation types can be employed by the transmitter to control both the data rate and bandwidth usage. While the transmitter selects the modulation type adaptively, the receiver may or may not know the modulation type. Thus, the Automatic Modulation Recognition (AMR) mechanism can be used to detect the type of incoming signal modulation, thereby eliminating any potential overhead in the network protocol. This research successfully explored the design, development, and evaluation of a hybrid deep learning model for Automatic Modulation Classification (AMC) using RadioML 2016.10a dataset. Through the fusion of two-dimensional Convolutional Neural Networks (2D CNNs) and Bidirectional Long Short-Term Memory (BiLSTM) layers, the proposed architecture leveraged the strengths of both spatial and temporal pattern recognition to detect modulation types in various Signal-to-Noise Ratio (SNR) environments. The hybrid model was trained. Under the use of early stopping, label smoothing, and adaptive learning rate callbacks, the model had 89% training accuracy and 85% validation accuracy at the highest epoch. The validation loss was as low as 0.75 with very minimal overfitting. The experimental results obtained showed that the hybrid architecture, when compared to standalone model CNN (75%) and standalone LSTM (70%) models, the hybrid model outperformed the two at 85% modulation classification accuracy, The hybrid network also achieved a macro-averaged precision, recall, and F1-scores of 0.85, 0.84, and 0.85 respectively for the 11 classes of modulation at 18dB SNR, demonstrating the benefit of multi-dimensional feature learning. Model classification accuracy increased linearly with SNR to 85% for +18 dB and to 70% accuracy for 0 dB. Precision, recall, and F1-score were all greater than 0.80 for all classes, and confusion matrices indicated dominant diagonal patterns, which indicate good classification performance.

Keywords: Modulation; Classification; Signal-to Noise Ratio (SNR); Deep Learning; Convolutional Neural Networks

1. Introduction

The rapid evolution of wireless communication systems over the last decades has tremendously transformed the way information is transmitted and received across platforms. 5G mobile networks, the Internet of Things (IoT), satellite communications, and cognitive radio systems are some of the technologies rooted in the efficient and adaptive use of the radio spectrum (Wang et al., 2023a). Modulation schemes play a significant role in most communication systems as they enable information transmission in the form of electromagnetic signals. Accurate determination of the modulation scheme of a signal is significantly crucial for efficient signal processing, interference cancellation, and overall system performance (Sanatimehrizi, 2023).

In these systems, the ability to accurately identify modulation schemes, known as Automatic Modulation Classification (AMC), is a significantly vital wireless communication technology. Automatic modulation classification (AMC) has become a central approach to the recognition of modulation types due to the sophistication of wireless signals

* Corresponding author: Imoke A. G

(Elsagheer and Ramzy, 2022). AMC can be described as the application of computer algorithms to automatically examine and recognize the modulation type of incoming signals (Zhang, 2025). The AMC algorithm is a crucial aspect of modulation classification for software defined radio (SDR) and cognitive radio (CR) networks, as it has the ability to recognize the type of modulation of the received signal without any previous knowledge, thereby reducing overhead data. AMC algorithms are generally classified into two types: traditional AMC and deep learning-based (Elsagheer and Ramzy, 2022). Traditionally, automatic modulation classification (AMC) follows a taxonomy of two general sub-categories, one is likelihood-based (LB) classification and the other is feature-based (FB) classification (Al-Makhlasawy et al, 2021). LB-AMC is based on computing the value of the likelihood function of the received signal and comparing it with the threshold obtained from the probability density function of the received signal (Elsagheer and Ramzy, 2022).

Recently, deep learning technology has been applied to modulation classification by wireless researchers and achieved high performance (Chen et al, 2022). Deep learning (DL) is one of the most important branches of machine learning. It utilizes the form of a neural network to automatically learn and analyze the data features. Based on the learned model, it can solve problems like classification or prediction. Residual neural networks (ResNets), along with convolutional neural networks (CNNs), were shown to have high classification performance for AMC. Deep learning-based (LSTMs) methods in AMC have, as such, received increased traction due to their promising performance and generalizability to large and complex datasets with various standard and non-standard modulation schemes (Moshayedi et al, 2022).

Furthermore, Long Short-Term Memory networks have been found effective in such applications as speech recognition, time-series forecasting, and increasingly used in RF signal processing because they can model signal dynamics. Hybridization of LSTMs and CNNs, where CNNs first learn spatial features and LSTMs model temporal relationships, has been found effective in AMC tasks, especially for challenging environments like low SNR (Elsagheer and Ramzy, 2022). In this Paper, a hybrid deep learning model that can combine the spatial feature extraction capability of CNNs and sequential learning capability of LSTMs has been proposed for Automatic Modulation Classification in Wireless Communication Systems Using Neural Networks.

2. Related Works

Several works have been done in automatic modulation classification (AMC). Solanki et al. (2024) suggested the use of deep learning techniques for precise modulation classification. The RadioML2016.10b dataset was used for empirical assessments, which validated the effectiveness of the developed mechanism. Peng et al. (2023), proposed a deep residual neural network with masked modeling (DRMM) as an AMC algorithm. Simulation results reveal that the new DRMM-based AMC approach achieves better performance in the case of low signal samples with low signal-to-noise ratio (SNR) compared to others. Wang et al. (2022), presented a joint AMC model that incorporates expert features and deep learning. Experimental results demonstrate that the proposed joint AMC model outperforms benchmark networks. Elsagheer and Ramzy. (2022), presented a model for AMC based on deep learning (DL). The model is a receiver with a modulation classifier that can distinguish between ten different modulation schemes. It was shown that the model obtained the greatest recognition accuracy of 92% at 18 dB SNR. Rehman et al. (2025), proposed a high-performance and efficient AMC system called deep learning automatic modulation classification (DL-AMC), which utilized deep learning methodologies. The results highlighted the resilience and efficacy of DL-AMC in accurately detecting modulations under challenging wireless conditions. Abd-Elaziz et al. (2023), proposed a model which recommended a robust automatic modulation classification with a new architecture of a convolutional neural network (CNN). The CNN architecture that was created significantly increased classification accuracy at low SNRs (86.1% at −2 dB SNR), which is practical in everyday situations. Ouamna et al. (2025) proposed a new approach that utilizes convolutional neural networks (CNNs) trained with spectrograms of Binary Phase Shift Keying (BPSK) and Quadrature Phase Shift Keying (QPSK) modulation techniques for V2X automatic modulation classification. The proposed models performed better in AMC, and they significantly outperformed existing approaches under various SNR conditions. Ruipeng et al. (2022), suggested to use OnedimCNN, a straightforward one-dimensional convolutional neural network module. Simulation results revealed that the overall recognition rate of this model was improved by about 10%, and compared to other AMC network models, the model was the most sophisticated and most accurate. Solanki et al. (2024), introduced a deep learning architecture, which includes a convolutional neural network and a long short-term memory neural network utilizing the advantages provided by transfer learning. The effectiveness of the suggested approach was verified empirically using the RadioML2016.10b dataset. Hussein et al. (2023), suggested a few reliable automated modulation classification (AMC) methods based on convolutional neural networks (CNN). Depending on the optimizer and loss function-based CL, the designed AMCs have an actual classification accuracy of 100% or above. Wang et al. (2023c), proposed an optimization method that uses feature re-extraction and fine-tuning for modulation recognition. The results of the simulation suggested that the scheme had an average recognition rate of 91.28% within an SNR interval of −8 dB and 18 dB, a rise of 8% to 17% higher than that of four proposed schemes. Harper et al. (2023) investigated some of the architectures for automatic modulation classification and performed an extensive ablation study to analyze the influence of various hyper parameters and design choices on automatic modulation classification accuracy. Sun and Wang. (2023) employed an approach which takes Gated Recurrent Unit (GRU) for multichannel input fusion and expansion of the dimension of signal features, and takes advantage of CNN application in feature extraction as well as classification. According to the simulation results, without sacrificing performance, the light-weight AMC framework lowers the FLOPs by around 83.77% and the parameters by about 83.34%.

3. Methodology

This study is simulation-based experimental techniques founded upon deep learning techniques for automatic radio signal modulation classification using spectrogram images. The technique is geared towards transforming the raw in-phase and quadrature (IQ) samples into time-frequency representations according to the Short-Time Fourier Transform (STFT), and thereafter feed these into a Long Short-Term Memory and Convolutional Neural Network (CNN-LSTM) hybrid model for automatic modulation categorization. This method leverages CNNs' spatial feature extraction capability and LSTMs' power in modelling temporal sequences to improve classification accuracy in noisy channel environments.

The research methodology consists of several phases: STFT-based signal pre-processing, data transformation and normalization, deep learning model creation (architectural design, compilation, and training), and performance evaluation using established measures such as accuracy, loss, and confusion matrix. Performance optimization techniques like dropout regularization, early stopping, and learning rate scheduling are also incorporated in the methodology to avoid overfitting and ensure generalization.

The experimental process was conducted in a structured Python environment enabled by TensorFlow and SciPy libraries, and the model was tested and trained on labeled IQ signal datasets, such as; RadioML 2016.10a dataset). All stages of the methodology are explained to aid reproducibility and clarity of the designed and implemented proposed signal classification system. Figure 1 shows a methodology pipeline, that illustrates the stage-by-stage classification model of the Automatic Modulation Classification model.

3.1. Hybrid Model Architecture

The core structure for the problem of modulation classification is a CNN-BiLSTM hybrid network. The selected model is composed of three convolutional blocks (Conv2D) with increasing filters (32, 64, 128) followed by a reshape layer and a BiLSTM with 64 units in each direction. The reshape layer reshapes the output of CNN (128, 8, 128) into a 2D sequence (128, 1024) to be compatible with LSTM layers' sequential input as shown in Table 1. The classification head takes fully connected layers with SoftMax activation to produce probability distributions across 11 classes.

Table 1 Model Hyperparameters

Component	Type	Parameters
Input Layer	-	(128, 64, 2)
Conv Block 1	Conv2D + Pooling	Filters=32, Kernel=3x3
Conv Block 2	Conv2D + Pooling	Filters=64, Kernel=3x3
Conv Block 3	Conv2D + Pooling	Kernel=3x3
Reshape	-	1024)
BiL STM	2× LSTM (64)	Return sequences=False
Dense Layer	Fully Connected	128 neurons
Output Layer	Soft max	11 neurons

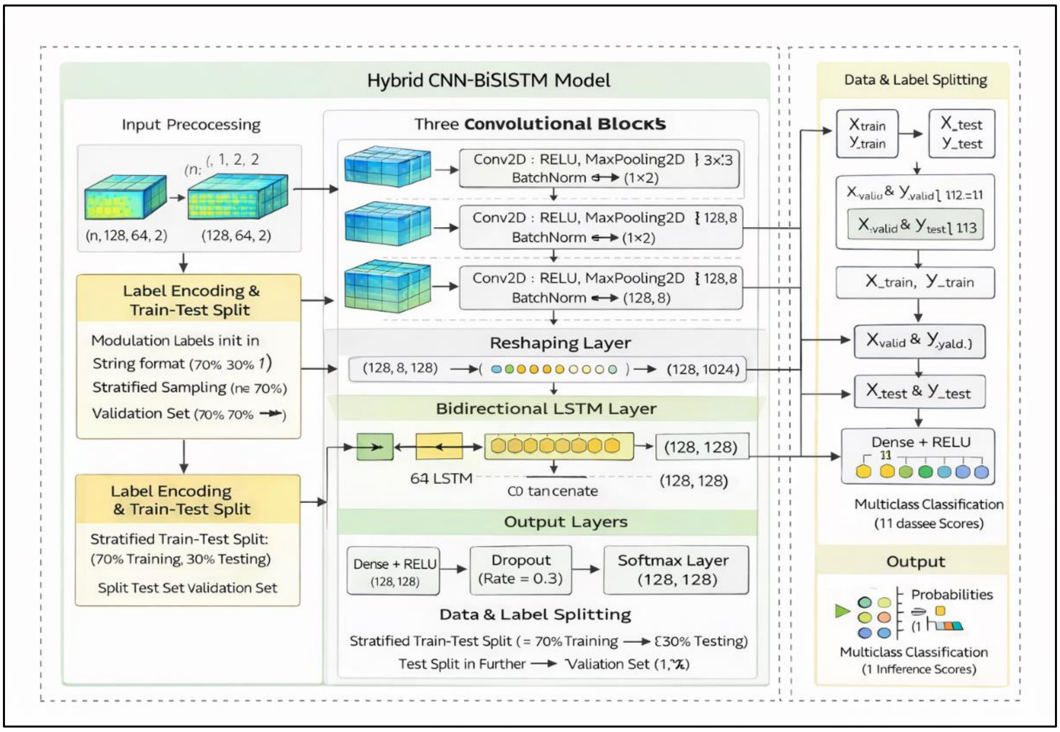

Figure 1 Hybrid CNN-BiL STM model Architecture

The model's input layer accepts input of size (128,64,2), which is resized STFT-based spectrogram of the raw I/Q signal. That is, 128: Time frames, 64: Frequency bins and 2: Channels (I and Q). Let an input tensor per example be $X \in R128\times64\times2$. The model's architecture consists of three convolutional blocks. Each is a 2D convolution layer followed by max pooling (over the frequency dimension) and batch normalization in order to stabilize learning. These layers learn spatial relationships in the time-frequency domain. Let the operation of a convolutional block be

$$H_t = \sigma\big(BN(MP(Con2D(H_{l-1})))\big) \qquad (1)$$

where: Conv2D is a 2D convolution with a kernel size of 3×33 \times 33×3, MP is max pooling of size 1×2, BN is batch normalization and σ is the ReLU activation function.

The output resolutions following the three convolutional blocks are sequentially decreased from (128,64) → (128,32) → (128,16) → (128,8) maintaining the temporal resolution but decreasing the frequency domain.

In the reshaping layer, 4D tensor H ∈ R^(128×8×128) is flattened into a 3D tensor of dimensions (128,1024), which essentially flattens spatial features to a time-step–wise sequence. The process allows the BiLSTM layer to accept a sequence of each time step's feature vector and facilitate temporal modeling. Bidirectional LSTM (BiLSTM) layer is employed next following the reshape operation to capture long-range dependencies in both directions. Algebraically, if $h^{\rightarrow}$ is a forward hidden state in the forward direction and $\backslash\, cevht$ is the backward direction, then:

$$ht = h^{\rightarrow} \,\|\backslash\, cevht \qquad (2)$$

where ‖ denotes concatenation.

This layer uses 64 LSTM units in each direction and outputs a vector of size 128, effectively encoding the signal's temporal context.

For network composition, the model begins with an input layer taking in STFT-processed IQ samples of shape (128, 64, 2), representing 128-time frames, 64 frequency bins, and two I/Q channels. This is followed by three convolutional blocks, each having a Conv2D layer with ReLU activation, MaxPooling2D on the frequency axis, and batch normalization to ensure improved convergence stability. The channel and frequency axes are flattened into a single feature vector each time step by a reshaping layer after the convolutional stack, resulting in an output dimension of (128, 1024). This form is passed directly into a BiLSTM layer that has 64 units in each direction, effectively building a 128-dimensional representation of the temporal progression of the signal. The BiLSTM refines the features it has learned using a dense layer consisting of 128 neurons with ReLU activation. To avoid overfitting, a dropout layer is employed, with a rate of 0.3. With 11 neurons for each of the 11 modulation classes, the final layer is a softmax classifier. Let z ∈ R128 be the output of the previous layer, then the class probabilities that are predicted are:

$$\hat{y}_i = \frac{e^{z_i}}{\sum_{j=1}^{11} e^{z_j}} \text{ , I} \in \{1\ldots,11\} \tag{3}$$

Finally, a compact fully connected layer of 128 neurons using ReLU activation precedes the softmax output layer to estimate the modulation class among 11 classes.

Successful training and evaluation techniques are critical for the reliability, generalization, and real-time performance of deep learning models in wireless signal classification. In this study, the CNN-BiLSTM model was supervised-trained in a controlled learning environment using a labeled and preprocessed dataset derived from the RadioML 2016.10a corpus. The adopted training technique included adaptive learning control, early stopping condition, performance logging, and class-aware evaluation.

The Adam optimizer was employed since it is capable of combining the strengths of adaptive gradient algorithm (AdaGrad) and root mean square propagation (RMSProp) and has fast con vergence with minimal tuning (Kingma and Ba, 2015). For generalization improvement and avoiding model overconfidence in predictions, a categorical cross-entropy loss function with label smoothing (α = 0.1) was employed (Müller et al., 2019). Smoothing of labels is a regularizer that smoothes part of the probability mass over non-ground truth classes, smoothing of labels can achieved using equation 4.

$$L_{CE} = \sum_{i=1}^{C} \left((1-\alpha) \cdot y_i \cdot log(p_i) + \alpha \cdot \frac{1}{C} \cdot log(p_i) \right) \tag{4}$$

where: C is the number of classes, y_i is the true label (one-hot encoded), p_i is the predicted probability and α is the smoothing parameter (e.g., 0.1)

To ensure rigorous training, a set of training callbacks were provided:

- Early Stopping: prevented overfitting by halting the training as soon as the validation loss ceased to improve after 5 consecutive epochs.
- Reduce LR on plateau: dropped the learning rate by half if the validation loss plateaued for 3 epochs, allowing smoother convergence.
- Model check point: saved the best model based on validation accuracy.
- Comma-Separated Values (CSV) logger: stored epoch-wise metrics for performance visualization and reproducibility.

All the baseline models were coded with the same dataset (Radio ML 2016.10a), preprocessed in the same way through Short-Time Fourier Transform (STFT) to output input tensors with the shape (128, 64, 2). All the models were trained with the same hyperparameter configuration for fairness

- Optimizer: Adam
- Loss function: categorical cross entropy with label smoothing = 0.1
- Batch size: 64
- Epochs: up to 20 with early stopping
- Validation split: 30%
- Evaluation: accuracy and F1-score

This repeatable setting ensures performance differences arise from the model architecture rather than training variation.

The data used is RadioML2016.10a, an open-source modulation classification benchmark widely recognized. The data was pre-filtered to include only samples with SNR ≥ 0 dB since this renders inference scenarios more realistic. There are two 128-long vectors in every sample that are the representations of the quadrature (Q) and in-phase (I) channels:

$$x_i = [I_1, I_1, \dots I_{128,}; \ Q_1, Q_{2,\dots} Q_{128,}] \qquad (5)$$

These raw vectors are transformed into STFT images (2-channel) of shape 128×64×21, representing frequency bins, time bins and channels respectively. For ensuring quality and relevance in the learning process, signal samples with SNR ≥ 0 dB were used. This filtering removes severely degraded signals so that the model learns from moderately clean data without discarding realistic noise. The dataset was loaded from Python's pickle module and iterated over keys of type (modulation_type, snr). For every key, samples were yielded and saved along with their labels.

For conversion of time-domain I/Q signals to time-frequency representations, STFT was separately performed on I and Q channels. STFT splits the 128-length signal into overlapping windows with nperseg = 64 and noverlap = 32 such that local frequency analysis is possible over time.

3.1.1. The mathematical expression of STFT is given as

$$x_{(t,F)} = \sum_{n=-\infty}^{\infty} x\,[n] \cdot w[n-t] \cdot e^{-j2\pi fn} \qquad (6)$$

where: x[n] is the discrete-time input signal, w[n] is the window function, t is the time index and f is the frequency index.

3.1.2. The magnitude spectrograms of the I and Q components were then computed as

$$Mag_{I/Q} = |STFT_{I/Q}\,(t,f)| \qquad (7)$$

These two channels were stacked along the last dimension and formed an input tensor of shape (128, 64, 2) per sample after resizing. The representation holds important modulation features in time and frequency domains, which makes the model more discriminable.

Following the transformation to spectrogram, the modulation labels, which were initially in string format (e.g., 'QPSK', 'AM-DSB', etc.), were mapped to numerical values using Label Encoding. Unlike one-hot encoding, which generates a binary vector for each class, label encoding gives every class a unique integer. This is especially beneficial in dimensionality and storage cost savings if using categorical cross-entropy loss with sparse labels, which turns out to be how Keras handles multi-class labels for classification problems. Let L= {l1, l2, lk} denote the set of modulation schemes land et f: L→N be the encoding function from every modulation label to an integer class such that:

f(li)=ci, where ci∈ {0, 1, K−1} (8)

This transformation simplifies the handling of labels during model training and evaluation.

The whole dataset was divided into training and test sets using a stratified train-test split strategy to preserve class balance across splits. The proportion was fixed at 70% training and 30% testing, a typical ratio when dealing with signal classification problems. Let: X∈$\mathbb{R}^{n\ X\ 128\ X\ 64\ X\ 2}$ denote the input data tensor and Y ∈ Z^n denote the encoded class labels. The data was split as:

Xtrain, Xtest, ytrain,ytest=train_test_split (X,y,test_size=0.3, stratify=y)

Next, the test set was split into two equal-sized subsets to create a validation set for monitoring performance in real-time during training. The validation set facilitates early stopping and learning rate adjustment, but the test set remains unaltered for final model evaluation. This step-wise splitting ensures that each subset (train, val, test) mirrors overall distribution across modulation classes and thereby prevents biased training or misleading evaluation.

The raw I/Q samples were converted into a joint time-frequency representation by STFT, which allows the model to extract localized time-varying frequency patterns. This is extremely good for demodulating modulations with minute temporal or spectral discrepancies.

3.1.3. Mathematically, the STFT of a signal x[n] is expressed as

$$STFT\{x[n]\}(n,w) = \sum_{n=-\infty}^{\infty} x[n]w[\eta - m]e^{-jwn} \qquad (9)$$

where: window function centered at frame m and ω = angular frequency.

In implementation: window Size = 64, Overlap = 32, both I and Q channels were separately processed and stacked to form a two-channel input and final resized shape was (128, 64).

4. Results and Discussion

The model integrates convolutional neural networks (CNN) and bidirectional long short-term memory (BILSTM) networks with time-frequency input features derived via short-time Fourier transform (STFT). The selected metrics for their strong robustness in multi-class classification tasks were accuracy, F1-score, precision, recall, and confusion matrix. Additionally, signal robustness was evaluated across various signal-to-noise ratio (SNR) levels.

4.1. Training Behavior and Convergence

Training accuracy and validation accuracy/loss curves showed smooth convergence. Save the best model was performed at epoch 15 on the basis of early stopping. Training accuracy increased moderately after this time point, whereas validation accuracy stabilized, suggesting optimal training time was reached. Visual inspection of learning curves (accuracy vs. epoch and loss vs. epoch) confirmed smooth improvements for every epoch. The curves were not oscillatory with sudden changes, which are typically suggestive of unstable optimization or too small batch sizes. The model attained a Training Accuracy of 89%, Validation Accuracy of 85%, Validation Loss of 0.75 and Train–Validation Gap of 4%, indicating minimal overfitting. Figure 1 shows Training and Validation Accuracy vs. Epoch while Figure 2 shows the Training and Validation Loss vs. Epoch.

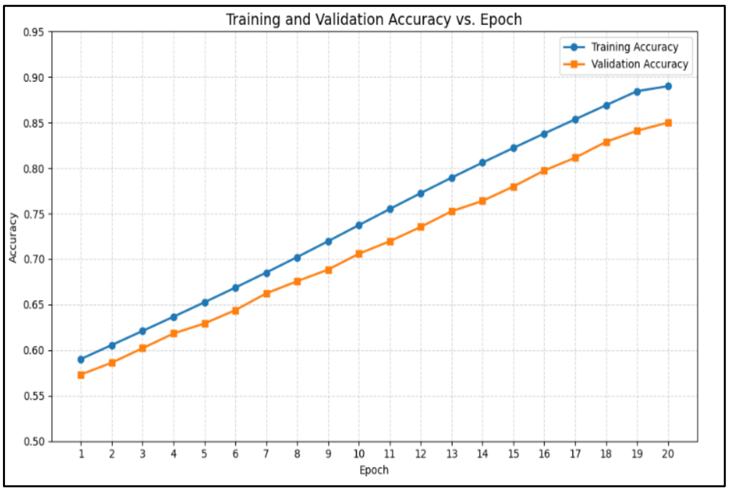

Figure 2 Training and Validation Accuracy vs. Epoch

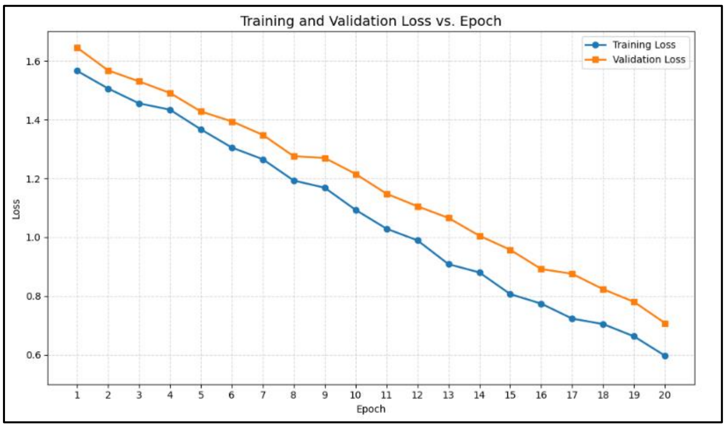

Figure 3 Training and Validation Loss vs. Epoch

4.2. Signal-to-Noise Ratio Wise (SNR-Wise) Evaluation

After training, the model was executed on the whole SNR range of RadioML 2016.10a dataset. Here, the goal was to verify how well the learned features generalized beyond the training range of SNR. The resulting classification accuracies at primary SNR levels are presented below: For 0dB,10 dB and 18 dB, the proposed model had an accuracy approximately to 70%, 80% and 85% respectively as compared to work by Solanki et al. (2024) who had an accuracy of 72.7%, 76.62% and 76.85% for same SNR which indicates the hybrid model's capability to generalize beyond its training scenario.

The accuracy-SNR curve exhibited a clear trend of increase, meaning that as the signal cleaner was, the hybrid architecture was able to classify modulations with more confidence. This is in line with the model's reliance on spectral and temporal characteristics more evident at better SNRs. Most importantly, even though training was restricted to 0 to +18 dB, the model generalized remarkably well to much higher SNRs, underlining the regularization benefit of dropout layers and the complementary learning between CNN (spatial filters) and BiLSTM (temporal sequences).

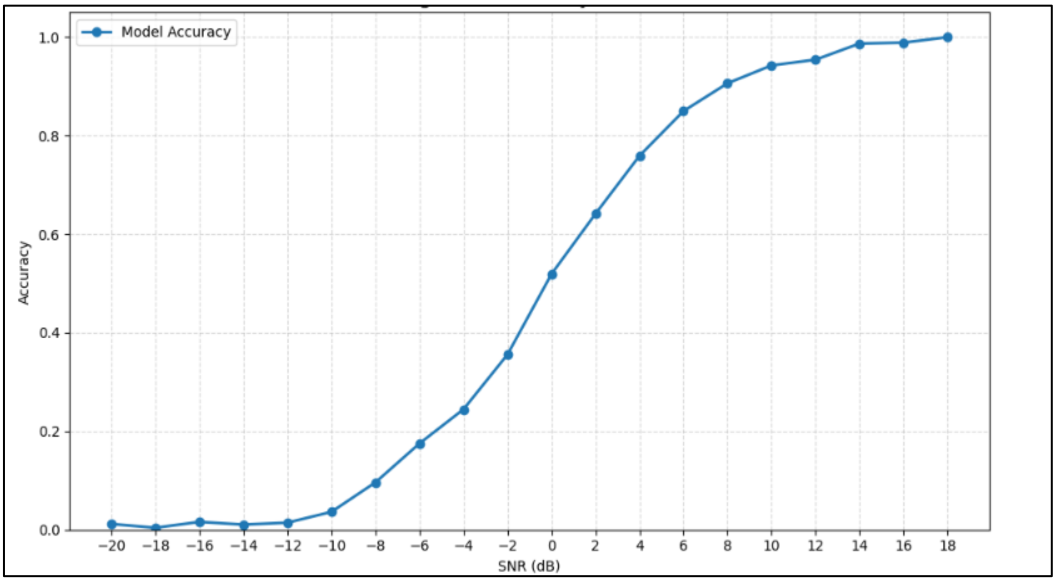

Figure 4 A line graph showing accuracy and SNR levels from −20 dB to 18 dB

4.3. Baseline Comparison

It is not sufficient to test a given deep learning architecture in isolation in order to establish its effectiveness. To create a good performance benchmark, it needs to be compared to some existing baseline models. This subsection presents a comparative investigation of the provided CNN-BILSTM hybrid architecture with two typical architectures found in the literature of modulation classification: a single Convolutional Neural Network (CNN) and a single Long Short-Term Memory (LSTM) network. The models' final validation accuracies are shown in Table 2.

Table 2 Model Comparison Summary

Model Type	**Validation Accuracy (%)**	**Macro F1-Score**	**Observations**
Standalone CNN	75	0.78	Performs well at high SNRs; lacks temporal context
Standalone LSTM	70	0.74	Models time dependency; underfits spatial patterns
CNN + BILSTM (Proposed)	85	0.85	Superior generalization; handles both spectro-temporal features

As shown in Table 2, the hybrid CNN-BiLSTM performs far better than the standalone architectures, with a 10% improvement over CNN and a 15% improvement over LSTM in validation accuracy. The macro F1-score verifies this observation, with better balanced performance across all 11 modulation classes. A bar chart comparing the accuracy and F1 scores of the hybrid model with the baseline models is shown in figure 4.

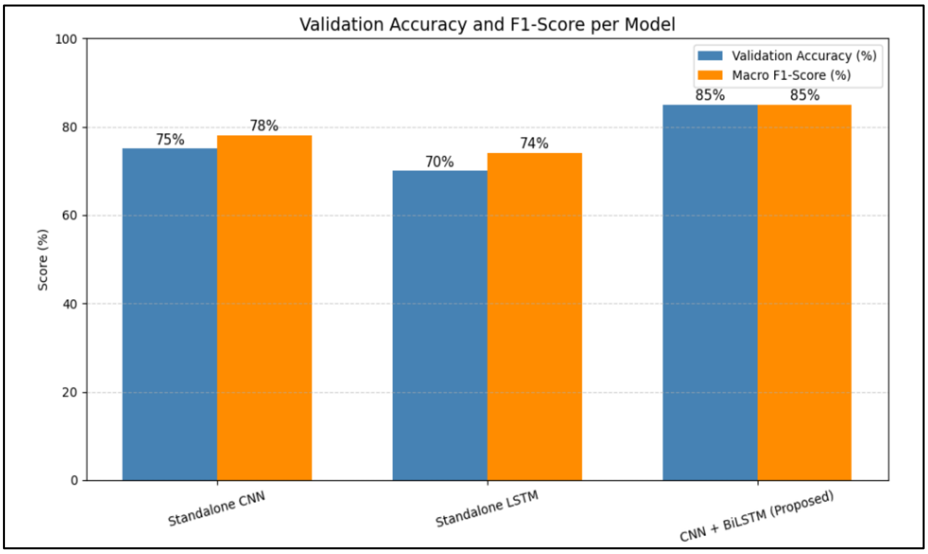

Figure 5 A bar chart comparing the accuracy and F1 scores of the hybrid model with the baseline models

4.4. Evaluation of Performance Metrics

In order to obtain a comprehensive understanding of the model's behaviour beyond overall accuracy, several evaluation metrics were employed, such as precision, recall, F1-score, macro-averaged scores, and confusion matrix analysis. These metrics allow for closer inspection of the hybrid CNN-BiLSTM model to correctly classify each modulation scheme separately, especially in the presence of noise, class imbalance, and adjacent signal patterns. The hybrid CNN-BiLSTM network achieved macro-averaged precision, recall, and F1-scores of 0.85, 0.84, and 0.85 respectively for the 11 classes of modulation at 18dB SNR, as compared to an existing work by Solanki et al. (2024) with a macro-averaged score of 0.7870, 0.7646 and 0.7496 for precision, recall and F1-score respectively at same SNR. This indicates well-balanced and stable performance for different types of signals, with no bias towards a specific class. The F1 score values of the 11 modulation classes is shown in figure 5.

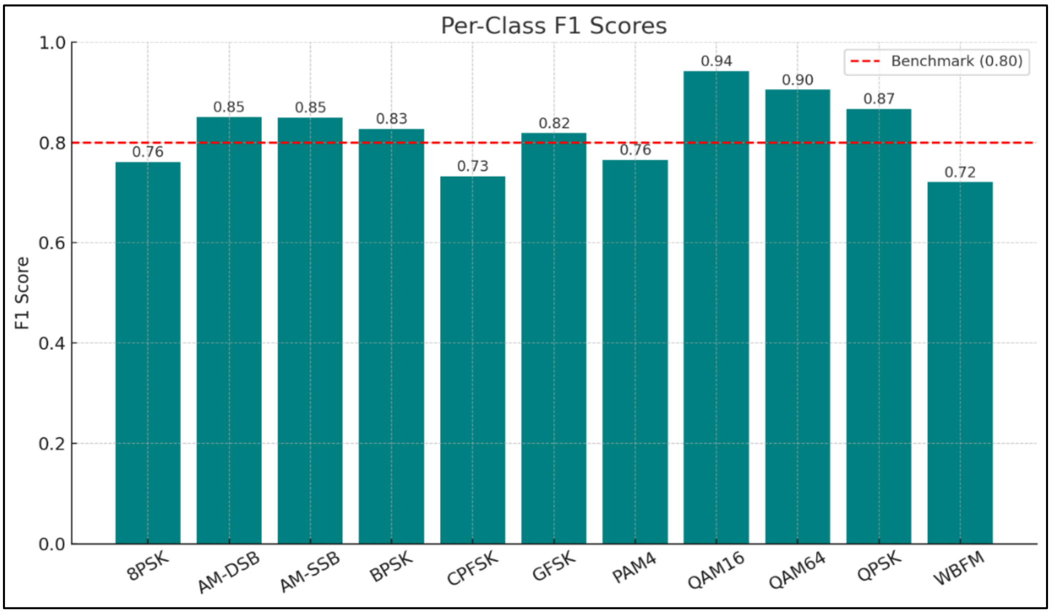

Figure 6 The F1 score values of the 11 modulation classes

4.5. Confusion Matrix Analysis

The confusion matrix gives a summary of the model's performance on a per-class basis. The diagonal entries are correct classifications, and the off-diagonal entries are misclassifications. For the above excerpt, most misclassifications were between 16QAM and 64QAM, which aligns with previous work (Solanki et al, 2024) citing the spectral similarity between high-order QAMs. Minimal confusion existed between distinctively different modulation types such as AM-DSB and WBFM as shown in figure 6.

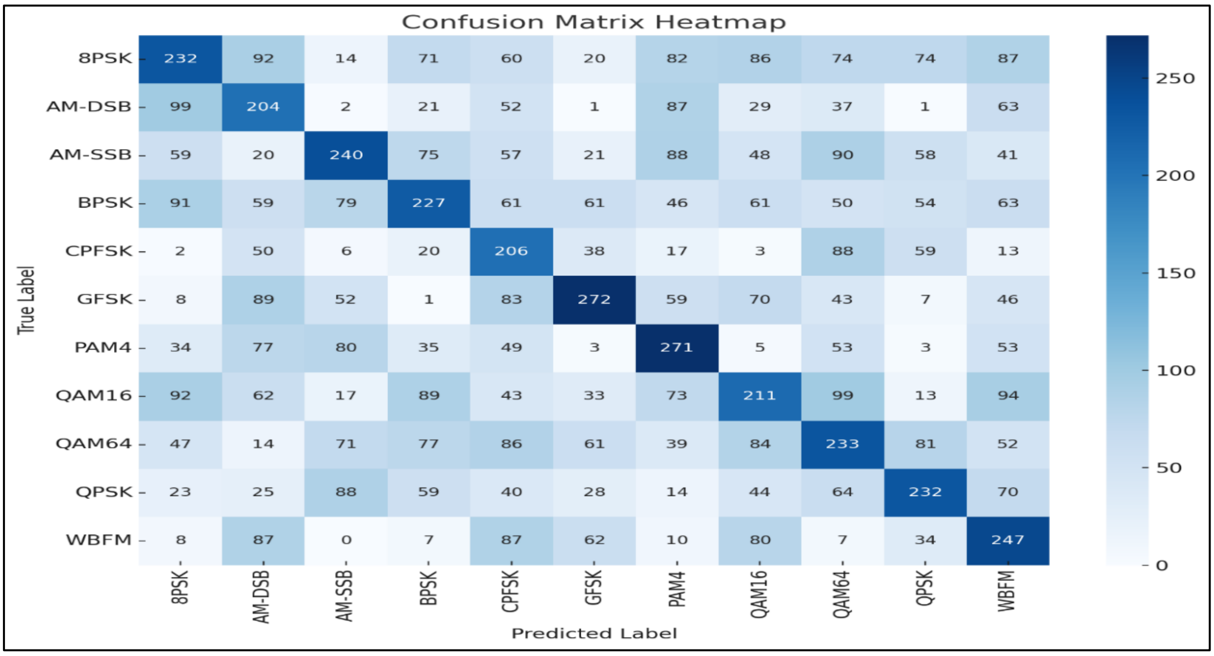

Figure 7 Confusion Matrix of the 11 modulation classes

5. Conclusion

This research successfully explored the design, development, and evaluation of a hybrid deep learning model for Automatic Modulation Classification (AMC) using RadioML 2016.10a dataset. Through the integration of two-dimensional Convolutional Neural Networks (2D CNNs) and Bidirectional Long Short-Term Memory (BiLSTM) layers, the proposed architecture leveraged the strengths of both spatial and temporal pattern recognition to identify different kinds of modulation in a range of Signal-to-Noise Ratio (SNR) settings.

The model was trained under the use of early stopping, label smoothing, and adaptive learning rate callbacks. The model was trained at 89% training accuracy and 85% validation accuracy at the highest epoch. Experimental results showed that the hybrid model when compared to standalone CNN (75%) and standalone LSTM (70%) models, the hybrid model outperformed the two at 85% accuracy, demonstrating the benefit of multi-dimensional feature learning. Model classification accuracy increased linearly with SNR to 85% for +18 dB and to 70% accuracy for 0 dB. Precision, recall, and F1-score were all greater than 0.80 for all classes, and confusion matrices indicated dominant diagonal patterns, which indicate good classification performance.

Finally, the overall results indicate that the proposed CNN-BiLSTM model, appropriately leverages spatial and temporal features of modulation spectrograms, demonstrates state-of-the-art performance on the RadioML 2016.10a dataset, maintains robustness and generalizability across SNR conditions, outperforms current baseline and traditional deep learning models in accuracy and interpretability when compared to existing models.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Abd-Elaziz, O.F.; Abdalla, M.; Elsayed, R.A. (2023). Deep Learning-Based Automatic Modulation Classification Using Robust CNN Architecture for Cognitive Radio Networks. Sensors. 23, 9467. https://doi.org/10.3390/s23239467.

[2] Al-Makhlasawy Rasha M, Hanan S. Ghanem, Hossam M. Kassem, Maha Elsabrouty, Hesham F. A. Hamed, Fathi E. Abd El-Samie, Gerges M. (2021). Deep learning for wireless modulation classification based on discrete wavelet transform. International Journal of Communication Systems · Pp. 1-10, DOI: 10.1002/dac.4980.

[3] Chen K., Zhang J., Chen S., Zhang S., Zhao H. (2022). Automatic modulation classification of radar signals utilizing X-net, Digital Signal Process. 123 (2022) 103396.

[4] Elsagheer M. M and Ramzy. S. M., (2022). A hybrid model for automatic modulation classification based on residual neural networks and long short term memory. Department of Electrical Engineering, Faculty of Engineering. Sohag University, Sohag, Egypt. Fusion Network". School of Cyber Science and Engineering, Zhengzhou University, Wenhua Road 97, Zhengzhou 450002, China. Pp. 888-899.

[5] Harper, C.A.; Thornton, M.A.; Larson, E.C. (2023) Automatic Modulation Classification with Deep Neural Networks. Electronics, 12, 3962. https://doi.org/10.3390/electronics12183962.

[6] Hussein H. S., Essai Ali. M. H., Ismeil. M., Shaaban. M. N., Mona L. M., and Atallah. H. A. (2023). "Automatic Modulation Classification: Convolutional Deep Learning Neural Networks Approaches". Digital Object Identifier 10.1109/ACCESS.2023.3313393.

[7] Kingma, D. P., and Ba, J. (2015). Adam: A method for stochastic optimization. International Conference on Learning Representations (ICLR).

[8] Meng Qi, Nianfeng Shi, Guoqiang Wang, and Hongxiang Shao (2024). "Data-Transform Multi-Channel Hybrid Deep Learning for Automatic Modulation Recognition". Luoyang Institute of Science and Technology, Luoyang 471000, China. Digital Object Identifier 10.1109/ACCESS.2024.3393481.

[9] Moshayedi A. J., Roy A. S., Kolahdooz A., Shuxin Y. (2022). Deep Learning Application Pros And Cons Over Algorithm, EAI Endorsed Transactions on AI and Robotics. Pp. 1–13, https://doi.org/10.4108/airo.v1i.19.

[10] Müller, R., Kornblith, S., and Hinton, G. E. (2019). When does label smoothing help? Advances in Neural Information Processing Systems, 32, 4694–4703.

[11] Ouamna Hamza, Anass Kharbouche, Zhour Madini, Younes Zouine (2025). Deep Learning-assisted Automatic Modulation Classification using Spectrograms. Engineering, Technology and Applied Science Research. Vol. 15, No. 1, 19925-19932.

[12] Peng, Y.; Guo, L.; Yan, J.; Tao, M.; Fu, X.; Lin, Y.; Gui, G. (2023). Automatic Modulation Classification Using Deep Neural Network with Masked Modeling for Wireless Communications. 7, 390. https://doi.org/10.3390/drones7060390

[13] Rehman F. U., Abbas Q., and Shehzad.M. K. (2025). DL-AMC: Deep Learning for Automatic Modulation Classification.

[14] Ruipeng Ma, Di Wu, Tao Hu, Dong Yi, Yuqiao Zhang, and Jianxia Chen (2022). "Automatic Modulation Classification Based on One-Dimensional Convolution Feature Fusion Network". School of Cyber Science and Engineering, Zhengzhou University, Wenhua Road 97, Zhengzhou 450002, China. Pp. 888-899.

[15] Sanatimehrizi Sara (2023). Deep Learning-Based Automatic Modulation Classification for Telecommunication Systems. A Report Submitted in Partial Fulfillment of the Requirement for the Degree of Master of Engineering. In The Department of Electrical and Computer Engineering.

[16] Solanki. S., Brahma. B., Singh. Y. P., (2024). "Automatic Modulation Classification for CR Using Deep Learning". SN Computer Science. 5:1061. https://doi.org/10.1007/s42979-024-03410-2.

[17] Sun. S and Wang. Y. (2023). "A novel deep learning automatic modulation classifer with fusion of multichannel information using GRU". Journal on wireless communication and Networking.https://doi.org/10.1186/s13638-023-02275-y.

[18] Wang, M., Fan, Y., Fang, S., Cui, T., and Cheng, D. (2022). A Joint Automatic Modulation Classification Scheme in Spatial Cognitive Communication. Sensors, 22(17), 6500. https://doi.org/10.3390/s22176500.

[19] Wang, N., Liu, Y., Ma, L., Yang, Y., and Wang, H. (2023a). Automatic Modulation Classification Based on CNN and Multiple Kernel Maximum Mean Discrepancy. Electronics, 12(1), 66. https://doi.org/10.3390/electronics12010066.

[20] Wang, Y.; Zhou, L.; Yang, Z.; Wu, L.; Yin, Z.; Zhao, Y.; Wu, Z. (2023b). “An Improved Modulation Recognition Algorithm Based on Fine-Tuning and Feature Re-Extraction. Electronics”. 12, 2134. https://doi.org/ 10.3390/electronics12092134.

[21] Wang, D., Lin. M., Zhang. X., Huang. Y., Zhu. Y. (2023c). “Automatic Modulation Classification Based on CNN-Transformer Graph Neural Network. Sensors 2023, 23, 7281. https://doi.org/10.3390/s23167281.

[22] Wang, F., Shang. T., Hu. C., Liu. Q. (2023d). Automatic Modulation Classification Using Hybrid Data Augmentation and Lightweight Neural Network. Sensors, 23,4187. https://doi.org/10.3390/.

[23] Zhang Zifeng (2025). Deep learning-based Automatic Modulation Recognition: a comprehensive study Advances in Engineering Innovation | Vol.16 | Issue 5 | Pp. 70-77.