First Libyan International Conference on Engineering Sciences & Applications (FLICESA_LA)
13 – 15 March 2023, Tripoli – Libya

Automatic Modulation Classification Using Convolutional Neural Networks with Feature Fusion Method

Mohamed Salem Elshebani
Department of Comm Eng.
Faculty Electrical Technology
Tripoli, Libya
Eldhebani@Libya.tripoli.mallinum

Yahya Ali
Department of Comm Eng.
university of Tripoli
Tripoli, Libya
Y.ali@uot.edu.ly

Nser Azroug
Department of Comm Eng.
Tripoli College for Sciences and Technology
Tripoli, Libya
maktoof73@gmail.com

Ramdan A. M. Khalifa
Department of Comm Eng.
High Institute of science and Technology-Suk Algumaa
Tripoli, Libya
ramadanamharee@gmail.com

***Abstract*— Automatic modulation classification (AMC) detects the modulation type of received signals to guarantee that the signals can be correctly demodulated and that the transmitted message can be accurately recovered. In DL-based modulation classification, one major challenge is to pre-process a received signal and represent it in a proper format before passing the signal into the neural network. However, most existing modulation classification algorithms neglect the concept of mixing features between different representations and the importance of features fusion. This paper attempts a Feature fusion scheme for AMC using convolutional neural networks (CNN). The approach attempts to fuse features extracted from In-phase & Quadrature (IQ) sequences, Amplitude & Phase (AP) Sequences and Constellation Diagram images. Finally, simulation results show that fusing features from different representations can incorporate the best accuracy achieved from each representation separately. Furthermore, our model achieves a classification accuracy of 84.68% at 0dB and 75.29% at -2 dB and over 90% accuracy for high SNRs with a maximum accuracy of 94.65%.**

Keywords— Modulation classification, deep learning, fusion, convolutional neural network, residual network, wireless communications, cognitive radio.

I. INTRODUCTION

Wireless communication plays an important role in modern communication. Modulation classification, as an intermediate process between signal detection and demodulation, is therefore attracting attention. Modulation recognition finds application in commercial areas such as space communication and cellular telecommunication in the form of Software Defined Radios (SDR). SDR uses blind modulation recognition schemes to reconfigure the system, reducing the overhead by increasing transmission efficiency. Furthermore, AMC serves an important role in the information context of a military field. The spectrum of transmitted signals spans a large range and the format of the modulation algorithm varies according to the carrier frequency. The detector needs to distinguish the source, property and content correctly to make the right processing decision without much prior information. Under such conditions, advanced automatic signal processing and demodulation techniques are required as a major task of intelligent communication systems. With the significant achievement of artificial intelligence, DL has aroused widespread concern and has been widely applied in various fields for its excellent data processing capability. Recently, many researchers applied DL to resolve AMC problems. One of the pioneers in the AMC field is O'Shea et al. [1], who proved that the performance of CNN trained on the baseband IQ data exceed those of the expert cyclic moment features based methods. Mend is et al. [2] proposed the AMC approach based on DL. However, they only considered a kind of time frequency distribution and neglected to combine with others to further improve performance. Similarly, Peng et al. [3] investigated an efficient AMC system that utilized CNN to learn features from constellation diagram of digital communication signals. Nevertheless, the method suffered from performance degradation in low SNR environments. A common limitation of the above-mentioned work is that they ignore the complementarities between different features. In [4], Fusion methods are explained and shows that feature-based fusion gives the best results, this paper proposes a 4-Layer CNN architecture with feature fusion utilizing three different representations of the received signal to further improve classification accuracy.

II. SIGNAL MODEL AND RADIOML DATASET

A. Problem Model

Modulation classification can be expressed as a classification problem with M modulations. The received signal can be expressed as:

$$r(t) = \alpha(t)e^{j(2\pi f_0 t + \theta_0(t))}s(t) + n(t) \qquad (1)$$

Where s(t) is the complex baseband envelope of the transmitted signal, α(t) is the impulse response of the transmitted wireless channel,$\theta_0(t)$ and f_0 are the carrier phase and frequency offsets due to disparate local oscillator and

Doppler effect caused by motion, *n(t)* is the additive white Gaussian noise (AWGN).

At the receiving side, the received signal, r(t), is first amplified, mixed, low-pass filtered and then sent to the analog to digital (A/D) converter, which samples the continuous-time signal at a rate $f_s = 1/T_s$ samples per second and generates the discrete version *r*[n]. In this paper the received signal is first represented in its IQ components. The received discrete signal r[n] is given as:

$$r[n] = r_I[n] + r_Q[n] \quad (2)$$

Where r_I[n] is the In-phase component, and r_Q[n] is the quadrature component.

The second representation is AP components, where r[n] is given as:

$$r[n] = r_A[n] \angle r_\phi[n] \quad (3)$$

Where rA[n] is the magnitude component and rϕ[n] is the phase component, We calculated $r_A[n]$, $r_\phi[n]$ from the IQ components as given below:

$$r_A[n] = \sqrt{r_i[n]^2 + r_Q[n]^2} \quad (4)$$

$$r_\emptyset[n] = \arctan\left(\frac{r_Q[n]}{r_i[n]}\right) \quad (5)$$

the aim of any modulation classifier is to give out *Pr* (s(t) ∈ M(i) | r(t)) given the received signal r(t), where *M(i)* represents the *i-th* modulation.

B. RADIOML 2016.10A Dataset

The RADIOML 2016.10A dataset [4] is one of 3 available datasets that are provided by DeepSig Inc, the dataset contains 11 different modulation schemes ranging across SNRs from -20dB to 18dB, with each data sample being an IQ time-series with 128 time-steps, represented as a 2x128 array. Realistic channel imperfections such as moderate local oscillator (LO) drift,light multipath fading and AWGN are included in the dataset that were generated by GNU radio. The dataset contains 8 digital modulation schemes (BPSK, QPSK, 8PSK, PAM4, GFSK, CPFSK, 16QAM, QAM64) and 3 Analog modulation schemes (WBFM, AM-DSB, AM-SSB), we will only focus on the 8 digital modulation schemes because it is almost impossible to differentiate between analog modulations because of pauses in the voice recording of the source dataset that analog modulations were generated from, the reason for using the RADIOML 2016.10A dataset is because of some flaws in the other 2 datasets, The description of each dataset is presented in Table 1.

Table 1: Summary of available RADIOML datasets

	2016.04C	2016.10A	2018.01A
Number of Modulations	11		24
Modulation Types	Digital: BPSK, QPSK, 8PSK,QAM16, QAM64, CPFSK,GFSK,PAM4. Analog:WBFM, AM-SSB,AM-DSB.		Digital: OOK, 4ASK, 8ASK, BPSK,QPSK, 8PSK, 16PSK, 32PSK,16APSK, 32APSK, 64APSK,128APSK, 16QAM, 32QAM,64QAM, 128QAM, 256QAM,GMSK, OQPSK. Analog: AM-SSB-WC, AM-SSB-SC, AM-DSB-WC, AM-DSB-SC, FM.
SNR Range	-20dB to 18dB		

To have a better understanding of the dataset we will try visualizing it in each representation utilized in this project. As mentioned before, the dataset is originally represented in IQ sequences, Fig.1 displays the dataset visualization in its raw IQ format in the time domain.

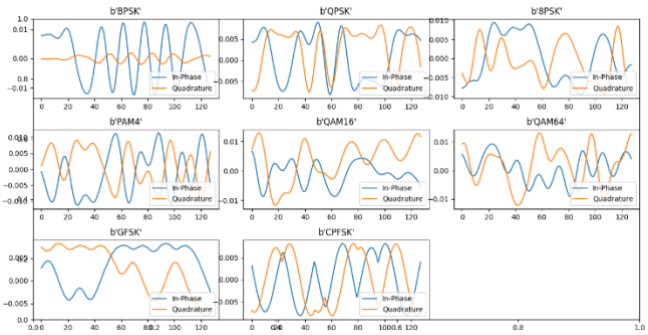


Figure 1: IQ signals time plot of RadioML2016.10A dataset at SNR=18dB.

For the second representation, we transformed the dataset to AP sequences using equations 4 & 5 and then the data was normalized. Fig..2 shows the dataset in its AP Representation in the time domain.

Finally, our last representation was constellation diagram images, Conversion to constellation image involved dividing the I and Q axes (in the desired region of the complex plane) into 48 bins and counting the number of points that fall into each bin, and then normalizing the counts to a range of -1 to 1. The constellation images investigated were only gray colored due to computational limitations. Fig..3 shows the dataset converted to constellation diagram images.

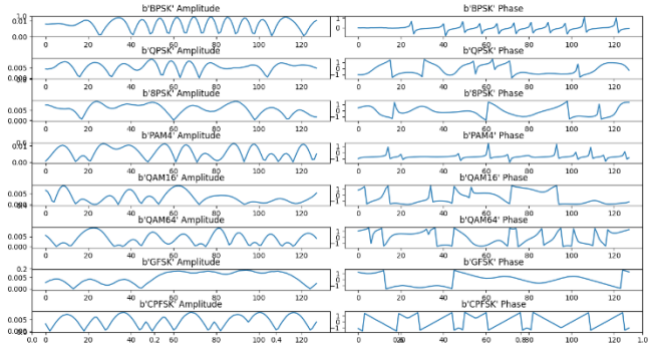


Figure 2: AP signals time plot of RadioML 2016.10A dataset at SNR=18dB

Figure .3: Constellation Images of Radio ML 2016.10A dataset at different SNRs.

III. IMPLEMENTED ARCHITECTURES

In this paper , four architectures were built, three of them were built for each of the three representations that were utilized in this project, and for the fourth architecture it is a concatenation of the previous architectures which applies the feature fusion method, all of the architectures are presented below.

A. CNN-IQ Model

This model was built for the IQ sequences, it contains 4 convolutional layers with activation function ReLu, and each convolutional layer is followed by a maxpooling layer for size reduction and a batch normalization layer to normalize the output, and a dropout layer to avoid overfitting. Furthermore, a flatten layer is used followed by two dense layers with ReLu activation and finally a dense layer with SoftMax activation for classification. Fig.4 shows the model architecture with the number of filters and each filter size in the convolutional layers and the number of neurons in the dense layers.

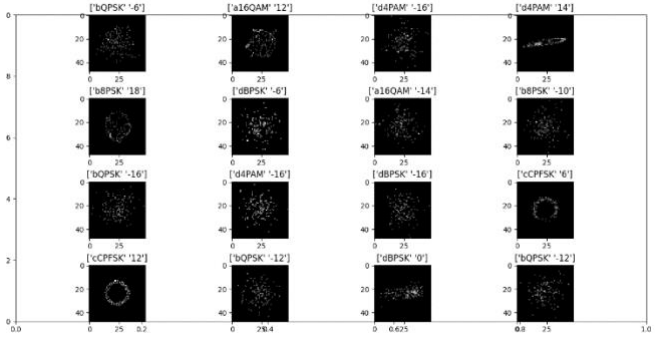
Figure 4: CNN-IQ Model Architecture.

B. CNN-AP Model

This model was built for the AP sequences, it is based on the same architecture as the CNN-IQ model, the difference is in the number of neurons in the dense layers and also a dropout layer is inserted between the two layers. Fig.5 shows the model architecture with the details of each layer.

C. CNN-CD Model

This model was built for the Constellation diagram images, it is based on the same architecture as the CNN-IQ model, the difference is in the input dimensions, the filter



sizes of the convolutional layer and the number of neurons in the second dense layer. Fig.6 shows the model architecture with the details of each layer.

Figure 5: CNN-AP Architecture.

Figure 6: CNN-CD Model Model Architecture.

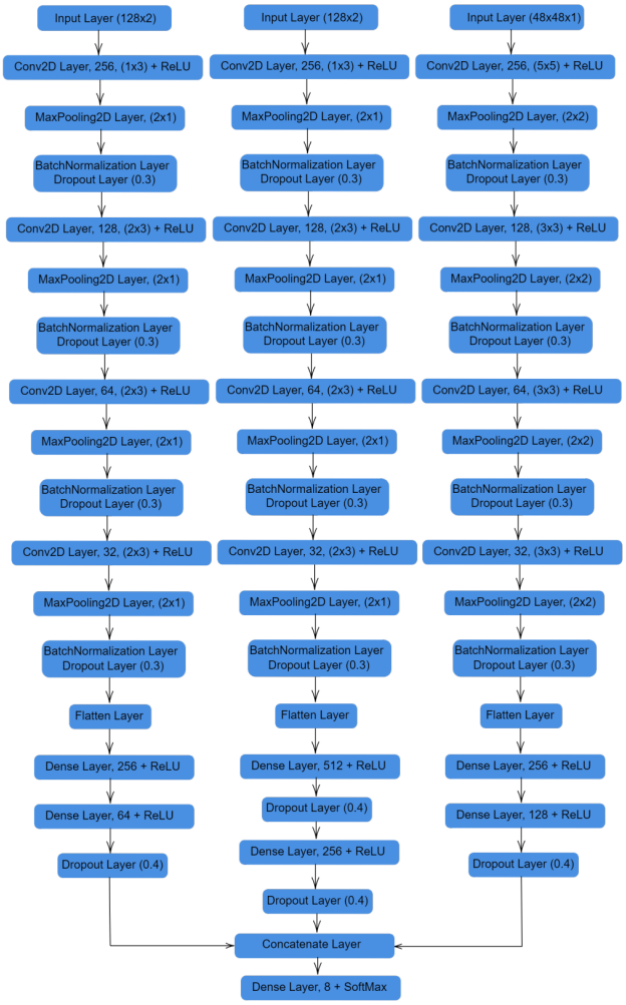


Figure 7: CNN-(IQ/AP/CD) Model Architecture.

accuracy is plotted against various SNRs. The Training was carried out using Microsoft Azure and evaluation was carried out using Visual Studio Code environment. The code was written in python language.

D. CNN-(IQ/AP/CD) Model

This model utilizes the feature-fusion based method, where each of the previous models is used for feature extraction and then all three models are concatenated and then a final dense layer with softmax activation is used. Fig.7 shows the model architecture with the details of each layer.

IV. SIMULATION RESULTS

Simulation results are presented to illustrate the classification accuracy of each model. First, the training phase is analyzed and then, the confusion matrix for each trained model is evaluated and finally, the classification

A. Neural Network Training

For the training phase, each of the used datasets were split into 80% training dataset, 10% test dataset, 10% validation dataset. The number of epochs used was 100 but in CNN-IQ and CNN-(IQ/AP/CD) models we stopped the training before 100 because of no further improvement in validation accuracy. The batch size used was 50 for all models. We also utilized the checkpoint function which saves the best model weights each time it improves in validation accuracy and after the end of the training we use the best model weights for evaluation. The longest training time of all models was three hours for CNN-(IQ/AP/CD) model because it was the most complex model, other models took around one and a half hour. During training we did not focus on the training accuracy because after a number of epochs the model just starts to memorize the dataset, instead, we focused on the validation accuracy because it provided the generalization of the model which means the higher the validation accuracy increases. The better performance we acquire. In Fig.8, a plot

of the validation accuracy of each model during the training phase is shown, it can be seen that CNN-(IQ/AP/CD) has the highest validation accuracy, and on the contrary, the CNN-CD has the lowest validation accuracy. We can also observe that CNN-IQ is the noisiest in validation
amongst all models. It should also be noted that the validation accuracy signifies the average classification accuracy of the model over all SNRs

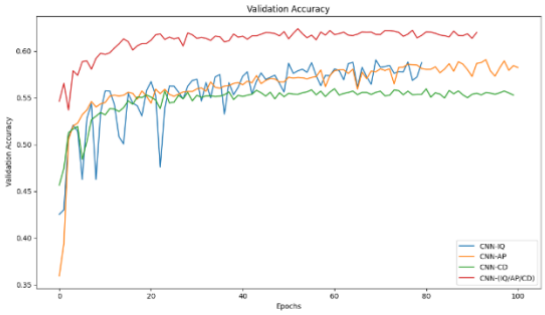


Figure 8: Validation Accuracy of the 4 Models

Fig.9shows the validation loss of each model during the training, CNN- (IQ/AP/CD) has the lowest validation loss, and CNN-CD has the highest validation loss.

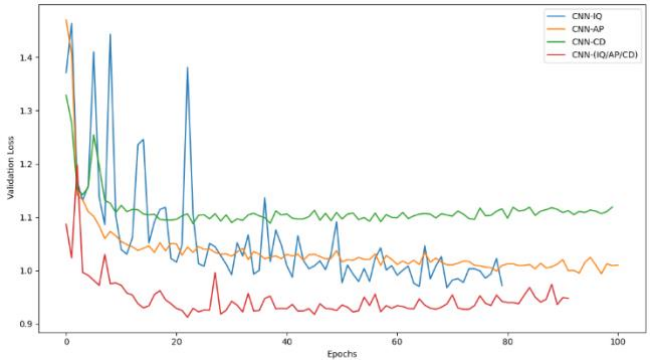


Figure 9: Validation Loss of the 4 Models.

V. EVALUATION RESULTS

After training each of the four models and saving the best weights for each model, we evaluated the classification accuracy by generating the confusion matrix, we should note that the SNR values were divided into three Ranges which are low SNR [-20dB,-8dB], medium SNR [-6dB,4dB], high SNR [6dB,18dB].

Starting with CNN-IQ model we generated the confusion matrix for all SNR Ranges as shown in Fig.10. We can see that the average accuracy is 59.19% in Fig.4.3a but this isn't very informing because of the effect of low SNR at the average accuracy. When we observe Fig.4.3b,the average accuracy is 19.28% because the model wasn't able to extract any desired signal features and is randomly guessing the labels. While Looking at Fig10 c, at medium SNR, there is a diagonal pattern in the confusion matrix, meaning that the model has learned better features leading to improved accuracy. As for Fig.10 d, at high SNR values, the signal power far exceeds the noise power, therefore, the model was able to learn better features. We can see from confusion matrix that most modulations are correctly recognized with an average accuracy of 87.3%. However, the model mistakes 16QAM for 64QAM, and slightly mixes between QPSK and 8PSK even at high SNR.

Figure 10: (a) CNN-IQ Confusion Matrix at All SNR Range, (b) Confusion Matrix at Low SNR, (c) Confusion Matrix at Medium SNR, (d) Confusion Matrix at High SNR.

Moving On to CNN-AP model, we will see if the model is able to perform better by learning more from AP sequences. We generated the four confusion matrices as shown in Fig.11. As for Fig.11a and Fig.11b, we mentioned that at low SNR the model is randomly guessing, therefore affecting the average accuracy across all SNRs. However, for medium SNR only, looking at Fig.11c, the average accuracy is 71.33% which is lower than the CNN-IQ model. The confusion matrix shows that CNN-AP is less accurate at recognizing some modulations compared to CNN-IQ at medium SNR. As for high SNR, looking at Fig.11d, the average accuracy is 93.27% and no mixing between QPSK and 8PSK occurs, meaning it outperforms CNN-IQ, however, it still suffers at classifying 16QAM and 64QAM at high SNR. model. The confusion

matrix shows that CNN-AP is less accurate at recognizing some modulations compared to CNN-IQ at medium SNR.

As for high SNR, looking at Fig.11d, the average accuracy is 93.27% and no mixing between QPSK and 8PSK occurs, meaning it outperforms CNN-IQ, however, it still suffers at classifying 16QAM and 64QAM at high SNR.

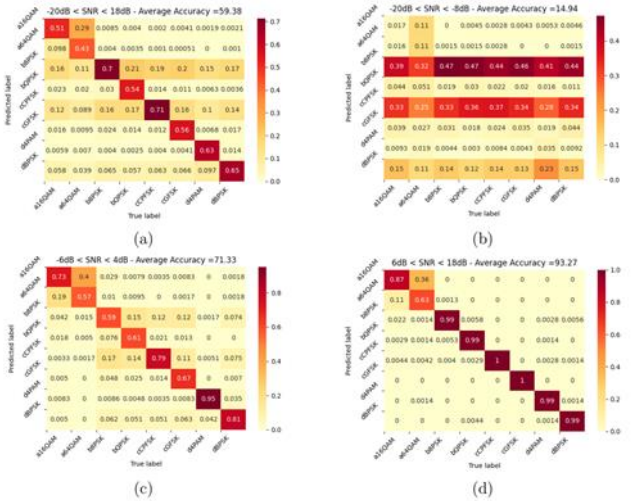


Figure 11: (a) CNN-AP Confusion Matrix at All SNR Range, (b) Confusion Matrix at Low SNR,
(c) Confusion Matrix at Medium SNR, (d) Confusion Matrix at High SNR Range.

For the third model CNN-CD, we trained it on constellation diagram images, this model suffers the most from the low number of points per sample in the dataset. Since 128 points aren't enough for a constellation diagram. The confusion matrix of each SNR range is shown in Fig.12.

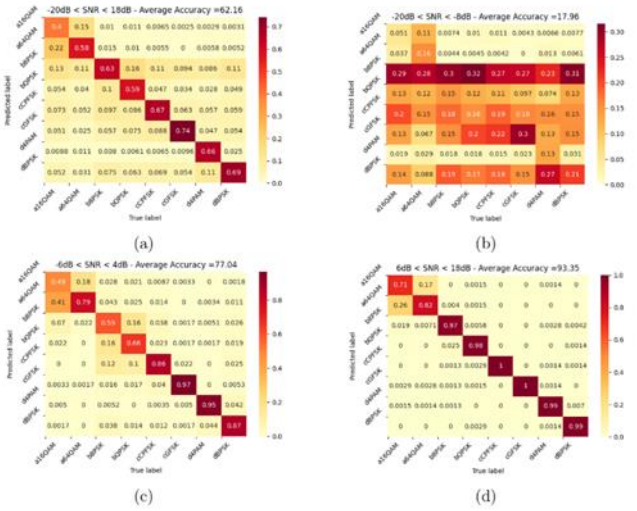
Figure 12: (a) CNN-CD Confusion Matrix at All SNR Range, (b) Confusion Matrix at Low SNR, (c) Confusion Matrix at Medium SNR, (d) Confusion Matrix at High SNR Range.

Looking at Fig.12a, the average accuracy being the lowest indicates that the performance of the model is not the best. From Fig.12c, At medium SNR the model suffers at classifying all modulations and gives the lowest average accuracy amongst the other models. As for Fig.12d, at high SNR the model gives better performance than CNN-IQ with an average accuracy of 90.92%. Overall the model extracts desirable features from the constellation images at high SNR range delivering good performance. However, it suffers at feature extraction and isn't suitable for medium SNR range.

For our fourth model CNN-(IQ/AP/CD), we used three separate models for feature extraction with each model extracting desired features from each representation (IQ,AP,CD). Furthermore, we concatenated those extracted features and then passed them to the classification layer, our goal was to fuse the best of each previous model performance into one model acquiring good performance at most SNR ranges. Fig.13 shows the confusion matrix of each SNR range.

Figure 13: (a) CNN-(IQ/AP/CD) Confusion Matrix at All SNR Range, (b) Confusion Matrix at Low SNR, (c) Confusion Matrix at Medium SNR, (d) Confusion Matrix at High SNR Range.

Looking at Fig.13a, the average accuracy is 62.16% which is higher than other model, therefore it is a good indicator of the models performance. Observing Fig13 c, the average accuracy is 77.04% which outperforms even CNN-IQ model but the model itself still suffers at recognizing certain modulations with good accuracy. Looking at Fig.13d, the average accuracy is 93.35% which is slightly above CNN-AP and it is evident that the model can recognize both 16QAM and 64QAM with good accuracy. Overall, this model



incorporated the best traits of the three models by delivering improved performance which shows the importance of incorporating different features that are extracted from different representations. The confusion matrices offered an insight on how accurate the proposed models are at recognizing each individual modulation for different SNR ranges. Furthermore, Table 2 summarizes the average classification accuracy of each proposed model at different SNR ranges.

Table 2: Average classification accuracy of all proposed models at different SNR ranges.

	Average Classification Accuracy (%)			
	Overall [-20dB,18dB]	High SNR [6dB,18dB]	Medium SNR [-6dB,4dB]	Low SNR [-20dB,-8dB]
CNN-IQ	59.16	87.3	72.55	19.28
CNN-AP	59.38	93.27	71.33	14.54
CNN-CD	55.3	90.92	61.29	14.15
CNN-(AP/IQ/CD)	62.16	93.35	77.04	17.96

Finally, we plotted the classification accuracy versus SNR for all proposed models as shown in Fig.14. We can see that the model trained on IQ sequences gives better performance below 0dB SNR, while training on AP sequences and constellation images gives better performance at higher than 0dB SNR. Furthermore, observing CNN-(IQ/AP/CD) which outperforms at most SNR values, shows that fusing the learned features from each representation incorporates the

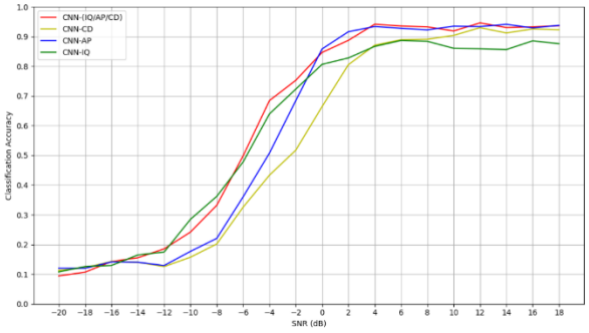


Figure 14: Classification accuracy comparison between different proposed models.

advantages of each representation. Table 3 compares the average classification accuracy at different SNR values.

We can see from Table 3 that CNN-(IQ/AP/CD) outperforms at -2dB, however, CNN-AP achieves the highest accuracy at 2dB, and finally, CNN- (IQ/AP/CD) achieves a decent accuracy of 94.65% at 12dB, we should note that not acquiring higher classification accuracy is mainly due to the signal length being short, it is believed that using a dataset with a longer signal length will achieve better results.

Table 3: Average classification accuracy comparison between proposed models.

		CNN-IQ	*CNN-CD*	*CNN-AP*	*CNN(IQ/AP/CD)*
SNR	*-2dB*	72.31%	51.7%	68.43%	75.29%
	2dB	82.82%	80.6%	91.66%	88.76%
	12db	85.94%	93%	93.4%	94.65%

VI. CONCLUSION

The proposed models functionality is composed of two stages, the first stage is feature extraction from input signal and the second is performing classification. The dataset is composed of 20000 samples per modulation scheme, we only used 8 modulations, therefore, 160000 samples in total. Furthermore, 80% was used for training, 10% for validation, 10% for testing. four models were built, three utilized only a single signal representation, and the last utilized the three representations. CNN-(IQ/AP/CD) generally outperformed the other models at medium and high SNR environments. There was a limitation to the classification accuracy due to the signal length in the used dataset being short (128 symbols).

REFERENCES

[1] Timothy J O'Shea, Johnathan Corgan, and T Charles Clancy. Convolutional radio modulation recognition networks. In International conference on engineering applications of neural networks, pages 213–226. Springer, 2016.

[2] Gihan J Mendis, Jin Wei, and Arjuna Madanayake. Deep learning-based automated modulationclassification for cognitive radio. In 2016 IEEE International Conference on Communication Systems(ICCS), pages 1–6. IEEE, 2016.

[3] Shengliang Peng, Hanyu Jiang, Huaxia Wang, Hathal Alwageed, and Yu-Dong Yao. Modulation classification using convolutional neural network based deep learning model. In 2017 26th Wireless and Optical Communication Conference (WOCC), pages 1–5. IEEE, 2017.

[4] Shilian Zheng, Peihan Qi, Shichuan Chen, and Xiaoniu Yang. Fusion methods for cnn-based automatic modulation classification. IEEE Access, 7:66496–66504, 2019

[5]Prokopios Panagiotou, Achilleas Anastasopoulos, and A Polydoros. Likelihood ratio tests for modulation classification. In MILCOM 2000 Proceedings. 21st Century Military Communications. Architectures andTechnologies for Information Superiority (Cat. No. 00CH37155), volume 2, pages 670–674. IEEE, 2000.

[6] King C Ho and Liang Hong. Likelihood method for bpsk and unbalanced qpsk modulation classification. In Digital Wireless Communication III, volume 4395, pages 155–162. SPIE, 2001.

[7] VD Orlic and ML Dukic. Multipath channel estimation algorithm for automatic modulation classification using sixth-order cumulants. Electronics letters, 46(19):1, 2010.

[8] KC Ho, W Prokopiw, and YT Chan. Modulation identification by the wavelet transform. In Proceedings of MILCOM'95, volume 2, pages 886–890. IEEE, 1995.