Low Complexity MIMO Modulation Classification via Distribution Test Ensembles

Zikang Gao, Lingjun Zhu, and Zhechen Zhu*

School of Electronic and Information Engineering

Soochow University

Suzhou, Jiangsu, China

Email: {zkgao, ljzhu66}@stu.suda.edu.cn, zczhu@suda.edu.cn

***Abstract*—Modulation classification in MIMO systems needs to overcome its complexity demand to be viable in a real world environment. The number of mathematics operations required by the state-of-the-art maximum likelihood classifier grows exponentially with the number of transmitting antennas and the orders of candidate modulations. In this paper, we propose a low complexity MIMO system modulation classifier by combining multiple magnitude-based distribution tests using a simple Multi-layer perceptron. The resulting solution provides good classification performance in AWGN channel while outperforming the maximum likelihood classifier in fading channels. Moreover, its computational complexity is much lower and does not scale with the number of transmitting antennas or the orders of candidate modulations.**

Keywords-MIMO; Classification; Modulation; Distribution Test; DTE

I. INTRODUCTION

In both electronic warfare and some advanced civilian communication systems, the modulation information of the transmitted signal is not pre-negotiated with or broadcast to the receiving end. Therefore, the receiver would require a Modulation Classification (MC) mechanism to recover the underling modulation scheme. In military scenarios, modulation information of the intercepted signals can be used to help recover the transmitted message, generate more effect radio frequency jamming signals or identify the unit transmitting the signal. In civilian scenarios, where link adaptation is used to improve link reliability and overall bandwidth, MC avoids the need for modulation information to be broadcast thus improves spectral efficiency.

Maximum likelihood (ML) classifiers [1], [2] has long been recognized to provide the optimal classification accuracy. However, the implementation of ML classifiers in MIMO systems [3] has proven to be rather costly in terms of computational complexity. The number of exponential operations required grows exponentially with the number transmitting antennas and the orders of candidate modulations. In the context of current mobile communication technology development (e.g. 5G mobile standards), massive MIMO and high order modulations are often desirable when pursuing higher throughput. Therefore, before MC algorithms could be employed alongside some of these aforementioned technologies, its computational complexity must be reduced drastically. Some effort has been made to achieve this goal. Kanterakis and Su proposed a low complexity likelihood based classifier for MIMO systems by modifying the likelihood function [4]–[6]. However, the level of complexity reduction is limited while the classification suffers significantly. Feature based methods are known to require less computational resource to complete MC tasks [7]–[10]. However, feature based methods have no intrinsic mechanism to overcome channel effects while relying only on the robustness of the selected expert features when dealing with unideal channel conditions [11]–[13].

In this paper, we propose classifier via Distribution Test Ensembles (DTE) to further reduce the computational complexity demanded in MIMO systems. While separate solutions based on different distribution tests have already been developed for MIMO systems [19], there is still room for improvement in terms of classification accuracy. Therefore, we propose to combine multiple classic distribution tests alongside a novel variance based distribution test to achieve better classification accuracy. The ensembles of these distribution tests is achieved via a simple Multi-layer perceptron. More details on the proposed implementation is given in the following sections.

II. SIGNAL MODEL

In this paper, we consider a MIMO system with N_T transmitting antennas and N_R receiving antennas. Under the assumption of a flat fading and time-invariant MIMO channel, the kth received signal vector at the instant k, denoted $\boldsymbol{r}_k = [r_k(1), r_k(2), \cdots, r_k(N_R)]^T$ can be expressed as

$$\boldsymbol{r}_k = \alpha e^{j(2\pi f_o nT + \theta_o)} \boldsymbol{H}\boldsymbol{x}_k + \boldsymbol{\omega}_k \ , \tag{1}$$

where $\boldsymbol{x}_k = [x_k(1), x_k(2), \cdots, x_k(N_T)]^T$ is the kth transmitted signal symbol vector $(N_T \times 1)$, which is assigned randomly from the alphabet set $\boldsymbol{A}_{\mathcal{M}}$ of modulation $\mathcal{M}$ with equal probability. The channel matrix $\boldsymbol{H}$ is a $N_R \times N_T$ complex matrix with the element $h_{j,i}$ representing the path gain between ith transmitting antenna and jth receiving antenna, having $N_R \geq N_T$. $\boldsymbol{\omega}_k = [\omega_k(1), \omega_k(2), \cdots, \omega_k(N_R)]^T$ is the additive noise observed at the kth signal sample. The additive noise is assumed to be white Gaussian with zero mean and variance σ_ω^2 which gives $\omega_k \in \mathcal{N}(0,\sigma_\omega^2 I_{N_R})$, where I_{N_R} is the identity matrix of size $N_R \times N_R$. Channel gain α is considered as constant in each signal realization, while varying uniformly between [0,1] among different signal realizations.

III. The Goodness-of-Fit Test

The Goodness-of-Fit (GoF) is often used to measure the consistency between observed samples and hypothesized statistical models. The test statistic for GoF tests can be applied to certain sequence of features $\{z_k\}_{k=1}^N$ extracted from received signal samples $\{r_k\}_{k=1}^N$. In this paper, we select the magnitude of $\{r_k\}$ for distribution test. z_k can be calculated as

$$z_k = |r_k| = \sqrt{(\Re\{r_k\})^2 + (\Im\{r_k\})^2} \quad k = 1,...,N. \tag{2}$$

Let $F_1(z)$ denote the empirical cumulative distribution function (CDF), and $F(z)$ can be represented as

$$F_1(z) = \frac{|\{k : z_k \le z, 1 \le k \le N\}|}{N} \triangleq \frac{1}{N}\sum_{k=1}^{N} \mathbb{I}(z_k \le z), \tag{3}$$

where $\mathbb{I}(\cdot)$ denotes the indicator function, which equals to one if the input is true, and equals to zero otherwise.

For transmitted signals x_k, since $\omega_k \in \mathcal{N}(0,\sigma_\omega^2 I_{N_R})$ and considering all the signal points in constellation as equiproable, the theoretical distribution of z_k under modulation candidate $\mathcal{M}_k$ is given by

$$F_0(z) = 1 - \frac{1}{|\mathcal{M}_k|}\sum_{x \in \mathcal{M}_k} Q\left(\frac{\sqrt{2}\,|x|}{\sigma}, \frac{\sqrt{2}z}{\sigma}\right), \tag{4}$$

where $Q(a,b)$ is the Marcum Q-function.

The decision in favor of modulation candidate which provides the minimum distance between empirical distribution and hypothesized distribution for modulation classification using GoF test can be represented as

$$\hat{\mathcal{M}} = \arg\min_{1 \le k \le M} D_{(\cdot)}^k, \tag{5}$$

where $D_{(\cdot)}^k$ denotes the distance evaluated for each modulation candidate $\mathcal{M}_k$.

A brief overview of steps involved in modulation classification using the GoF test are given as Algorithm 1.

Algorithm 1 Framework of GoF test in MIMO system

Input: Received signals $\{r_k\}_{k=1}^N$ and noise variance σ_ω^2

1: **for** each receiving antenna n_r **do**
2: Obtain the sequence of signal features $\{z_k\}_{k=1}^N$;
3: Sort the signal samples with sort $(\{z_k\}_{k=1}^N)$;
4: Calculate the emperical distribution using (3);
5: **for** each modulation candidate $\mathcal{M}_k$ **do**
6: Generate modulation symbols $S_m = \mathbf{H}A_m$;
7: **for** each modulation symbol s_i **do**
8: Calculate the theoretical CDF;
9: **end for**
10: Calculate the mean of the theoretical CDF;
11: Evaluate the distance between emperical CDF and theoretical CDF;
12: **end for**
13: **end for**
14: Calculate the mean of the test statistics $D_{(\cdot)}^k$;

Output: Choose $\hat{\mathcal{M}}_{(\cdot)} = \arg\min_{1 \le k \le M} D_{(\cdot)}^k$

A. Kolmogorov-Smirnov Test

In the Kolmogorov-Smirnov (K-S) test [20], the test statistic for K-S test is given as

$$D_{KS} \triangleq \sup_{z \in \mathbb{R}} |F_1(z) - F_0(z)|, \tag{6}$$

and in practice, it is calculated by

$$\hat{D}_{KS} = \max_{1 \le n \le N} |F_1(z_n) - F_0(z_n)|. \tag{7}$$

B. Cramer-von Mises Test

The Cramer-Von Mises test (C-v-M) is also used for measuring the goodness of fit in one-sample scenarios [21]. The test statistics is defined as the integral of the squared difference between the empirical CDF $F_1(x)$ and the theoretical CDF $F_0(x)$, i.e.,

$$D_{CvM} \triangleq \int_{-\infty}^{\infty} \left[F_1(z) - F_0(z)\right]^2 dF_0(z). \tag{8}$$

Then the decision statistics is defined in practice as

$$\hat{D}_{CvM} = nD_{CvM} = \frac{1}{12n} + \sum_{i=1}^{n}\left[\frac{2i-1}{2n} - F(z_i)\right]^2, \tag{9}$$

where $F(z_i)$ is the theoretical CDF value of the ith signal sample and n is the total number of the observed samples.

C. Anderson-Darling Test

The Anderson-Darling (A-D) test is also a statistical test model with no parameters needed [22]. Compared with the C-v-M test, the test statistics is the C-v-M test with an added weight function $w(z)$. It is defined as

$$D_{AD} \triangleq \int_{-\infty}^{\infty} \left[F_1(z) - F_0(z)\right]^2 w(z) dF_0(z), \tag{10}$$

where the weight function is defined by

$$w(z) = \frac{1}{F_0(z)\left(1 - F_0(z)\right)}. \tag{11}$$

Then the decision statistics is defined in practice as

$$\hat{D}_{AD} = nD_{AD} = -n - \frac{1}{n}\sum_{i=1}^{n}[(2i-1)\ln F(z_i) + (2n+1-2i)\ln[1 - F(z_i)]], \tag{12}$$

where $F(z_i)$ is the theoretical CDF value of the ith signal sample and n is the total number of the observed samples.

D. Variance test

Inspired by the classic distribution tests, we propose a method of Variance test (Var) to determine the format of the received signals. In this paper, Var test is used to calculate the variance $(\hat{\sigma}^2)$ of the difference (d_i) between empirical CDF and theoretical CDF, as given by

$$d_i = \hat{F}_1(z_i) - F_0(z_i), \tag{13}$$

$$\hat{D}_{Var} = \hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{N}(d_i-\mu)^2 = \left(\frac{1}{N}\sum_{i=1}^{N}d_i^2\right) - \mu^2, \tag{14}$$

where the mean (μ) of the difference is given by

$$\mu = \frac{1}{N}\sum_{i=1}^{N}d_i. \tag{15}$$

IV. Proposed Classification Strategy

Multi-layer perceptron (MLP) is a class of feed-forward artificial neural network. A MLP consists of at least three layers of nodes: an input layer, a hidden layer and an output layer. Use for feature combination the MLP can be expressed as equation

$$y_k = \phi\left(\sum_{i=1}^{q}\omega_{ki}\phi\left(\sum_{j=1}^{p}\omega_{ij}x_j\right)\right), \tag{19}$$

where y_k is the output of the MLP network, w_{ij} is the weight value from neuron j of input layer to neuron i of hidden layer, w_{ki} is the weight value from neuron i of hidden layer to neuron k of output layer, x_j is the jth input feature, and $\phi(\cdot)$ is the activation function.

1) *Feature Combination:* In this paper, the test statistics $D_{(\cdot)}^k$ based on GoF test are utilized to generate new features, $t_{mod_n}^*$.

Firstly, we should obtain the matrix $\boldsymbol{t}_{mod}^*$ of the test statistics from the process of GoF test and Var test.

$$\boldsymbol{t}_{mod}^* = \begin{bmatrix} t_{mod_1}^{KS} & t_{mod_2}^{KS} & \cdots & t_{mod_n}^{KS} \\ t_{mod_1}^{CVM} & t_{mod_2}^{CVM} & \cdots & t_{mod_n}^{CVM} \\ t_{mod_1}^{AD} & t_{mod_2}^{AD} & \cdots & t_{mod_n}^{AD} \\ t_{mod_1}^{Var} & t_{mod_2}^{Var} & \cdots & t_{mod_n}^{Var} \end{bmatrix}, \tag{20}$$

where n represents the number of the candidate modulation formats.

Then, the input features can be calculated by equation (21),

$$t_{mod_i}^* = t_{mod_i}^{KS} + t_{mod_i}^{CVM} + t_{mod_i}^{AD} + t_{mod_i}^{Var},\ i = 1, 2, \cdots, n. \tag{21}$$

Since modulation classification in our work is a supervised learning process, the labels corresponding to the new features are given as $Lable$,

$$Lable = \{0, 1, 2, \cdots, n-1\}. \tag{22}$$

2) *MLP structure:* The parameters of the MLP classifier including activation function and weight optimization are optimized by cross-validated grid-search over a parameter grid. Particularly, the number of hidden layer neurons is designed based on an experience formula

$$h = \sqrt{m+n} + a, \tag{23}$$

where h is the number of hidden layer neurons, m is the number if input layer neurons, n is the number of output layer neurons, a is an integer between 1 and 10. Specific parameters for MLP Structure are listed in Table I.

TABLE I. MLP Structure

Parameters	Values
Number of Network Layers	3
Number of Input Neurons	4
Number of Hidden Neurons	8
Number of Output Neurons	1
Activation Function	tanh
Weight Optimization	lbfgs

V. Experiments and Results

To evaluate the proposed modulation classification methods, we set up the simulation tests in the PYTHON environment to investigate the classifier performance under AWGN channel and fading channel conditions. In the simulations, the number of transmitting antennas and receiving antennas was set to $N_T = 2$, $N_R = 4$. Unless otherwise mentioned, the observed signal length was $N = 128$. Under each channel condition, 1,000 signal classifications for each signal modulation are generated using ML, K-S test, C-v-M test, A-D test, Var test, and the proposed classifier via DTE. Moreover, the computational complexity of is also discussed.

The following modulation schemes are considered in all of our simulations $\mathfrak{M} = \{BPSK, 8-PSK, 4-QAM, 16-QAM\}$. Other digital modulations can be classified in the same procedure with little modification. Other parameter configurations of the simulations are summarized in Table II.

A. Performance in AWGN Channel

In channel with AWGN noise, two sets of experiments are conducted.

In the first set of experiments, the experiment simulation results are provided to compare the performance of six different types of modulation classifiers with 1,000 realizations of testing signals for each modulation candidate and each SNR varying from 0 dB and 20 dB. In Figure 1, it is clean that the the proposed DTE classifier based on magnitude feature maintains the advant-

TABLE II. TEST PARAMETER

Parameter	Notation	Value
Candidate Modulation Pool	$\mathcal{M} \in \mathfrak{M}$	$\{BPSK,\ 8\text{-}PSK\}$ $\{4\text{-}QAM,\ 16\text{-}QAM\}$
Transmitting Antennas	N_T	2
Receiving Antennas	N_R	4
AWGN Noise Level	SNR	$0dB, 1dB \ldots 20dB$
Signal Length	N	128 & 50,100...1000
Frequency Offset (ratio)	f_o	$1E-3, 1.1E-3 \ldots 2E-3$
Phase Offset	θ_0	$6^\circ, 7^\circ \ldots 20^\circ$
Maximum Number of Iterations	n	300
Training Samples for DTE Classifier	n_1	1000
Testing Samples for DTE Classifier	n_2	1000

age over GoF test especially in the higher and lower range of SNRs. The biggest difference is exhibited at 16 dB where DTE offers an accuracy of 93.5% and A-D test offers 91.4%. On the other hand, the accuracy of the modulation classifier based on Var test is gradually increased along with the increasing SNR and also shows a good performance in high SNRs.

In the second set of experiments, keeping majority of the settings in the previous experiment, different signal lengths N between 50 and 1000 are used. From Figure 2, although the ML classifier has a classification accuracy of almost 100% at low SNR, the difference begins to decrease when more signal samples is introduced and at $N = 550$ DTE provides performance similar to ML classifier. The biggest advantage of DTE is observed at $N = 50$, where DTE obtains a classification accuracy of 89.5%, which is 4.1% over C-v-M test having a classification accuracy of 81.8%. Excluding ML classifier, the results also show that the other classifiers suffer from reduced number of samples and the classification performance declines almost linearly with signal length below 200. However, when sufficient statistics is available (signal length above 500), all methods are able to achieve a relatively stable performance.

Classification accuracy of proposed DTE for individual modulation in AWGN channel are shown in Table III. Both BPSK and 16-QAM modulations are able to receive the higher classification rate by the DTE classifier using magnitude feature at higher SNRs.

B. Performance in Fading Channel

In order to understand the effect of different channel conditions, we test phase offset and frequency offset separately but both with certain level of AWGN noise. Phase offset is simulated in slow fading scenario, and the phase offset is assumed to be a constant throughout a signal realization. Moreover, frequency offset is added to the signal separately from the phase offset.

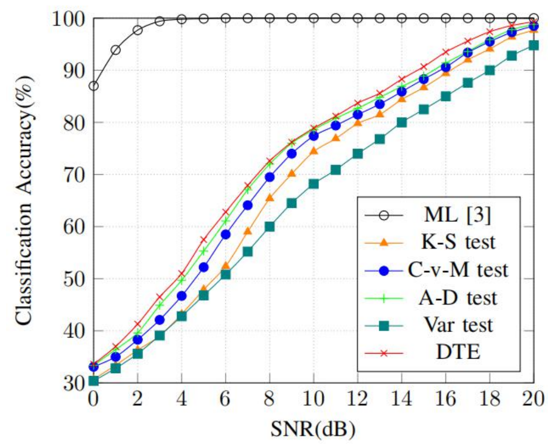

Figure 1. Classification accuracy of Magnitude-based classifier with varying noise levels.

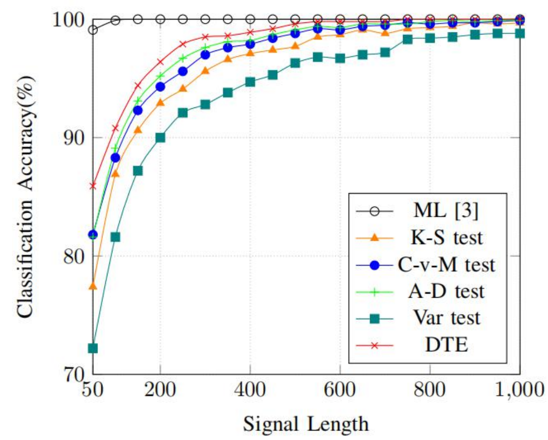

Figure 2. Classification accuracy of Magnitude-based classifier with varying signal length with SNR of 16dB.

TABLE III. CLASSIFICATION ACCURACY OF PROPOSED DTE FOR INDIVIDUAL MODULATION IN AWGN CHANNEL

Magnitude	0dB	5dB	10dB	15dB	20dB
BPSK	32.6%	62.0%	97.5%	99.9%	99.9%
8-PSK	31.7%	47.9%	59.4%	83.2%	98.8%
4-QAM	30.8%	42.5%	59.4%	81.6%	98.6%
16-QAM	37.2%	71.8%	98.5%	100.0%	100.0%

In the frequency offset channel, it is evident from Figure 3 that the proposed DTE classifier excels all other classifiers benchmarked in the tests. The performances of ML classifier are significantly affected by the frequency offset due to the sever mismatching between received signals and ideal models. While the ML classifier starts with the 100%classification accuracy, the performance starts to decline sharply when more frequency offset is introduced. However, the K-S test, C-v-M test, A-D test, Var test and the proposed DTE classifier are all able to maintain a constant level of performance in the given frequency offset

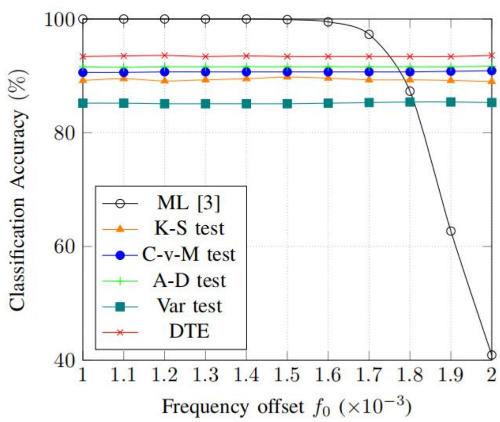

Figure 3. Classification accuracy of Magnitude-based classifier with frequency offset from 1×10^{-3} to 2×10^{-3} at SNR of 16 dB.

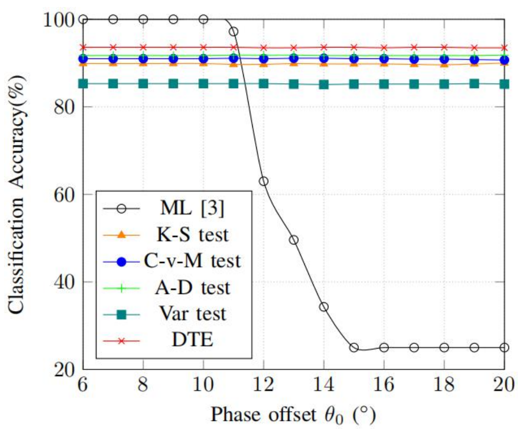

Figure 4. Classification accuracy of Magnitude-based classifier with phase offset from $6°$ to $20°$ at SNR of 16 dB.

range. Particularly, the DTE classifier achieves a classification accuracy of 93% having an advantage of 2% as compared to the 91% classification accuracy of the A-D test classifier. In the phase offset channel, it can be observed from Figure 4 that the proposed DTE classifier achieves a consistent classification accuracy throughout the tested phase offset range. While the K-S test, C-v-M test, A-D test, Var test and the proposed DTE classifier all achieve robust performance, the average classification accuracy of the DTE classifier reaches 93%, which has a more than 2% higher accuracy as compared to the 91% classification accuracy of the AD test classifier. In contrast, the ML classifier is rather sensitive to the slow phase offset. From Figure 4, it can be seen that, despite being more accurate with little phase offset, the ML classifier's performance starts to degrade when more than 11° of phase offset is introduced. Distinctly, when there is more than 11° of phase offset, the proposed DTE classifier become the best option among the benchmarked methods.

C. *Complexity Analysis*

To analyze and compare the computational complexity of the various classifiers, the number of mathematical operations required for each classifier are calculated and presented in Table IV. Prior to calculating the operation numbers, we first make the following hypotheses: the number of samples of each receiving antenna is N, the modulation candidate pool has M number of modulation candidates, the number of receiving antennas is given by N_R, the number of symbols of the *m*th candidate modulation is represented by I_m, l is the Number of Network Layers, n_k is the number of the *k*th layer neurons.

From the six tested methods in Table IV, we can obviously see that the ML classifier has highest-level of computational complexity compared with the other four distribution test methods and DTE method, which is attributed to the large number of exponential and logarithm operations needed. Notably, the number of exponential operation scale exponentially with the number of transmitting antennas and modulation order. For the Var test classifier, it is clear that this has a much lower the computational complexity and requires fewer additions compared with K-S test method. Although there is some complex computation involved in the training of weights for DTE, MLP in this paper is a simple network requiring fewer addition and multiplier operations. Moreover, it is worth clarifying that it is done offline beforehand and will not be repeated for every classification task. Without considering the complexity of MLP training, the requirement on computational power and memory of DTE classifier is only the sum of the complexity of distribution test. Furthermore, compared with distribution test classifiers, while DTE classifier exhibits a higher computational complexity, the results show that the proposed approach via distribution test ensembles outperforms the single distribution test in all cases. In terms of the system memory, three distribution tests and Var test need identical memory and have a much higher memory space than do the ML classifier when received signals N is bigger.

VI. Conclusion

A novel AMC classifier, namely DTE, using statistics combination has been proposed for the purpose of classifying M-QAM and M-PSK modulations in a robust manner with little computational complexity. The performance of the proposed algorithms was evaluated in a 2x4 MIMO system with perfect channel knowledge. It has been demonstrated from the classification performance that the proposed DTE classifier outperforms distribution test based classifier in different channel conditions including AWGN channel and Fading channel. In addition, while the

TABLE IV. NUMBER OF OPERATORS NEEDED FOR DIFFERENT CLASSIFIERS USED IN EXPERIMENT

Classifiers	Addition	Multiplier	Logarithm	Exponential	Memory
ML[3]	$6NM^{N_T}N_R \cdot \sum_{m=1}^{M^{N_T}} I_m$	$5NM^{N_T}N_R \cdot \sum_{m=1}^{M^{N_T}} I_m$	$NM^{N_T}N_R$	$NM^{N_T}N_R \cdot \sum_{m=1}^{M^{N_T}} I_m$	$M^{N_T}N_R$
K-S test	$MN_R(\log_2 N + 2N)$	0	0	0	MNN_R
C-v-M test	$MN_R(\log_2 N + 3N)$	NMN_R	0	0	MNN_R
A-D test	$MN_R(\log_2 N + 3N)$	$2NMN_R$	0	0	MNN_R
Var test	$MN_R(\log_2 N + N)$	0	0	0	MNN_R
DTE	$MN_R(4\log_2 N + 9N)$	$3NMN_R$	0	0	MNN_R
MLP	$\sum_{k=2}^{l} n_{k-1}n_k$	$\sum_{k=2}^{l} n_{k-1}n_k$	0	0	$\sum_{k=2}^{l} n_k$

performance of the DTE classifiers is inferior to the popular ML classifier at high noise level, it is of much greater computationally efficient and applicable in systems with higher number of transmitting antennas and modulation order.

ACKNOWLEDGMENT

The work is supported by the Natural Science Foundation of Jiangsu Province under grant BK20170344.

REFERENCES

[1] W. Wei and J. Mendel, "Maximum-Likelihood Classification for Digital Amplitude-Phase Modulations," IEEE Trans. Commun., vol. 48, no. 2, pp. 189-193, 2000.

[2] F. Hameed, O. A. Dobre, and D. Popescu, "On the likelihood-based approach to modulation classification," IEEE Trans. on Wire. Commu., vol. 8, no. 12, pp. 5884-5892, 2009.

[3] V. Choqueuse, S. Azou, K. Yao, L. Collin, and G. Burel, "Blind Modulation Recognition for MIMO Systems," Milit. Tech. Acad. Rev., vol. XIX, no. 2, pp. 183-196, 2009.

[4] E. Kanterakis and W. Su, "Modulation Classification in MIMO Systems," in 2013 IEEE Milit. Commun. Conf. (MILCOM), San Diego, CA, pp. 35-39, 2013.

[5] P. Urriza, E. Rebeiz, P. Pawelczak, and D. Cabric, "Computationally Efficient Modulation Level Classification Based on Probability Distribu- tion Distance Functions," IEEE Commu. Lett., vol. 15, no. 5, pp. 476-478, 2011.

[6] V. Zarzoso and A. K. Nandi, "Blind separation of independent sources for virtually any source probability density function," IEEE Trans. on Signal Proc., vol. 47, no. 9, pp. 2419-2432, 1999.

[7] M. S. Muhlhaus, M. Oner, O. A. Dobre, and F. K. Jondral, "A Low Complexity Modulation Classification Algorithm for MIMO Systems," in IEEE Commun. Lett., vol. 17, no. 10, pp. 1881-1884, 2013.

[8] S. Huang, Y. Yao, Z. Wei, Z. Feng, and P. Zhang, "Automatic Modulation Classification of Overlapped Sources Using Multiple Cumulants," IEEE Trans. Vehi.Tech., vol. 66, no. 7, pp. 48827-48839, 2017.

[9] S. Huang, Y. Jiang, X. Qin, Y. Gao, Z. Feng, and P. Zhang, "Automatic Modulation Classification of Overlapped Sources Using Multi-Gene Genetic Programming With Structural Risk Minimization Principle," IEEE Access, vol. 6, pp. 48827-48839, 2018.

[10] M. Abu-Romoh, A. Aboutaleb, and Z. Rezki, "Automatic Modulation Classification Using Moments and Likelihood Maximization," in IEEE Commun. Lett., vol. 22, no. 5, pp. 938-941, 2018.

[11] K. Zhang, E. L. Xu, Z. Feng, and P. Zhang, "A Dictionary Learning Based Automatic Modulation Classification Method," in IEEE Access, vol. 6, pp. 5607-5617, 2018.

[12] A. Ali and F. Yangyu, "Automatic Modulation Classification Using Deep Learning Based on Sparse Autoencoders With Nonnegativity Constraints," in IEEE Signal Process. Lett., vol. 24, no. 11, pp. 1626-1630, 2017.

[13] Z. Wu, S. Zhou, Z. Yin, B. Ma, and Z. Yang, "Robust Automatic Modulation Classification Under Varying Noise Conditions," in IEEE Access, vol. 5, pp. 19733-19741, 2017.

[14] F. Wang and X. Wang, "Fast and robust modulation classification via Kolmogorov-Smirnov test," IEEE Trans. Commun., vol. 58, no. 8, pp. 2324-2332, 2010.

[15] Z. Zhu, M. W. Aslam, and A. K. Nandi, "Genetic algorithm optimized distribution sampling test for M-QAM modulation classification," Signal. Proc., vol. 94, pp. 264-277, 2014.

[16] F. Gao, T. Cui, A. Nallanathan, and C. Tellambura, "Maximum Like-lihood Detection for Differential Unitary Space-Time Modulation with Carrier Frequency Offset," IEEE Trans. on Commu., vol. 56, no. 11, pp. 1881-1891, 2008.

[17] Q. Shi and Y. Karasawa, "Automatic modulation identification based on the probability density function of signal phase," IEEE Trans. on Commu., vol. 60, no. 4, pp. 1-5, 2012.

[18] A. Tomasoni and S. Bellini, "Efficient OFDM channel estimation via an information criterion," IEEE Trans. on Wir. Commu., vol. 12, no. 3, pp. 1352-1362, 2012.

[19] Z. Gao and Z. Zhu, "Distribution Test Based Low Complexity Modulation Classification in MIMO Systems," in Int. Conf. Wire. Commun. and Signal Proc.(WCSP), 2018, no. 1, pp. 1-5.

[20] J. Frank and J. Massey, "The Kolmogorov-Smirnov Test for Goodness of Fit," J. Am. Stat. Assoc., vol. 46, No. 253, pp. 68-78, 1951.

[21] H. Cramer, "On the Composition of Elementary Errors," Sc. Act. J., vol. 1, pp. 13-74, 1928.

[22] T. W. Anderson and D. A. Darling, "A Test of Goodness of Fit," J. Am. Stat. Assoc., vol. 49, No. 268, pp. 765-769, 1954.

[23] F. Laio, "Cramer--von Mises and Anderson-Darling goodness of fit tests for extreme value distributions with unknown parameters," Water. Res. R., vol. 40, no. 9, pp. 1-10, 2004.