SIGNAL DETECTION IN MIMO COMMUNICATIONS SYSTEM WITH NON-GAUSSIAN NOISES BASED ON DEEP LEARNING AND MAXIMUM CORRENTROPY CRITERION

Mohammad Reza pourmir[1], Reza Monsefi[1] and Ghosheh Abed Hodtani[2]

[1]Computer Department, Engineering Faculty Ferdowsi University of Mashhad (FUM)
[2]Electrical engineering Department,
Engineering Faculty Ferdowsi University of Mashhad (FUM)

ABSTRACT

In this paper, we study signal detection in multi-input-multi output (MIMO) communications system with non-Gaussian noises such as Middleton Class A noise, Gaussian mixtures and alpha stable distributions, using several deep neural network-based detector models such as FULLYCONNECTED and DETNET detector. By applying information theoretic criterion of Maximum Correntropy , SVD analysis on the channel matrix and reducing network complexity, the suggested deep neural network detector performs well in environments with non-Gaussian noises and, compared to the deep neural network-based detector with MSE loss function, achieves better performance.

KEYWORDS

Signal Detection, MIMO, Deep Learning, Information theory.

1. INTRODUCTION

Machine learning (ML) based wireless communications is a new research topic, and although with various theoretical difficulties due to wireless network heterogeneity, varying service demands, a huge number of connections, it has become a helpful tool with potential capabilities. Wireless channel estimation in non-Gaussian noises is a difficult problem. In this study, we look at receiver design for non-Gaussian noise-interrupted transmission channels. One of the most difficult tasks in MMO communication systems is to reduce bit error rate (BER) without increasing the complexity of the inclusive detector [1].

Multiple antenna communications known as multi-input multi-output (MIMO) technology, incorporated in several wireless standards such as 802.11 and LTE, has two important advantages, i.e., multiplexing gain (increasing transmission rate) and diversity gain (decreasing error probability); however, in high Signal to Noise Ratios, and hence, necessitating its application with orthogonal frequency division multiplexing (OFDM). A MIMO transmitter sends multiple data streams per antenna, while a MIMO receiver receives multiple copies of multiple data streams with noise on each antenna. A MIMO detector detects multiplexed data from all receiver antennas, in the presence of noise and interferences [2-5].

In general, a maximum a posterior (MAP) detector gives excellent detection performance but has a limited detection range. Iterative detection algorithms [6-7], successive detection (SD) [8-9] and successive interference cancellation (SIC) [10] show high performance with medium

DOI:10.5121/ijwmn.2022.14501

complexity. All of these detectors need comprehensive knowledge of channel state information (CSI). The development of machine learning (ML) techniques is one possible strategy to reducing detector complexity. They may offer solutions working well in addition to being simple to be deployed [11]. Deep learning, in particular, when improved by large data, may record complicated associations; the name "deep learning" refers to the number of network layers and includes more than two depths, and yet bigger networks can accept information better than shallow models [12]. Several researches have looked at the use of ML for signal recognition, which has resulted in a number of detectors. DNN is widely utilized because of its great learning capabilities [13 -14]. However, because of the enormous number of neurons and layers, particularly those built for large-scale systems, its computational cost and energy consumption are often high [15-35]. Other ML methods for signal recognition include CNN [16-19], RNN [20] and ELM [21-22].

Low-grade matrix analysis and core learning are two approaches that may be used to develop sophisticated learning systems. The computational cost of processing large of core matrices can be considerably reduced by using low-grade decomposition. Existing techniques, on the other hand, are generally unsupervised and do not incorporate supplementary information such as class labels, making parsing less effective for a specific learning objective. Kernel learning approaches, on the other hand, aim to construct nuclear matrices whose structure is well matched with the learning goal, hence improving the overall performance of kernel methods.

In this study, we leverage the advantages of matrix analysis in addition to the Gaussian kernel to achieve the benefits of both methodologies. The following are the key points of this paper: We discuss RESNET-based detectors and FULLY CONNECTED networks for MIMO communications, by exploring maximum Correntropy criterion in the presence of non-Gaussian noises such as Middleton Class A models, Gaussian mixtures and alpha stable distributions. Based on numerical data, we illustrate which model outperforms the others in terms of performance using the suggested cost function.

In section II, preliminaries are explained. In Section III, the proposed model is described, and in Section IV, we have numerical results. Section V concludes the paper.

2. PRELIMINARIES

In this section, we review briefly the MIMO channel, non-Gaussian noises, signal detection and maximum Correntropy criterion (MCC)

2.1. MIMO Model

As said in Introduction, MIMO technology uses multiple antennas in both the transmitter and receiver, leading to improvement in data transmission rate and or detection error probability in high signal to noise ratios (SNRs), necessitating the usage of OFDM (orthogonal frequency division multiplexing) signalling [36]. In a MIMO system with N_T Transmitter antennas and N_R Receiver antennas, the received baseband signal is as follows:

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{z} \tag{1}$$

$$\begin{bmatrix} \boldsymbol{y_1} \\ \vdots \\ \boldsymbol{y_{N_R}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{h_{11}} & \cdots & \boldsymbol{h_{1N_T}} \\ \vdots & \ddots & \vdots \\ \boldsymbol{h_{N_R1}} & \cdots & \boldsymbol{h_{N_RN_T}} \end{bmatrix} \begin{bmatrix} \boldsymbol{x_1} \\ \vdots \\ \boldsymbol{x_{N_T}} \end{bmatrix} + \begin{bmatrix} \boldsymbol{z_1} \\ \vdots \\ \boldsymbol{z_{N_R}} \end{bmatrix}$$

The $\boldsymbol{y}$ and $\boldsymbol{x}$ are the received and transmitted signal vectors, respectively. And $\boldsymbol{H}$ represents the wireless channel coefficients matrix between the transmitter and receiver, whereas $\boldsymbol{z}$ represents the noise vector (Fig. one) to be independent of noise.

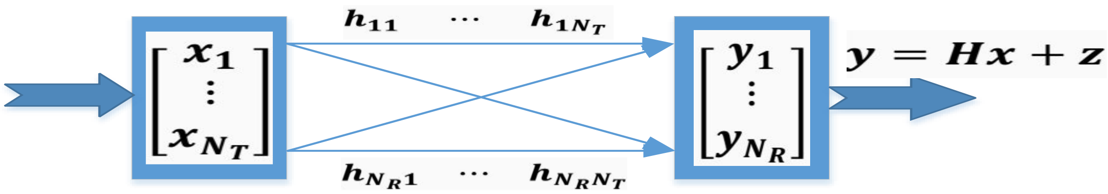

Figure 1-MIMO model

One of the most critical difficulties in wireless communication is the reduction of noise effects. As a result, using deep learning to detect the signal can be an effective method.

2.2. Non-Gaussian Noises in Wireless Channel

The Gaussian distribution for noise is an assumption based on central limit theorem, leading to easily obtain mathematical results about channel estimation, channel rate and capacity and signal detection. However, in many practical situations, the Gaussian distribution is not adequate to model the underlying noise phenomenon, and various non-Gaussian noise models are used to design and analyze the performance of communication systems. Non-Gaussian noise models such as Middleton Class A noise, Gaussian mixtures and alpha stable distributions, on the other hand, makes difficult to solve communication problems, e.g., to estimate channel model parameters prior to signal detection.

Many of the optimization problems are solved by minimizing the known mean squared error (MSE). The MSE second order standard is computationally simple and simple to implement, but it does not work well in non-Gaussian noises. To deal with non-Gaussian noises, the application of information-theoretic criteria provides an efficient approach to dealing with nonlinear, non-static, and non-Gaussian problems.

Gaussian mixtures (GM) have been used to model various non-Gaussian noises [23-24]. Probability density function of a Gaussian mixture $f_M(x)$ is given by the sum of the weighted density functions of N Gaussian distribution probabilities $f_i\,(x)$ with means and variances

$$(\mu_1 ، \dots ،\mu_N ،\sigma_1{}^2 ، \dots ،\sigma_N{}^2)$$

as:

$$\boldsymbol{f}_{\boldsymbol{M}}(\boldsymbol{x}) = \sum_{\boldsymbol{i}=\boldsymbol{1}}^{\boldsymbol{N}} \boldsymbol{\lambda}_{\boldsymbol{i}} \boldsymbol{f}_{\boldsymbol{i}}\,(\boldsymbol{x}) \tag{2}$$

Where, the weights are positive and add up to one. The expectation is:

$$\boldsymbol{\mu} = \sum_{\boldsymbol{i}=\boldsymbol{1}}^{\boldsymbol{N}} \boldsymbol{\lambda}_{\boldsymbol{i}} \boldsymbol{\mu}_{\boldsymbol{i}} \tag{3}$$

This model is general, and it can be used to approximate various symmetric and mean densities, such as the Laplacian and alpha stable distributions. Impact noise is a type of acoustic noise caused by instantaneous sharp sounds such as clicks and pops, known as impulses. Middleton first described the phenomenon of impact noise in detail in the 1960s, when he proposed a model for impulse noise in communication. To create the model, he described impulsive noise in a system as a series of pulses that occur at random impulses of varying duration and intensity.

$$z(t) = \sum_{i=1}^{I} a_i \, \delta(t - t_i)$$

Many authors investigated percussion noise modelling in the wake of Middleton noise models. Some percussion noise models from the literature can be expressed as follows:

I. Impulse noise models without memory

- Middleton Class A
- Bernoulli-Gaussian
- Symmetric Alpha-Stable distribution(SαS)

II. Impulse noise models with memory

- Markov-Middleton
- Markov-Gaussian

To model the phenomena encountered in practice, $S\alpha S$ distributions are used. These phenomena do not have a Gaussian distribution; instead, their possible distributions may have fat tails when compared to Gaussian distribution sequences. The following parameters describe these distributions:

- α: is the characteristic exponent, and describes the tail of the distribution$(1 < \alpha \leq 2)$
- β: describes the skewness of the distribution $(-1 \leq \beta \leq +1)$
- γ: is the scaling parameter $\gamma > 0$
- δ:is a real number that gives the location of the distribution

Despite being such an appealing model for impulsive noise, the alpha stable distribution family has received little attention in the literature because, with the exception of a few special cases, there are no explicit compact expressions for the probability density function. Although this method is efficient for a large number of samples, it does not provide an analytic form and is not suitable for real-time applications due to the extensive numerical integrations involved. Only for the Gaussian $(\alpha = 2)$, Cauchy $(\alpha = 2; \beta = 0)$, and Pearson $(\alpha = \frac{1}{2}; \beta = -1)$ distributions there exist closed form expressions for the probability distribution function (P.D.F.) Figure below (Fig. 2) compares various noise models.

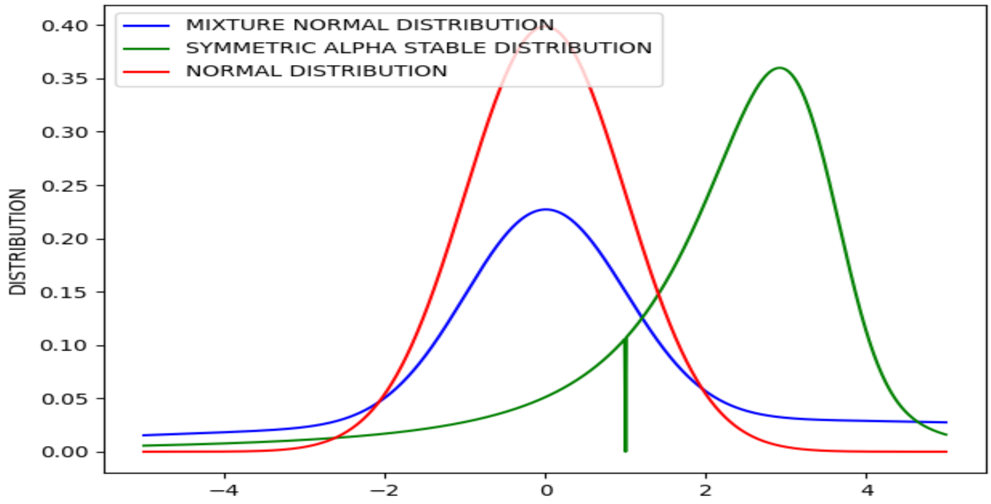

Figure 2-comparison of additive noise models

2.3. Signal Detection

Signal recognition and classification are two of the most fundamental signal processing issues, with numerous applications in domains such as communication, voice recognition, biomedicine, image processing, and many more [25]. The difficulty of recognizing the existence of a favored signal from noisy data is referred to as signal detection. A significant variety of techniques for detecting communication signals have been developed in the literature. [26-27-28]

The great majority of these algorithms assume that the noise in the radio channel has a Gaussian distribution. Many non-Gaussian statistical phenomena exist [29]. We employ the maximum Correntropy criterion-based cost function in the DETNET-based deep neural network model as well as fully connected models to increase signal identification performance in the presence of non-Gaussian noise.

2.4. The Maximum Correntropy Criterion (MCC)

The Maximum Correntropy criterion (MCC) has found effective applications in the fields of signal processing and machine learning in recent years, which is especially beneficial when signals are tainted by heavy-duty impact noise [30-33]. The best model is obtained using the MCC between the model's output variable and the TARGET variable:

$$\boldsymbol{M}^* = \boldsymbol{arg}\,\underset{\boldsymbol{M}\in\boldsymbol{\mathcal{M}}}{\boldsymbol{max}}(\boldsymbol{V}_{\boldsymbol{\sigma}}(\boldsymbol{T}\text{،}\,\boldsymbol{Y})) \tag{4}$$

Where M^* is the optimal model, $\boldsymbol{Y}$ is the model output variable and $\boldsymbol{T}$ is the target variable, $\boldsymbol{V}_{\boldsymbol{\sigma}}$ Cross-Correntropy for two random $\boldsymbol{T}$ and $\boldsymbol{Y}$, When sampling from the densities, cross-Correntropy can be estimated with PARZEN estimator:

$$\widehat{\boldsymbol{V}}_{\boldsymbol{\sigma}}(\boldsymbol{e}) = \frac{\boldsymbol{1}}{\boldsymbol{N}}\sum_{\boldsymbol{i=1}}^{\boldsymbol{N}} \boldsymbol{G}_{\boldsymbol{\sigma}}\,(\boldsymbol{t}_{\boldsymbol{i}} - \boldsymbol{y}_{\boldsymbol{i}}) = \frac{\boldsymbol{1}}{\boldsymbol{N}}\sum_{\boldsymbol{i=1}}^{\boldsymbol{N}} \boldsymbol{G}_{\boldsymbol{\sigma}}\,(\boldsymbol{e}_{\boldsymbol{i}}) \tag{5}$$

and $\boldsymbol{G}_{\boldsymbol{\sigma}}(\boldsymbol{e})$ is a Gaussian kernel function as follows:

$$G_\sigma(e) = \frac{1}{\sigma\sqrt{2\pi}}\exp(-\frac{e^2}{2\sigma^2})$$

Where $e = T - Y$ is the error between T and Y and σ is the core bandwidth. Since the Gaussian kernel function $G_\sigma(e)$ is a local function of the error variable , Correlation may be utilized as a strong error measure in signal processing and machine learning and the maximum Correntropy criterion (MCC) algorithm is

$$MCC = \max_{w} \hat{V}_{\boldsymbol{\sigma}}\ (e)$$

Where the parameters w control the error PDF $e = T - Y$.

Assume the following is the sample set:

$$\mathcal{D} = \{(t_i \text{،} y_i)\}_{i=1}^{N}$$

In this case, the experimental form $\hat{V}_\sigma(e)$ Based on the equation (5) is as follows:

$$\hat{\boldsymbol{V}}_{\boldsymbol{\sigma}}(\boldsymbol{e}) = \frac{\mathbf{1}}{\boldsymbol{N}}\sum_{\boldsymbol{i=1}}^{\boldsymbol{N}} \boldsymbol{G}_{\boldsymbol{\sigma}}\left(\frac{\left(\boldsymbol{t}_{\boldsymbol{i}} - \boldsymbol{g}(\boldsymbol{x}_{\boldsymbol{i}})\right)^2}{\mathbf{2}\boldsymbol{\sigma}^{\mathbf{2}}}\right) \quad \textbf{(6)}$$

Where the function g is the mapping between the input and output of the model and σ scaling parameter. In view of the above considerations, in this paper, our main concerns are the following two aspects:

- We are concerned with the connections between the Correntropy loss and the least squares loss when they are employed in MIMO detection problems.
- We make a comparison between detector models based on Correntropy loss criterion.

3. THE PROPOSED MODEL

In general, the primary notion in detectors based on model machine learning is based on a learning algorithm such that the output $\hat{\boldsymbol{s}}_{\boldsymbol{ML}}$ of the model is a high-accuracy estimate of the transmitted signal vector and may be expressed as follows:

$$\hat{\mathbf{s}}_{\boldsymbol{ML}} = \boldsymbol{\mathcal{T}}(\boldsymbol{f}(\boldsymbol{s} \text{ ، } \boldsymbol{\mathcal{P}})) \quad \textit{(7)}$$

This comprises $\mathcal{P}$ a set of learnable parameters and $\boldsymbol{s}$ as well as any incoming information and channel state information (CSI); f is a nonlinear function and $\mathcal{T}$ that has been mapped to function f. Unlike traditional detectors, which have a high level of complexity, detectors inspired by machine learning approaches have been developed. These detectors are known as unfolding detectors. Deep neural networks are used in these detectors because neural networks employ nonlinear functions for training. Furthermore, in Figure 3, we may employ fully connected FC-DNN neural networks also for detection although DETNET models require a layer-by-layer structure.

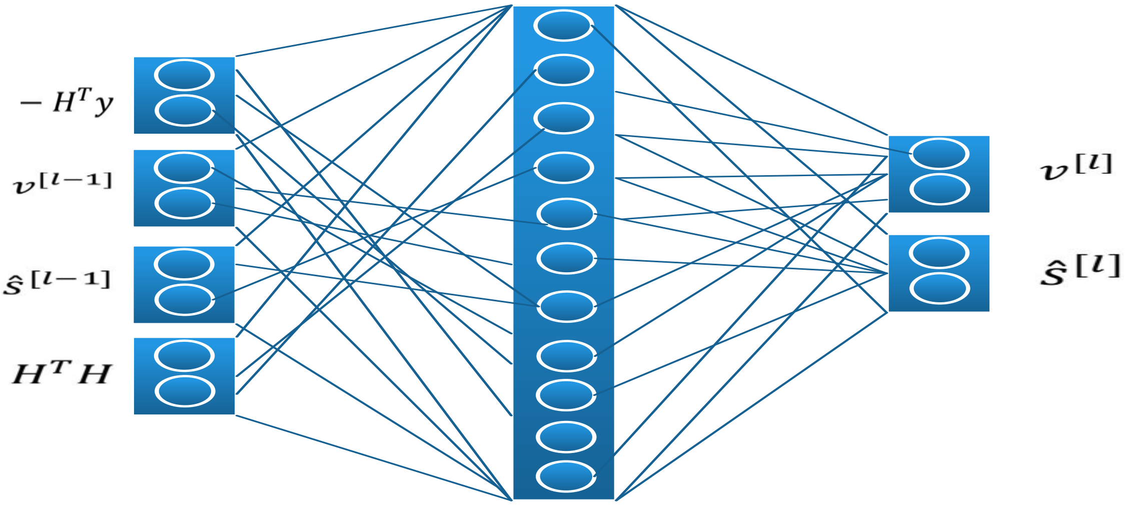

Figure 3. Fully connected Detector

DETNET was first proposed in "Deep MIMO detection" [34]. Based on the iterative optimization method, projected gradient descent(PGD) solves the problem of signal detection. DETNET is a multi-layered neural network for signal detection in a MIMO system and has a structure based on Figure 4. In the DETNET network, all layers have the same structure and each layer has 4 inputs $\left(v^{(l-1)}، H^T H، \hat{s}^{(l-1)}، -H^T y\right)$. where, y wireless channel output vector and H channel matrix and $H^T y$ and $H^T H$ common inputs are all layers .DETNET is iterative network, The output of each unit can be used as the overall network output, and as the number of network units increases, the output of each unit gets closer to transmission signals based on Euclidean distance. We should make the network as deep as possible for better performance. Network output $\hat{s}_{ML}$ is the result of update on the L layer of the network and the output of each layer is as follows:

$$\hat{\boldsymbol{s}}^{[l]} = \boldsymbol{f}[\boldsymbol{s} - \boldsymbol{\delta}^{[l]} \frac{\partial\|\boldsymbol{y} - \boldsymbol{Hs}\|^2}{\partial \boldsymbol{s}}]_{\boldsymbol{s}=\hat{\boldsymbol{s}}^{[l-1]}} = \boldsymbol{f}[\hat{\boldsymbol{s}}^{[l-1]} - \boldsymbol{\delta_1}^{[l]} \boldsymbol{H}^T \boldsymbol{y} + \boldsymbol{\delta_2}^{[l]} \boldsymbol{H}^T \boldsymbol{H} \hat{\boldsymbol{s}}^{[l-1]}] \tag{8}$$

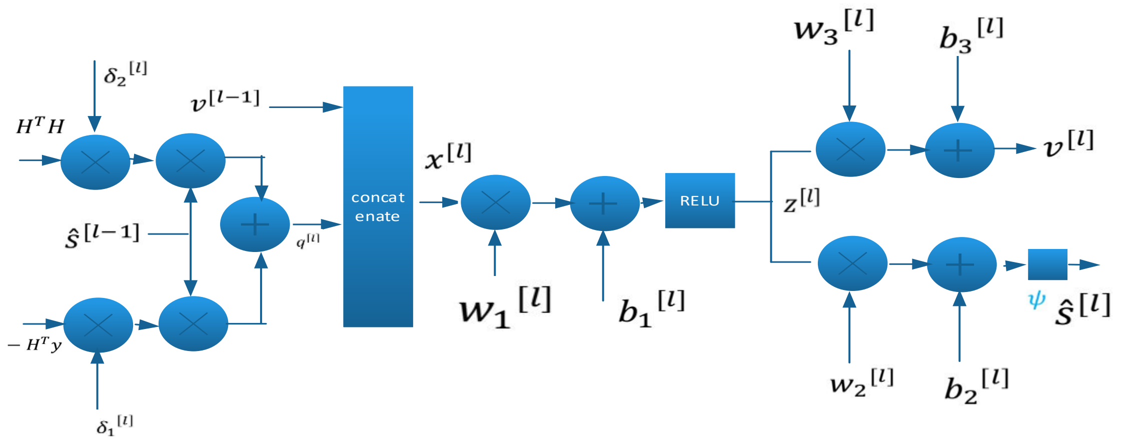

Figure 4. DETNET model

Where, $f(\cdot)$ is a nonlinear mapping (RELU) and δ_1 and δ_2 are step sizes and $\hat{s}^{[l]}$ is network output. The loss function of the network is as follows:

$$Loss = \sum_{l=1}^{L} \log(l) \frac{\left\| x - \hat{s}^{[l]} \right\|^2}{\| x - (H^T\, H)^{-1} H^T\, y \|^2}$$

In summary, the input-output relationships in the DETNET are as follows:

$$q^{[l]} = \hat{s}^{[l-1]} - \delta_1{}^{[l]} H^T y + \delta_2{}^{[l]} H^T H \hat{s}^{[l-1]}$$
$$x^{[l]} = \left[v^{[l-1]} q^{[l]} \right]^T$$
$$z^{[l]} = f(w_1{}^{[l]} x^{[l]} + b_1{}^{[l]})$$
$$\hat{s}^{[l]} = \psi(w_2{}^{[l]} z^{[l]} + b_2{}^{[l]})$$
$$v^{[l]} = w_3{}^{[l]} z^{[l]} + b_3{}^{[l]}$$

Where the initial values are $v^{[0]} = 0$ and $\hat{s}^{[0]} = 0$ and the network parameters are:

$$\left\{ w_1{}^{[l]}، w_2{}^{[l]}، w_3{}^{[l]}، b_1{}^{[l]}، b_2{}^{[l]}، b_3{}^{[l]}، \delta_1{}^{[l]}، \delta_2{}^{[l]} \right\}$$

Finally, the output of the L layer is equal to

$$s_{ML} = Q[\hat{s}^{[L]}]$$

Where Q is the sign function. As can be seen, Although DetNet is a high-performance MIMO detection neural network model, there is still room for improvement. We simplify the network in this section. SCNET is a sparsely connected neural network that has been simplified. Although there are two outputs in Fig.4, only one is used as an approximation of the transmitted signal x, therefore, parameter $v^{[l]}$ can be deleted, reducing network complexity (SCNET). When the number of transmitter and reception antennas on a DETNET network is equal, the network's performance suffers. DETNET networks with Rayleigh channel models and modulations such as BPSK and QPSK have an excellent record. SCNET networks outperform DETNET networks in terms of performance and are less complicated due to a reduced number of parameters. Furthermore. The loss function for SCNET based on MSE is as follows:

$$\boldsymbol{LOSS_{MSE}} = \sum_{\boldsymbol{l=1}}^{\boldsymbol{L}} \mathbf{log}(\boldsymbol{l}) \left\| \boldsymbol{x} - \left(\boldsymbol{H^T\, H} \right)^{\mathbf{-1}} \boldsymbol{H^T\, z} \right\|^{\mathbf{2}} \tag{9}$$

Where H is the channel matrix , z is the nonlinear mapping of the received signal vector, and x is the sent signal vector . Although the DETNET model performs well, there is still opportunity for development. We utilize metrics based on information theory such as MCC which outperform MSEs in non-Gaussian noises, and compare the models' performances. Based on MCC criteria, we define the loss function as follows:

$$\boldsymbol{LOSS_{MCC}} = \boldsymbol{\sigma^2} \left[\mathbf{1} - \boldsymbol{\eta} \mathbb{E}[\boldsymbol{G_\sigma(e)}] \right] = \boldsymbol{\sigma^2} \left[\mathbf{1} - \boldsymbol{\eta} \frac{\mathbf{1}}{\boldsymbol{N}} \sum_{\boldsymbol{i=1}}^{\boldsymbol{N}} \boldsymbol{G_\sigma(e_i)} \right] \tag{10}$$

Where $\sigma > 0$ denotes a scale parameter and e denotes the error signal vector, and is defined as follows:

$$e = x - H^T z^{[l]}$$

The vector $z^{[l]}$ is a nonlinear mapping of the received signal of the wireless channel in l layer and $G_\sigma(e)$ is a Gaussian kernel function with scale parameter $\sigma > 0$ and η is learning coefficient. Next, in order to reduce the complexity of the model, we remove the inverse matrix calculation from Equation (9) and examine the performance of the model. Secondly, we perform a pre processing on the model using SVD or QR analysis on the channel matrix H to separate and eliminate the interference caused by the transmitter antennas, and in this case, we also examine the performance of the model. The proposed MCC metric based on the vector $z^{[l]}$ and x and based on Equation (10) is as follows:

$$\boldsymbol{LOSS} = \boldsymbol{\sigma^2}\left[\mathbf{1} - \boldsymbol{\eta}\mathbb{E}\left[\boldsymbol{G_\sigma}\left(\boldsymbol{x} - (\boldsymbol{H^T H})^{-1}\boldsymbol{H^T z^{[l]}}\right)\right]\right] = \boldsymbol{\sigma^2}[\mathbf{1} - \boldsymbol{\eta}\frac{\mathbf{1}}{\boldsymbol{N}}\sum_{\boldsymbol{i=1}}^{\boldsymbol{N}}\left[\boldsymbol{G_\sigma}\left(\boldsymbol{x_i} - (\boldsymbol{H^T H})^{-1}\boldsymbol{H^T z_i}^{[l]}\right)\right]] \tag{11}$$

Our goal is to make maximum similarity between network output and send signals as close as possible. Next by removing the inverse matrix calculation from Equation (11), the proposed metric is as follows:

$$\boldsymbol{LOSS} = \boldsymbol{\sigma^2}[\mathbf{1} - \boldsymbol{\eta}\frac{\mathbf{1}}{\boldsymbol{N}}\sum_{\boldsymbol{i=1}}^{\boldsymbol{N}}\left[\boldsymbol{G_\sigma}\left(\boldsymbol{x_i} - \boldsymbol{H^T z_i}^{[l]}\right)\right]] \tag{12}$$

SVD analysis is a matrix factorization method used in many numerical applications of linear algebra such as PCA. This technique enhances our understanding of key components and provides a robust computational framework that allows us to achieve higher accuracy for datasets. These are important because they help to find methods for actually calculating and estimating results for the models and algorithms we use. In this case, we use the following loss function (Eq.13)

$$H = U\Sigma V^T$$
$$\boldsymbol{LOSS} = \boldsymbol{\sigma^2}\left[\mathbf{1} - \boldsymbol{\eta}\mathbb{E}\left[\boldsymbol{G_\sigma}\left(\boldsymbol{x} - \boldsymbol{\Sigma^T z^{[l]}}\right)\right]\right] = \boldsymbol{\sigma^2}[\mathbf{1} - \boldsymbol{\eta}\tfrac{\mathbf{1}}{\boldsymbol{N}}\textstyle\sum_{\boldsymbol{i=1}}^{\boldsymbol{N}}\boldsymbol{G_\sigma}\left(\boldsymbol{x_i} - \boldsymbol{\Sigma^T z_i}^{[l]}\right)] \tag{13}$$

In the following, based on the DETNET model, considering a fully connected neural network (Fig.3) and the proposed loss function, we examine the model and compare it with DETNET, by investigating completely connected models, deleting parameter $v^{[l]}$, and using the proposal loss function. The input and output of the first layer of the network are as follows:

$$x_i{}^{[l]} = \left[H^T y \;\; \hat{s}^{[l-1]} H^T H\,\right]^T$$
$$\hat{s}^{[l]} = f(w^{[l]} x_i{}^{[l]} + b^{[l]})$$

And hence, the loss function can be written as follows (Eq. 14):

$$\boldsymbol{LOSS} = \boldsymbol{\sigma^2}[\mathbf{1} - \boldsymbol{\eta}\mathbb{E}\left[\boldsymbol{G_\sigma}\left(\boldsymbol{x} - \hat{\boldsymbol{s}}^{[l]}\right)\right]] \tag{14}$$

In the following section, we will compare the performance of the models.

4. NUMERICAL RESULTS

In this section, we present some simulation results to confirm the effectiveness of the proposed DETNET detector based on the changes expressed on the non-Gaussian noise model. We express the results based on the fixed channel model, assuming 4 transmitter antennas and 8 receiver antennas. We explore BPSK modulation and evaluate the results in the Gaussian mixed noise model as well as the alpha stable distribution. We also use FULLY CONNECTED models and express the analysis based on the comparison of models. In each model and each repetition, we produce 5000 educational data and use 2000 repetitions in the model training stage.

We then obtain the bit error rate (BER) based on 1000 test data and repeat it 200 times. Using the fully connected model as well as the 90-layer DETNET model, we obtain the results based on the range of 1 to 20 for the SNR as the ratio of signal power to noise power, that the signal power is in the form of the following relation (noting to $y = Hx + z$ and independence $\boldsymbol{x}$ and $\boldsymbol{H}$):

$$\mathbb{E}\left[|\boldsymbol{H}|^{\mathbf{2}}\right]\mathbb{E}[|\boldsymbol{x}|^{\mathbf{2}}] \tag{15}$$

Where, $\boldsymbol{H}$ is the channel matrix and $\boldsymbol{x}$ is the transmitted signal. The noise is in the form of a Gaussian mixture based on Equation (2) and as follows (Equation (16)):

$$\boldsymbol{z} = \sum\nolimits_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{N}} \boldsymbol{\lambda}_{\boldsymbol{i}}\, \boldsymbol{\mathcal{N}}(\boldsymbol{\mu}_{\boldsymbol{i}}\text{‘}\boldsymbol{\sigma}_{\boldsymbol{i}}{}^{\mathbf{2}}\ ) \tag{16}$$

Based on this and using Equation (3), the power of noise can be expressed as follows (Equation (17)):

$$\boldsymbol{\sigma}_{\boldsymbol{z}}{}^{\mathbf{2}} = \sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{N}} \boldsymbol{\lambda}_{\boldsymbol{i}}\boldsymbol{\sigma}_{\boldsymbol{i}}{}^{\mathbf{2}} + \sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{N}} \boldsymbol{\lambda}_{\boldsymbol{i}}\,(\boldsymbol{\mu}_{\boldsymbol{i}} - \boldsymbol{\mu}) \tag{17}$$

Based on Equations (16) and (17), the SNR ratio can be expressed as follows:

$$SNR = \frac{\mathbb{E}[|H|^2]\mathbb{E}[|x|^2]}{\boldsymbol{\sigma}_{\boldsymbol{z}}{}^{\mathbf{2}}}$$

To obtain different SNR values, we assume that the values λ_i are constant and follow the following condition:

$$\lambda_1 > \lambda_2 > \cdots > \lambda_N$$

By making changes on the mean variance of Gaussian distributions and assuming the transmitted signal strength is constant, we calculate different SNR values, or assuming that the mean variance of the Gaussian distributions is constant and the transmitted signal power changes, we gain access to different SNR values. In this article, we use the second method and examine non-Gaussian noise in two cases. In the first case, we assume that the noise model of a Gaussian mixture includes a standard Gaussian distribution with a colored Gaussian distribution, and in the second case, we consider an impulse noise, which we estimate the noise model using Scale mixtures of the Gaussian to express the noise model.

4.1. Mixture Guassian Noise

The Gaussian mixture model can be considered as follows:

$$\boldsymbol{f}(\boldsymbol{z}) = (\mathbf{1} - \boldsymbol{\varepsilon})\boldsymbol{\eta}(\boldsymbol{z}) + \boldsymbol{\varepsilon}\boldsymbol{h}(\boldsymbol{z}) \tag{18}$$

where ε is some small positive constant,η is a Gaussian density function, and h is some other density function with heavier tails. Clearly, f defined by (Eq.18) is a valid density function as long as ε lies in the interval [0,1]. For small enough values of ε the behavior of f near the origin is dominated by that of η, assuming that h is a bounded function. For large values of $|z|$, however, h dominates the behavior of f since its tails decay at a slower rate than do those of η. In this case, we consider the added noise as follows:

$$z = \lambda_1 \mathcal{N}(0،1) + \lambda_2 \mathcal{N}(2،6) + \lambda_3 \mathcal{N}(4،8)$$

Based on the added noise model and SNR selection, the results in the added noise distribution and the two detector models FULLY CONNECTED and DETNET PRE SVD based on the proposed loss function (Equation [12]), are reviewed and compared with MSE-based loss function.

The results based on Table 1 show that the DETNET network achieves better results with SVD preprocessing in the proposed Gaussian noise. In addition, the BER TEST results and the comparison of FULLY CONNECTED and DETNET PRE SVD models based on the two loss functions are shown in Figures 5,6 .

As can be seen, firstly, in FULLY CONNECTED models and based on CORRENTROPY criterion for $Le\ parameter\ =\ 1.4$, the best results are obtained in high SNRs, but in low SNRs, FULLY CONNECTED model based on CORRENTROPY and $scale\ parameter =\ 2.1$ leads to better results, and secondly DETNET PRE SVD offers the best results based on CORRENTROPY criterion in high and low SNRs.

The results based on two loss functions based on MSE and CORRENTROPY show that BER values are almost the same in both models and the use of CORRENTROPY leads to better BER but the results are slightly different. Gaussian mixed models that can be expressed based on a Gaussian distribution Using CORRENTROPY leads to better results than MSE, but due to the small difference between the results of the two loss functions, each of these loss functions can be used.

Table1 - Compare Ber Test for Fully Connected Detector and Detnet Pre Svd with Correntropy Loss Function and Mse Loss Function

DETECTOR MODEL	***SNR / NOISE***	***SNR=4***	***SNR=8***	***SNR=12***	***SNR=16***	***SNR=20***
Fully connected $\sigma = 2.1\ \eta = 1$	Gaussian mixed	*0.10281375*	*0.03356875*	*0.0063675*	*0.00128875*	*0.000635*
Fully connected $\sigma = 1.4\ \eta = 1$	Gaussian mixed	*0.1031825*	*0.03397*	*0.00631625*	*0.00112625*	*0.00058625*
Fully connected MSE	Gaussian mixed	*0.103755*	*0.03386625*	*0.00639475*	*0.00117875*	*0.00065625*
DETNET PRE SVD $\sigma = 1.3\ \eta = 1$	Gaussian mixed	***9.93150e03***	***3.344875e03***	***7.1125e-04***	***1.30375e-04***	***4.5125e-05***
DETNET PRE SVD MSE	Gaussian mixed	*9.93775e03*	*3.3535e-03*	*7.17375e-04*	*1.425e-04*	*5.0125e-05*

In the following and in the form of seven training errors and TRAIN BER in FULLY CONNECTED and DETNET PRE SVD models and based on the LOSS FUNCTIONS studied, as can be seen, all models converge to close error values. Also, TEST time is examined in different models and the results are expressed in Table 2.

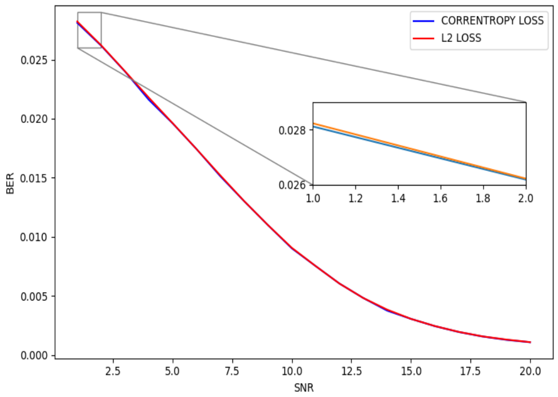

Figure 5- Comparison of fully connected model performance based on LOSS FUNCTION

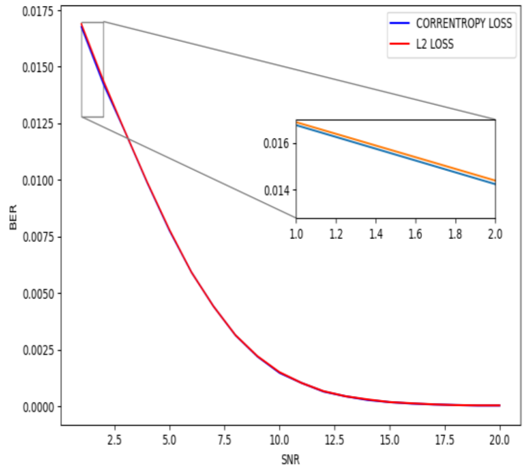

Figure 6- -Comparison of PRE SVD DETNET model performance based on LOSS FUNCTION

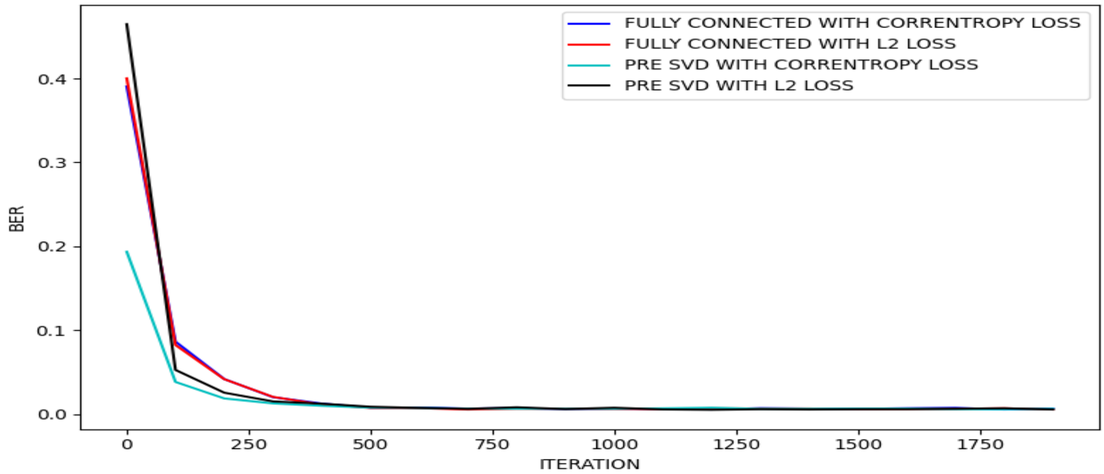

Figure 7- Comparison of train BER detector performance based on LOSS FUNCTION

Table 2 - Compare Test Time for Fully Connected Detector and Detnet Pre SVD with Correntropy Loss Function and Mse Loss Function

DETECTOR MODEL	***SNR*** / ***NOISE***	***SNR=4***	***SNR=8***	***SNR=12***	***SNR=16***	***SNR=20***
Fully connected $\sigma = 2.1$ $\eta = 1$	Gaussian mixed	*4.3743e-05*	*4.34597e-05*	*4.38334e-05*	*4.36815e-05*	*4.3354e-05*
Fully connected $\sigma = 1.4$ $\eta = 1$	Gaussian mixed	*4.1989e-05*	*4.42998e-05*	*4.42923e-05*	*4.32603e-05*	*4.4293e-05*
Fully connected MSE	Gaussian mixed	*4.2731e-05*	*4.23451e-05*	*4.20554e-05*	*4.18524e-05*	*4.1950e-05*

DETNET PRE SVD $\sigma = 1.3$ $\eta = 1$	Gaussian mixed	***2.6792e-05***	***2.74955e-05***	***2.66355e-05***	***2.68854e-05***	***2.7027e-05***
DETNET PRE SVD MSE	Gaussian mixed	*3.1325e-05*	*3.27833e-05*	*3.47213e-05*	*2.89999e-05*	*2.8199e-05*

4.2. Symmetric alpha stable NOISE

In this case, we assume that the model of noise added to the signal in the wireless channel is as follows:

$$z = S\alpha S(\alpha = 1.2 ، \ \beta = \ 0.7 \ )$$

Based on what is stated in section one, the $S\alpha S$ distribution can be estimated based on a Gaussian mixed distribution ,so we assume that the noise distribution is in the form and to calculate the SNR, we use the noise distribution estimation based on a Gaussian mixed distribution as follows:

$$z = \sum_{i=1}^{N} \lambda_i f(x; \mu_i ، \sigma_i{}^2 \ )$$

The important point in this estimation is the calculation of coefficients, so we assume that the number and average and variance of Gaussian distributions are known and the coefficients of these distributions are unknown, and we use a Kernel density estimation model to calculate the coefficients and obtain the best estimate of noise distribution.

Kernel density estimation (KDE) is a non-parametric method for estimating the probability density function of a given random variable. It is also referred to by its traditional name, the Parzen-Rosenblatt Window method, after its discoverers. Using this estimation, the noise power can be calculated and different values of SNR coefficient can be obtained, and using these coefficients, the detector models based on loss functions in the presence of non-Gaussian noise can be analysed . The results are shown in Table 3 and Figures 8,9 .

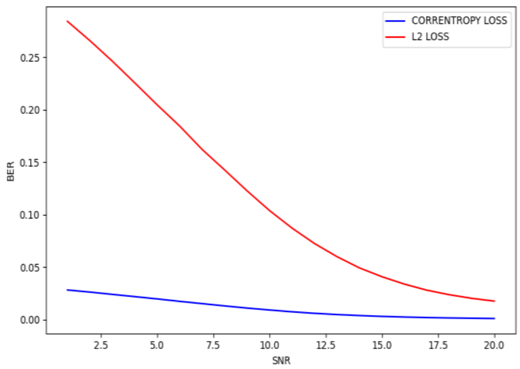

Figure 8- Comparison BER of DETNET without v in the presence of impact noise

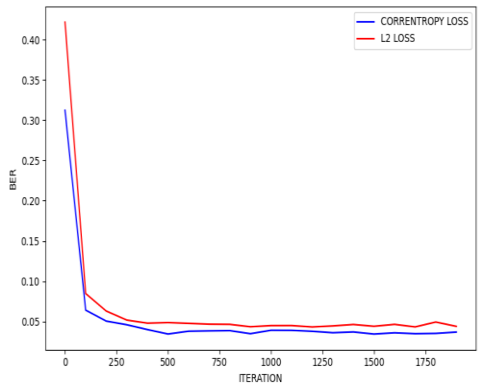

Figure 9- Comparison train BER of DETNET without v in the presence of impact noise

Table 3 - Compare Test BER for Fully Connected Without V Detector and DETNET without V with Correntropy Loss Function and MSE Loss Function

DETECTOR MODEL	*SNR* / *NOISE*	*SNR=1*	*SNR=4*	*SNR=8*	*SNR=12*	*SNR=16*	*SNR=20*
DETNET WITHOUT V $\sigma = 1.4$ $\eta = 1$	Levy distribution	*0.02185525*	*0.0218552*	*0.01301225*	*0.005966*	*0.00244275*	*0.00103275*
DETNET WITHOUT V MSE	Levy distribution	*0.284045*	*0.2253862*	*0.14246125*	*0.07240625*	*0.0337925*	*0.01758125*

Based on the data in Table 3, it can be seen that the detector model based on the CORRENTROPY criterion has given much better results in impulse noise.

5. CONCLUSIONS AND FUTURE WORK

In this paper, a better in-depth learning model for detection in MIMO systems with non-Gaussian environments is proposed based on DETNET model, by exploring MCC criterion. Calculations on BER and SNR performances show that DETNET PRE SVD not only simplifies complexity, but also improves the performances. The obtained results can be studied for other kinds of noise models and loss functions.

REFERENCES

[1] L. Azzam and E. Ayanoglu, "Reduced Complexity Sphere Decoding for Square QAM via a New Lattice Representation." IEEE GLOBECOM 2007-2007 IEEE Global Telecommunications Conference, 2007, doi: 10.1109/glocom.2007.807.

[2] S. Yang and L. Hanzo, "Fifty Years of MIMO Detection: The Road to Large-Scale MIMOs." IEEE Communications Surveys & Tutorials, vol. 17, no. 4, pp. 1941-1988, 2015, doi: 10.1109/comst.2015.2475242.

[3] D. Gesbert, S. Hanly, H. Huang, S. Shamai Shitz, O. Simeone, and W. Yu, "Multi-Cell MIMO Cooperative Networks: A New Look at Interference." IEEE Journal on Selected Areas in Communications, vol. 28, no. 9, pp. 1380-1408, 2010, doi: 10.1109/jsac.2010.101202.

[4] F. Sacuto, Modeling of the Impulsive Noise in the Power Substation Environment and Its Application to Receiver Design. 2016.

[5] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing." Proceedings of the National Academy of Sciences, vol. 106, no. 45, pp. 18914-18919, 2009, doi: 10.1073/pnas.0909892106.

[6] M. Bayati and A. Montanari, "The Dynamics of Message Passing on Dense Graphs, with Applications to Compressed Sensing." IEEE Transactions on Information Theory, vol. 57, no. 2, pp. 764-785, 2011, doi: 10.1109/tit.2010.2094817.

[7] A. Askri and G. R.-B. Othman, "DNN assisted Sphere Decoder." 2019 IEEE International Symposium on Information Theory (ISIT), 2019, doi: 10.1109/isit.2019.8849786.

[8] M. Mohammad karimi, M. Mehrabi, M. Ardakani, and Y. Jing, "Deep Learning-Based Sphere Decoding." IEEE Transactions on Wireless Communications, vol. 18, no. 9, pp. 4368-4378, 2019, doi: 10.1109/twc.2019.2924220.

[9] Won-Joon Choi, Kok-Wui Cheong, and J. Cioffi, "Iterative soft interference cancellation for multiple antenna systems." 2000 IEEE Wireless Communications and Networking Conference. Conference Record (Cat. No.00TH8540), doi: 10.1109/wcnc.2000.904647.

[10] T. O'Shea and J. Hoydis, "An Introduction to Deep Learning for the Physical Layer." IEEE Transactions on Cognitive Communications and Networking, vol. 3, no. 4, pp. 563-575, 2017, doi: 10.1109/tccn.2017.2758370.
[11] S. Liang and R. Srikant, "Why Deep Neural Networks for Function Approximation?" 5th International Conference on Learning Representations (ICLR), March 2017
[12] C. Zhang, P. Patras, and H. Haddadi, "Deep Learning in Mobile and Wireless Networking: A Survey." IEEE Communications Surveys & Tutorials, vol. 21, no. 3, pp. 2224-2287, 2019, doi: 10.1109/comst.2019.2904897.
[13] X. Li, F. Dong, S. Zhang, and W. Guo, "A Survey on Deep Learning Techniques in Wireless Signal Recognition." Wireless Communications and Mobile Computing, vol. 2019, pp. 1-12, 2019, doi: 10.1155/2019/5629572.
[14] S. Xue, Y. Ma, N. Yi, and T. E. Dodgson, A Modular Neural Network Based Deep Learning Approach for MIMO Signal Detection. arXiv, 2020. doi: 10.48550/ARXIV.2004.00404.
[15] M.-S. Baek, S. Kwak, J.-Y. Jung, H. M. Kim, and D.-J. Choi, "Implementation Methodologies of Deep Learning-Based Signal Detection for Conventional MIMO Transmitters." IEEE Transactions on Broadcasting, vol. 65, no. 3, pp. 636-642, 2019, doi: 10.1109/tbc.2019.2891051.
[16] X. Liu and Y. Li, "Deep MIMO Detection Based on Belief Propagation." 2018 IEEE Information Theory Workshop (ITW), 2018, doi: 10.1109/itw.2018.8613336.
[17] N. Samuel, T. Diskin, and A. Wiesel, "Learning to Detect." IEEE Transactions on Signal Processing, vol. 67, no. 10, pp. 2554-2564, 2019, doi: 10.1109/tsp.2019.2899805.
[18] V. Corlay, J. J. Boutros, P. Ciblat, and L. Brunel, "Multilevel MIMO Detection with Deep Learning." 2018 52nd Asilomar Conference on Signals, Systems, and Computers, 2018, doi: 10.1109/acssc.2018.8645519.
[19] N.Farsad and A. Goldsmith, ''Detection algorithms for communication Systems using deep learning,''2017,arXiv:1705.08044. [Online].Available :http://arxiv.org/abs/1705.08044
[20] X. Yan, F. Long, J. Wang, N. Fu, W. Ou, and B. Liu, "Signal detection of MIMO-OFDM system based on auto encoder and extreme learning machine." 2017 International Joint Conference on Neural Networks (IJCNN), 2017, doi: 10.1109/ijcnn.2017.7966042.
[21] J. Liu, K. Mei, X. Zhang, D. Ma, and J. Wei, "Online Extreme Learning Machine-Based Channel Estimation and Equalization for OFDM Systems." IEEE Communications Letters, vol. 23, no. 7, pp. 1276-1279, 2019, doi: 10.1109/lcomm.2019.2916797.
[22] D. Stein, "Detection of random signals in Gaussian mixture noise." IEEE Transactions on Information Theory, vol. 41, no. 6, pp. 1788-1801, 1995, doi: 10.1109/18.476307.
[23] B. Sadler and R. Kozick, "Maximum-likelihood array processing in non-Gaussian noise with Gaussian mixtures." IEEE Transactions on Signal Processing, vol. 48, no. 12, pp. 3520-3535, 2000, doi: 10.1109/78.887045.
[24] S. M. Kay, Fundamentals of Statistical Signal Processing: Detection theory. Prentice Hall, 1998.
[25] T. Yucek and H. Arslan, "A survey of spectrum sensing algorithms for cognitive radio applications." IEEE Communications Surveys & Tutorials, vol. 11, no. 1, pp. 116-130, 2009, doi: 10.1109/surv.2009.090109.
[26] L. Garth and H. Poor, "Detection of non-Gaussian signals: a paradigm for modern statistical signal processing." Proceedings of the IEEE, vol. 82, no. 7, pp. 1061-1095, 1994, doi: 10.1109/5.293163.
[27] O. A. Dobre, A. Abdi, Y. Bar-Ness, and W. Su, "Survey of automatic modulation classification techniques: Classical approaches and new trends," IET Communications, vol. 1, no. 2, pp. 137–156, 2007, doi: 10.1049/iet-com:20050176.
[28] Xiaodong Wang and H. Poor, "Robust multiuser detection in non-Gaussian channels." IEEE Transactions on Signal Processing, vol. 47, no. 2, pp. 289-305, 1999, doi: 10.1109/78.740103.
[29] Ran He, Bao-Gang Hu, Wei-Shi Zheng, and Xiang-Wei Kong, "Robust Principal Component Analysis Based on Maximum Correntropy Criterion." IEEE Transactions on Image Processing, vol. 20, no. 6, pp. 1485-1494, 2011, doi: 10.1109/tip.2010.2103949.
[30] E. Hasanbelliu, L. S. Giraldo, and J. C. Principe, "Information Theoretic Shape Matching." IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 36, no. 12, pp. 2436-2451, 2014, doi: 10.1109/tpami.2014.2324585.
[31] W. Ma, H. Qu, G. Gui, L. Xu, J. Zhao, and B. Chen, "Maximum correntropy criterion based sparse adaptive filtering algorithms for robust channel estimation under non-Gaussian environments." Journal of the Franklin Institute, vol. 352, no. 7, pp. 2708-2727, 2015, doi: 10.1016/j.jfranklin.2015.03.039.

[32] B. Chen, X. Wang, N. Lu, S. Wang, J. Cao, and J. Qin, “Mixture correntropy for robust learning.” Pattern Recognition, vol. 79, pp. 318-327, 2018, doi: 10.1016/j.patcog.2018.02.010.
[33] A. Singh, R. Pokharel, and J. Principe, “The C-loss function for pattern classification.” Pattern Recognition, vol. 47, no. 1, pp. 441-453, 2014, doi: 10.1016/j.patcog.2013.07.017.
[34] N. Samuel, T. Diskin, and A. Wiesel, “Deep MIMO detection.” 2017 IEEE 18th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), 2017, doi: 10.1109/spawc.2017.8227772.
[35] Jefferson Jesus Hengles Almeida, P. B. Lopes, Cristiano Akamine, and Nizam Omar. AN APPLICATION OF NEURAL NETWORKS TO CHANNEL ESTIMATION OF THE ISDB-TB FBMC SYSTEM. International Journal of Wireless & Mobile Networks (IJWMN) Vol. 10, No. 1, February 2018, DOI: 10.5121/ijwmn.2018.10107.
[36] A. Omri and R. Bouallegue . NEW TRANSMISSION SCHEME FOR MIMO-OFDM SYSTEM. International Journal of Next-Generation Networks (IJNGN) Vol.3, No.1, March 2011, DOI : 10.5121/ijngn.2011.3102.

AUTHORS

Mohammad reza pourmir received the B.Sc. degree in electronics engineering from Sistan and Baluchistan University, Zahedan, Iran, in 1994, and the M.Sc. degree in computer engineering from Iran University of Science and Technology ,in 2012, and he is now a Ph.D. student in computer engineering in Mashhad Ferdowsi University. . His research interests are in information theory, communication theory, wireless communications and signal processing. He can be contacted at email: mo.pourmir@mail.um.ac.ir

Reza monsefi received the B.Sc. degree in electronics engineering from Manchester University, Manchester ,U.K, in 1979 and received the M.Sc. and Ph.D. degree from Salford University, U.K, in 1981 and 1983;and he is now a professor in computer engineering. His research interests are in information theory and signal processing. He can be contacted at email: monsefi@um.ac.ir.

Ghosheh abed hodtani received the B.Sc. degree in electronics engineering and the M.Sc. degree in communications engineering, both from Isfahan University of Technology, Isfahan, Iran, in 1985, 1987, respectively. He joined Electrical Engineering Dept., at Ferdowsi University of Mashhad, Mashhad, Iran, in 1987. He decided to pursue his studies in 2005 and received the Ph.D. degree (with excellent grade) from Sharif University of Technology, Tehran, Iran, in 2008;and he is now a professor in electrical engineering. His research interests are in multi-user information theory, communication theory, wireless communications and signal processing. Prof. Hodtani is the author of a textbook on electrical circuits and is the winner of the best paper award at IEEE ICT -2010. He can be contacted at email: hodtani@um.ac.ir.