IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 12, 2020, accepted June 28, 2020, date of publication July 10, 2020, date of current version July 23, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3008531

Modulation Classification in MIMO Systems With Distribution Test Ensemble

ZIKANG GAO[1], ZHECHEN ZHU[1], (Member, IEEE), AND ASOKE K. NANDI[2], (Fellow, IEEE)

[1]School of Electronic and Information Engineering, Soochow University, Suzhou 215006, China
[2]Department of Electronic and Computer Engineering, Brunel University London, London UB8 3PH, U.K.

Corresponding author: Zhechen Zhu (zczhu@suda.edu.cn)

This work was supported in part by the Natural Science Foundation of Jiangsu Province under Grant BK20170344, and in part by the Youth Program of National Natural Science Foundation of China under Grant 61901290.

ABSTRACT In classification of signal modulation types in MIMO systems, it is difficult to achieve both high accuracy and high computational efficiency at the same time. State-of-the-art likelihood based methods incur massive increase in computational complexity when the number of transmitting antennas and the order of modulation increase. To make modulation classification feasible in time critical systems, we propose a low complexity algorithm with an ensemble of distribution tests. Three goodness-of-fit and a novel variance based distribution tests are employed to examine the mismatch between unknown signal and different modulation hypotheses. The results from all tests are combined by a multilayer perceptron classifier for improved robustness under a variety of channel conditions including AWGN channel and slow fading channels. The resulting solution achieves performance close to the maximum likelihood classifier at high SNR. Yet, it requires much lower computational complexity in all cases.

INDEX TERMS Modulation classification, modulation recognition, MIMO systems, distribution test, low complexity.

I. INTRODUCTION

In communication systems where modulation information is unknown to the receiving end, modulation classification algorithm can be employed to provide such information by analyzing statistical characteristics of the receiving signal [1]. Despite having been studied extensively in SISO systems [2]–[6], the implementation of modulation classification in MIMO systems faces many new challenges [7], [8]. Due to the increase in the number of transmitting antennas, received signals exhibits much more complex characteristics, especially combined with higher-order modulations.

Choqueuse *et al.* first formulated the likelihood based method for modulation classification in MIMO systems [9]. The proposed method provided optimal classification accuracy when perfect channel-state-information is available to the receiver. However, the required number of operations grows exponentially with the number of transmitting antennas and the modulation order. Thus the optimal accuracy can only be achieved when sufficient computational power is provided or there is no constraint on processing time. In reality, most computation systems are unable to dedicate enough computational power to such tasks. Such problem is worsened when the number of antennas and the order of modulations increase to facilitate higher data throughput.

Kanterakis and Su tried to reduce the computational complexity of the likelihood based classifier by modifying the likelihood function [10]. However, the level of complexity reduction is limited and classification accuracy is significantly affected. Dulek proposed a likelihood based method in combination with an online channel estimator [11]. It significantly improves the computational complexity for the estimation process of channel-state-information. However, the classification process remains largely the same as the one developed by Choqueuse. Mühlhaus *et al.* developed a feature based MIMO modulation classifier using high-order statistics [12]. Similar feature based methods have also be explored by Han *et al.* [13]. While the complexity of feature based classifiers is often unaffected by the number of transmitting antennas or the modulation order, they are unable to fully utilized the channel-state-information to achieve optimized classification accuracy. A combination of moment features and likelihood maximization has been experimented for a balance between classification accuracy and computational complexity in [20]. However, its performance has not been validated in MIMO systems.

Recently, research on machine learning aided modulation classification has been developing rapidly. Zhang *et al.*

The associate editor coordinating the review of this manuscript and approving it for publication was Mostafa M. Fouda.

proposed a dictionary learning based modulation classifier by training with known signal and classifying unknown signals with their sparse representation on the trained dictionary [14], [15]. Peng *et al.* adopted convolutional neural network (CNN) for the task of modulation classification [16]. Similar deep learning based methods have also been proposed by Ali and Yangyu [21], Zheng *et al.* [17] and Yang *et al.* [18]. Most of these machine learning based methods claim to have low computational complexity. However, such claim is only valid when ignoring the length of the training process. In addition, none of these methods was developed for MIMO systems.

To address the computational complexity issues of modulation classification in MIMO systems, the authors propose a distribution based classifier with significantly lower complexity and the ability to utilize channel-state-information for improved classification accuracy. Distribution test based methods have proven to be computationally efficient in SISO systems [19], [23]–[26]. As distribution tests can be vulnerable against abnormality under noisy conditions, we propose to combine three classic distribution tests together with a novel variance distribution test for improved robustness in AWGN channel and slow fading channels. The main contributions of the paper can be summarized as follows: (1) deriving the Kolmogorov–Smirnov, Cramer-Von Mises, and Anderson-Darling tests for modulation classification in MIMO systems; (2) proposing the variance distribution test; and (3) developing a MLP mechanism to provide classification decision by considering results from different distribution tests jointly.

The paper is organized as follows. The signal model is established in Section II. The distribution tests employed are presented in Section III. The classification strategy is discussed in Section IV. Section V includes the simulated experiment setup, numerical results as well as analysis of both classification and computational complexity based on the results.

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 $N_R \geq N_T$, i.e.,

$$\boldsymbol{H} = \begin{pmatrix} h_{1,1} & \cdots & h_{1,N_T} \\ \vdots & \ddots & \vdots \\ h_{N_R,1} & \cdots & h_{N_R,N_T} \end{pmatrix}. \tag{2}$$

$\boldsymbol{\omega}_k = [\omega_k(1), \omega_k(2), \cdots, \omega_k(N_R)]^T$ 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.

Prior to classification, the received signal samples are first normalized to zero mean and unit power on their in-phase and quadrature components respectively. The normalization formula is defined by

$$r_k = \frac{\Re(r_k) - \overline{\Re(r)}}{\sigma(\Re(r))} + j\frac{\Im(r_k) - \overline{\Im(r)}}{\sigma(\Im(r))}, \tag{3}$$

where $\overline{\Re(r)}$ and $\overline{\Im(r)}$ indicate the means of the real and imaginary components of the complex signal separately, and $\sigma(\Re(r))$ and $\sigma(\Im(r))$ represent the standard deviations of the real and imaginary components of the complex signal separately.

III. THE GOODNESS-OF-FIT TEST

The Goodness-of-Fit (GoF) is often used to measure the consistency between observed samples and hypothesised 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 $\{r_k\}_{k=1}^N$, where the real and imaginary components, the phase, or the magnitude of the received signal $\{r_k\}$ can be used. In order to compute the empirical cumulative distribution, consider $\{z_k\}_{k=1}^N$ containing N number of feature values of the received signal organized in order. Let $F_1(z)$ denote the empirical cumulative distribution obtained from the sequence of features extracted from received symbols. $F_1(z)$ can be represented as:

$$F_1(z) = \frac{|\{k : z_k \leq z, 1 \leq k \leq N\}|}{N} \triangleq \frac{1}{N}\sum_{k=1}^{N} \mathbb{I}(z_k \leq z), \tag{4}$$

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 equiprobable, the theoretical distribution $F_0(z)$ of z_k under modulation candidate $\mathcal{M}_k$ is given by

$$F_0(z) = 1 - \frac{1}{|\mathcal{M}_k|}\sum_{x \in \mathcal{M}_k} Q\left(\frac{\sqrt{2}|x|}{\sigma}, \frac{\sqrt{2}z}{\sigma}\right), \tag{5}$$

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

Assuming that the hypothesized distribution is $F_0(z)$, the modulation classification problem can be expressed as a hypothesis testing problem with a null hypothesis H_0, and the general alternative hypothesis H_1, i.e.,

$$H_0 : F_1(z) = F_0(z), \tag{6}$$
$$H_1 : F_1(z) \neq F_0(z). \tag{7}$$

The intuitive understanding of the hypothesis H_0 is that $\{z_k\}_{k=1}^N$ is an i.i.d. sequence generated by the empirical distribution $F_1(z)$ against alternative H_1 which states that $\{z_k\}_{k=1}^N$ is not an i.i.d. sequence generated by the empirical distribution function $F_1(z)$. Under the null hypothesis, $F_1(z)$ will be close to $F_0(z)$ when N is large enough.

The modulation classification decision can be achieved by finding the the modulation candidate which provides the minimum distance between empirical distribution and hypothesized distribution using the GoF test. This process can be described with the following equation,

$$\hat{\mathcal{M}} = \arg \min_{1 \le k \le M} D_{(.)}^k, \tag{8}$$

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

Different GoF tests have been proposed in mathematical statistics based on the definition of distance between the two distributions $F_1(z)$ and $F_0(z)$. The distance between the two distributions indicates the fit between the empirical distribution and the hypothesized distribution. There are different types of non-parametric distribution tests to calculate such difference including Kolmogorov-Smirnov test, Cramer-von Mises test, and Anderson-Darling test.

A brief overview of steps involved in modulation classification using the GoF test is 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

for each receiving antenna n_r **do**
  Obtain the sequence of signal features $\{z_k\}_{k=1}^N$;
  Sort the signal samples with sort ($\{z_k\}_{k=1}^N$);
  Calculate the empirical distribution using (4);
  for each modulation candidate $\mathcal{M}_k$ **do**
    Generate modulation symbols $S_m = HA_m$;
    for each modulation symbol s_i **do**
      Calculate the theoretical CDF;
    end for
    Calculate the mean of the theoretical CDF;
    Evaluate the distance between empirical
    CDF and theoretical CDF;
  end for
end for
Calculate the mean of the test statistics $D_{(.)}^k$;

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

A. KOLMOGOROV-SMIRNOV TEST

In the Kolmogorov-Smirnov (KS) test [27], firstly, a sequence of feature $\{z_k\}_{k=1}^N$ is obtained from the received signals $\{r_k\}_{k=1}^N$ using any of the features such as magnitude, phase, or real and imaginary components. Secondly, empirical distribution is evaluated from the features obtained from received samples using (24). Thirdly, for each modulation candidate $\mathcal{M}_k$ we obtain the hypothesized distribution. The test statistic for KS test is given as

$$D_{KS} \triangleq \sup_{z \in \mathbb{R}} |F_1(z) - F_0(z)|, \tag{9}$$

and in practice, it is calculated by

$$\hat{D}_{KS} = \max_{1 \le n \le N} |F_1(z_n) - F_0(z_n)|. \tag{10}$$

B. CRAMER-VON MISES TEST

The Cramer-von Mises test (CvM) is also a statistical test model same as KS test [28], used for measuring the goodness-of-fit of the theoretical CDF $F_0(z)$ compared to the empirical CDF $F_1(z)$ in one-sample scenarios [30]. The test statistics is defined as the integral of the squared difference between the empirical CDF $F_1(z)$ and the theoretical CDF $F_0(z)$, i.e.,

$$D_{CvM} \triangleq \int_{-\infty}^{\infty} [F_1(z) - F_0(z)]^2 \, dF_0(z). \tag{11}$$

Then the decision statistics is defined in practice as

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

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

C. ANDERSON-DARLING TEST

The Anderson-Darling (AD) test is also a statistical test model with no parameters needed [29]. AD test is based on empirical distribution function. Compared with the Cramer-von Mises test, the test statistics is the Cramer-von Mises test with an added weight function $w(z)$. It is defined as

$$D_{AD} \triangleq \int_{-\infty}^{\infty} [F_1(z) - F_0(z)]^2 \, w(z) dF_0(z), \tag{13}$$

where the weight function is defined by

$$w(z) = \frac{1}{F_0(z)(1 - F_0(z))}. \tag{14}$$

Then the decision statistics is defined in practice as

$$\begin{aligned} \hat{D}_{AD} &= ND_{AD} \\ &= -n - \frac{1}{n} \sum_{i=1}^{n} [(2i-1) \ln F(z_i) \\ &\quad + (2n + 1 - 2i) \ln[1 - F(z_i)]], \end{aligned} \tag{15}$$

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

The aforementioned tests measure distribution mismatch with maximum or accumulated distribution difference in essence. However, when signals are not properly equalized or the samples are not equally distributed among all modulation symbols, the mismatch can be exaggerated even if the observed signal matches the hypothesized distribution. In such cases, the consistency of the difference over different locations can provide useful information to compensate for such effects.

In order to utilize this information, we propose a Variance (Var) test to determine how the distribution mismatch varies over the different signal components. The Var test results is obtained by calculating the variance ($\hat{\sigma}^2$) of the difference (d_i) between empirical CDF and theoretical CDF, as given by

$$d_i = F_1(z_i) - F_0(z_i), \tag{16}$$

$$\hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{N}(d_i - \mu)^2 = \left(\frac{1}{N}\sum_{i=1}^{N} d_i^2\right) - \mu^2, \tag{17}$$

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

$$\mu = \frac{1}{N}\sum_{i=1}^{N} d_i. \tag{18}$$

Notably, the purpose of the Var test is not to provide accurate classification independently. Instead, it is meant to be combined with other GoF tests to provide improved overall robustness.

IV. PROPOSED CLASSIFICATION STRATEGY

Given results from four different distribution tests, the classifier needs to combine these results and make a final classification decision. In this paper, a multi-layer perceptron (MLP) network is used for this purpose. Similar approaches have been adopted in modulation classification with good effect [31], [32]. A MLP network consists of at least three layers of nodes: an input layer, a hidden layer, and an output layer. Fig. 1 shows the structure of multilayer perceptron network with a single hidden layer. Except for the input nodes, each node is a neuron that uses a nonlinear activation function which maps the input space non-linearly into an output space. The MLP process can be expressed using the following 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 a nonlinear activation function.

A. TEST RESULTS COMBINATION

In this paper, the test statistics $D^k_{(.)}$ based on GoF tests and Var test are first combined to create a matrix $t^*_{mod_n}$ that matches the MLP input layer.

Firstly, the test results are grouped into the matrix $\boldsymbol{t}^*_{mod}$ in the following format,

$$\boldsymbol{t}^*_{mod} = \begin{pmatrix} t^{KS}_{mod_1} & t^{KS}_{mod_2} & \cdots & t^{KS}_{mod_L} \\ t^{CvM}_{mod_1} & t^{CvM}_{mod_2} & \cdots & t^{CvM}_{mod_L} \\ t^{AD}_{mod_1} & t^{AD}_{mod_2} & \cdots & t^{AD}_{mod_L} \\ t^{Var}_{mod_1} & t^{Var}_{mod_2} & \cdots & t^{Var}_{mod_L} \end{pmatrix}, \tag{20}$$

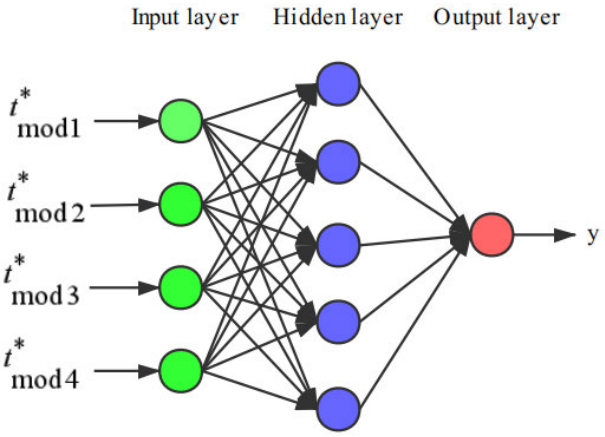


FIGURE 1. Multilayer perceptron.

TABLE 1. MLP classifier structure.

Parameters	values
Network Layers	3
Input Neurons	4(3)
Hidden Neurons	8
Output Neurons	1
Activation Function	tanh
Weight Optimization	lbfgs

where L represents the total number of the candidate modulation types. Each element in the matrix corresponds to the result from an individual distribution test when tested against a candidate modulation. For example, $t^{KS}_{mod_1}$ is equivalent to $\hat{D}_{KS}$ when tests against the first modulation candidate. The other elements can be interpreted in a similar fashion.

Secondly, each column in $\boldsymbol{t}^*_{mod}$ are combined using weighted sum to obtain $t^*_{mod_l}$ for the lth modulation candidate. The process can be depicted in the following equation,

$$t^*_{mod_l} = w_k t^{KS}_{mod_l} + w_c t^{CvM}_{mod_l} + w_a t^{AD}_{mod_l} + w_v t^{Var}_{mod_l}. \tag{21}$$

In this paper, the weights are equally distributed among different test results. However, further optimization can be performed to achieve improved performance.

B. MLP CLASSIFIER STRUCTURE

The MLP network consists of three or more layers (an input and an output layer with one or more hidden layers) of nonlinear activating nodes. 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

$$h = \sqrt{m+n} + a, \tag{22}$$

where h is the number of hidden layer neurons, m is the number of input layer neurons, n is the number of output layer neurons, a is an integer between 1 and 10. Specific parameters for MLP classifier structure are listed in TABLE 1. Fig. 2 shows the classifier strategy of proposed method for MIMO system.

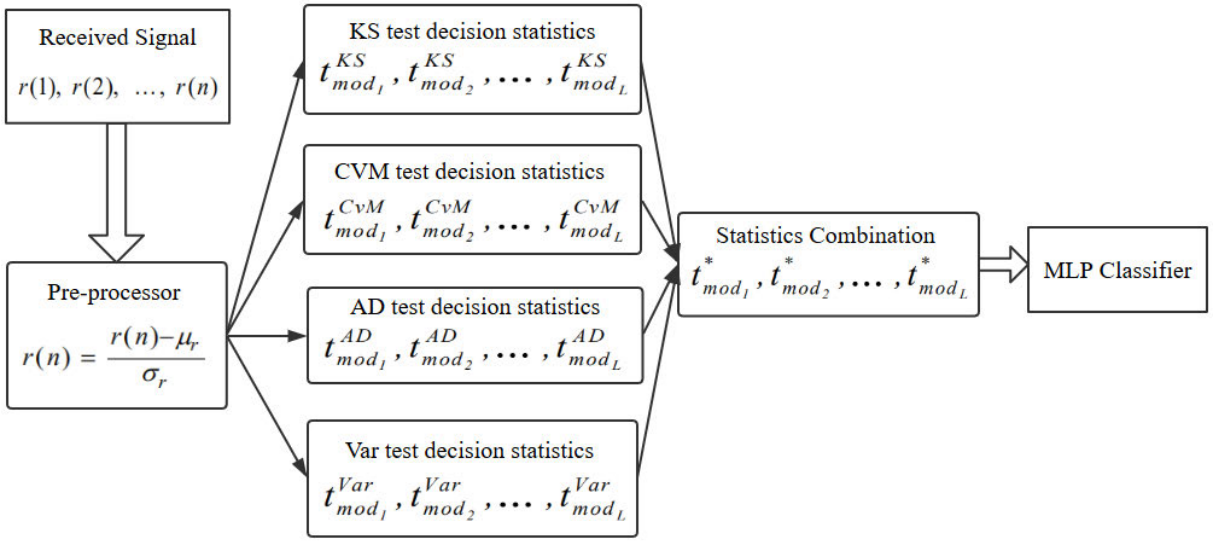


FIGURE 2. Classifier strategy of proposed method for MIMO system.

V. EXPERIMENTS AND RESULTS

To evaluate the proposed modulation classification methods, simulations have been conducted in the PYTHON environment to investigate the classifier performance under AWGN channels and fading channel conditions. In the simulations, the number of transmitting antennas and receiving antennas are set to $N_T = 2$ and $N_R = 4$. Unless otherwise stated, there are $N = 128$ samples in each signal realization. Under each channel condition, 1,000 signal realizations are generated for each signal modulation.

The following modulation schemes are considered in all of our simulations $\mathfrak{M} = \{BPSK, QPSK, 8-PSK, 4-QAM, 16-QAM\}$. Other digital modulations can be classified in the same procedure with very little modification. Other parameters of the simulations are summarized in TABLE 2.

The classification accuracy P_{acc} for each modulation candidate is given in percentage and estimated by

$$P_{acc} = \frac{N_c}{N_{total}} \times 100\%, \tag{23}$$

where N_c is the number of a modulation correctly classified, and N_{total} is the total number of test realizations.

A. BENCHMARKING CLASSIFIER

To test the proposed method against other state-of-the-art classifiers, The Maximum Likelihood (ML) classifier is selected as the benchmarking classifier [9]. The likelihood equation in the ML classifier as given in the following equation

$$\mathcal{L}(r_k|\mathcal{M},\sigma) = \prod_{k=1}^{N}\sum_{m=1}^{M}\frac{1}{M}\frac{1}{(2\pi\sigma^2)^{N_R}}e^{-\frac{||r_k - \boldsymbol{H}A_m||_F^2}{2\sigma^2}}, \tag{24}$$

TABLE 2. Test parameters.

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

where $||\cdot||_F^2$ represents the Frobenius norm. K denote the number of states of given a modulation signal, and then $M = K^{N_T}$ represents the number of transmitted symbol vectors. A_m is the mth possible transmitted symbol set.

With the consideration of analytical convenience in practical cases, we adopt the natural logarithm of the likelihood function $\mathcal{L}$ as the likelihood value, i.e.,

$$\begin{aligned}&\log\mathcal{L}(r_k|\mathcal{M},\sigma)\\ &= \log\left(\prod_{k=1}^{N}\sum_{m=1}^{M}\frac{1}{M}\frac{1}{(2\pi\sigma^2)^{N_R}}e^{-\frac{||r_k - \boldsymbol{H}A_m||_F^2}{2\sigma^2}}\right)\\ &= \sum_{k=1}^{N}\log\left(\sum_{m=1}^{M}\frac{1}{M}\frac{1}{(2\pi\sigma^2)^{N_R}}e^{-\frac{||r_k - \boldsymbol{H}A_m||_F^2}{2\sigma^2}}\right).\end{aligned} \tag{25}$$

The modulation classification decision can be obtained by the one which maximizes the Likelihood Function, i.e.

$$\hat{\mathcal{M}} = \arg\max_{\mathcal{M}_i\in\mathfrak{M}} \log\mathcal{L}(r_k|\mathcal{M},\sigma). \tag{26}$$

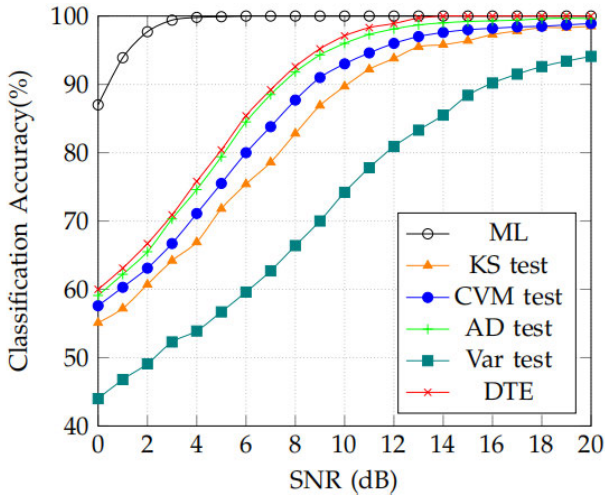


FIGURE 3. Classification accuracy of Quadrature-based classifier (BPSK, 8-PSK, 4-QAM, 16-QAM) with varying noise levels.

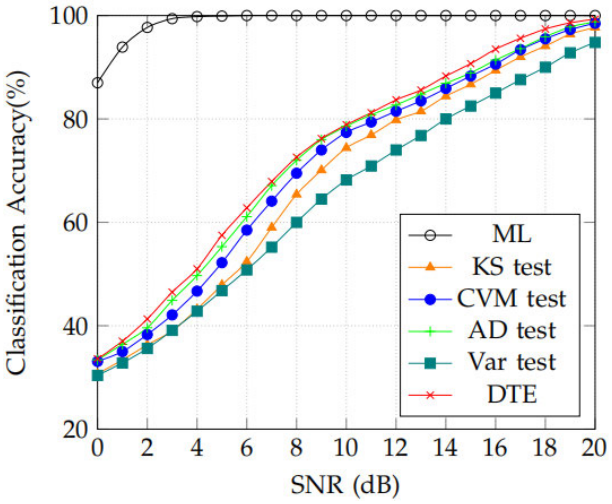


FIGURE 5. Classification accuracy of Magnitude-based classifier (BPSK, 8-PSK, 4-QAM, 16-QAM) with varying noise levels.

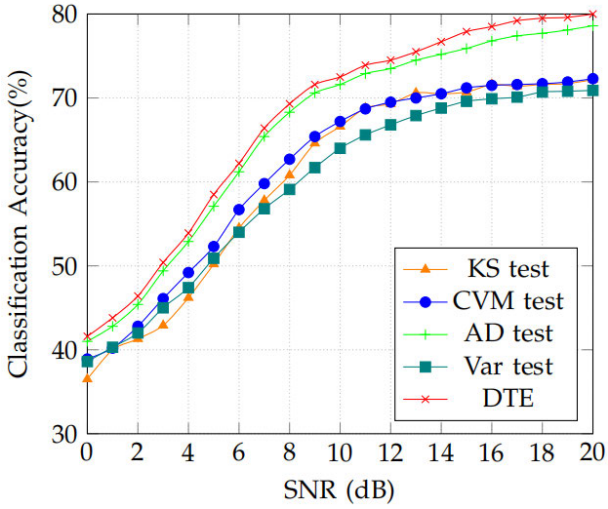


FIGURE 4. Classification accuracy of Quadrature-based classifier (4-QAM, 16-QAM, 64-QAM) with varying noise levels.

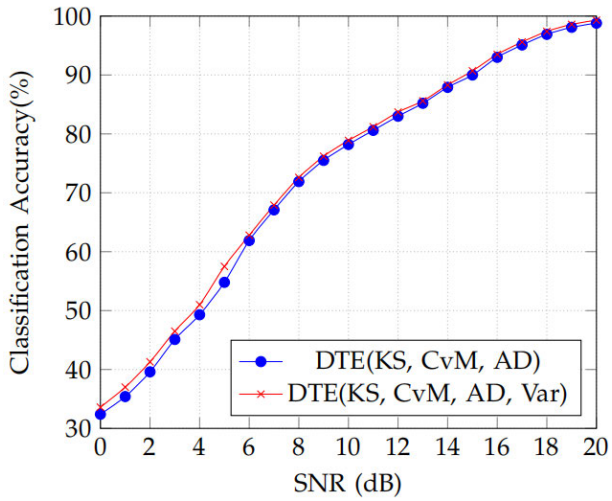


FIGURE 6. Classification accuracy of Magnitude-based classifier (BPSK, 8-PSK, 4-QAM, 16-QAM) with varying noise levels.

B. PERFORMANCE IN AWGN CHANNEL

In the AWGN channel, two sets of experiments are conducted to evaluate performance under varying noise levels and varying signal lengths respectively. Since distribution tests can be performed on different signal components, they are applied to signal quadrature, magnitude and phase components separately in different tests.

In the first set of experiments, the simulation results are provided to compare the performances of the ML classifier, four classifiers each based on a single distribution test, and the proposed DTE classifier. As demonstrated in Fig. 3, the ML classifier achieves highest accuracy at all SNR levels as expected. When applied to signal quadrature, the DTE classifier is able to achieve good classification accuracy at SNR $\geq$ 10 dB where classification accuracy is at 97.1%. Perfect classification can be achieved by DTE at SNR $\geq$ 14 dB. Compared against other distribution test based classifier, the proposed DTE classifier achieves between 0.5% and 1.5% higher accuracy than the best signal test classifier over the tested SNR range. The biggest performance improvement is seen between SNR of 10 dB and 14 dB. Fig. 4 shows the classification performance of the proposed DTE classifier for higher-order modulation modes. The results show that the DTE classifier is better than any single distributed test classifier in the entire SNR range. Under high SNR conditions, the difference in recognition accuracy between DTE and distribution test is more obvious. In Fig. 5, it is clear that the the proposed DTE classifier based on signal magnitude maintains the advantage over single test classifiers especially in the higher and lower range of SNRs. The biggest difference is observed at 16 dB where DTE offers an accuracy of 93.5% and AD test based classifier

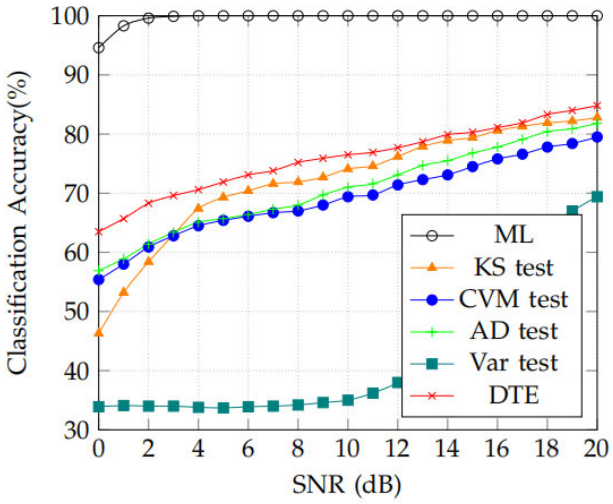


FIGURE 7. Classification accuracy of Phase-based classifier (BPSK, QPSK, 8-PSK) with varying noise levels.

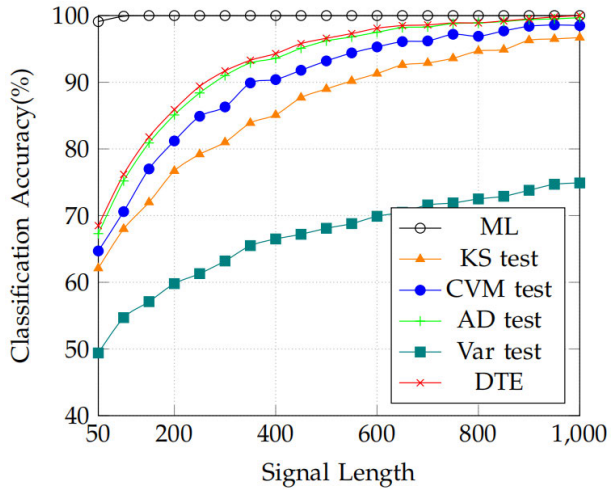


FIGURE 8. Classification accuracy of Quadrature-based classifier (BPSK, 8-PSK, 4-QAM, 16-QAM) with varying signal length with SNR of 5 dB.

offers 91.4%. Fig. 6 demonstrates performance difference between DTE without the proposed Var test and DTE with the Var test. With the introduction of Var test, the DTE classifier achieves improved classification accuracy over the entire SNR range tested. When the SNR is lower than 6 dB, the advantage of DTE with Var test over DTE without Var test is more obvious and the biggest difference is 2.7% at SNR $=$ 5 dB. In addition to the quadrature and magnitude component, signal phase is easily recognized as an intuitive object for analysis when classifying M-PSK modulations. According to the classification performance in Fig. 7, it is distinct that the proposed DTE classifier outperforms the single test classifiers throughout the SNR range. In the mid and lower range of SNRs, DTE classifier provides greater advantage and achieves a classification accuracy of 68.3% offering the largest performance improvement of 6.9%.

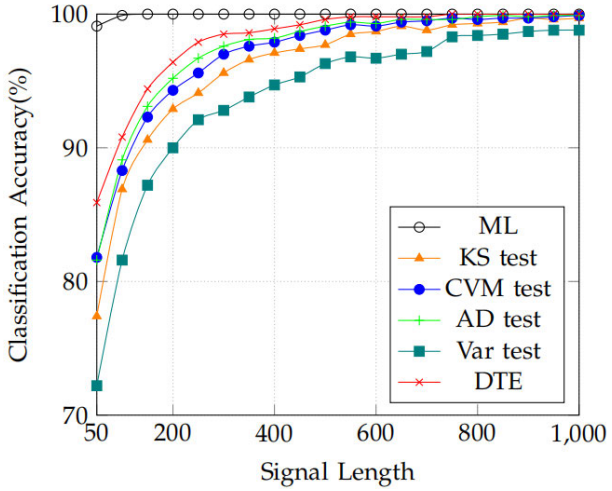


FIGURE 9. Classification accuracy of Magnitude-based classifier (BPSK, 8-PSK, 4-QAM, 16-QAM) with varying signal length with SNR of 16 dB.

In the second set of experiments, with majority of the experiment setup unchanged, different signal lengths N between 50 and 1000 are tested. According to the results shown in Fig. 8, the ML classifier excels in all signal lengths from $N = 50$ to $N = 1000$. Its performance advantage reduces when signal length increases. When compared with the AD test classifier, the DTE method shows a superior robustness especially when signal length is below 600, while the improvement is limited when the number of samples is more than 650. In Fig. 9 where signal magnitude is used for analysis, the biggest advantage of DTE is observed at $N = 50$, where DTE obtains a classification accuracy of 85.9%, which is 4.1% over CvM 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 are available (signal length above 500), all methods are able to achieve a relatively stable performance. Fig. 10 shows the accuracy of different classifier when tested under a fading channel with varying signal length. Under the fixed SNR (16 dB) and frequency offset (2×10^{-3} of the carrier frequency), the performance of the DTE classifier gradually improves when the signal length increases and is hardly affected by the frequency offset. Overall, the distribution test classifiers maintain an upward trend when the signal length gets larger, but the recognition performance partially decreases, especially when the signal length is large. The cause of the decrease may be that signal realizations is insufficient. As presented in Fig. 11 where signal phase is considered for the classification of M-PSK modulations, the performance of the DTE classifier is superior to KS test for all the signal length ranges, and DTE gains more advantage with signal length below 400.

Classification accuracy of proposed DTE for individual modulation in AWGN channel are shown in TABLE 3.

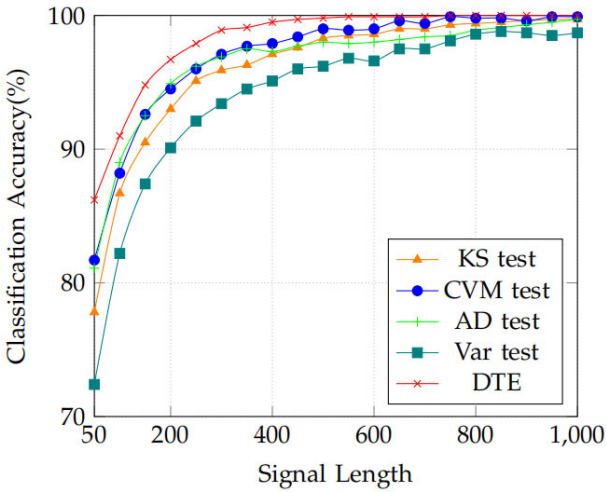


FIGURE 10. Classification accuracy of Magnitude-based classifier (BPSK, 8-PSK, 4-QAM, 16-QAM) with varying signal length with SNR of 16 dB and frequency offset of 2×10^{-3} of the carrier frequency.

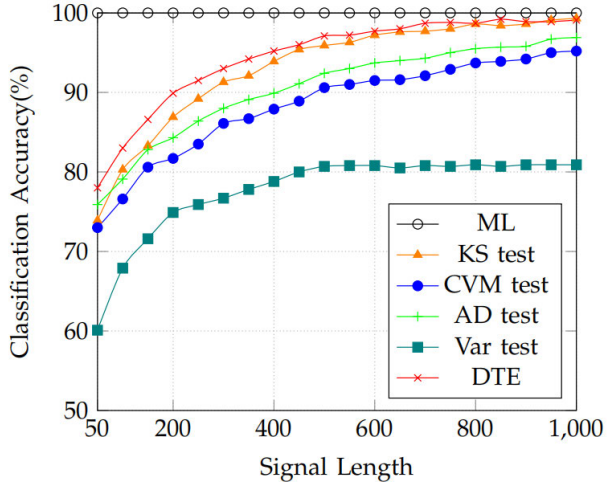


FIGURE 11. Classification accuracy of Phase-based classifier (BPSK, QPSK, 8-PSK) with varying signal length with SNR of 20 dB.

TABLE 3. Classification accuracy of proposed DTE for individual modulation in AWGN channel.

Quadrature	0dB	5dB	10dB	15dB	20dB
BPSK	99.8%	100.0%	100.0%	100.0%	100.0%
8-PSK	48.4%	72.6%	93.1%	100.0%	100.0%
4-QAM	41.9%	70.2%	97.2%	100.0%	100.0%
16-QAM	46.0%	76.0%	96.5%	100.0%	100.0%

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%

Phase	0dB	5dB	10dB	15dB	20dB
BPSK	85.4%	99.4%	99.9%	100.0%	100.0%
QPSK	51.5%	58.1%	63.1%	65.7%	71.3%
8-PSK	50.3%	58.0%	65.2%	76.3%	87.3%

Firstly, it can be concluded that the BPSK modulation can be recognized most easily with Quadrature-based classifier. The modulations of 8-PSK, 4-QAM, 16-QAM, however,

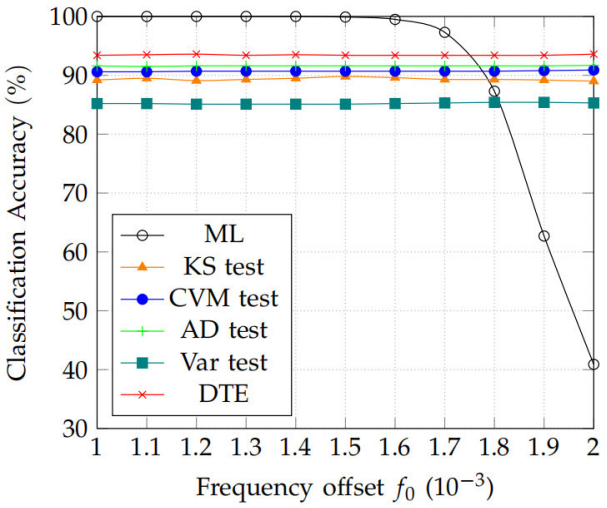


FIGURE 12. Classification accuracy of Magnitude-based classifier (BPSK, 8-PSK, 4-QAM, 16-QAM) with frequency offset from 1×10^{-3} to 2×10^{-3} of the carrier frequency.

also have a great performance at SNR $>$ 10 dB. Secondly, Both BPSK and 16-QAM modulations can be classified with higher accuracy by the DTE classifier using magnitude feature at higher SNRs. Thirdly, though the numerical results show that the proposed algorithm with phase feature exhibits good performance for BPSK modulation, QPSK and 8-PSK modulations in classification process reflect poor performance for all tested values of SNRs.

C. PERFORMANCE IN FADING CHANNEL

In order to test the proposed method in more complex channel conditions, phase offsets and frequency offsets are applied to the signal separately along with fixed level of AWGN noise at 16 dB.

It must be stated that the magnitude-based DTE calssifier is selected as the proposed solution in this scenario. As the signal magnitude does not distort under phase offset and frequency offset, the magnitude of the received signals is selected for robust performance in these scenarios. Slow fading is assumed in the following tests. Thus the phase offset is assumed to be a constant throughout a signal realization while a range of offset values between 6° and 20° are tested. Frequency offset is added to the signal separately from the phase offset, in which case the amount of offset tested is set to a range between 1×10^{-3} and 2×10^{-3} of the carrier frequency. Reminding set up follows the ones in AWGN channel with only the SNR fixed at 16 dB.

In the frequency offset channel, it is evident from Fig. 12 that the proposed DTE classifier excels over other benchmarking classifiers. The performances of ML classifier are significantly affected by frequency offset due to the severe mismatch between received signals and the assumed models. while the ML classifier starts with the 100% classification accuracy with no offset, the performance starts to decline sharply when more frequency offset is introduced. However,

TABLE 4. Number of operators needed for different classifiers used in experiments.

Classifiers	Addition	Multiplier	Logarithm	Exponential	Memory
ML [9]	$6NM^{N_T}N_R \cdot \sum_{m=1}^{M^{N_T}} I_m$	$5NM^{N_T}N_R \cdot \sum_{m=1}^{M^{N_T}} I_m$	$NM^{N_T}N_R$	$NM^{N_T}N_R \cdot \sum_{m=1}^{M^{N_T}} I_m$	$M^{N_T}N_R$
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	$4MNN_R$
MLP	$\sum_{k=2}^{l} n_{k-1}n_k$	$\sum_{k=2}^{l} n_{k-1}n_k$	0	0	$\sum_{k=2}^{l} n_k$

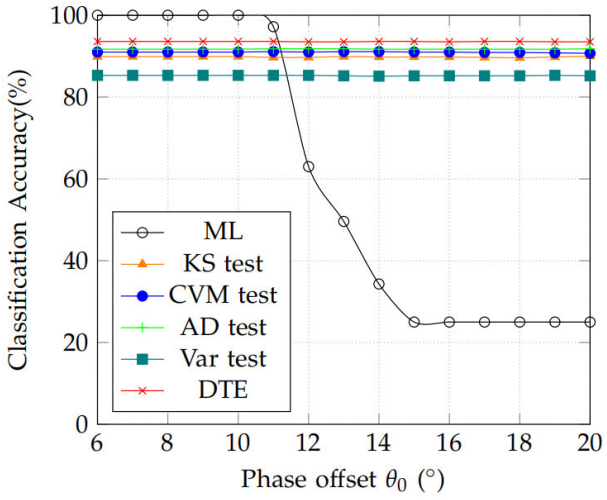


FIGURE 13. Classification accuracy of Magnitude-based classifier (BPSK, 8-PSK, 4-QAM, 16-QAM) with phase offset from 6° to 20° at SNR of 16 dB.

the KS test, CvM test, AD test, Var test and the proposed DTE classifier are all able to maintain a constant level of performance in the given frequency offset range. Particularly, the DTE classifier achieves a classification accuracy of 93% with an advantage of 2% as compared to the 91% classification accuracy of the AD test classifier.

In the phase offset channel, it can be observed from Fig. 13 that the proposed DTE classifier achieves a consistent classification accuracy throughout the tested phase offset range. while the KS test, CvM test, AD test, Var test and the proposed DTE classifier all achieve the robust performance, the average classification accuracy of the DTE classifier remains at 93% and is 2% higher than the 91% classification accuracy observed for the AD test classifier. In contrast, the ML classifier is rather sensitive to the slow phase offset. From Fig. 13, it can be seen that, despite being more accurate with little phase offset, the ML classifier become less accurate when there is more than 11° of phase offset introduced. When there is more than 11° of phase offset, the proposed DTE classifier become the best option among the benchmarked methods.

D. COMPLEXITY ANALYSIS

To analyze and compare the computational complexity of the proposed method against other classifiers, the number of mathematical operations required for each classifier are worked out. The numerical results are presented in Table 4, where 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 mth candidate modulation is represented by I_m, l is the number of network layers, n_k is the number of the kth layer neurons.

From the six tested methods in Table 4, 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 scales exponentially with the number of transmitting antennas and modulation order. For the Var test classifier, it is clear that it has a much lower the computational complexity and requires fewer additions compared against the KS test based 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 for the DTE classifier is only the sum of the complexity of distribution test. Furthermore, compared with distribution test classifiers, while the 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 the Var test need identical memory and have a much higher memory space than the ML classifier when received signals N is larger.

VI. CONCLUSION

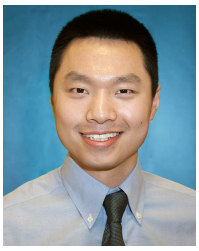
A novel modulation classifier, namely DTE, has been proposed for the purpose of classifying M-QAM and M-PSK

modulations in MIMO systems with low computational complexity. It has been demonstrated by numerical results that the proposed DTE classifier outperforms other distribution test based classifier in different channel conditions. While the performance of the DTE classifiers is inferior to the ML classifier in AWGN channel, DTE can offer superior performance in slow fading channels when appropriate signal component is chosen for analysis. In addition, the computational complexity of the proposed method is much lower than the ML classifier. In massive MIMO systems with high-order modulations, the ML classifier would be impractical due to its high computational complexity. On the other hand, the proposed DTE classifier offers good classification accuracy, requiring much lower complexity.

REFERENCES

[1] Z. Zhu and A. K. Nandi, *Automatic Modulation Classification: Principles, Algorithms and Applications*, 1st ed. New York, NY, USA: Wiley, 2015.
[2] E. E. Azzouz and A. K. Nandi, *Automatic Modulation Recognition of Communication Signals*. Amsterdam, The Netherlands: Kluwer Academic, 1996.
[3] Z. Zhu and A. K. Nandi, "Blind digital modulation classification using minimum distance centroid estimator and non-parametric likelihood function," *IEEE Trans. Wireless Commun.*, vol. 13, no. 8, pp. 4483–4494, Aug. 2014.
[4] M. Waqar Aslam, Z. Zhu, and A. Kumar Nandi, "Automatic modulation classification using combination of genetic programming and KNN," *IEEE Trans. Wireless Commun.*, vol. 11, no. 8, pp. 2742–2750, Aug. 2012.
[5] A. K. Nandi and E. E. Azzouz, "Algorithms for automatic modulation recognition of communication signals," *IEEE Trans. Commun.*, vol. 46, no. 4, pp. 431–436, Apr. 1998.
[6] E. E. Azzouz and A. K. Nandi, "Automatic identification of digital modulation types," *Signal Process.*, vol. 47, no. 1, pp. 55–69, Nov. 1995.
[7] Y. A. Eldemerdash, O. A. Dobre, and M. Oner, "Signal identification for multiple-antenna wireless systems: Achievements and challenges," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 3, pp. 1524–1551, 3rd Quart., 2016.
[8] M. Marey and O. A. Dobre, "Blind modulation classification algorithm for single and multiple-antenna systems over frequency-selective channels," *IEEE Signal Process. Lett.*, vol. 21, no. 9, pp. 1098–1102, Sep. 2014.
[9] V. Choqueuse, S. Azou, K. Yao, L. Collin, and G. Burel, "Blind Modulation Recognition for MIMO Systems," *Milit. Tech. Acad. Rev.*, vol. 12, no. 2, pp. 183–196, 2009.
[10] E. Kanterakis and W. Su, "Modulation classification in MIMO systems," in *Proc. MILCOM - IEEE Mil. Commun. Conf.*, San Diego, CA, USA, Nov. 2013, pp. 35–39.
[11] B. Dulek, "Online hybrid likelihood based modulation classification using multiple sensors," *IEEE Trans. Wireless Commun.*, vol. 16, no. 8, pp. 4984–5000, Aug. 2017.
[12] M. S. Muhlhaus, M. Oner, O. A. Dobre, and F. K. Jondral, "A low complexity modulation classification algorithm for MIMO systems," *IEEE Commun. Lett.*, vol. 17, no. 10, pp. 1881–1884, Oct. 2013.
[13] L. Han, F. Gao, Z. Li, and O. A. Dobre, "Low complexity automatic modulation classification based on order-statistics," *IEEE Trans. Wireless Commun.*, vol. 16, no. 1, pp. 400–411, Jan. 2017.
[14] K. Zhang, E. L. Xu, Z. Feng, and P. Zhang, "A dictionary learning based automatic modulation classification method," *IEEE Access*, vol. 6, pp. 5607–5617, 2018.
[15] K. Zhang, E. L. Xu, H. Zhang, Z. Feng, and S. Cui, "Data driven automatic modulation classification via dictionary learning," *IEEE Wireless Commun. Lett.*, vol. 7, no. 4, pp. 586–589, Aug. 2018.
[16] S. Peng, "Modulation Classification Based on Signal Constellation Diagrams and Deep Learning," *IEEE Trans. Neural Net. Learn. Syst.*, vol. 30, no. 3, pp. 718–727, Oct. 2019.
[17] S. Zheng, P. Qi, S. Chen, and X. Yang, "Fusion methods for CNN-based automatic modulation classification," *IEEE Access*, vol. 7, pp. 66496–66504, 2019.
[18] C. Yang, Z. He, Y. Peng, Y. Wang, and J. Yang, "Deep learning aided method for automatic modulation recognition," *IEEE Access*, vol. 7, pp. 109063–109068, 2019.
[19] Z. Zhu, M. Waqar Aslam, and A. K. Nandi, "Genetic algorithm optimized distribution sampling test for M-QAM modulation classification," *Signal Process.*, vol. 94, pp. 264–277, Jan. 2014.
[20] M. Abu-Romoh, A. Aboutaleb, and Z. Rezki, "Automatic modulation classification using moments and likelihood maximization," *IEEE Commun. Lett.*, vol. 22, no. 5, pp. 938–941, May 2018.
[21] A. Ali and F. Yangyu, "Automatic modulation classification using deep learning based on sparse autoencoders with nonnegativity constraints," *IEEE Signal Process. Lett.*, vol. 24, no. 11, pp. 1626–1630, Nov. 2017.
[22] Z. Wu, S. Zhou, Z. Yin, B. Ma, and Z. Yang, "Robust automatic modulation classification under varying noise conditions," *IEEE Access*, vol. 5, pp. 19733–19741, 2017.
[23] F. Wang and X. Wang, "Fast and robust modulation classification via kolmogorov-smirnov test," *IEEE Trans. Commun.*, vol. 58, no. 8, pp. 2324–2332, Aug. 2010.
[24] M. Mohammadkarimi and O. A. Dobre, "Blind identification of spatial multiplexing and alamouti space-time block code via kolmogorov-smirnov (K-S) test," *IEEE Commun. Lett.*, vol. 18, no. 10, pp. 1711–1714, Oct. 2014.
[25] F. Wang, O. A. Dobre, C. Chan, and J. Zhang, "Fold-based kolmogorov–smirnov modulation classifier," *IEEE Signal Process. Lett.*, vol. 23, no. 7, pp. 1003–1007, Jul. 2016.
[26] X. Lin, Y. A. Eldemerdash, O. A. Dobre, S. Zhang, and C. Li, "Modulation classification using received Signal's amplitude distribution for coherent receivers," *IEEE Photon. Technol. Lett.*, vol. 29, no. 21, pp. 1872–1875, Nov. 1, 2017.
[27] F. J. Massey, "The kolmogorov-smirnov test for goodness of fit," *J. Amer. Stat. Assoc.*, vol. 46, no. 253, pp. 68–78, Mar. 1951.
[28] H. Cramér, "On the composition of elementary errors," *Scandin. Actuarial J.*, vol. 1928, no. 1, pp. 13–74, Jan. 1928.
[29] T. W. Anderson and D. A. Darling, "A test of goodness of fit," *J. Amer. Stat. Assoc.*, vol. 49, no. 268, pp. 765–769, 1954.
[30] F. Laio, "Cramer-von mises and anderson-darling goodness of fit tests for extreme value distributions with unknown parameters," *Water Resour. Res.*, vol. 40, no. 9, pp. 1–10, Sep. 2004.
[31] M. L. D. Wong and A. K. Nandi, "Automatic digital modulation recognition using artificial neural network and genetic algorithm," *Signal Process.*, vol. 84, no. 2, pp. 351–365, Feb. 2004.
[32] K. Hassan, I. Dayoub, W. Hamouda, C. N. Nzeza, and M. Berbineau, "Blind digital modulation identification for spatially-correlated MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 11, no. 2, pp. 683–693, Feb. 2012.

ZIKANG GAO received the B.S. degree in communication engineering from the Luoyang Institute of Science and Technology, Luoyang, China, in 2017. He is currently pursuing the M.S. degree with Soochow University, Suzhou, China. His research interests include optimization of computational complexity of modulation classification algorithms and machine learning in MIMO systems.



ZHECHEN ZHU (Member, IEEE) received the Ph.D. degree in electronic and computer engineering from Brunel University London, U.K., in 2014.

Since 2015, he has been with the School of Electronic and Information Engineering, Soochow University, Suzhou, China, where he is currently an Associate Professor. He has authored a research monograph entitled *Automatic Modulation Classification: Principles, Algorithms and Applications* (Wiley, 2015). His current research interests include signal processing in communication systems, signal classifications, data mining, and machine learning.

ASOKE K. NANDI (Fellow, IEEE) received the Ph.D. degree in physics from the Trinity College, University of Cambridge, Cambridge, U.K.

He held academic positions in several universities, including Oxford, U.K., Imperial College London, U.K., Strathclyde, U.K., and Liverpool, U.K., as well as a Finland Distinguished Professorship, Jyvaskyla, Finland. He was a Chair and the Head of Electronic and Computer Engineering with Brunel University London, U.K., in 2013. Additionally, he is a Distinguished Visiting Professor with Tongji University, China, and an Adjunct Professor with University of Calgary, Canada. He co-discovered the three fundamental particles known as W^+, W^- and Z^0 (by the UA1 team at CERN), in 1983, providing the evidence for the unification of the electromagnetic and weak forces, for which the Nobel Committee for Physics, in 1984, awarded the prize to his two team leaders for their decisive contributions. He has made many fundamental theoretical and algorithmic contributions to many aspects of signal processing and machine learning. He has much expertise in Big Data, dealing with heterogeneous data. He has authored over 600 technical publications, including 240 journal articles as well as five books, entitled *Condition Monitoring with Vibration Signals: Compressive Sampling and Learning Algorithms for Rotating Machines* (Wiley, 2020), *Automatic Modulation Classification: Principles, Algorithms and Applications* (Wiley, 2015), *Integrative Cluster Analysis in Bioinformatics* (Wiley, 2015), *Blind Estimation Using Higher-Order Statistics* (Springer, 1999), and *Automatic Modulation Recognition of Communications Signals* (Springer, 1996). The H-index of his publications is 75 (Google Scholar) and his ERDOS number is two. His current research interests include signal processing and machine learning, with applications to communications, image segmentations, biomedical data, and so on.

Prof. Nandi is a Fellow of the Royal Academy of Engineering (U.K.) as well as a Fellow of seven other institutions, including the IEEE and IET. He received many awards, including the Institute of Electrical and Electronics Engineers (USA) Heinrich Hertz Award, in 2012, the Glory of Bengal Award for his outstanding achievements in scientific research, in 2010, the Water Arbitration Prize of the Institution of Mechanical Engineers (U.K.), in 1999, and the Mountbatten Premium, Division Award of the Electronics and Communications Division, of the Institution of Electrical Engineers (U.K.), in 1998. From 2018 to 2019, he was the IEEE EMBS Distinguished Lecturer.