

Ensemble of Convolution Neural Networks for Improving Automatic Modulation Classification Performance

Ha-Khanh Le, Van-Sang Doan*, Van-Phuc Hoang

Abstract—This paper investigates convolutional neural networks (CNN) to classify 26 types of signal modulation under the influence of five different fading channels and Gaussian noise with SNR from -20 dB to +18 dB. Specifically, five CNN models, including ResNet18, SqueezeNet, GoogleNet, MobileNet, and RepVGG, are taken into account for an accuracy competition to discover the best one. As a result, the SqueezeNet model achieves the highest accuracy of 97.5% for the SNR value of +8 dB. Based on the evaluation results of the single models, we propose an ensemble learning approach, which integrates some robust networks to improve classification accuracy. The numerical results show that ensemble learning can improve the automatic modulation classification accuracy compared to those single models. Specifically, the ensemble learning model gains the accuracy of 52.7% at the SNR of -20 dB and 77% at the SNR of -2 dB. In addition, three types of ensemble methods are considered for analysis and comparison. Consequently, the weighted ensemble provides a better performance in terms of accuracy than unweighted one.

Index Terms—Automatic modulation classification; Deep learning; Convolutional neural network; ensemble learning.

1. Introduction

AUTOMATIC modulation classification (AMC) is an important task that holds the opportunity for many different applications in civil and military scenarios. It is an intermediate step between signal detection and demodulation for determining the modulation scheme of a radio signal [1]. However, designing a classifier that works well in noisy environments and other conditions is very challenging. Generally, AMC algorithms can be divided into two categories: Likelihood-based (LB) and Feature-based (FB). The LB methods can achieve an optimal classification accuracy when they apply in perfect channel models and priority-known parameters. Fundamentally, the LB methods compare the probability ratio of the received signal in the group of considered modulations. However, the LB methods require the priority knowledge of channel parameters, which make the computational burden heavier [2]. In contrast, the FB methods can also obtain high accuracy but less computational complexity, especially, they are independent with the

channel information. The performance of FB methods mainly depends on extracted features, which are regularly handcrafted by experts to suit the channel environments. Furthermore, searching the efficient features requires a significant concern in terms of data [3]. In recent years, deep learning has evolved continuously and outstripped other approaches. It can be employed with both manual and automatic processes to learn representative features of entities from the raw data [4]. Current models, such as ResNet [5], DenseNet [6], CLDNN [7], etc, demonstrate that deep learning is a new modern approach for an efficient signal processing. These models do not even need pre-processing and denoising for classifying the signals. However, they still exist some limitations with low SNR signals or under multi-path fading conditions.

In order to improve the accuracy of modulation classification, the requirement for CNN models should consider higher computational complexity, larger structures such as deeper and wider models, which consequently results in the increase in the learnable parameters. As a result, the trade-off for training would cost a lot of mathematical operations, computing memories, hardware and time delay [8]. It is reported that an ensemble learning method can help to leverage the advantages of single models. Indeed, it can combine several simple algorithms but still improves modulation classification accuracy compared to a large model. In the ensemble approach, CNN models are trained for an identical task and dataset; then, their final prediction is judged by some specific rules of output scores. Some advantages of ensemble learning method in increasing signal classification performance can be listed as followed:

Ha-Khanh Le is with the Institute of System Integration, Le Quy Don Technical University, Hanoi 11300, Vietnam (e-mail: lhkhanhhvk-tqs@gmail.com).

Van-Sang Doan is with the Faculty of Communication and Radar, Naval Academy, Nha Trang 650000, Vietnam (e-mail: doansang.g1@gmail.com).

Van-Phuc Hoang is with the Institute of System Integration, Le Quy Don Technical University, Hanoi 11300, Vietnam (e-mail: phuchv@lqdtu.edu.vn).

*Corresponding author: Van-Sang Doan (e-mail: doansang.g1@gmail.com)

Manuscript received December 6, 2021; revised December 31, 2021; accepted March 07, 2022.

Digital Object Identifier 10.31130/ud-jst.2022.293E

- *Avoid over-fitting*: A CNN model can achieve high accuracy when training on a small dataset because its weights over-fit the dataset, but the trained model provides inaccurate results on unseen new data. This phenomenon is called an over-fitting problem. By combining the scores of different algorithms for final judgement, the ensemble method facilitates reducing the over-fitting level; therefore, it can improve the overall predictive performance.
- *Avoid the local minimum trap*: Combining different models can reduce the risk of reaching the local minimum due to the diversity of learning processes. In addition, there are fewer local minima in learning of small models than that of large models. Therefore, it can reduce the probability of falling into the local minimum traps.

In some previous works, the dataset with Gaussian noise is often used for the automatic modulation classification, such as RadioML 2018 in [9]. Nevertheless, the signals in practice are often affected by different types of noise, especially multi-path fading channels. For that reason, we use another dataset, namely HisarMod2019.1 in [10], in this research work for a more comprehensive evaluation of the actual signal classifiers. In this work, we assess different networks for the modulation classification task on the mentioned dataset. Afterwards, several ensemble learning methods are taken into account to compare with the single models to exhibit the robustness of ensemble learning in the signal modulation classification. Simulation shows that ensemble learning models obtain accuracy higher than the individual ones at low SNR, for example about 52.7% for SNR = -20 dB, or 77% at SNR = -2 dB.

The rest of this paper is organized as follows. Section II briefly describes the considered neural networks and proposes ensemble learning techniques for signal modulation classification. Then, the comparison results in terms of classification performance between models using the single and ensemble learning approaches are discussed in Section III. Finally, Section IV will conclude the results and directions of future works.

2. Deep Learning Networks for Automatic Modulation Classification

Recent studies of deep learning networks are focusing on improving signal classification accuracy using the state-of-the-art models, such as ResNet [5], SqueezeNet [11], MobileNet [12], GoogleNet [13], and VGG [14]. Therefore, this work studies these models in terms of classification accuracy to reveal their advantages and disadvantages. Then, the ensemble model is proposed and applied to improve the accuracy of automatic modulation classification.

2.1. Overview of Existing Networks

2.1.1. ResNet

The ResNet, a CNN model, is presented in ImageNet and COCO 2015 [5]. The network has resolved the

vanishing and over-fitting problems by creating skip-connections between different layers. The key of ResNet is residual blocks as shown in Fig. 1, where we can see that the feature map from the output of weight layers is combined with input feature map (from skip-connection) via an addition layer. This idea is to re-use the former feature map, which can help to improve the classification accuracy.

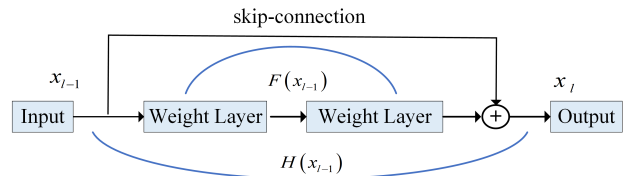


Fig. 1: Residual Block.

2.1.2. SqueezeNet

SqueezeNet is a network proposed in 2016 [11] to classify images. Despite occupying fewer parameters, SqueezeNet gains the same accuracy compared to some other well-known models. The structure of SqueezeNet is shown in Fig. 2a, which consists of a convolutional layer (Conv1), 8 blocks (from fire2 to fire9) and a convolutional layer (Conv10). The number of filters in convolutional layers gradually increases from 16 filters at the beginning to 256 filters at the end of the network. It can be observed from Fig. 2b that the fire block is constructed by a “squeeze” layer, which is the convolutional layer of 1×1 filters, and two “extended” layers, which are convolutional layers of 1×1 and 3×3 filters. Reducing the number of filters in the “squeeze” layer helps to decrease the model learnable parameters.

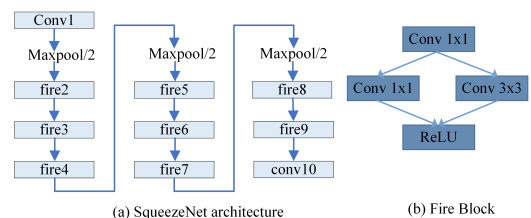


Fig. 2: SqueezeNet Block.

2.1.3. MobileNet

MobileNet is a lightweight network model designed by a Google team [12] for applying on compact devices with limited resources. Not only on optimising latency, but the MobileNet model also focuses on small structures for increasing speed. The structure of MobileNet is built by a convolution method, so-called Depthwise Separable Convolution (DSC), to reduce the model size and computational complexity as presented in [12].

2.1.4. GoogLeNet

The GoogLeNet is a network model launched in 2014 [13]. The model focuses on the problem of finding which size of the convolutional filter is the best. Some good

results can sometimes be achieved by combining filters of different sizes. The basic convolution block in the GoogleNet model is called Inception, as shown in Fig. 3, where we can see that there are four parallel branches in the Inception block.

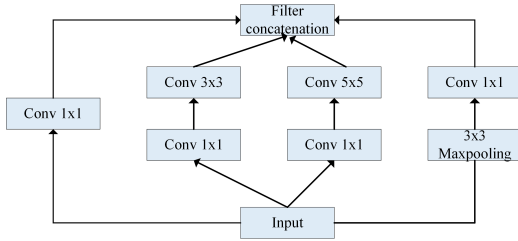


Fig. 3: Inception Block.

2.1.5. RepVGG

RepVGG network is developed from the VGG network to have a simple structure that ensures efficiency, where the max-pooling layer is not used [15]. It used method called reparameterisation. This is a technique to transform a set of parameters from one architecture to another, so even though the two architectures are different, they can still share the weights. The RepVGG model comprises five stages; each stage will include structurally similar blocks, as shown in Fig. 4.

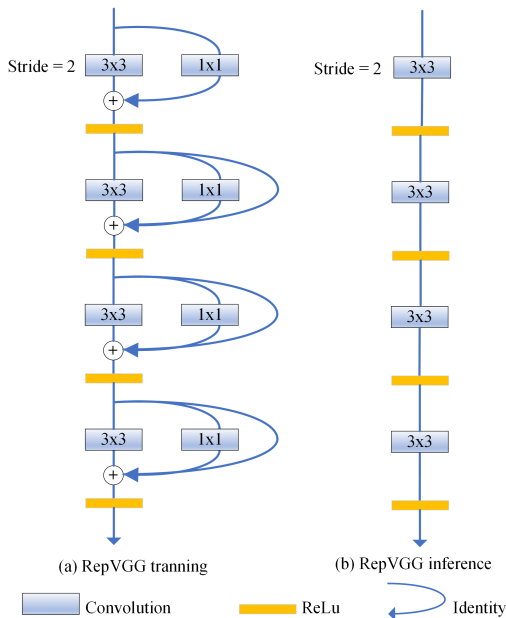


Fig. 4: RepVGG Block.

2.2. Proposed Ensemble Method

The ResNet, MobileNet, GoogleNet, SqueezeNet, and RepVGG models are popularly applied in many fields. However, when using these models for the modulation signal classification task, the signal classification accuracy at the low SNRs is low, for instance, less than 60% accuracy for $\text{SNR} \leq -2$ dB of all models. Meanwhile,

with high SNRs, some model achieves much better accuracy; concretely, the ResNet18 model achieves more than 92.67% for $\text{SNR} \geq +6$ dB, as shown in Fig. 5. In order to eliminate the above-mentioned issue, we use ensemble learning to improve the accuracy of models on the same data set because of its diversity. Indeed, models in an ensemble one will have different predictive abilities, so that a good combination will be more efficient than an single model. Therefore, it can improve overall performance compared to using the models individually [8], [16]. Ensemble learning is a method that incorporates the predictions from all the base learners or creates an ensemble of well-chosen strong and diverse models. Ensemble models gain more accuracy and robustness by combining data from numerous modeling approaches.

In this section, we propose a method that combines a number of CNN models to provide a higher AMC accuracy. With performing an ensemble classifier, we assess how the classification capability of an ensemble classifier outperforms each individual classifier. Accordingly, three ensemble techniques, including majority voting and two mean probability (weighted and unweighted) ones are utilized, as follows:

2.2.1. Unweighted Average

In this method, the Softmax function is applied to calculate the predicted probability value at the output of the final class of CNN models. Then, an unweighted average of the probability values of the models is computed. As a result, the decision for the highest probability will subsequently be made [17]. The unweighted average formula is defined as follows:

$$p_j = \frac{1}{n} \sum_{i=1}^n y_{ij} \quad (1)$$

where y_{ij} is the score vector of the j^{th} modulation class and n is the number of CNN models.

2.2.2. Unweighted majority vote

In this method, instead of averaging, the highest probabilities of all CNN models is firstly taken at the output. Then, they are voted by counting the majority from all of the predicted labels and the final decision is made afterward. The unweighted majority vote formula is expressed as follows:

$$\hat{y}_{ij} = \begin{cases} y_{ij} & \text{for } y_{ij} = \max(\mathbf{y}_i) \\ 0 & \text{for } y_{ij} \neq \max(\mathbf{y}_i) \end{cases} \quad (2)$$

$$p_j = \frac{1}{n} \sum_{i=1}^n \hat{y}_{ij} \quad (3)$$

where \hat{y}_{ij} the highest probability value of the score vector y_{ij} and n is the number of CNN models.

2.2.3. Weighted average

The weighted average method is implemented by multiplying the different weighted values at the CNN outputs, supposed that the sum of the weighted values

is equal to one. The equation of the weighted average method is given:

$$p_j = \frac{1}{n} \sum_{i=1}^n \alpha_{ij} y_{ij} \quad (4)$$

where y_{ij} is the score vector of j^{th} modulation class and n is the number of CNN models.

3. Results

3.1. Dataset

In some previous works, the dataset with Gaussian noise is often used for the automatic modulation classification, such as RadioML 2018 in [9]. Nevertheless, the signals in practice are often affected by different types of noise, especially multi-path fading channels. For that reason, we use another dataset, namely HisarMod2019.1 in [10], in this research work for a more comprehensive evaluation of the actual signal classifiers. The dataset includes 26 modulation types from 5 different modulation families: analogue modulations, frequency shift keying (FSK) modulations, pulse amplitude modulations (PAM), phase shift keying (PSK) modulations, and quadrature-phase modulations (QAM). All modulation types can be summarized as follows:

- Analog modulation: AM-DSB, AM-SC, AM-USB, AM-LSB, FM, PM.
- FSK modulation: 2FSK, 4FSK, 8FSK, 16FSK.
- PAM modulation: 4PAM, 8PAM, 16PAM.
- PSK modulation: BPSK, QPSK, 8PSK, 16PSK, 32PSK, 64PSK.
- QAM modulation: 4QAM, 8QAM, 16QAM, 32QAM, 64QAM, 128QAM, 256QAM.

The dataset provides wireless signals under ideal, static, Rayleigh, Rician with $k = 3$ and Nakagami- m with $m = 2$ channel conditions. Channels that have additive white Gaussian noise (AWGN) only are called ideal channels. For a static channel, its coefficients are set randomly at the beginning and remain unchanged during the propagation time. In Rayleigh fading channel, there are reflected, scattered, and diffracted components of the incoming signal at the receiver without line-of-sight signal. In contrast, for Rayleigh fading channel, the distribution is Rice with shape parameter of $k = 3$ including the light of sight signal to the receiver. Furthermore, the distribution of received power is selected as Nakagami- m with a shape parameter of $m = 2$. As a result, all of the signals in the dataset are set up with different fading models. Therefore, more realistic channel conditions can be learned by the DL-based AMC methods. In the dataset, each modulation type has 1500 signals of 1024×2 I/Q sample length. In total, the dataset has 780,000 signals covering 26 modulation types with the signal to noise ratios from -20 dB to +18 dB, steps of 2 dB. The dataset is divided into 520,00 signals (occupy 80%) for training and the rest (20%) for testing. The input data to the model is an I/Q data array structure of size 1024×2 . The AMC accuracy is measured on the test dataset. The models are trained

with 10 epochs, a mini-batch size of 64, and an initial learning rate of 0.001. The device used for the simulation is a computer with a 3.70 GHz CPU, 2x16GB RAM, and an NVIDIA GeForce RTX 3060ti GPU.

3.2. Network Performance Metrics

The performance metrics, including Precision, Recall, F1-Score, and Accuracy, are used to evaluate and compare different models in this paper [18]. While Precision is the ratio of correct modulation classification of a class to the total number observations of that class, Recall is the ratio of correct classification of a class to the all observations in actual class. F1-Score is the weighted average of Precision and Recall. Accuracy is the most intuitive performance measure of deep neural networks and it is a ratio of correct modulation prediction to the total observations. These metric parameters are computed as follows:

$$\begin{aligned} \text{Accuracy} &= \frac{TP+TN}{TP+TN+FP+FN} \\ \text{Precision} &= \frac{TP}{TP+FP} \\ \text{Recall} &= \frac{TP}{TP+FN} \\ \text{F1 - score} &= 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \end{aligned} \quad (5)$$

where TP is the True Positive, FP is the False Positive, TN is the True Negative, and FN is the False Negative.

3.3. Discussion of Results

3.3.1. Comparison between different CNN models

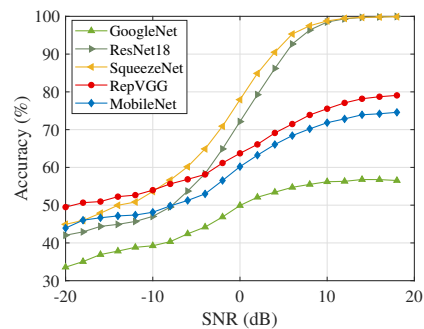


Fig. 5: Comparison of the classification accuracy of different single CNN models.

In this section, we use Matlab as a simulation tool to evaluate the effectiveness of ResNet18, SqueezeNet, MobileNet, GoogleNet, and RepVGG networks. The comparison result in terms of AMC accuracy of the different models is shown in Fig. 5, where we observe that the GoogleNet model obtains the lowest accuracy. Its highest accuracy is only 56.76% at +18 dB SNR. Two models MobileNet and RepVGG gain significantly higher accuracy than GoogleNet by about 10% and 15% for SNR from -20 dB to +18 dB, respectively. ResNet and SqueezeNet achieve very good modulation classification accuracy for the high SNRs from +2 dB to

+18 dB. Especially, the SqueezeNet model yields the highest correct modulation classification rate compared to other models, specifically 97.5% for $\text{SNR} \geq +8$ dB.

3.3.2. Analysis of ensemble models and their comparison with single models

From the evaluation results of single models, we propose to combine the models to improve the modulation classification accuracy. Accordingly, we combine the considered models to build the following ensemble models: EnCNN5 (combined by ResNet18, SqueezeNet, GoogleNet, MobileNetV2, and RepVGG), EnCNN3 (combined by ResNet18, SqueezeNet, and RepVGG), and EnCNN2 (combined by RepVGG and SqueezeNet). These ensemble models aim to improve modulation classification accuracy, especially at the low SNRs from -20 dB to +2 dB.

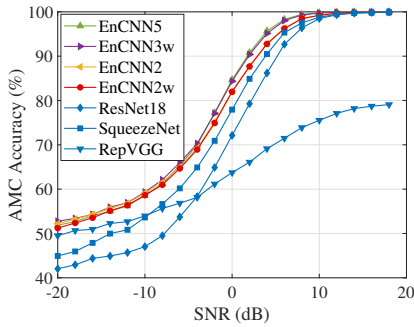


Fig. 6: Accuracy of different ensemble and single models.

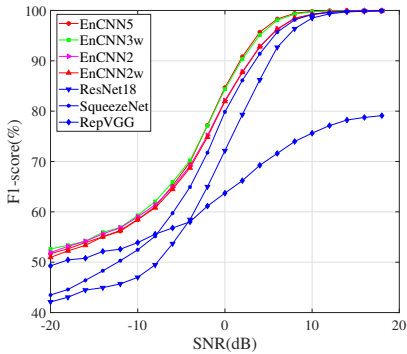


Fig. 7: F1-score of different ensemble and single models.

Experimental results in Fig. 6 show that the three ensemble methods can increase the AMC accuracy. Specifically, four models EnCNN5, EnCNN2 (using unweighted average), EnCNN3w, and EnCNN2w (using weighted average) obtain higher accuracy than single individual ones at SNR from -20 dB to +18 dB. With SNRs from -20 dB to +2 dB, the ensemble models significantly outperform the ResNet18 and RepVGG models about 10% and 4% accuracy, respectively. With the high SNRs, the ensemble models gain about 20% higher accuracy than RepVGG, and 8% higher than SqueezeNet for SNRs from +2 dB to +8 dB. From the combination of CNN models, we find that the model that combines the

SqueezeNet, ResNet18 and RepVGG using the weighted average method yields the highest AMC accuracy compared to other considered ones. Specifically, its accuracy is achieved 52.2% at $\text{SNR} = -20$ dB and 99% at $\text{SNR} = +18$ dB.

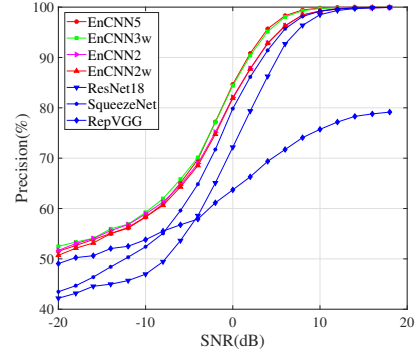


Fig. 8: Precision of different ensemble and single models.

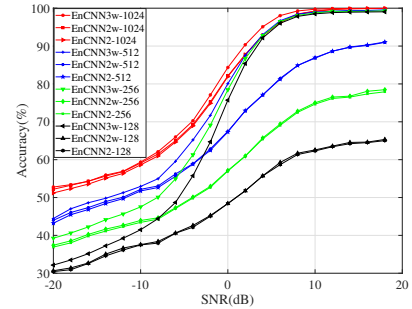


Fig. 9: Performance of ensemble models with different lengths of signal.

Besides comparing the average accuracy, we also use the F1-score and Accuracy to measure the model performance. The results in Fig. 7 and Fig. 8 indicate that the EnCNN3w model has the highest overall micro-averaged F1-scores and Precision. This result is reasonable because the SqueezeNet, ResNet18 and RepVGG models have remarkably better performance than other considered ones. Fig. 6, Fig. 7, and Fig. 8, show that the ensemble models improve the signal classification accuracy significantly, especially at the low SNR segment. Specifically, the accuracy yield 51.2% at -20 dB SNR and over 75% for $\text{SNR} \geq -2$ dB. Meanwhile, the single models provide lower modulation accuracy than the ensemble ones, as shown in Fig. 5. Concretely, the RepVGG model provide only 49% correct modulation classification at $\text{SNR} = -20$ dB; and the SqueezeNet model obtain only 68% accuracy at $\text{SNR} = -2$ dB. The classification results of the single models indicate that SqueezeNet and RepVGG have higher signal classification accuracy at low SNR than other ones, in which RepVGG is the highest. In contrast, ResNet and SqueezeNet achieve the highest correct classification probability at high SNRs. The above evaluation reveals that the ensemble learning method can help improving the classification performance in terms of accuracy.

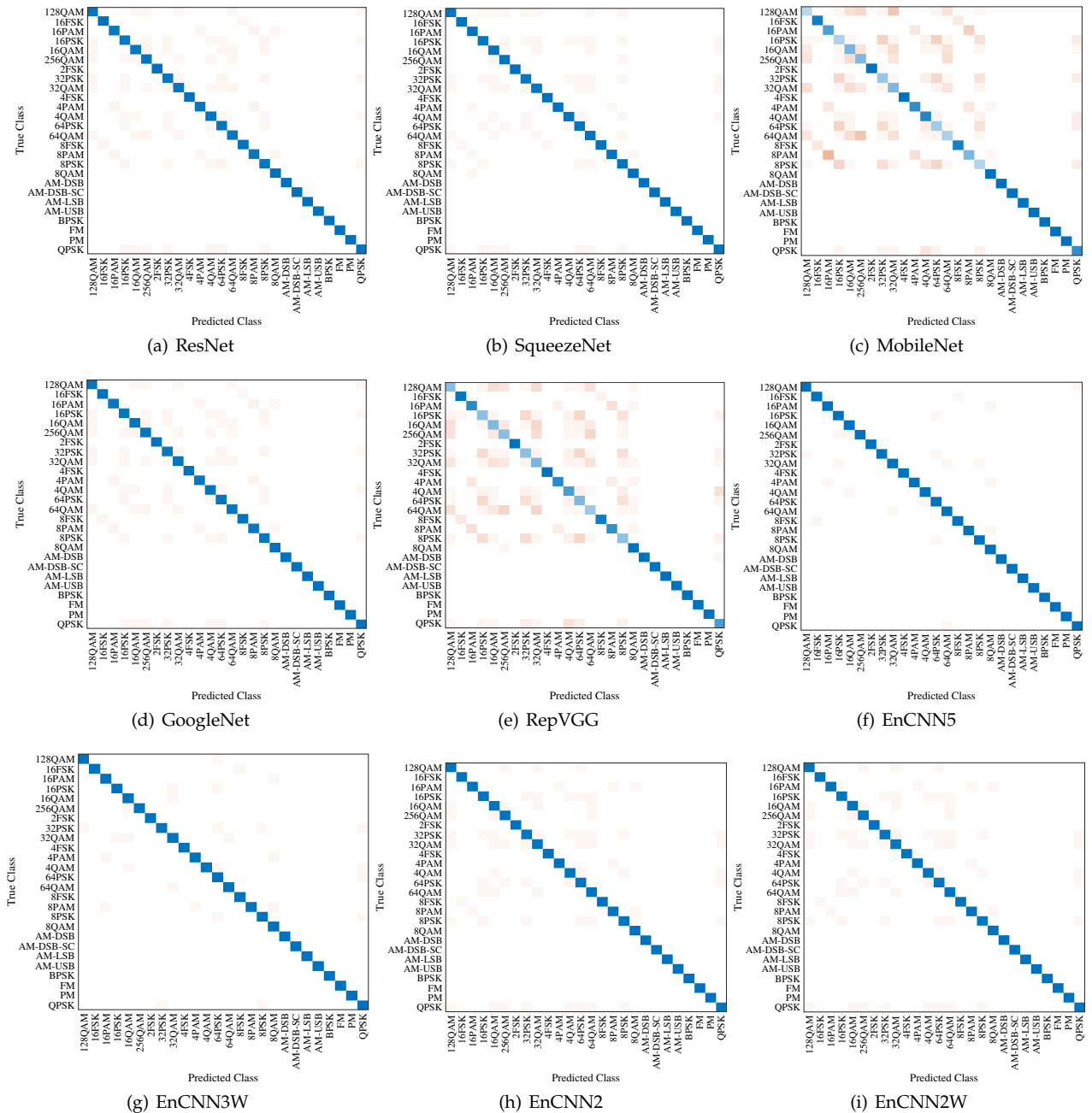


Fig. 10: Matrix confusion different single models at the SNR of +10 dB.

TABLE 1: Computational speed.

Model	Time (ms)
EnCNN5	3.22
EnCNN3w	2.15
EnCNN2	0.62
EnCNN2w	0.62

Accordingly, this paper compares different models through computational speed as shown in Table 1. Here, it can be seen that the EnCNN5 model has the most significant number of parameters include five models, so it also has an enormous processing time (3.22 ms). Meanwhile, the EnCNN2 and EnCNN2w models have

the same number of parameters and fastest processing time (0.62 ms) because they only use two models to combine.

In the next simulation, we build three types of ensemble model, including EnCNN3w, EnCNN2, and EnCNN2w, from the aforementioned CNN models to analyze the performance of these three ensemble models with varying signal lengths, such as 128, 256, 512, and 1024. The results in Fig. 9 show that changing the signal length at the input of the models causes the significant change of the signal classification accuracy. When the signal length is 1024, we can see that the ensemble models have similar accuracy and are not

significantly different. However, as the input length signal is reduced, the classification accuracy gap of the three EnCNN3, EnCNN2, EnCNN2w models changes significantly. Specifically, the signal length is 128, 256, and 512, the EnCNNw model always has the highest accuracy and is higher than the other two models by 28%, 22%, and 12%, respectively. It can be concluded that the longer the signal is, the more representative information for individual modulation is extracted. Consequently, the higher modulation classification accuracy can be obtained. Specifically, the models with length 1024 give the highest classification accuracy, whereas the models with length 128 give the lowest classification accuracy.

In addition, confusion matrices of 26 modulations of the single models and ensemble models at the SNR of +10 dB are shown in Fig. 10. From the confusion matrices, it can be seen that some lower-order modulation types, such as AM and FM have less confusing classification results, specifically more than 85% and 99% accuracy at +10 dB SNR for the single modes and ensemble models. With higher-order modulations, such as PSK and QAM, the accuracy of single models is less than 70% at +10 dB SNR; meanwhile, ensemble models have an accuracy of around 95.4%. It can be seen that although the high-order modulated signals give faster transmission speed, they cannot be used for the far distance of communication because the AMC accuracy is quite low. Still, the modulation classification changes a lot as the error rate increases because the signal constellation distribution is close together. When affected by noise, the modulation classification efficiency will be degraded. Therefore, the combination of single models has significantly improved the signal classification accuracy, especially for high-order modulation signals such as PSK or QAM.

4. Conclusion

This paper has demonstrated that the ensemble deep learning approach can help to improve the accuracy of modulation classification. Specifically, we have analyzed three types of ensemble methods and compared them with each other and with other single models. Simulation results indicate that the ensemble models have remarkably outperformed the single ones, especially for the low SNR values. In addition, the weighted ensemble provided better performance in terms of accuracy than unweighted one. In the future work, we will develop the ensemble method for lightweight models, which can ensure a high speed and a high AMC accuracy. Moreover, the experimental measurements will be performed to verify the proposed method.

Acknowledgment

This work is funded by Vietnam - Czech bilateral project "NEO classification of signals (NEOCLASSIG) for radio surveillance systems" under grant number NDT/CZ/22/12.

References

- [1] J. Ma, S.-C. Lin, H. Gao, and T. Qiu, "Automatic modulation classification under non-Gaussian noise: A deep residual learning approach," *Proc. IEEE Int. Conf. Commun. (ICC)*, pp. 1–6, May 2019.
- [2] A. Maxwell, R. Li, B. Wang et al., "Deep learning architectures for multi-label classification of intelligent health risk prediction," *BMC Bioinformatics*, vol. 18, no. 14, pp. 121–131, Mar. 2017.
- [3] J. L. Xu, W. Su, and M. Zhou, "Likelihood-ratio approaches to automatic modulation classification," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 41, no. 4, pp. 455–469, Jul. 2011.
- [4] T. Huynh-The, C. H. Hua, V. S. Doan, and D. S. Kim, "Accurate modulation classification with reusable-feature convolutional neural network," *Proc. 2020 IEEE Eighth International Conference on Communications and Electronics (ICCE)*, Phu Quoc Island, Vietnam, pp. 12–17, Jun. 2021.
- [5] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Proc. 2016 IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, USA, pp. 770–778, Jun. 2016.
- [6] G. Huang, Z. Liu, and K. Q. Weinberger, "Densely connected convolutional networks," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2261–2269, Nov. 2017.
- [7] T.J. O'Shea, J. Corgan, T. C. Clancy, "Convolutional Radio Modulation Recognition Networks," *Communications in Computer and Information Science*, vol. 629, pp. 213–226, Jun. 2016.
- [8] O. Sagi and L. Rokach, "Ensemble learning: A survey," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 8, pp. 1–18, Jun. 2018.
- [9] S. Ramjee, S. Ju, D. Yang, X. Liu, A. E. Gamal, and Y. C. Eldar, "Robust and Fast Automatic Modulation Classification," *IEEE Machine Learning for Communications Emerging Technologies Initiatives (MLCETI)*, Jan. 2019.
- [10] K. Tekbiyik, A. R. Ekti, A. Görçin, G. K. Kurt, C. Keçeci, "Robust and Fast Automatic Modulation Classification with CNN under Multipath Fading Channels," *2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring)*, pp. 1–6, May. 2020.
- [11] F. N. Iandola, H. Song, M.W. Moskewicz, K. Ashraf, W. J. Dally, and K. Kurt, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size," *arXiv preprint arxiv.1602.07360*, Nov. 2016.
- [12] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottleneck," *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4510–4520, Jun. 2018.
- [13] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolution," *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1–9, Jun. 2015.
- [14] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, Apr. 2015.
- [15] X. Ding, X. Zhang, N. Ma, J. Han, G. Ding, J. Sun, "RepVGG: Making VGG-style ConvNets Great Again," *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 13728–13737, Mar. 2021.
- [16] G. Tsoumakas and I. Vlahavas, "Random k-labelsets: An ensemble method for multilabel classification," *European conference on machine learning*, vol. 4701 pp. 406–417, Sep. 2007.
- [17] J. M. Moyano, E. L. Gibaja, J. C. Krzysztof, and S. Ventura, "An evolutionary approach to build ensembles of multi-label classifiers," *Information Fusion*, pp. 168–180, October. 2019.
- [18] H. Gu, Y. Wang, S. Hong, and G. Gui, "Blind channel identification aided generalized automatic modulation recognition based on deep learning," *IEEE Access*, vol. 7, pp. 110722–110729, August. 2019.



Ha-Khanh Le received the B.Sc. and M.Sc. degrees in electronic and telecommunication from Le Quy Don Technical University, Hanoi, Vietnam. He is now a PhD candidate at Institute of System Integration, Le Quy Don Technical University, Hanoi, Vietnam. His current research interest includes communication systems, radio signal processing, and deep learning.



Van-Phuc Hoang received PhD degree in Electronic Engineering from The University of Electro-Communications, Tokyo, Japan in 2012. He has worked as postdoc researcher, visiting scholar at The University of ElectroCommunications, Tokyo, Japan, Telecom Paris, France and University of Strathclyde, Glasgow, UK during the period of 2012-2018. He is working as an Associate Professor, Director with Institute of System Integration, Le Quy Don Technical University, Hanoi, Vietnam. He is also serving as Vice Chair in International Affairs and Conferences, Radio Electronics Association of Vietnam (REV). His research interests include advanced signal processing, hardware security, digital circuits and systems, embedded systems for Internet of Things, and VLSI architecture for digital signal processing. He was the PI of 02 NAFOSTED funded projects and one World Bank funded project in hardware security. He was the Technical Program Chair of several IEEE international conferences such as ICDV 2017, MCSoc 2018, SigTelCom 2019, APCCAS 2020 and ATC 2020. He is a member of IEEE.



Van-Sang Doan received the M.Sc. and Ph.D. degrees in electronic systems and devices from the Faculty of Military Technology, the University of Defence in Brno, Czech Republic in 2013, and 2016, respectively. He was three times awarded the Honors degrees by Faculty of Military Technology, the University of Defence in Brno, in 2011, 2013, and 2016. He received a Post-Doctoral Research Fellow with ICT Convergence Research Center at Kumoh National Institute of Technology, Republic of Korea from 2019 to 2020. He is currently working at the Faculty of Communication and Radar, Vietnam Naval Academy in Nha Trang City, Khanh Hoa Province, Vietnam. His current research interest includes radar, sonar, and communication systems, signal processing, and deep learning.