

# Automatic Modulation Classification for MIMO Systems via Deep Learning and Zero-Forcing Equalization

Yu Wang, Jie Gui, *Senior Member, IEEE*, Yue Yin, Juan Wang, Jinlong Sun, Guan Gui, *Senior Member, IEEE*, Haris Gacanin, *Senior Member, IEEE*, Hikmet Sari, *Fellow, IEEE*, and Fumiyuki Adachi, *Life Fellow, IEEE*

**Abstract**—Automatic modulation classification (AMC) is one of the most critical technologies for non-cooperative communication systems. Recently, deep learning (DL) based AMC (DL-AMC) methods have attracted significant attention due to their preferable performance. However, the study of most of DL-AMC methods are concentrated in the single-input and single-output (SISO) systems, while there are only a few works on DL-based AMC methods in multiple-input and multiple-output (MIMO) systems. Therefore, we propose in this work a convolutional neural network (CNN) based zero-forcing (ZF) equalization AMC (CNN/ZF-AMC) method for MIMO systems. Simulation results demonstrate that the CNN/ZF-AMC method achieves better performance than the artificial neural network (ANN) with high order cumulants (HOC)-based AMC method under the condition of the perfect channel state information (CSI). Moreover, we also explore the impact of the imperfect CSI on the performance of the CNN/ZF-AMC method. Simulation results demonstrated that the classification performance is not only influenced by the imperfect CSI, but also associated with the number of the transmit and receive antennas.

**Index Terms**—Automatic modulation classification, deep learning, zero-forcing equalization, channel state information, multiple-input and multiple-output systems.

## I. INTRODUCTION

Automatic modulation classification (AMC) is one of the most critical techniques for non-cooperative communication for analyzing the unauthorized users [1], [2]. AMC is generically modeled as a pattern recognition problem, and the traditional AMC methods are based on efficient classifier designs [3]. Specifically, signal features are extracted from the signal, and then apply support vector machine (SVM)

This work was supported by the Project Funded by the National Science and Technology Major Project of China under Grant TC190A3WZ-2, the National Natural Science Foundation of China under Grant 61671253, the Jiangsu Specially Appointed Professor under Grant RK002STP16001, the Innovation and Entrepreneurship of Jiangsu High-level Talent under Grant CZ0010617002, the Six Top Talents Program of Jiangsu under Grant XYDXX-010, the 1311 Talent Plan of Nanjing University of Posts and Telecommunications. (*Corresponding author: Guan Gui*)

Y. Wang, Y. Yin, J. Wang, J. Sun and G. Gui, and H. Sari are with the College of Telecommunications and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China (e-mails: {1018010407, 1018010408, 1219012920, sunjinlong, guiguan, hikmet}@njupt.edu.cn).

J. Gui is with the Department of Computational Medicine and Bioinformatics, University of Michigan, MI 48109, USA (e-mail: guijie@umich.edu).

H. Gacanin is with the Nokia Bell Labs, 2018 Antwerp, Belgium (e-mail: harisg@ieee.org).

F. Adachi is with the Research Organization of Electrical Communication (ROEC), Tohoku University, Sendai 980-8577 Japan (e-mail: adachi@ecei.tohoku.ac.jp).

or traditional artificial neural network (ANN) to classify the modulation types [4]. In addition, these features can represent different modulation types. The modern features includes high order cumulants (HOC), instantaneous frequency features, wavelet transformation (WT) features, and so on. The most common combination of the traditional AMC method is ANN with HOC [4], [5], which is applied into both single-input and single-output (SISO) systems and multiple-input and multiple-output (MIMO) systems.

Recently, deep learning (DL) has emerged as one of the most powerful tools for classification [6]–[10]. Thus, DL has been applied into various communication technologies [11]–[13], e. g., beam management [14], resource allocation [15], [16], non-orthogonal multiple access (NOMA) [17], traffic control [18]–[20], to enhance physical layer and network layer communication [21]. DL can be divided in two different categories. The first one is based on the in-phase and quadrature (IQ) components of signals. T. J. O Shea *et al.*, firstly proposed a convolutional neuron network (CNN)-based AMC method, trained on a large number of IQ samples, and achieved outstanding performances [22]. Then, various neural networks, such as long short-term memory network (LSTM) [24] and convolutional long short-term deep neural networks (CLDNN) [23], were proposed for AMC under various noise conditions. The other is the constellation diagram-based AMC methods, where the trimmed CNN-based supervised AMC method and generative adversarial network (GAN)-based semi-supervised AMC method have been proposed [25], [26], respectively.

In this correspondence, we propose a CNN-based zero-forcing (ZF) equalization AMC (CNN/ZF-AMC) method for MIMO systems. The ZF equalization technology is adopted to enhance the classification performance under the perfect CSI and the imperfect CSI. In the former case, we compare the perfect CSI-aided CNN/ZF-AMC method with the traditional methods. Our results reveal the huge advantage of the CNN/ZF-AMC method. In the latter case, the imperfect CSI is generated by the channel error model rather than the estimated CSI to study the factors affecting the classification performance.

## II. SYSTEM MODEL

Assuming that the MIMO channel is a time-invariant complex-valued MIMO channel, the received signal at the  $n$ -th

sampling time can be presented as follows.

$$\mathbf{R}(n) = \mathbf{H}\mathbf{T}(n) + \mathbf{G}(n), \quad (1)$$

where  $\mathbf{H}$  is the MIMO channel matrix of size  $N_r \times N_t$ , ( $N_r \geq N_t$ ), and it obeys the circular symmetric complex normal distribution with zero mean and unit variance;  $\mathbf{R}(n) = [R_1(n), R_2(n), \dots, R_{N_r}(n)]^T$  is the ( $N_r \times 1$ ) received signal vector, obtained perfectly by Nyquist sampling without phase offset and frequency offset;  $\mathbf{T}(n) = [T_1(n), T_2(n), \dots, T_{N_t}(n)]^T$  is the ( $N_t \times 1$ ) transmitted signal vector, and  $E[\mathbf{T}(n)\mathbf{T}^H(n)] = E_T \mathbf{I}_{N_t}$ , where  $[\cdot]^H$  is the conjugate transpose operation;  $\mathbf{G}(n)$  is the additive white Gaussian noise (AWGN), the element of which obeys the circular symmetric complex normal distribution with zero mean and  $E_G$  variance.

Equalization is applied to reveal the ambiguity of the received signal sequence [4], and ZF equalization is considered. The received signal via ZF equalization can be written as

$$\hat{\mathbf{R}}(n) = \text{ZF}(\hat{\mathbf{H}})\mathbf{R}(n), \quad (2)$$

where  $\text{ZF}(\hat{\mathbf{H}}) = \hat{\mathbf{H}}^\dagger = (\hat{\mathbf{H}}^H \hat{\mathbf{H}})^{-1} \hat{\mathbf{H}}^H$  is the equalization matrix, where  $(\hat{\mathbf{H}}^H \hat{\mathbf{H}})^{-1} \hat{\mathbf{H}}^H$  is denoted as the pseudo inverse operation of  $\hat{\mathbf{H}}$ . In addition,  $\hat{\mathbf{H}}$  is the estimated channel matrix. In this paper, we consider perfect CSI case ( $\hat{\mathbf{H}} = \mathbf{H}$ ) and imperfect CSI case ( $\hat{\mathbf{H}} \neq \mathbf{H}$ ).

Assuming the perfect CSI, our proposed CNN/ZF-AMC method is compared with other AMC methods, while we adopt a channel error model to generate the channel matrix as the imperfect CSI. There are two typical channel error models, which are shown as

$$\hat{\mathbf{H}}_1 = \sqrt{1 - \sigma_e} \mathbf{H} + \sqrt{\sigma_e} \mathbf{D}, \quad \hat{\mathbf{H}}_2 = \mathbf{H} + \sqrt{\sigma_e} \mathbf{D}, \quad (3)$$

where  $\sigma_e$  is the channel error coefficient, and  $\mathbf{D}$  is the error matrix, which is independent of  $\mathbf{H}$  and each element obeys the zero-mean and unit-variance circular symmetric complex normal distribution. Below, we choose  $\hat{\mathbf{H}}_1$  rather than  $\hat{\mathbf{H}}_2$  as  $\hat{\mathbf{H}}$  by analyzing four metrics for the channel error models, which are mean  $\mu$ , variance  $\sigma^2$ , normalized mean square error (NMSE)  $NMSE_{\hat{\mathbf{H}}, \mathbf{H}}$  and correction coefficient between the actual channel and the estimated channel  $\rho_{\hat{\mathbf{H}}, \mathbf{H}}$  [27]. In addition to the common metrics of  $\mu$  and  $\sigma^2$ , the other two are given by

$$NMSE_{\hat{\mathbf{H}}, \mathbf{H}} = \frac{E(|h_{ij} - \hat{h}_{ij}|^2)}{E(|h_{ij}|^2)}, \quad (4)$$

$$\rho_{\hat{\mathbf{H}}, \mathbf{H}} = \frac{E(h_{ij} \hat{h}_{ij}^*)}{\sqrt{E(|h_{ij}|^2)E(|\hat{h}_{ij}|^2)}}, \quad (5)$$

where  $h_{ij}$  and  $\hat{h}_{ij}$  are the  $(i, j)$ -th element of  $\mathbf{H}$  and  $\hat{\mathbf{H}}$ , respectively. The four metrics are related with  $\sigma_e$ , and their theoretical values are listed in Tab. I.

The same characteristics of these two channel error models are the NMSE and the channel correction coefficient. Specifically, according to Taylor expansion when  $\sigma_e \ll 1$ , the absolute difference of two channel error models about  $NMSE_{\hat{\mathbf{H}}, \mathbf{H}}$  and  $\rho_{\hat{\mathbf{H}}, \mathbf{H}}$  is approximately equal to  $\sigma_e^2/4 +$

TABLE I  
THE THEORETICAL VALUES OF THE FOUR METRICS.

Model	$\hat{\mathbf{H}}_1$	$\hat{\mathbf{H}}_2$
Metrics		
$\mu$	0	0
$\sigma^2$	1	$1 + \sigma_e$
$NMSE_{\hat{\mathbf{H}}, \mathbf{H}}$	$2 - 2\sqrt{1 - \sigma_e}$ $= \sigma_e + \sigma_e^2/4 + o(\sigma_e^2)$	$\sigma_e$
$\rho_{\hat{\mathbf{H}}, \mathbf{H}}$	$\sqrt{1 - \sigma_e} = 1 - \sigma_e/2$ $-\sigma_e^2/8 + o(\sigma_e^2)$	$1/\sqrt{1 + \sigma_e} = 1 - \sigma_e/2$ $-3\sigma_e^2/8 + o(\sigma_e^2)$

$o(\sigma_e^2)$ , which can be ignored. It means that  $\hat{\mathbf{H}}_1$  and  $\hat{\mathbf{H}}_2$  can be considered to be almost independent from the aspects of the NMSE and the channel correction coefficient. In addition, their means are also the same as the mean of  $\mathbf{H}$ .

The variance of  $\hat{\mathbf{H}}_1$  is consistent with that of  $\mathbf{H}$ , but the variance of the  $\hat{\mathbf{H}}_2$  is changed. Thus,  $\hat{\mathbf{H}}_1$  is applied to analyze the influence for the CNN/ZF-AMC method under different channel estimation errors, though the different between these two models is too limited, and  $\hat{\mathbf{H}}$  refers to  $\hat{\mathbf{H}}_1$  in the remaining content.

### III. THE PROPOSED CNN/ZF-AMC METHOD

In this section, we introduce the CNN/ZF-AMC method, whose structure is shown in Fig. 1(a). The CNN/ZF-AMC method consists of three main parts: channel estimation, ZF equalizer, and CNN applied for identifying modulation types. In order to make understanding easier, we mainly introduce the part from three aspects: dataset generation, CNN for the ZF-AMC method, and ANN and HOC for the traditional AMC method.

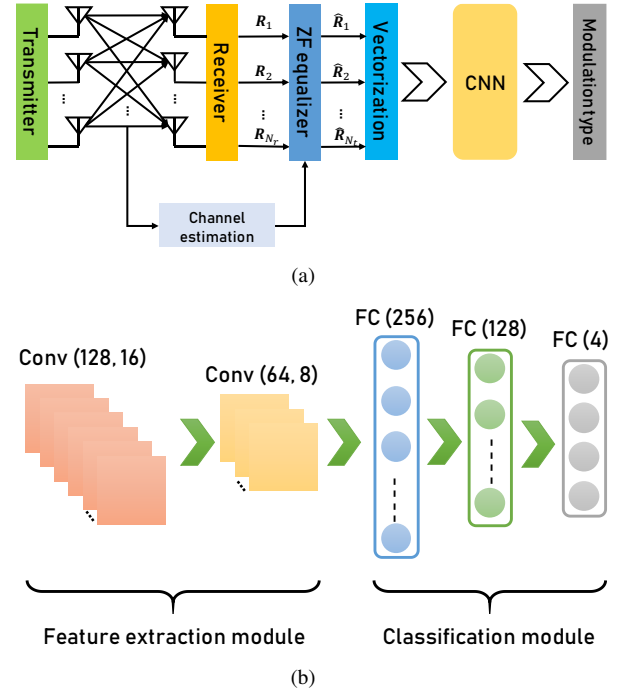


Fig. 1. The structure of the CNN/ZF-AMC method with the specified structure of CNN, where “Conv” is the convolutional layer and “FC” represents the fully-connected layer.

### A. Dataset Generation

Here, a complex-baseband equivalent multi-antenna system model is considered and the process of dataset generation is shown in Fig. 2. Specifically, random data are modulated with different modulation types, including binary phase shift keying (BPSK), quadrature phase shift keying (QPSK), eight phase shift keying (8PSK), and sixteen quadrature amplitude modulation (16QAM). The modulation signal vector can be denoted as  $\mathbf{X}$ , the size of which is  $1 \times N$  ( $N$  is the number of symbols, and here  $N=128$ ). In addition, for a fair comparison,  $\mathbf{X}$  is normalized with unit power, i. e.,  $\|\mathbf{X}\|_2^2 = 1$ . Then,  $\mathbf{X}$  is reshaped into a  $N_t \times N/N_t$  matrix, and it can be represented as  $[\mathbf{T}_1; \mathbf{T}_2; \dots; \mathbf{T}_{N_t}]$ , where  $\mathbf{T}_i = [T_i(1), T_i(2), \dots, T_i(N/N_t)]$ ,  $i \in [1, N_t]$  is the transmitted signal vector at the  $i$ -th antenna.

When passing through the MIMO channel, the received signal vector at the  $j$ -th receive antenna is denoted as  $[\mathbf{R}_1; \mathbf{R}_2; \dots; \mathbf{R}_{N_r}]$  with size  $N_r \times N/N_t$ , and  $\mathbf{R}_j = [R_j(1), R_j(2), \dots, R_j(N/N_t)]$ ,  $j \in [1, N_r]$ . Next, the received signal matrix can be equalized by ZF equalizer, and the equalized signal sequence is  $[\hat{\mathbf{R}}_1, \hat{\mathbf{R}}_2, \dots, \hat{\mathbf{R}}_{N_r}]^T$  with size  $N_t \times N/N_t$ , which is vectorized into a  $1 \times N$  vector  $\hat{\mathbf{R}}$ .

The training and test samples are extracted from  $\hat{\mathbf{R}}$ . Specifically, the real part and imaginary part of  $\hat{\mathbf{R}}$ :  $\mathcal{R}(\hat{\mathbf{R}})$  and  $\mathcal{I}(\hat{\mathbf{R}})$  are separated and then they are combined into a  $2 \times N$  matrix  $[\mathcal{R}(\hat{\mathbf{R}}); \mathcal{I}(\hat{\mathbf{R}})]$ , which is a sample for training or test. It is noted that we prepare 20000 samples for training, and 10000 samples for testing for each SNR value.

### B. The Proposed CNN/ZF-AMC Method

1) *CNN structure*: In this correspondence, we adopt a simple CNN with one feature extraction module with two convolutional layers and one classification module with three fully-connected layers, the structure of which is shown in Fig. 1(b). What's more, rectified linear unit (ReLU), batch normalization (BN), and dropout follow behind each available layer except the last fully-connected layer, and the former one is as activation function, while the latter two are to prevent overfitting and slightly accelerate the training process. In addition, Softmax is chosen as the activation function of the last layer.

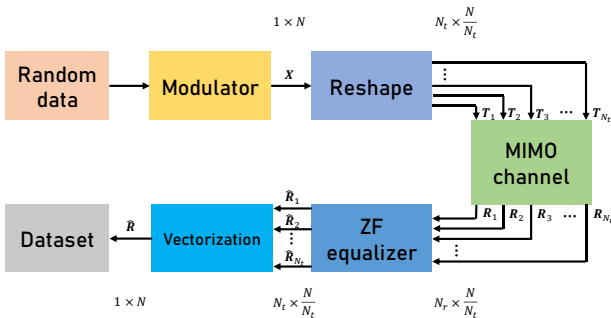


Fig. 2. The process of dataset generation.

2) *Training and Test Phase*: Before training, the training dataset is divided into training part and validation part for cross-validation, which are applied to update the trainable

TABLE II  
THE THEORETICAL VALUES OF  $C_4$  [3], [4].

Type	BPSK	QPSK	8SPK	16QAM
$C_{40}$	-2	-1	0	-0.68
$C_{41}$	-2	0	0	0
$C_{42}$	-2	-1	-1	-0.68

parameters of CNN, and choose the best trained model or parameters, respectively. We choose an adaptive learning rate optimizer of ADAM [28], and select the classification cross entropy function as the object function. Other parameters, including the maximum epoch, early-stopping epoch, batch size are set as 100, 20 and 500, respectively. After training, the test samples are fed into the trained CNN for the predicted labels.

### C. Review of Traditional AMC Method

Here, ANN and HOC-based traditional AMC method, which is a classical combination of classifier and feature [4], is as a comparison for highlighting the superior performance of the CNN/ZF-AMC method. The structure of the traditional method is similar to the CNN/ZF-AMC methods in Fig. 1(a), “CNN” in the CNN/ZF-AMC method is replaced with “ANN+HOC” in the traditional AMC method. Specifically, the fourth order HOC features are applied, which is denoted as  $C_4$  and is shown in Tab. II, and the feature vector is extracted from the dataset for CNN in the CNN/ZF-AMC method. In addition, the ANN structure has the same structure as the classification module in Fig. 1(b).

## IV. RESULTS AND DISCUSSIONS

In this section, we show two sets of simulation results which are respectively in the perfect CSI and imperfect CSI cases. In the former case, the CNN/ZF-AMC method and the ANN and HOC-based traditional AMC method are compared, while the impact of the channel estimation error on the classification performance is investigated in the latter case. Here, the correct classification probability is adopted as the evaluation metric, and it can be represented as  $P_{cc} = S_c/S \times 100\%$ , where  $S_c$  is the number of correctly classified samples, and  $S$  is the number of the total samples for the given SNR.

### A. Performance Comparison in the Perfect CSI Case

The classification performances are shown in Fig. 3. It can be obviously observed that the CNN/ZF-AMC method has a great advantage over the ANN and HOC-based traditional AMC-method, where “CNN ( $N_r, N_t$ )” represents the former one and “ANN with  $C_4$  ( $N_r, N_t$ )” is the latter one. In addition, the fewer transmitter antennas, the better performance, when the number of the receive antennas is fixed. For explaining this result, we perform some analysis as follows.

The received signal sequence via ZF equalization with perfect CSI can be written as follows.

$$\hat{\mathbf{R}}(n) = ZF(\mathbf{H})\mathbf{R}(n) = \mathbf{T}(n) + \mathbf{H}^\dagger \mathbf{G}(n), \quad (6)$$

and the post-processing noise can be represented as  $\hat{\mathbf{G}}(n) = \hat{\mathbf{H}}^\dagger \mathbf{G}(n)$ . Thus, the post-processing SNR [4] can be written as

$$\tilde{\gamma}_i = \frac{\left[ \frac{E[\mathbf{T}(n)\mathbf{T}^H(n)]}{E(\hat{\mathbf{G}}(n)\hat{\mathbf{G}}^H(n))} \right]_{ii}}{\gamma} = \frac{E_T}{E_G[\mathbf{H}^\dagger(\mathbf{H}^\dagger)^H]_{ii}}, \quad (7)$$

$$= \frac{\gamma}{[(\mathbf{H}^H \mathbf{H})^{-1}]_{ii}}, i \in [1, N_t],$$

where  $[\cdot]_{ii}$  is the  $i$ -th diagonal element of a matrix, and  $\gamma = E_T/E_G$  is actual SNR. In addition,  $1/[(\mathbf{H}^H \mathbf{H})^{-1}]_{ii}$  is known as a chi-square distributed random variable with  $2(N_r - N_t + 1)$  degrees of freedom [27], i. e.,  $1/[(\mathbf{H}^H \mathbf{H})^{-1}]_{ii} \sim \chi^2[2(N_r - N_t + 1)]$ . The exception of  $\tilde{\gamma}_k$  is

$$E[\tilde{\gamma}_i] = 2(N_r - N_t + 1)\gamma = 2(\Delta + 1)\gamma, \quad (8)$$

and the SNR gain is determined by the difference,  $\Delta = N_r - N_t$  between the number of the receive antennas and that of the transmit antennas  $\Delta = N_r - N_t$ , which means that the larger  $\Delta$ , the more performance improvement, but the performance gap with different transmit antennas in the CNN/ZF-AMC method is more limited than that in the traditional method.

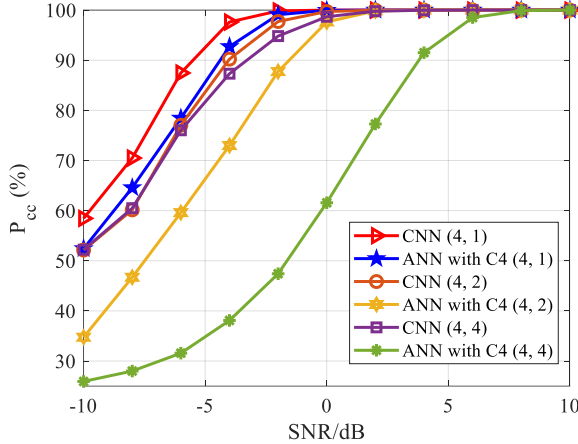
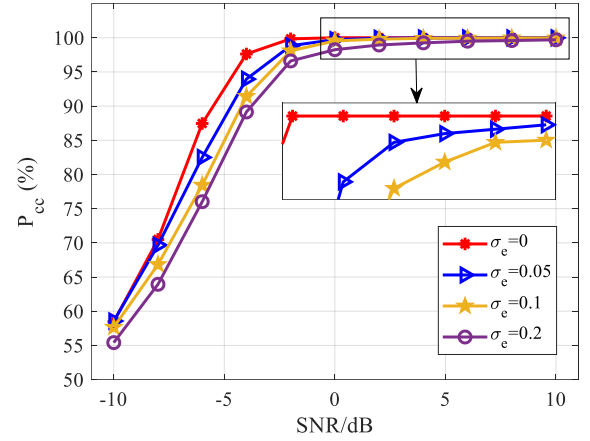


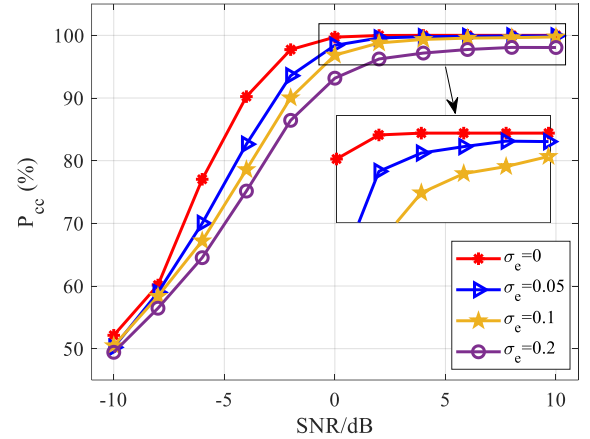
Fig. 3. The performance comparison between the the proposed CNN/ZF-AMC method and the ANN-based traditional AMC method.

### B. Performance Comparison vs. Channel Error Coefficient

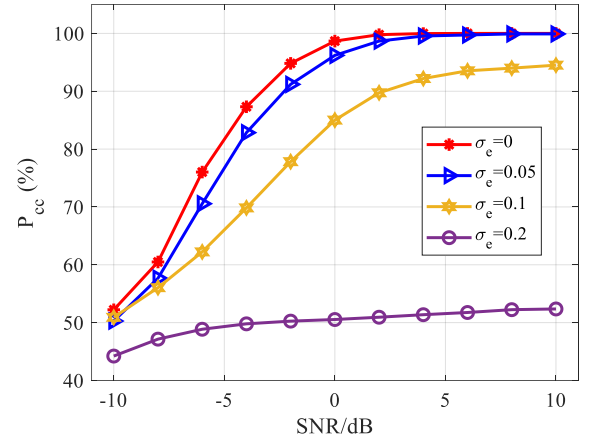
The perfect CSI is hardly obtained in the actual communication systems. Thus, we focus on the CNN/ZF-AMC method in the imperfect CSI case. It can be observed that with the increasing of  $\sigma_e$ , the classification performance is gradually decreasing, which is shown in Fig. 4. However, there is a huge differences in the classification performance for a different combination of receive and transmit antennas, when  $\sigma_e$  is the same. Specifically, when  $\sigma_e = 0.2$  and SNR = 10 dB, the correct classification probability of the MIMO system with  $N_r = 4$  and  $N_t = 1$  can reach up to nearly 100%, but that of the MIMO system with  $N_r = 4$  and  $N_t = 4$  barely exceed 50%, which are shown in Fig. 4(a) and Fig. 4(c). We give some analysis for the detailed factors that lead to the above performance difference, which are shown as follows.



(a)  $N_r = 4, N_t = 1$



(b)  $N_r = 4, N_t = 2$



(c)  $N_r = 4, N_t = 4$

Fig. 4. The classification performances of the ZF-AMC method with different antennas under different  $\sigma_e \in \{0, 0.05, 0.1, 0.2\}$ .

Based on the channel error model  $\hat{\mathbf{H}}$ , the ZF equalization-based received signal can be given by

$$\hat{\mathbf{R}}(n) = \text{ZF}(\hat{\mathbf{H}})\mathbf{R}(n) = \hat{\mathbf{H}}^\dagger[\mathbf{H}\mathbf{T}(n) + \mathbf{G}(n)], \quad (9)$$

where  $\hat{\mathbf{H}}^\dagger = (\sqrt{1-\sigma_e}\mathbf{H} + \sqrt{\sigma_e}\mathbf{D})^\dagger$ , which can be approximated to  $\frac{1}{\sqrt{1-\sigma_e}}\mathbf{H}^\dagger(\mathbf{I}_{N_r} - \sqrt{\frac{\sigma_e}{1-\sigma_e}}\mathbf{D}\mathbf{H}^\dagger)$  by Taylor

expansion when  $\sigma_e \ll 1$ . Thus, (9) can be approximated as

$$\hat{\mathbf{R}}(n) \approx \frac{1}{\sqrt{1-\sigma_e}} [\mathbf{T}(n) + \mathbf{H}^\dagger \mathbf{G}(n) - \sqrt{\frac{\sigma_e}{1-\sigma_e}} \mathbf{H}^\dagger \mathbf{D} \mathbf{T}(n) - \sqrt{\frac{\sigma_e}{1-\sigma_e}} \mathbf{H}^\dagger \mathbf{D} \mathbf{H}^\dagger \mathbf{G}(n)], \quad (10)$$

and the post-processing transmitted signal and noise can be represented by  $\hat{\mathbf{T}}(n) = \mathbf{T}(n)/\sqrt{1-\sigma_e}$ , and  $\hat{\mathbf{G}}(n) = \frac{1}{\sqrt{1-\sigma_e}} \mathbf{H}^\dagger \mathbf{G}(n) - \frac{\sqrt{\sigma_e}}{1-\sigma_e} \mathbf{H}^\dagger \mathbf{D} \mathbf{T}(n) - \frac{\sqrt{\sigma_e}}{1-\sigma_e} \mathbf{H}^\dagger \mathbf{D} \mathbf{H}^\dagger \mathbf{G}(n)$ , respectively. Then, the post-processing SNR can be expressed as

$$\tilde{\gamma}_i = \frac{E[\hat{\mathbf{T}}(n) \hat{\mathbf{T}}^H(n)]}{E[\hat{\mathbf{G}}(n) \hat{\mathbf{G}}^H(n)]} \Bigg|_{ii} = \frac{\gamma}{[1 + \frac{\sigma_e}{1-\sigma_e} N_t \gamma + \frac{\sigma_e}{1-\sigma_e} \text{tr}((\mathbf{H}^H \mathbf{H})^{-1})][(\mathbf{H}^H \mathbf{H})^{-1}]_{ii}}, \quad (11)$$

where  $i \in [1, N_t]$  and  $\text{tr}(\cdot)$  is the matrix trace operation, and  $\text{tr}((\mathbf{H}^H \mathbf{H})^{-1})$  can be ignored [27]. The exception of  $\tilde{\gamma}_i$  is

$$E(\tilde{\gamma}_i) = \frac{2(N_r - N_t + 1)\gamma}{1 + \frac{\sigma_e}{1-\sigma_e} N_t \gamma} = \frac{2(\Delta + 1)\gamma}{1 + \frac{\sigma_e}{1-\sigma_e} N_t \gamma}, \quad (12)$$

when  $\gamma$  is very high (e.g.,  $\gamma \rightarrow \infty$ ), (12) can be approximated by  $E(\tilde{\gamma}_i) \rightarrow 2(\Delta + 1)/(\frac{N_t \sigma_e}{1-\sigma_e})$ . It is obvious that the classification performances of the CNN/ZF-AMC method depend not only on  $\sigma_e$ , but also on  $\Delta$  for the given value of  $N_r$ . Hence, when  $\sigma_e$  and  $N_r$  are fixed, the more transmitted antennas will lead to the worse identification performances.

## V. CONCLUSION

In this correspondence, we proposed an effective CNN/ZF-AMC method for MIMO systems. Specifically, ZF equalization technique was applied to reveal the ambiguity of the received signal with the aid of CSI for the improvement of various AMC methods. We considered the perfect CSI and the imperfect one. In the perfect CSI case, the CNN/ZF-AMC method can achieve much better performances than the traditional ANN and HOC-based AMC method. We also explored the classification performances of the CNN/ZF-AMC method in the case the imperfect CSI. Then, we demonstrated that the classification performance of the proposed method is not only influenced by channel error coefficient, but also related to the number of the transmit antennas and receive antennas. The proposed CNN/ZF-AMC method in the MIMO systems with more transmit antennas has worse classification performance under the same receive antennas and error coefficient, and vice versa.

## REFERENCES

- [1] F. Meng, P. Chen, *et al.*, "Automatic modulation classification: a deep learning enabled approach," *IEEE Trans. Veh. Technol.*, vol. 67, no. 11, pp. 10760–10772, 2018.
- [2] Y. Wang, M. Liu, J. Yang and G. Gui, "Data-driven deep learning for automatic modulation recognition in cognitive radios," *IEEE Trans. Veh. Technol.*, vol. 68, no. 4, pp. 4074–4077, 2019.
- [3] A. Swami and B. Sadler, "Hierarchical digital modulation classification using cumulants," *IEEE Trans. Commun.*, vol. 48, no. 3, pp. 416–429, 2000.
- [4] K. Hassan, I. Dayoub, *et al.*, "Blind digital modulation identification for spatially-correlated MIMO systems," *IEEE Trans. Wirel. Commun.*, vol. 11, no. 2, pp. 683–693, 2011.
- [5] S. Kharbech, *et al.*, "Blind digital modulation identification for time-selective MIMO channels," *IEEE Wirel. Commun. Lett.*, vol. 3, no. 4, pp. 373–376, 2014.
- [6] B. Mao, *et al.*, "Routing or computing? The paradigm shift towards intelligent computer network packet transmission based on deep learning," *IEEE Trans. Comput.*, vol. 66, no. 11, pp. 1946–1960, Nov. 2017.
- [7] G. Gui, F. Liu, J. Sun, J. Yang, Z. Zhou, and D. Zhao, "Flight delay prediction based on aviation big data and machine learning," *IEEE Trans. Veh. Technol.*, in press, doi: 10.1109/TVT.2019.2954094.
- [8] H. Gacanin, "Autonomous wireless systems with artificial intelligence: A knowledge management perspective," *IEEE Veh. Technol. Mag.*, vol. 14, no. 1, pp. 51–59, 2019.
- [9] B. Mao, F. Tang, Z. Md. Fadlullah, and N. Kato, "An intelligent route computation approach based on real-time deep learning strategy for software defined communication systems," *IEEE Trans. Emerg. Topics Comput.*, in press, doi: 10.1109/TETC.2019.2899407.
- [10] J. Sun, W. Shi, Z. Han, J. Yang, G. Gui, "Behavioral modeling and linearization of wideband RF power amplifiers using BiLSTM networks for 5G wireless systems," *IEEE Trans. Veh. Technol.*, vol. 68, no. 11, pp. 10348–10356, 2019.
- [11] H. Huang, S. Guo, *et al.*, "Deep learning for physical-layer 5G wireless techniques: Opportunities, challenges and solutions," *IEEE Wirel. Commun.*, in press, doi: 10.1109/MWC.2019.1900027.
- [12] Z. Md. Fadlullah, *et al.*, "State-of-the-art deep learning: evolving machine intelligence toward tomorrow's intelligent network traffic control systems," *IEEE Commun. Surveys and Tuts.*, vol. 19, no. 4, pp. 2432–2455, 2017.
- [13] N. Kato, *et al.*, "The deep learning vision for heterogeneous network traffic control: Proposal, challenges, and future perspective," *IEEE Wirel. Commun. Mag.*, vol. 24, no. 3, pp. 146–153, 2016.
- [14] H. Huang, *et al.*, "Fast beamforming design via deep learning," *IEEE Trans. Veh. Technol.*, in press, doi: 10.1109/TVT.2019.2949122.
- [15] M. Liu, *et al.*, "Deep learning-inspired message passing algorithm for efficient resource allocation in cognitive radio networks," *IEEE Trans. Veh. Technol.*, vol. 69, no. 1, pp. 641–653, Jan. 2019.
- [16] L. Liang, H. Ye, G. Yu, and G. Y. Li, "Deep-learning-based wireless resource allocation with application to vehicular networks," *Proc. IEEE*, in press, doi: 10.1109/JPROC.2019.2957798.
- [17] G. Gui, *et al.*, "Deep learning for an effective nonorthogonal multiple access scheme," *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 8440–8450, 2018.
- [18] Z. Md. Fadlullah, *et al.*, "On intelligent traffic control for large-scale heterogeneous networks: A value matrix-based deep learning approach," *IEEE Commun. Lett.*, vol. 22, no. 12, pp. 2479–2482, 2018.
- [19] F. Tang, *et al.*, "An intelligent traffic load prediction-based adaptive channel assignment algorithm in SDN-IoT: A deep learning approach," *IEEE Internet Things J.*, vol. 5, no. 6, pp. 5141–5154, 2018.
- [20] B. Mao, *et al.*, "A novel non-supervised deep learning based network traffic control method for software defined wireless networks," *IEEE Wirel. Commun. Mag.*, vol. 25, no. 4, pp. 74–81, 2018.
- [21] Z. Md. Fadlullah, *et al.*, "Value iteration architecture based deep learning for intelligent routing exploiting heterogeneous computing platforms," *IEEE Trans. Comput.*, vol. 68, no. 6, pp. 939–950, 2019.
- [22] T. O'Shea and J. Hoydis, "An introduction to deep learning for the physical layer," *IEEE Trans. Cogn. Commun. Netw.*, vol. 3, no. 4, pp. 563–575, 2017.
- [23] N. E. West and T. O'Shea, "Deep architectures for modulation recognition," *IEEE DySPAN*, Baltimore, MD, USA, Mar. 6–9, 2017, pp. 1–6.
- [24] S. Hu, Y. Pei, P. Lian and Y. Liang, "Deep neural network for robust modulation classification under uncertain noise conditions," *IEEE Trans. Veh. Technol.*, in press, doi: 10.1109/TVT.2019.2951594.
- [25] Y. Tu and Y. Lin, "Deep neural network compression technique towards efficient digital signal modulation recognition in edge device," *IEEE Access*, vol. 7, no. 1, pp. 58113–58119, 2019.
- [26] Y. Tu, *et al.*, "Semi-supervised learning with generative adversarial networks on digital signal modulation classification," *Comput. Mater. Continua*, vol. 55, no. 2, pp. 243–254, May. 2018.
- [27] C. Wang, *et al.*, "On the performance of the MIMO zero-forcing receiver in the presence of channel estimation error," *IEEE Trans. Wirel. Commun.*, vol. 6, no. 3, pp. 805–810, 2007.
- [28] D. P. Kingma and J. Lei Ba, "ADAM: a method for stochastic optimization," *arXiv:1412.6980v9*, 2014.