Hybrid Deep Learning for Signals Automatic Modulation Classification

Muhammad Moinuddin[1*], Hitham K. Alshoubaki[2], Omar Ayad Alani[3], Ubaid M. Al-Saggaf[4], Karim Abed-Meraim[5]
Department of Electrical & Computer Engineering, King Abdulaziz University, Jeddah 21589, Saudi Arabia [1, 2, 3, 4]
Center of Excellence in Intelligent Engineering Systems (CEIES), King Abdulaziz University, Jeddah 21589, Saudi Arabia [1, 2]
PRISME Lab. (IUF member), Orléans Univ., France[5]
Institut Universitaire de France (IUF), France[5]

***Abstract*—Classifying signals or modulation classification is a crucial step in developing communication receivers. A common practice is to extract features before categorizing the signal, which requires implementing long preprocessing techniques. Due to breakthroughs in neural network topologies, machine learning (ML) algorithms, and optimization techniques, referred to as "deep learning" (DL), we have witnessed a vast degree of change over the previous five years. Advanced deep learning algorithms can be applied to the same automatic modulation classification problem and generate excellent outcomes without requiring time-consuming, manual, and complex feature extraction methods. In recent years, various DL techniques have been explored for automatic modulation classification (AMC). However, it has been observed that these techniques are effective only for higher Signal-to-Noise-Ratio (SNR) values. To overcome this challenge, we proposed a hybrid DL-based AMC technique by combining a customized EfficientNet with a customized Transformer Block. The transformer block is used to enhance the DL performance for the lower SNR values. The performance of the proposed hybrid model is tested on a benchmark dataset, RadioML2018.01A, and compared with the state-of-the-art existing DL method which shows the supremacy of the proposed hybrid model.**

Keywords—Automatic modulation classification; deep learning; machine learning; EfficientNet; Transformer Network

I. INTRODUCTION

Independent radio spectrum interpretation is becoming increasingly important in various applications due to the rapid development of different evolved standards and best practices for wireless communications. These applications include electronic warfare and vulnerability assessments in military scenarios, dynamic channel access, spectrum interference detection, and monitoring in civil proceedings [1]. Encoder, modulation, and multiplexing are all necessary components in a wireless transmission channel. First, the signal from the source is transformed into a format suitable for signal processing techniques such as noise reduction and interference reduction by a Source Encoder. Adding a bit of redundancy to the signal during this process might help it withstand noise. Once the signal has been modulated with an appropriate modulation technique, such as Phase Shift Keying (PSK), Frequency Shift Keying (FSK), or Quadrature Phase Shift Keying (QPSK), it is delivered into the channel as rapidly as possible using antennas. To retrieve the sent data, the receiver side employs the same techniques as the transmitter side, including de-multiplexing, demodulation, and decoding, among other techniques. However, there are several drawbacks, including co-channel interference and signal distortion over spectral channels due to aggressive spectrum usage in massive wireless communication systems. To deal with these issues, one effective solution is Adaptive Modulation (AM), whose aim is to encode radio signals utilizing a variety of modulation forms from a predetermined candidate pool depending on channel conditions and system specifications. Thus, the AM enables intelligent spectrum management in modern communication systems, as shown in Fig. 1 [2].

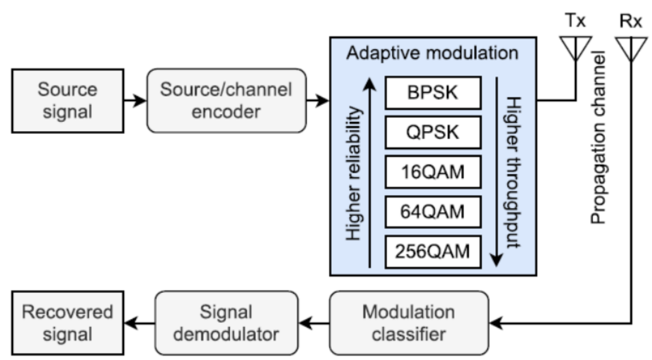

Fig. 1. Communication system with adaptive modulation [2].

However, a major challenge for implementing the AM in wireless communications is the dynamic estimation of channel conditions and data characteristics. Hence, automatic modulation classification (AMC) is introduced, which is a technique that improves spectrum utilization efficiency by applying it to a detected signal at the receiver's physical layer without knowing the data sent or channel characteristics.

A. Task of AMC

Consider a communication system sending a noise-free signal $x(t)$ (whose sampled version is denoted as $x[k]$) through some channel (with discrete version denoted as $h[k]$), then the received signal for the nth symbol at time k can be expressed as Eq. (1) [2]:

$$y[n] = x[n, h_k] + w[n] \tag{1}$$

where, $x[n, h_k]$ is the modulated input data passed through the channel h_k and $w[n]$ is the additive noise at the receiver with variance σ_w^2. The task of an AMC in a communication system is

*Corresponding author.

to predict the modulation type using only input-output data without estimating the channel impulse response.

B. AMC Approaches

Both probabilistic and conventional machine learning frameworks are used in the traditional AMC approaches. Based on the assumption that all signal and channel models are fully understood, likelihood-based systems have the best level of accuracy since they employ probability theory and hypothetical models to solve classification difficulties. Machine learning (ML) features are more likely to be adopted because of their simplicity. However, this comes at the expense of a few drawbacks, such as restricted learning capacity and a poor discriminative experience of handmade features.

Likelihood-Based (LB) Approach: The optimum approach is to employ the LB approach, which minimizes the possibility of incorrect categorization. This assumption is correct if a waveform's density function (PDF) contains all relevant information. Using a decision threshold, the probability ratio from the PDF is used to categorize the modulation type [3]. The drawback of this approach is its computational complexity. Real-time classification requires a thorough inspection of multiple pulses and emitters. It is possible to separate a single emitter by using other ways of locating the Primary Rat Interface (PRI). Due to this, it is possible to forgive a few wrong classifications in modulation categorization if the majority of the categories are suitable. Algorithm strength must also be considered while extracting data. A lack of inherent robustness makes LB approaches susceptible to noise, timing issues, and phase offsets. Given this and the necessity of real-time operation, a more accessible, feature-based approach may be adopted.

Feature-Based (FB) Approach: The FB approach employs representative features to identify and categorize the differences between signals. The typical characteristics are the time domain, signal changes, zero crossings, and statistics. In addition, the FB AMC technique often uses a hierarchical approach, first identifying modulation and defining its exact kind [3]. Then, PDF-based, Euclidean distance, and artificial intelligence classifiers are utilized to make decisions. FB algorithms are generally preferred because of their simple complexity and acceptable performance, even if their performance is poor.

C. Related Works: Deep Learning in AMC

In recent years, the DL techniques have been successfully applied in various fields such as computer vision [4], [5], wireless communications [6], [7], bioinformatics [8]-[13], and signal processing [14], [15], etc. Motivated by the power of DL, several state-of-the-art DL techniques have been employed in the field of AMC, which resulted in improved modulation classification performance of ACM. For example, convolutional neural networks (CNNs) based AMC [16], [17], [18], long short-term memory networks (LSTMs) based AMC [19], [20], and recurrent neural networks (RNNs) based AMC [21]. One major advantage of DL-based AMC over conventional ML-based AMC is that DL provides automatic feature extraction with higher classification accuracy [22]. Another important fact about the DL-based AMC is that it can enhance the classification accuracy of higher-order modulation in the presence of a synthetic channel impairment [23]. Finally, it is found that the DL-based AMC has the capability to effectively process big data, and thus it is well-suited for deploying AMC in Internet-of-Things (IoT) systems [24].

D. Challenges in AMC

A major challenge in ML-based AMC is the requirement of designing efficient feature extraction methods for each dataset. On the other hand, although the DL-based AMC can provide automatic feature extraction, with improved modulation classification accuracy, there are still many open challenges in this field. One big challenge is that all the existing DL-based AMC methods proved to be better in terms of classification accuracy only for higher values of signal-to-noise ratio (SNR). Their performance degrades drastically at lower SNR values, which implies that if the noise level is higher, the model would most likely fail to categorize the signal. In the existing literature, there is no work addressing this issue. In this work, we mainly focus on this challenge.

E. Our Contributions

As outlined in the previous subsection, the DL-based AMC has one major challenge of poor performance at lower SNR values, we focused on solving this issue by designing a hybrid DL-based AMC solution. More specifically, the proposed method is capable of providing reasonably better modulation accuracy at lower SNR values. For this purpose, we developed a hybrid DL-based AMC solution in which we utilized the combination of the EfficientNet and the Transformer Block via a switching mechanism. The transformer block is used to enhance the DL performance for the lower SNR values. The performance of the proposed hybrid model is tested on a benchmark dataset, RadioML2018.01A [25].

The study is organized as follows: Following this introduction, the baseline methods are discussed in Section II. In Section III, the proposed hybrid CNN model is presented. An overview of the RadioML 2018 dataset is provided in Section IV. Next, the experiment on AMC using the proposed model is described in Section V. Results and discussion are provided in Section VI. Finally, the concluding remarks are given in Section VII.

II. Baseline Methods

In the context of AMC, various DL architectures have been explored in the past [25]. Among these methods, the best performance is achieved by the modified ResNet proposed in [25]. In this section, we provided an overview of both the standard and the modified ResNet architectures.

A. Standard ResNet [26]

ResNet is a deep residual network that has all the elementary parts in its structure as existed in any deep CNN architecture [26]. The only unique characteristic of the ResNet is the use of identity connection or mapping between the levels. In order to understand this Identity Mapping in a ResNet, see the block diagram presented in Fig. 2, where F(x) is a function called the Residual Function that operates between two convolutional weight layers. More precisely, it is the difference between the input x and the output H(x) of the residual block, as shown in Fig. 2. Thus, the F(x) can be formulated as Eq. (2):

$$F(x) = H(x) - x \tag{2}$$

Therefore, the major purpose of including this function in ResNet is to utilize the stacked layers to estimate the function F(x) instead of estimating the function H(x). Consequently, training the deep residual network tries to learn F(x), which results in improving the overall accuracy. Moreover, it is straightforward to conclude that the network's accuracy will be improved by increasing the depth, provided that the issue of over-fitting is resolved. However, the difficulty with the increase in depth is that the signal necessary to modify the weights, which emanates from the end of the network by comparing ground truth and forecast, gets extremely little at the initial layers as depth increases. It indicates that the initial levels are learned to a minor extent. The term for this phenomenon is vanishing gradient. The second issue with training deeper networks is that optimization is performed on a large parameter space, resulting in naively adding layers, which increases training error. As illustrated in the figure, residual networks permit the training of deep networks by creating the network using modules termed residual models.

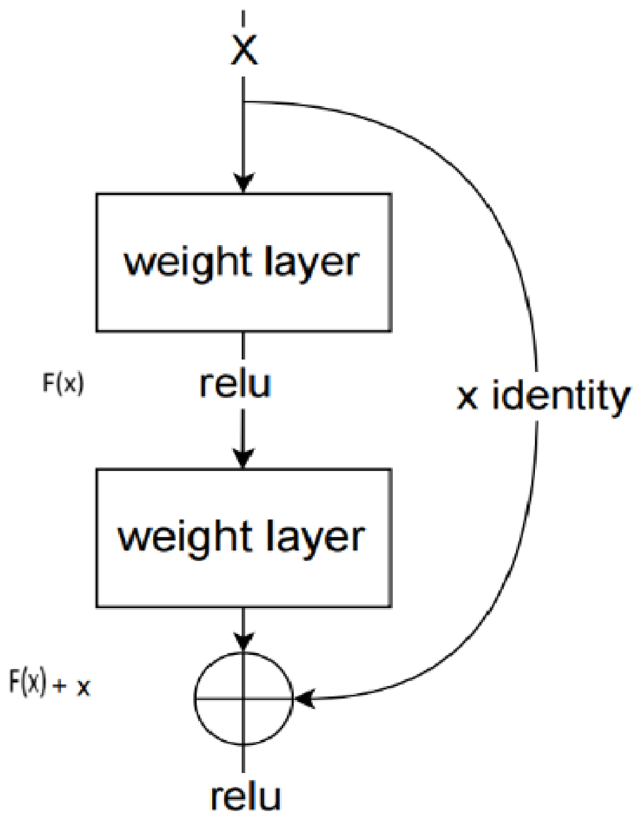

Fig. 2. Residual learning: a building block.

B. Modified ResNet [25]

In [25], a modified ResNet was proposed for implementing the AMC. In this architecture, the residual unit and stack of residual units are used, as shown in Fig. 3, and its network layout is shown in Table I. In addition, self-normalizing neural networks, the activation function of scaled exponential linear unit (SELU), mean response scaled initialization (MRSA), and Alpha Dropout were used and resulted in a marginally better performance than standard ReLU.

TABLE I. MODIFIED RESNET NETWORK LAYOUT [25]

Layer	Output Dimension
Input	2 x 1024
Residual Stack	32 x 512
Residual Stack	32 x 256
Residual Stack	32 x 128
Residual Stack	32 x 64
Residual Stack	32 x 32
Residual Stack	32 x 16
FC/SeLU	128
FC/SeLU	128
FC/Softmax	24

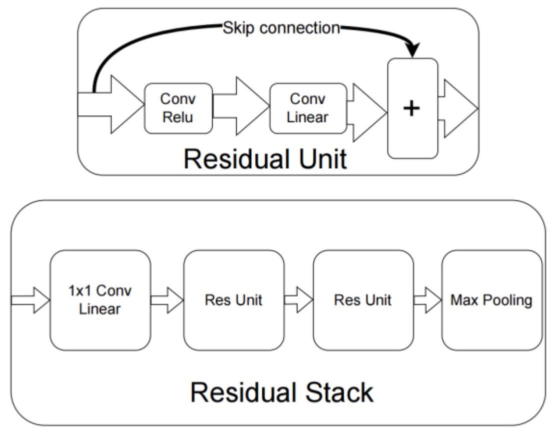

Fig. 3. Hierarchical layers used in modified ResNet

III. PROPOSED HYBRID CNN MODEL

In this work, we proposed a hybrid model of CNN which utilizes a combination of EfficientNet and Transformer Network. The architectures of both the networks and the proposed hybrid model are presented in the ensuing.

A. EfficientNet Network

Developing convolutional neural networks are done at a set cost. These networks may be expanded to achieve higher accuracy, when additional resources are available. For example, to increase the size of a ResNet 18 model to a ResNet 200 model, more layers can be added to the initial model. Mostly, this scaling strategy has improved the accuracy of many benchmarking datasets. However, the initial methods of model scaling are quite unpredictable. Some models are scaled horizontally, while others are scaled vertically. Some models simply capture a higher-resolution image to get better outcomes. Manual tweaking and numerous person-hours are required when using this strategy of randomly scaling models, and it frequently results in little or no improvement in performance. EfficientNet's authors suggested that CNN models be scaled up to improve accuracy and efficiency more ethically [27].

EfficientNet uses a method known as compound coefficients to scale up models in an easy yet effective manner. Compound scaling, instead of random scaling up of width, depth, or resolution, uniformly scales each dimension with a predetermined set of scaling coefficients. The efficient developers used the scaling approach and AutoML to construct seven models of varied dimensions that outperformed the current state-of-the-art accuracy of most convolutional neural networks and were significantly more efficient.

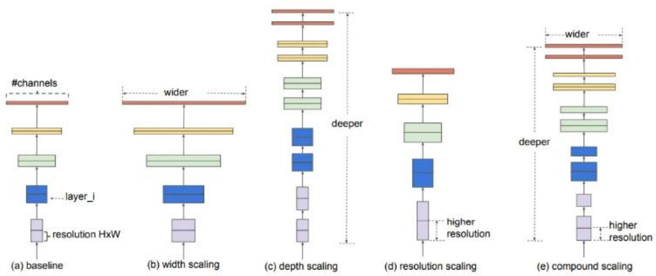

Fig. 4. Various types of scaling used: (a) is a baseline network; (b) to (d) conventional scaling via width, or depth, or resolution dimension. (e) is compound scaling used in EfficientNet [27].

The authors evaluated the effects of each scaling methodology on the model's performance and efficiency to develop the compound scaling method. In their opinion, scaling single dimensions helps in enhancing model performance, balancing the scale in all three dimensions: width, depth, and picture resolution, considering the varied resources available, which best improves the overall model performance. The compound scaling method proposed for EfficientNet is shown in Fig. 4. The goal of the compound scaling method is to balance the width, depth, and resolution measurements achieved by scaling with a constant ratio.

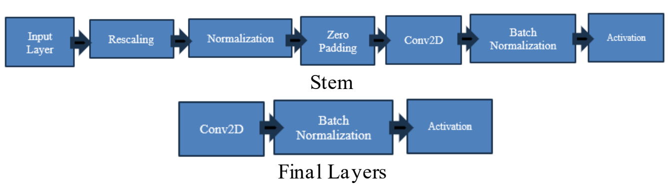

Fig. 5. Stem and final layers of EfficientNet.

A neural architecture search utilizing the AutoML MNAS framework created the baseline network used in EfficientNet. For optimum accuracy, the network is fine-tuned and punished if it is computationally intensive. It is also punished for sluggish inference times when the network takes time to create predictions. The system employs a mobile inverted bottleneck convolution similar to MobileNet V2; however, it is substantially larger owing to the rise in FLOPS. This basic model is then built to provide the networks' EfficientNets family. The stem and final layers of EfficientNet are presented in Fig. 5.

Further, each of the modules has seven blocks. These blocks are further subdivided into a different number of sub-blocks. These numbers increase when moving from EfficientNet-B0 to EfficientNet-B7.

B. Transformer Network

The encoder-decoder architecture has evolved into the "Transformer model" [28, 29]. Unlike the encoder-decoder design, the Transformer doesn't employ recurrent neural networks (RNNs) to acquire sequential data. Transformer-based models have replaced Long-Short Term Memory (LSTM) networks [30] in many sequence-to-sequence situations and are superior in quality. Conducting some diggings to discover the best time-series model that has just been released on the internet and in journals.

Recently, a solution using the newest state-of-the-art time series model called Transformer is devised [29]. Compared to the prior models, this model has various advantages, such as parallelization processing, which allows it to use contemporary graphics processing units (GPUs), which were created specifically for parallel calculation. Moreover, the vanishing gradient problem can be solved in Transformer Network (TN) as the input size can be adjustable, and the network can execute simultaneously rather than sequentially. Furthermore, the TN is capable of capturing detailed data context due to its advantage of the network modules' positional encoding and self-attention mechanisms [31]. The main crucial network module within the TN is provided in [31].

C. Switching-Based Proposed Hybrid Model

Our experimental analysis showed that the EfficientNet and Transformer Net are good in predicting signals with high SNR and low SNR, respectively. We proposed to combine the two network models of EfficientNet and Transformer Net effectively, such that the classification accuracy of the AMC should be enhanced for both low and high SNR values. For this purpose, an SNR threshold-based switching is proposed as shown in Fig. 6. This SNR threshold value is set according to experimental findings. In our experiments, we have used a 0 dB SNR value as a switching threshold; that is, the proposed model will utilize EfficientNet if SNR is above 0 dB, and it will switch to TN otherwise.

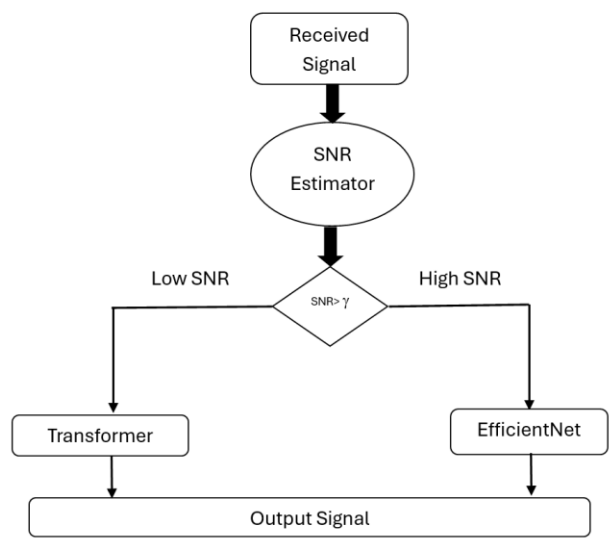

Fig. 6. SNR switch between EfficientNet and the transformer block.

To implement the proposed SNR threshold-based switching, instantaneous SNR values are estimated using the received samples. Knowing the fact that there are well-established techniques for estimating noise variance estimation ($i.e., \hat{\sigma}_w^2$) [32, 33], the received SNR can be estimated using an average over window size N_w as follows [see Eq. (3)]:

$$\widehat{\text{SNR}}[n] = \frac{1}{\hat{\sigma}_w^2}\left\{\left(\frac{1}{\text{N}_w}\sum_{k=n-\text{N}_w+1}^{n}|y[k]|^2\right) - \hat{\sigma}_w^2\right\} \quad (3)$$

Finally, switching is performed between EfficientNet and Transformer block by inspecting the estimated SNR against a defined threshold value (γ) that can be set by the designer.

IV. Overview of RadioML 2018 Dataset

The majority of existing AMC strategies have been assessed on the reenactment datasets produced by programming. According to the point of view of reasonableness, producing tweaked signs to benchmark the execution of profound models ought to consider two essential worries: modulation technique (such as modulation types and modulation count) and channel condition. For models, RadioML2018.01A, a presently accessible and broadly utilized dataset of regulation grouping, conceals 24 modulations (counting analogue and digital procedures and a few testing high-request designs), where the regulation transmissions are engendered in a multipath Rician blurring with carrier frequency offset, symbol rate offset, delay spread, and AWGN to almost acquire true peculiarities in remote correspondences. On the other hand, various AMC techniques have produced easy datasets with few basic digital modulations. Thus, the adequacy and effect of exploration commitment in view of execution assessment in reproductions can be unconvincing other than the unsubstantial dependability of grouping models to carry out in practical frameworks. For example, a plain-design profound organization [32] accomplished high exactness with three modulation candidates (BPSK, QPSK, and 8-PSK). In different works [33], the considered spread channel is less difficult with level blurring and time-invariant rather than recurrence, particularly multipath blurring and testing time float. The modulation classification task turns out to be easier with a few given classes, where the between-class and intra-class segregation issues are not thought rigorously.

Even though DL-AMC can give huge execution, it requires a lot of training data. In reality, obtaining a sufficient number of valid training samples is typically expensive and challenging. Hence, the employed huge scope preparing dataset ought to be developed cautiously enough with more reasonable signs under a wide scope of SNR. Besides, data expansion techniques can arrange radio regulation plans more effectively by utilizing more limited radio examples [34], which will furnish an improved DL model with a more modest order reaction time. Nonetheless, the Signs dataset is especially needed for preparing, approving, and testing the networks in ML-AMC and DL-AMC models. A few analysts utilized their simulated datasets, and others like to utilize those presented in the journal. A portion of the datasets is introduced as follows.

GNU Radio datasets are utilized in this research. The RML2018 dataset is one of the most complex modulation classification datasets published in [25]. Twenty-four digital and analog modulation schemes are measured over the air and distributed across a wide range of SNR values. In addition, it comprises almost 2.5 million signals with artificially created channel distortions. "Over-the-air deep learning-based radio signal categorization" was published in the IEEE Journal of Selected Topics in Signal Processing in 2017 and gives further information and explanation of the dataset. In hdf5 format, around 1.5 million instances are recorded as complex floating-point values with a sample length of 1024 samples, respectively.

The updated techniques [35] were used to create the new dataset (RadioRML 2018). 24 modulators, both analog and digital, are utilized to modulate a wide variety of single carriers. Fig. 7 demonstrates OTA transmission channels containing clean signals without any synthetic channel impairments. Digital signals are formed with a root-raised cosine beat modulating filter [36] with a scope of roll-off values.

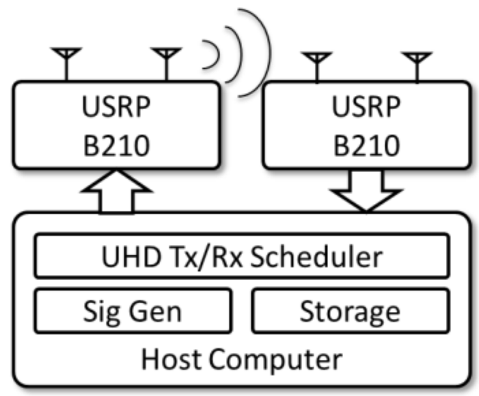

Fig. 7. Over-the-air test configuration [25].

V. Experiment on AMC

In this study, Python is used to perform the AMC task, which was done using Google Colaboratory (called Colab). Google Colab allows users to take advantage of backend hardware such as GPUs and TPUs to accelerate their computations. Thus, it facilitates performing all of the tasks in a Jupyter notebook hosted on your local system without installing or configuring any additional software.

A. *EfficientNet (for High SNR)*

First, we import the EfficientNet for Tenserflow-Keras Library, which is the Conventional EfficientNet. However, we chose the B0 Architecture, since it has the lowest processing complexity SNR switch.

As part of our suggested strategy, we attempt to employ two network models, one effective for predicting signals with high SNR and the other effective at predicting signals with low SNR. The transformer neural network will be used for low SNR models because it is a time series neural network that may be used to forecast complicated data, such as signals with low SNR rates. Our improved EfficientNet, on the other hand, is employed for categorizing signals with greater signal-to-noise ratios. Following the input of our signal into two models, we will ensemble them to obtain the highest accuracy output from the two models. For this purpose, the SNR value of 0 dB is selected as the switching threshold among the eight possible types of EfficientNet. The original EfficientNet B0 has an input size of 224x224x3; therefore, we must adapt the input size to match that of RML 2018, which is 1024x2, so we must alter the input layer appropriately. To accommodate the adaptation, the reshape layer has been introduced in the first layer before the modified EfficientNet, as shown in Fig. 8.

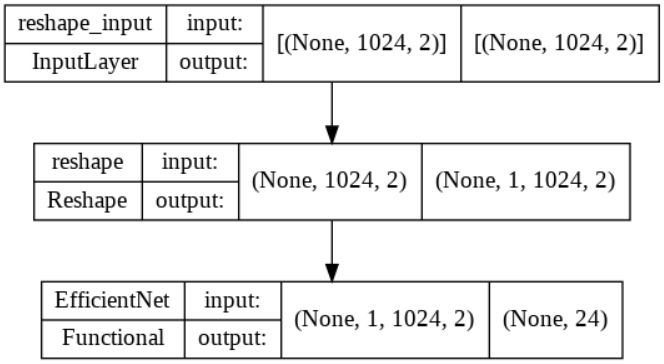

Fig. 8. Rescale layer and the customized EfficientNet.

The training process for our network uses a batch size of 1024 and a validation testing dataset comprised of 20% of the total training dataset.

B. Transformer Block (for Low SNR)

As shown in Fig. 9, the architecture being utilized is designed primarily to deal with noisy signal modulation / low SNR signal modulation. Two fully linked networks and Alpha Dropout are used in conjunction with batch normalization to minimize overfitting. After an Alpha dropout, the mean and variance of inputs retain their original values, ensuring that the self-normalizing characteristic continues to hold even after the dropout has occurred. In addition, we selected SeLU, which stands for Scaled Exponential Linear Unit, for the activation function to support the self-normalizing feature of the dropout layer, which is supported by the dropout layer itself. Lastly, we continue to use Lazy Adam as our model enhancer. The details of layer type, output type, and number of parameters are provided in Table II.

TABLE II. LAYERS OF THE TRANSFORMER BLOCK USED IN AMC

Layer (type)	**Output Shape**	**Param #**
input_1 (InputLayer)	[(None, 1024, 2)]	0
reshape (Reshape)	(None, 2, 1024)	0
transformer_block (TransformerBlock)	(None, 2, 1024)	4728064
Total params: **Trainable params:** **Non-trainable params:**		**4,882,968** **4,880,920** **2,048**

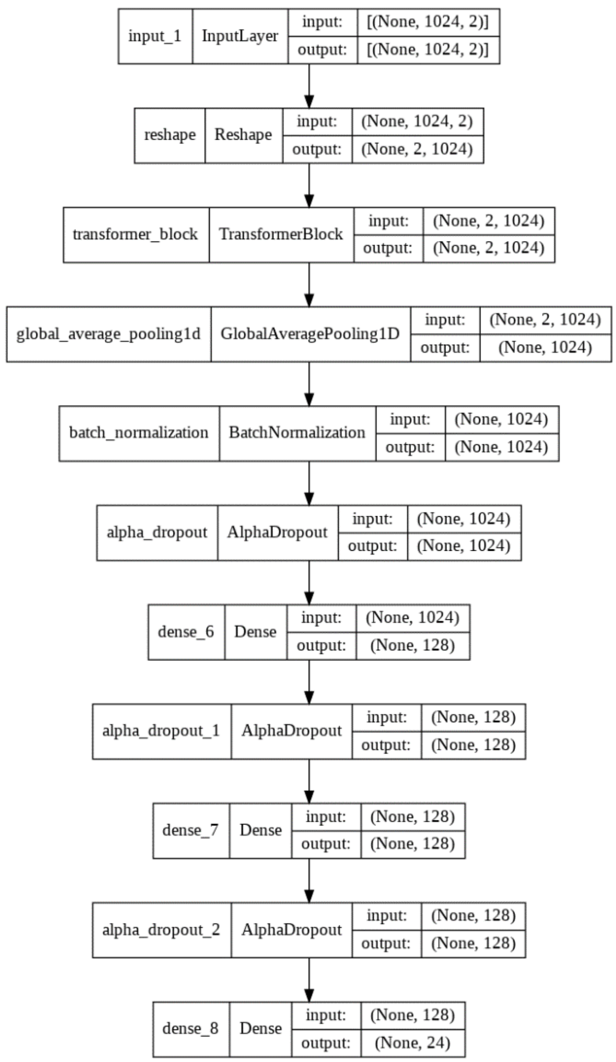

Fig. 9. Architecture of the transformer block used in AMC.

VI. Results and Discussion

This section will demonstrate the results of applying the proposed method to create the hybrid model. These results are reported in Table III in terms of overall Accuracy of the proposed model at different SNR values. Accuracy is calculated using the following relation, Eq. (4):

$$Accuracy = \frac{TP+TN}{P+N} \quad (4)$$

where, *TP, TN, P*, and *N* represent True Positive, True Negative, Total Positive, and Total Negative, respectively.

Our model consists of two parts; the first part deals with the clean signals that have bigger than 0 SNR values; our Customized EfficientNet shows outstanding performance in the modulation classification for such values, and the maximum accuracy of the modulation classification is 92% at 30 SNR average. However, the EfficientNet has the same characteristics as the other convolutional neural networks (CNNs), and the accuracy is degraded, especially for the signals under 0 SNR values. On the other hand, Transformer Block is working on modulation classification for the lower SNR range of the signals. The Transformer shows an average of 70% accuracy of the modulation classification in the signals with lower SNR values than zero. By applying the parts (Customized EfficientNet and Transformer Block), we have a wider modulation classification range that works from -20 SNR up to 30 SNR with average accuracy (80.74%) among all the SNR ranges. Compared to any solo system, our hybrid model has superior performance considering the full range of SNR. The following subsections in this section illustrate the accuracy results of each part individually; then, the final results will be shown for the combined parts of the hybrid system.

Furthermore, both parts of the hybrid system (Customized EfficientNet and Transformer Block) are trained to utilize the raw data directly, without any feature extraction or preprocessing stage. However, the training stage took a long time for training of the two models, around one hour for each epoch, and it took 36 epochs to complete the training; this long time is due to the huge size of the used dataset. Noting that, 10% of the training data set was considered a validation split to determine the optimum point to stop the training process and have the best generalization characteristics for the system.

A. Performance of the Customized EfficientNet

By implementing the customized EfficientNet discussed in Section V(A), the overall accuracy of the model on the AMC task is reported in Table III and displayed graphically in Fig. 10. It can be observed that the customized EfficientNet alone performs very good for positive or higher values of SNR. However, it has a poor performance for negative and lower SNR values.

TABLE III. EfficientNet Accuracy

SNR (dB)	Overall Accuracy	SNR (dB)	Overall Accuracy
-20	4.34%	2	68.73%
-18	4.57%	4	79.58%
-16	4.87%	6	86.08%
-14	5.46%	8	89.15%
-12	7.81%	10	90.30%
-10	12.43%	12	91.49%
-8	16.16%	14	91.12%
-6	24.11%	16	91.80%
-4	31.65%	18	92.22%
-2	42.32%	20	92.01%
0	57.06%	22	91.93%
		24	92.32%
		26	91.89%
		28	92.05%
		30	92.03%

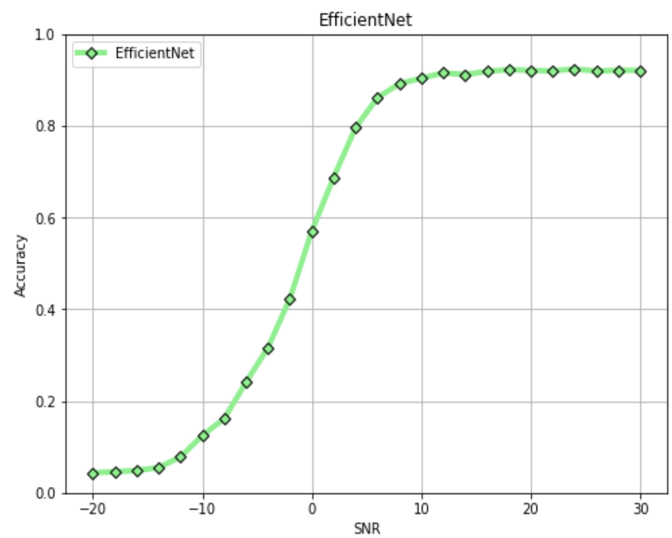

Fig. 10. AMC accuracy for customized EfficientNet alone.

B. Performance of the CustomizedTransformer Block

By implementing the customized Transformer block discussed in Section V(B), the overall accuracy of the model on the AMC task is reported in Table IV and displayed graphically in Fig. 11. It can be observed that the customized Transformer block alone performs very good for negative and lower values of SNR. However, it has a poor performance for positive or higher SNR values.

TABLE IV. TRANSFORMERS' ACCURACY

SNR (dB)	Overall Accuracy	SNR (dB)	Overall Accuracy
-20	69.79%	2	52.34%
-18	68.97%	4	45.72%
-16	69.85%	6	44.84%
-14	70.13%	8	41.26%
-12	70.08%	10	40.88%
-10	71.76%	12	41.95%
-8	72.16%	14	42.25%
-6	72.32%	16	42.36%
-4	71.08%	18	42.27%
-2	69.09%	20	41.96%
0	61.18%	22	42.84%
		24	42.46%
		26	42.49%
		28	42.29%
		30	42.47%

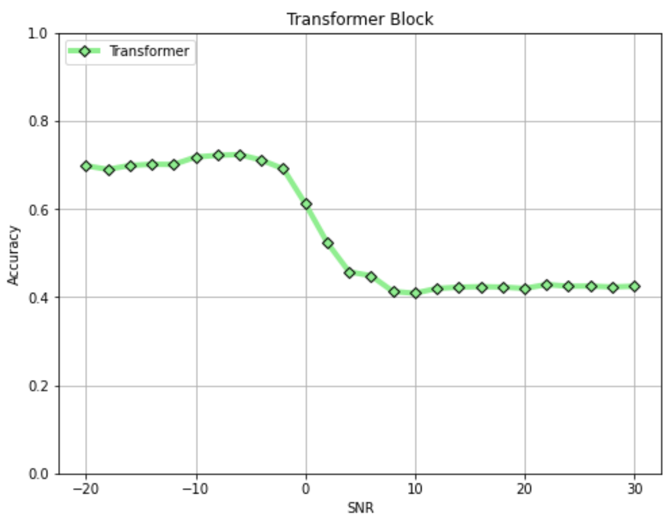

Fig. 11. AMC accuracy for the customized transformer block alone.

C. Performance of the Proposed Hybrid Model

By implementing the proposed hybrid model for the AMC task using the switching mechanism discussed in Section V, the overall accuracy of the proposed model on the AMC task is reported in Table V and displayed graphically in Fig. 12. Moreover, the performance of the proposed hybrid model is also compared to the one achieved by the modified ResNet of [25]. It can be observed that the proposed hybrid model initially performed bad at 0 dB SNR, and later it performed very well for the larger SNR values. The reason for this bad performance is due to abrupt switching via the imperfection in SNR estimation using Eq. (3). This is one of the limitations of the proposed method that needs to be improved in future research. Moreover, the proposed hybrid model outperformed the modified ResNet of [25] for lower SNR values (10 dB and below). For the case of higher SNR values (above 10 dB), its performance is almost identical to that of the modified ResNet of [25]. This is important to highlight that the AMC at lower SNR is more crucial and hence, our proposed model has attained this goal.

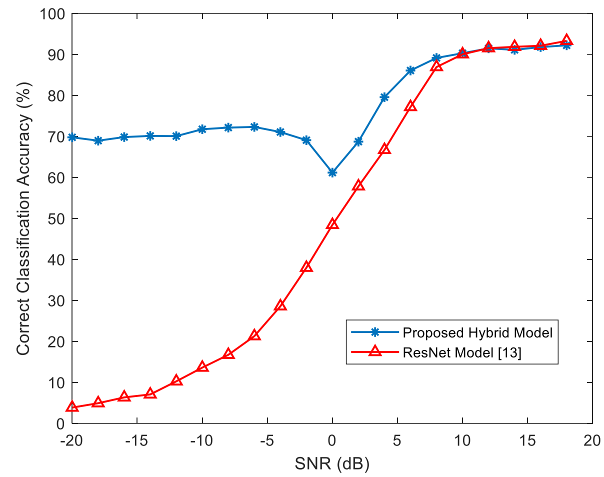

Fig. 12. AMC accuracy comparison for the proposed model.

TABLE V. AMC ACCURACY COMPARISON FOR THE PROPOSED MODEL

SNR(dB)	Accuracy of Proposed Hybrid Model	Accuracy of Modified ResNet Model [25]
-20	69.79%	3.87%
-18	68.97%	4.94%
-16	69.85%	6.36%
-14	70.13%	7.08%
-12	70.08%	10.27%
-10	71.76%	13.62%
-8	72.16%	16.74%
-6	72.32%	21.29%
-4	71.08%	28.58%
-2	69.09%	37.98%
0	61.18%	48.43%
2	68.73%	57.84%
4	79.58%	66.72%
6	86.08%	77.18%
8	89.15%	86.93%
10	90.30%	90.04%
12	91.49%	91.53%
14	91.12%	91.86%
16	91.80%	92.12%
18	92.22%	93.31%
20	92.01%	92.87%
Avg.	**78.04%**	**49.50%**

VII. Conclusion

The capacity to categorize signals is a critical skill that may be used in various applications. In this work we have developed a hybrid CNN model by intelligently combining the EfficientNet and the Transformer Block for the AMC application. For this purpose, customized architectures of both the EfficientNet and the Transformer Block are developed. A switching mechanism is opted to develop the proposed hybrid model such that the EfficientNet is used for higher SNR values, while the transformer block is used to enhance the DL performance for the lower SNR values. The performance of the proposed hybrid model is examined on a benchmark dataset, RadioML2018.01A. The results show that the proposed hybrid model has an average classification accuracy of 78.04% in contrast to 49.5% achieved by its state-of-the-art counterpart. Thus, confirming the supremacy of the proposed model for the AMC application. In the future, more accurate methods of SNR switching can be explored. Moreover, the AMC task can be expanded to other types of modulation and channel models.

Acknowledgment

This research work was funded by the Institutional Fund Project under grant no. IFPHI-123-135-2020. Therefore, the authors gratefully acknowledge technical and financial support from the Ministry of Education and King Abdulaziz University, DSR, Jeddah, Saudi Arabia. The Deanship of Scientific Research (DSR) at King Abdulaziz University (KAU), Jeddah, Saudi Arabia, has funded this project under Grant No. (KEP-MSc: 63-135-1443).

Conflicts of Interest

The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] T. Huynh-The, C.-H. Hua, V.-S. Doan, Q.-V. Pham, and D.-S. Kim, “Accurate deep CNN-based waveform recognition for intelligent radar systems," *IEEE Communication Letters*, vol. 25, no. 9, pp. 2938-2942, Sep. 2021.

[2] T. Huynh-The, Q. Pham, T. Nguyen, T. Nguyen, R. Ruby, M. Zeng, D. Kim, “Automatic Modulation Classification: A Deep Architecture Survey”, *IEEE Access,* vol. 9, pp. 142950 – 142971, 2020.

[3] [3] O. A. Dobre, A. Abdi, Y. Bar-Ness, and W. Su, “Survey of automatic modulation classification techniques: classical approaches and new trends,” *IET Communications*, vol. 1, no. 2, pp. 137-156, April 2007.

[4] S. M. S. Islam, S. Rahman, M. M. Rahman, E. K. Dey, and M. Shoyaib, "Application of deep learning to computer vision: A comprehensive study," *2016 5th International Conference on Informatics, Electronics and Vision (ICIEV)*, 2016, pp. 592-597, doi: 10.1109/ICIEV.2016.7760071.

[5] U. Shah and A. Harpale, "A Review of Deep Learning Models for Computer Vision," *2018 IEEE Punecon*, 2018, pp. 1-6, doi: 10.1109/PUNECON.2018.8745417.

[6] L. Dai, R. Jiao, F. Adachi, H. V. Poor, and L. Hanzo, "Deep Learning for Wireless Communications: An Emerging Interdisciplinary Paradigm," in *IEEE Wireless Communications*, vol. 27, no. 4, pp. 133-139, August 2020, doi: 10.1109/MWC.001.1900491.

[7] W. Yu, F. Sohrabi, and T. Jiang, "Role of Deep Learning in Wireless Communications," in *IEEE BITS the Information Theory Magazine*, 2022, doi: 10.1109/MBITS.2022.3212978.

[8] C.-H. Hua, T. Huynh-The, and S. Lee, ``Retinal vessel segmentation using round-wise features aggregation on bracket-shaped convolutional neural networks," in *Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Berlin, Germany, Jul. 2019, pp. 36_39.

[9] C.-H. Hua, T. Huynh-The, K. Kim, S.-Y. Yu, T. Le-Tien, G. H. Park, J. Bang, and W. Ali, ``Bimodal learning via trilogy of skip-connection deep networks for diabetic retinopathy risk progression identi_cation," *Int. J. Med. Informat.*, vol. 132, Dec. 2019, Art. no. 103926.

[10] T. Huynh-The, C.-H. Hua, and D.-S. Kim, ``Visualizing inertial data for wearable sensor based daily life activity recognition using convolutional neural network," in *Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2019, pp. 2478_2481.

[11] C.-H. Hua, T. Huynh-The, and S. Lee, ``DRAN: Densely reversed attention based convolutional network for diabetic retinopathy detection," in *Proc. 42nd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Montreal, QC, Canada, Jul. 2020, pp. 1992_1995.

[12] C.-H. Hua, K. Kim, T. Huynh-The, J. I. You, S.-Y. Yu, T. Le-Tien, S.-H. Bae, and S. Lee, ``Convolutional network with twofold feature augmentation for diabetic retinopathy recognition from multimodal images," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 7, pp. 2686_2697, Jul. 2021.

[13] T. Huynh-The, C.-H. Hua, N. A. Tu, and D.-S. Kim, ``Physical activity recognition with statistical-deep fusion model using multiple sensory data for smart health," *IEEE Internet Things J.*, vol. 8, no. 3, pp. 1533_1543, Feb. 2021.

[14] Y. -W. Chen et al., "CITISEN: A Deep Learning-Based Speech Signal-Processing Mobile Application," in *IEEE Access*, vol. 10, pp. 46082-46099, 2022, doi: 10.1109/ACCESS.2022.3153469.

[15] Y. Liu, Y. Li, Y. Zhu, Y. Niu and P. Jia, "A Brief Review on Deep Learning in Application of Communication Signal Processing," *2020 IEEE 5th International Conference on Signal and Image Processing (ICSIP)*, 2020, pp. 51-54, doi: 10.1109/ICSIP49896.2020.9339345.

[16] S. Peng, H. Jiang, H. Wang, H. Alwageed and Y. -D. Yao, "Modulation classification using convolutional Neural Network based deep learning model," *2017 26th Wireless and Optical Communication Conference (WOCC)*, 2017, pp. 1-5, doi: 10.1109/WOCC.2017.7929000.

[17] Z. Zhang, C. Wang, C. Gan, S. Sun and M. Wang, "Automatic Modulation Classification Using Convolutional Neural Network With Features Fusion of SPWVD and BJD," in *IEEE Transactions on Signal and Information Processing over Networks*, vol. 5, no. 3, pp. 469-478, Sept. 2019, doi: 10.1109/TSIPN.2019.2900201.

[18] D. A. Jariwala and K. M. Captain, "Automatic Modulation Classification: A Novel Convolutional Neural Network Based Approach," *2021 IEEE 18th India Council International Conference (INDICON)*, 2021, pp. 1-6, doi: 10.1109/INDICON52576.2021.9691687.

[19] Q. Zhou, X. Jing, Y. He, Y. Cui, M. Kadoch and M. Cheriet, "LSTM-based Automatic Modulation Classification," *2020 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, 2020, pp. 1-4, doi: 10.1109/BMSB49480.2020.9379677.

[20] Y. Chen, W. Shao, J. Liu, L. Yu and Z. Qian, "Automatic Modulation Classification Scheme Based on LSTM With Random Erasing and Attention Mechanism," in *IEEE Access*, vol. 8, pp. 154290-154300, 2020, doi: 10.1109/ACCESS.2020.3017641.

[21] D. Hong, Z. Zhang and X. Xu, "Automatic modulation classification using recurrent neural networks," *2017 3rd IEEE International Conference on Computer and Communications (ICCC)*, 2017, pp. 695-700, doi: 10.1109/CompComm.2017.8322633.

[22] Clement, J.C., Indira, N., Vijayakumar, P. et al. Deep learning based modulation classification for 5G and beyond wireless systems. *Peer-to-Peer Netw. Appl.* 14, 319–332 (2021). https://doi.org/10.1007/s12083-020-01003-3.

[23] S. Zhou, Z. Yin, Z. Wu, Y. Chen, N. Zhao, and Z. Yang, ``A robust modulation classi_cation method using convolutional neural networks," *EURASIP J. Adv. Signal Process.*, vol. 2019, no. 1, pp. 1_15, Dec. 2019.

[24] S. Zhou, Z. Yin, Z. Wu, Y. Chen, N. Zhao, and Z. Yang, ``A robust modulation classi_cation method using convolutional neural networks," *EURASIP J. Adv. Signal Process.*, vol. 2019, no. 1, pp. 1_15, Dec. 2019.

[25] T. J. O'Shea, T. Roy, and T. C. Clancy, “Over-the-air deep learning based radio signal classification,” *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 1, pp. 168-179, Feb. 2018, doi: 10.1109/JSTSP.2018.2797022.

[26] K. He, X. Zhang, S. Ren, J Sun, "Deep Residual Learning for Image Recognition", *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[27] M. Tan, and Q. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks", in *proceedings International Conference on Machine Learning (ICML)*, pp: 6105-6114, 2019.

[28] V. Agarwal, "Complete Architectural Details of all EfficientNet Models", [Online]. Available: https://medium.com/data-science/complete-architectural-details-of-all-efficientnet-models-5fd5b736142

[29] H. Jing, "Illustrated Guide to Transformer", 2020. [Online]. Available: https://jinglescode.github.io/2020/05/27/illustrated-guide-transformer/

[30] Colah's blog, "Understanding LSTM Networks", 2015. [Online]. Available: https://colah.github.io/posts/2015-08-Understanding-LSTMs/

[31] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. Gomez, Ł. Kaiser, "Attention Is All You Need", in *Advances in Neural Information Processing Systems*, 2017. [Online]. Available: https://arxiv.org/pdf/1706.03762.pdf

[32] M. Komatsu, N. Tanabe and H. Matsumoto, "A Consideration on Noise Variance Estimation for Direct Blind Equalization with OFDM Environment," *2024 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)*, Kaohsiung, Taiwan, 2024, pp. 1-4, doi: 10.1109/ISPACS62486.2024.10868237.

[33] Szymon Greś, Michael Döhler, Vasilis K. Dertimanis, Eleni N. Chatzi, "Subspace-based noise covariance estimation for Kalman filter in virtual sensing applications," *Mechanical Systems and Signal Processing*, Volume 222, 2025, 111772.

[34] 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*, vol. 8, pp. 42841-42847, 2020.

[35] Y. Wang, J. Yang, M. Liu, and G. Gui, "LightAMC: Lightweight automatic modulation classification via deep learning and compressive sensing," *IEEE Trans. Veh. Technol.*, vol. 69, no. 3, pp. 3491-3495, Mar. 2020.

[36] S. Zheng, P. Qi, S. Chen, and X. Yang, "Fusion methods for CNNbased automatic modulation classification," *IEEE Access*, vol. 7, pp. 66496-66504, 2019.