Automatic Modulation Classification Based on Deep Feature Fusion for High Noise Level and Large Dynamic Input

Hui Han 1, Zhiyuan Ren 2, Lin Li 2,* and Zhigang Zhu 2

Abstract: Automatic modulation classification (AMC) is playing an increasingly important role in spectrum monitoring and cognitive radio. As communication and electronic technologies develop, the electromagnetic environment becomes increasingly complex. The high background noise level and large dynamic input have become the key problems for AMC. This paper proposes a feature fusion scheme based on deep learning, which attempts to fuse features from different domains of the input signal to obtain a more stable and efficient representation of the signal modulation types. We consider the complementarity among features that can be used to suppress the influence of the background noise interference and large dynamic range of the received (intercepted) signals. Specifically, the time-series signals are transformed into the frequency domain by Fast Fourier transform (FFT) and Welch power spectrum analysis, followed by the convolutional neural network (CNN) and stacked auto-encoder (SAE), respectively, for detailed and stable frequency-domain feature representations. Considering the complementary information in the time domain, the instantaneous amplitude (phase) statistics and higher-order cumulants (HOC) are extracted as the statistical features for fusion. Based on the fused features, a probabilistic neural network (PNN) is designed for automatic modulation classification. The simulation results demonstrate the superior performance of the proposed method. It is worth noting that the classification accuracy can reach 99.8% in the case when signal-to-noise ratio (SNR) is 0 dB.

Keywords: automatic modulation classification; stacked auto-encoder; convolutional neural network; probabilistic neural network

1. Introduction

Automatic modulation classification is a technology for the automatic classification of signal modulation types [1], which is widely used in interference identification, spectrum sensing, electronic countermeasures and other fields. With the significant development of radio communication, the signal modulation and electromagnetic environment are becoming increasingly complex and diverse, and result in a high noise level, usually greater than −100 dBHz, viz. SNR is extremely low. Meanwhile, due to the low probability of the intercept (LPI), the SNR of the intercepted signal is constantly changing in the dynamic range, and therefore the range of SNR becomes much larger. These new circumstances make AMC tasks more difficult. Hence, it is crucial to explore more effective techniques for AMC.

Generally, AMC approaches can be classified into two categories: likelihood-based (LB) methods and feature-based (FB) methods. The likelihood-based method, based on a likelihood function and combined with a hypothesis test, can obtain the highest classification performance in theory. The likelihood function can be divided into three groups: hybrid likelihood ratio test [2], average likelihood ratio test [3], and generalized likelihood ratio test [4]. However, such methods suffer from high computational complexity due...
to too many unknown parameters. On the contrary, feature-based methods consisting of feature extraction and classifier design, have been widely applied in the AMC field to recognize more modulation types with lower algorithm complexity (see [5]). Consequently, one or more stable and distinguishing features can improve the classification accuracy significantly. In recent years, researchers have studied various features such as high-order cumulants [6–8], instantaneous features [9], cyclostationary features [10], etc., and a variety of methods for classification. Besides the decision tree model, the existing classification methods are mainly based on machine learning algorithms, including the artificial neural network (ANN) [11], support vector machine (SVM) [12], and SAE [13].

With the rapid development of artificial intelligence (AI) technology, deep learning (DL), with its excellent data-processing ability, has attracted widespread attention and been used in various fields. In recent years, researchers have applied related algorithms in AMC, and achieved expected results. (see [14–16]). Experiments in [14] show that the classification accuracy of a convolutional neural network trained by in-phase/quadrature (I/Q) data is superior to that trained by cyclic-moment features. Reference [15] proposed a two-step training method based on transfer learning. The results show that the training speed of this method is from 40 to 1700 times faster than that of single-step training. In [16], an AMC algorithm using CNN to extract constellation features of digital communication signals is studied, which combines image classification and deep learning.

In this paper, an AMC scheme based on deep learning and feature fusion is proposed, which can achieve high classification accuracy in a low SNR environment, and within a large range of SNR. In [17], a CNN-based fusion scheme is proposed to deal with the problem of signal duration burst. In [18], an image fusion algorithm is proposed for modulation classification, which considers the images from different time-frequency methods. Recently, the multimodal deep learning approaches that proposed image and speech recognition have also been considered for radio signals. Differing from the existing feature fusion approaches, we utilized the complementarity among features in different domains, and especially, for high noise level and large dynamic input. Specifically, the contributions of the paper are as follows:

(i) For frequency features, we consider the differences between spectrum obtained by FFT and power spectrum by Welch approach (windowed smoothing). The spectrum with high-frequency resolution preserves the detailed information of the modulation characteristics. On the other hand, the windowed smoothing power spectrum with low-frequency resolution presents the stable feature, which is suitable for low SNR;

(ii) CNN and SAE are proposed for feature extraction. A new CNN scheme is constructed to extract sparse features from the spectrum. Meanwhile, the unsupervised SAE is proposed for feature extraction from the Welch power spectrum;

(iii) The statistical feature is proposed for complementarity, which stands for the information at the time domain. Specifically, we extract the statistical features from instantaneous amplitude and instantaneous phase, and combine the higher-order cumulants;

(iv) A PNN scheme is proposed for modulation classification, which provides a much higher classification accuracy than the existing methods.

The rest of this paper is arranged as follows. Section 2 introduces the signal model and dataset used in this paper. Section 3 describes the new framework of the proposed method, including feature fusion and classification model. Section 4 shows simulation results and analysis. A brief conclusion is given in Section 5.

2. Signal Model

In this paper, we consider the additive white Gaussian noise (AWGN) channel for modulation signals. Formally, an intercepted (received) radio signal $r(t)$ can be expressed as

$$r(t) = s(t) + n(t),$$

where $s(t)$ and $n(t)$ denote the transmitted signal and the Gaussian noise, respectively; see details in [12].
When transmitted signal is M-ASK, M-FSK or M-PSK (see Table 1 for the modulation types), \( s(t) \) can be represented as

\[
s(t) = A_m \sum_n a_n g(t - nT_s) \cos(2\pi(f_c + f_m)t + \phi_0 + \phi_m),
\]

(2)

where \( A_m \) and \( a_n \) represent the modulation amplitude and symbol sequence, respectively. Note that \( a_n \) is modulated by the window function \( g(t) \), defined as

\[
g(t) = \begin{cases} 
1, & 1 \leq t \leq T_s \\
0, & \text{otherwise}
\end{cases}
\]

(3)

\( T_s \) represents the symbol period, \( f_c \) and \( f_m \) denote carrier frequency and modulation frequency.

For signal with M-QAM, there are two orthogonal carriers, modulated by the amplitudes of \( a_n \) and \( b_n \), written as

\[
s(t) = \left[ A_m \sum_n a_n g(t - nT_s) \cos(2\pi f_c t + \phi_0) \right] + \left[ A_m \sum_n b_n g(t - nT_s) \sin(2\pi f_c t + \phi_0) \right].
\]

(4)

In this paper, we mainly consider common digital modulations, such as those described above. We define the modulation dataset \( S \), used to test our method, where \( S = \{ S^c, c = 1, 2, \ldots, C \}, \) \( C \) is the number of the modulation types. Table 1 shows the detailed dataset \( S \) used in our paper, where \( S = \{ 2\text{PSK}, 4\text{PSK}, 2\text{ASK}, 4\text{ASK}, 8\text{ASK}, 2\text{FSK}, 4\text{FSK}, 8\text{FSK}, 32\text{QAM}, 64\text{QAM} \} \). Note that we consider 10 modulation types.

### Table 1. Modulation types in dataset \( S \).

| Class | Name      | Modulation                      |
|-------|-----------|---------------------------------|
| S^1   | 2PSK      | 2-ary Phase Shift Keying        |
| S^2   | 4PSK      | 4-ary Phase Shift Keying        |
| S^3   | 2ASK      | 2-ary Amplitude Shift Keying    |
| S^4   | 4ASK      | 4-ary Amplitude Shift Keying    |
| S^5   | 8ASK      | 8-ary Amplitude Shift Keying    |
| S^6   | 2FSK      | 2-ary Frequency Shift Keying Keying |
| S^7   | 4FSK      | 4-ary Frequency Shift Keying Keying |
| S^8   | 8FSK      | 8-ary Frequency Shift Keying Keying |
| S^9   | 32QAM     | 32-ary Quadrature Amplitude Modulation |
| S^10  | 64QAM     | 64-ary Quadrature Amplitude Modulation |

### 3. The Proposed Framework

#### 3.1. Overview of the Proposed Framework

It can be observed from existing works on AMC, that features in different domains, such as time, frequency, cyclostationary, high-order cumulants, and time-frequency representation, have been studied, analyzed and tested broadly. However, few researchers have attempted to explore the differences among features from different domains, and integrate different features with complementary advantages. In this section, first we present the proposed feature-fusion framework for modulation classification; see Figure 1. The framework contains three steps: signal processing, feature extraction and classification.
The FFT and Welch power spectrum [19] are used to convert all signals into the frequency-domain, but preserving different information. Then, we introduce a self-designed CNN model to learn the features after FFT. SAE is used to extract unsupervised features from Welch spectrum data. Meanwhile, features based on instantaneous characteristics and higher-order cumulants are used as statistical features. After that, in order to obtain better feature representation, all the extracted features are combined to form a joint feature vector. Finally, to further improve the classification performance, we utilize a PNN for final classification, which has a higher classification accuracy and better fitting effect than BP neural network [20]. Figure 2 shows the sketch map of the proposed method.

3.2. Features Extraction

3.2.1. CNN-Based Feature Extraction

Deep learning, as a novel development method, can extract more meaningful features through its hierarchical learning process. For modulation classification, DL-based methods can automatically learn distinctive representations of high-dimensional data, such as the received radio signals [21].

CNN is composed of a convolution layer and pooling layer, so as to extract effective features. The basic operation of the convolution layer and pooling layer is shown in Figure 3. There are usually two kinds of pooling operations: one is max-pooling, where
the maximum value is selected as the output value in the size coverage area of the pooling filter. The other is mean-pooling, which pools the average value of the coverage area as the output value. Mean-pooling and max-pooling correspond to mean filtering and maximum filtering in image processing, and can simultaneously reduce the number of features.

The activation function in the convolution layer is the rectified linear unit (ReLU), which can be expressed as

$$ReLU(x) = \max\{0, x\}$$  \hspace{1cm} (5)

The neuron will enter a “dead” state and is insensitive to other gradients when a large gradient flows through a neuron, which is called the “dying ReLU” [22]. To avoid the “dying ReLU”, Swish function [23] and SeLU function [24] are selected as the activation function in the convolution layer and the fully connected layer, which are expressed as (6) and (7), respectively.

$$Swish(x) = x \cdot \text{sigmoid}(x)$$  \hspace{1cm} (6)

$$SeLU(x) = \lambda \begin{cases} x, & x > 0 \\ ae^x - a, & x \leq 0 \end{cases}$$  \hspace{1cm} (7)

Due to the dimension of the input data being different from that of the traditional images or videos, we set up a new CNN structure. This structure consists of 16 layers, including a convolution layer, pool layer, full connection layer and Softmax layer. Table 2 shows the detailed model configuration.

| Layer (Type)            | Output Dimensions | Parameters Number |
|-------------------------|-------------------|-------------------|
| Input                   | $2 \times 16,384$ | 0                 |
| Conv2- Pool2/Swish     | $1 \times 8192 \times 1$ | 48               |
| Conv2- Pool2/Swish     | $1 \times 4096 \times 4$ | 48               |
| Conv2- Pool2/Swish     | $1 \times 2048 \times 4$ | 48               |
| Conv2- Pool2/Swish     | $1 \times 1024 \times 4$ | 48               |
| Conv2- Pool2/Swish     | $1 \times 512 \times 4$ | 48               |
| Conv2- Pool2/Swish     | $1 \times 256 \times 4$ | 48               |
| Conv2- Pool2/Swish     | $1 \times 128 \times 4$ | 48               |
| Full Connected/SeLU    | 64                | 32,768            |
| Softmax                 | 10                | 640               |

In this paper, we use a $1 \times 3$ convolution kernel for feature extraction. We also introduce a $1 \times 2$ pool core to reduce the parameters. In order to avoid over-fitting of the network, a dropout operator with ratio of 0.5 is introduced into the convolution layer. Moving average and $L_2$ regularization are introduced into the model. Finally, the last layer of the network is connected to the Softmax classifier to output the prediction score of the modulation signal types.

3.2.2. SAE-Based Feature Extraction

The Welch power spectra of M-PSK and M-QAM signals are relatively close to each other [19], which means that supervised learning may cause feature confusion. Hence,
we introduce the SAE method, a simple neural network for feature extraction, which is an unsupervised learning approach.

SAE first takes a non-linear mapping of the input $x \in \mathbb{R}^{d_x}$ into a hidden representation $y \in \mathbb{R}^{d_y}$, where $d_x$ and $d_y$ are the dimensions of $x$ and $y$, respectively. Then, map $y$ back to the input space as $x'$. These two steps are called the encoding and decoding, respectively [25],

$$y = h(Wx + b), \quad (8)$$

$$x' = h'(W'y + b'). \quad (9)$$

Here, $h$ stands for the encoder transfer function, where the commonly used function is Log-sigmoid function (Logsig), written as

$$\text{Logsig}(x) = \frac{1}{1 + e^{-x}}. \quad (10)$$

$h'$ is the decoder transfer function, which can be linear (Purelin). $W$ and $W'$ are the weight matrices of the encoder and decoder, respectively. $b$ and $b'$ are the bias units.

The pre-training of each auto-encoder aims to adjust the weight matrices by minimizing the cost function in (11) with respect to the weight matrices $W$ and $W'$ and biases $b$ and $b'$.

$$\text{Cost} = \|x - x'\|^2_2 + \frac{\lambda}{2} \sum_{l=1}^{L} \sum_{i=1}^{n} \sum_{j=1}^{k} ((W)^l_{ij})^2 + \beta \psi_s, \quad (11)$$

where $L$ is the number of the hidden layers, $n$ is the number of data samples, $k$ is the number of input features, $\lambda$ and $\beta$ are coefficients. $\psi_s$ is the sparsity regularizing term given by the Kullback–Leibler divergence.

The Cost function includes three terms, where the Cross Entropy function (the first term) fits the input data, the weight-regularizing term (the second term) is controlled by the coefficient $\lambda$ and the sparsity-regularizing term (the third term) is controlled by the coefficient $\beta$.

In order to reveal the unsupervised learning of SAE, the Welch power spectrum, extracted features and decoding outputs of four digital modulation signals, 2PSK, 32QAM, 4FSK and 8FSK are shown in Figure 4. The top row of Figure 4 shows the input of the SAE, the Welch spectrum of the signal in (1) with SNR = 0 dB. The middle row shows the SAE feature vectors used for classification. The bottom row shows the decoding of unsupervised learning. Observe that the output of SAE is very close to the input. Actually, the root-mean-square error (RMSE) between the input and the output is negligible, which indicates that the loss of feature extracted by the self-encoder can be ignored when SNR = 0 dB.

3.2.3. Statistical Feature Extraction

Considering the complementary information of time-domain, the instantaneous amplitude (phase) statistics and higher-order cumulants are extracted as the statistical features for fusion, which can reveal statistical time-varying information [26]. Meanwhile, the higher-order cumulants are insensitive to Gaussian noise and robust to phase rotation, see [27] and [28].

Suppose the received signal $x(n)$ (after the quadrature demodulation and normalization, with complex value and zero-mean) is written as

$$x(n) = x_I(n) + jx_Q(n) = a(n)e^{j\phi(n)}, n = 1, 2, \ldots, N_s, \quad (12)$$

where $x_I(n), x_Q(n) \in [-1, 1]$ are in-phase and quadrature components, and $a(n) \in [0, 1], \phi(n) \in [0, 2\pi]$ are called the instantaneous amplitude and phase, respectively. $N_s$ is the number of samples. Note that the means of $x_I(n)$ and $x_Q(n)$ are equal to 0, due to the fact that $x(n)$ is zero-mean.
Let \( \tilde{a}(n) \) and \( \tilde{\phi}(n) \) be the centered instantaneous amplitude and phase, respectively, defined as
\[
\tilde{a}(n) = a(n) - \bar{a}(n),
\]
\[
\tilde{\phi}(n) = \phi(n) - \bar{\phi}(n),
\]
where \( \bar{x} \) denotes the mean of \( x \). Then, we define the following statistics.

(a) The periodic characteristic of \( \tilde{a}(n) \)
\[
\gamma_{\text{max}} = \frac{\max|\mathcal{F}[\tilde{a}(n)]|^2}{N_s},
\]
where \( \mathcal{F} \) stands for the \( N_s \)-point discrete Fourier transform of \( \tilde{a}(n) \).

(b) The standard deviation of \( \tilde{a}(n) \)
\[
\sigma_a = \|\tilde{a}(n)\|_2.
\]

(c) The standard deviations of \( \phi(n) \) and \( \tilde{\phi}(n) \)
\[
\sigma_\phi = \|\phi(n)\|_2, \quad \sigma_{\tilde{\phi}} = \|\tilde{\phi}(n)\|_2.
\]

(d) The standard deviation of the derivative of \( \phi_u(n) \)
\[
\sigma_f = \|\phi_d(n)\|_2,
\]
where \( \phi_d(n) = \phi_u(n) \), \( \phi_u(n) \) is the unwrapped \( \phi(n) \).

Next, we introduce the 4th-, 6th-, and 8th-order cumulants of the received signal \( x(n) \) in (12). First, define the \( p \)th-order moment as
\[
M_{pq} = E\left[x(n)^p \cdot x^*(n)^q\right].
\]
Then we have the following cumulants.
\[
C_{40} = M_{40} - 3M_{20}^2,
\]
\[
C_{41} = M_{41} - 3M_{20}M_{21},
\]
Finally, we construct the features mentioned above to a vector, as follows

\[
\zeta = [\gamma_{\max}, \sigma_a, \sigma_\phi, \sigma_{\tilde{\phi}}, \sigma_f, C_{40}, C_{41}, C_{42}, C_{60}, C_{63}, C_{80}], \tag{26}
\]

where \(\zeta\) denotes the statistical feature space.

### 3.3. Classification with PNN

In order to obtain a better signal representation, all the extracted features in Section 3.2 are combined to form a joint feature vector. To further improve the classification performance of AMC, we introduce a PNN network for final classification, see (Figure 5). The PNN network in Figure 5 is a feed-forward network [29], which realizes the prediction classification of models through Bayesian decision theory. In our proposed framework, joint features are first input into the input layer. Then, the samples are trained by the RBF neurons in the pattern layer. The similar output values are given by the summation layer, and finally, the competing neurons of output layer are judged to obtain the types of digital modulation.

![Figure 5. Structure of the PNN network.](image)

### 4. Experiment Results and Analysis

In this section, we will use a series of simulations to evaluate the performance of our proposed framework. Recalling the dataset \(S\) (modulation types) given by Table 1, here, we set up the signal parameters as follows; see Table 3. The bandwidths of all the signals vary from 4 to 6 MHz randomly. During training, we generated 84,000 signals for all the modulation types, with SNR continuously randomly chosen from \(-20\) to \(20\) dB. In addition, 21,000 examples are randomly generated with SNR ranging from \(-20\) to \(20\) dB at an interval of 2 dB (with random selection of modulation type and bandwidth) to test the performance of our method. Considering the background noise \(-100\) dBHz (dBHz is the unit of power spectral density), bandwidth 5 MHz and the sampling rate 50 MHz, the input SNR from \(-20\) to \(20\) dB is equivalent to the input signal power of from \(-70\) to \(-30\) dBm (here, dBm is the unit of power) approximately. Hence, we consider the input dynamic range to be about 40 dB. Such a large dynamic input for training is scarcely considered in the existing works about AMC.
Table 3. Signal parameters.

| Parameter               | Value/Range | Description       |
|-------------------------|-------------|-------------------|
| Sampling frequency      | 50 MHz      | The sampling rate |
| Signal length           | 16,384      | The number of sampling points |
| Bandwidth               | [4, 6] MHz  | Randomly selection |
| SNR for training        | [−20, 20] dB| The dynamic range |
| SNR for test            | {−20, −18, ..., 20} dB | For evaluation |
| Training samples        | 84,000      | The total samples for training |
| Test samples            | 21,000      | The total samples for test |
| Test samples for each SNR | 100         | For each signal type |

Firstly, in order to verify the effectiveness of feature fusion, we compare the performances of different methods; see the results in Figure 6. Note that “F”, “W” and “S” are the initials of the three features, which are CNN-based features (frequency features), SAE-based features (Welch power spectrum features) and statistical features, respectively. FW, FS and WS stand for the fusions of two types of features, chosen from “F”, “W” and “S”, with different combinations. Note that FWS stands for the fusion of all the three types of feature, which is the method proposed in this paper. We use PNN for classification, but with different schemes to fit the different sizes of input feature.

For single-type features, the classification performance of CNN-based features is much better than other two types under high SNR (when SNR ≥ −2 dB), due to the global optimization of model training. However, the SAE-based features are better for low SNR (when SNR ≤ −4 dB). By feature fusing, the methods FW, FS and FWS can improve the overall classification accuracy. However, the WS method is best for the case of extremely low SNR (when SNR ≤ −14 dB). The results show there is a significant complementarity among features, namely, the complementarities of different domains and different SNRs (low SNR and high SNR, since we consider the large dynamic range input). It is worth noting that the total classification accuracy of FWS is the highest, which proves the validity of our feature fusion frame.

In addition, Figure 7 shows the confusion matrices of the feature fusion methods above, under the condition of low SNR, namely SNR = −6 dB. Obviously, the two M-PSK signals and the three M-ASK signals can not be distinguished by the WS, respectively. This is due to the fact that the Welch spectra of these signals are close to each other. However, our proposed method, the FWS, achieves the highest classification accuracy. This is because the combination of features in different domains achieves complementary advantages.
To demonstrate the superiority of our proposed algorithm, Figure 8 shows the classification accuracy of different methods, including the feature fusion method proposed in this paper (FWS), CNN-LSTM with the signal waveform in [14] and the three convolutional neural networks for feature and decision fusion (CNN3) in [17]. Obviously, the classification performance of FWS is superior to the other two methods at low SNRs, which proves the superiority of the proposed method for large dynamic input.

We also provide the classification accuracy of each signals by the proposed FWS method, see Figure 9. The classification accuracy of most signals is above 96% when SNR = 0 dB, and some of them even reach 100%. Although the noise has a direct impact on the signal, FSK signals also show excellent performance at low SNR (SNR = −16 dB). The average classification accuracy of M-ASK signals is comparatively lower than the other signals, but it is still greater than 70% when SNR = −8 dB.

Finally, in order to demonstrate the superiority of our classification scheme, we compare PNN with SAE [13], SVM [12] and ANN [11]. As expected, we can see from Figure 10 that the classification accuracy based on PNN is higher than the other three classifiers, especially at low SNRs, which is mainly due to the strong nonlinear approximation ability of PNN. Note that SVM is close to SAE, and the ANN model has the lowest classification rate. The results show that the proposed framework with PNN has high robustness in a complex noise environment.

Figure 7. Confusion matrix of different methods when SNR = −6 dB.
Moreover, we evaluate the computational complexity of the four classifiers above by the average training time. Note that all methods are performed on the same hardware system. As shown in Table 4, the training time of SVM is much longer than other methods, which shows that SVM is not suitable for training large-scale datasets. Meanwhile, the training time of PNN is less than SAE, ANN and SAE, which is due to that PNN is easier to converge than other classifiers.
Table 4. Training times of different classifiers.

| Models   | PNN    | ANN    | SAE    | SVM    |
|----------|--------|--------|--------|--------|
| Training Time (s) | 13.711 | 18.233 | 34.249 | 231.178 |

5. Conclusions

This paper deals with the problem of feature fusion for the complex electromagnetic environment, especially for a high noise level and large dynamic input. An AMC framework based on deep learning is proposed, which aims to fuse different features of the input signals to obtain more differentiating features. For the CNN-based and SAE-based features, the statistical features are fused and a PNN is designed for automatic modulation classification. The simulation results demonstrate the superiority of the proposed method, especially in a complex noise environment, with SNR ranges from −20 to 20 dB. In future works, more modulation types should be considered to further improve the generalization and robustness of the proposed method.

Author Contributions: Conceptualization, H.H., Z.R. and Z.Z.; methodology, Z.R. and L.L.; writing—original draft preparation, Z.R.; writing—review and editing, L.L.; funding acquisition, H.H. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the Open Project of State Key Laboratory of Complex Electromagnetic Environment Effects on Electronics and Information System (No. CEMEE2019K0101A) and the National Natural Science Foundation of China (No. 62071349).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All modulation data and code will be made available on request to the correspondent author’s email with appropriate justification.

Conflicts of Interest: The authors declare no conflict of interest.

References
1. Dobre, O.A.; Abdi, A.; Bar-Ness, Y.; Su, W. Survey of automatic modulation classification techniques: Classical approaches and new trends. IEEE Commun. 2007, 1, 137–156. [CrossRef]
2. Panagiotou, P.; Anastasopoulos, A.; Polydoros, A. Likelihood ratio tests for modulation classification. In Proceedings of the IEEE Military Communications Conference, Los Angeles CA, USA, 22–25 October 2000; pp. 670–674.
3. Wei, W.; Mendel, J.M. Maximum-likelihood classification for digital amplitude-phase modulations. IEEE Trans. Commun. 2000, 48, 189–193. [CrossRef]
4. Ho, K.C.; Hong, L. Likelihood method for BPSK and unbalanced QPSK modulation classification. Proc. SPIE Int. Soc. Opt. Eng. 2001, 4395, 155–162.
5. Majhi, S.; Gupta, R.; Xiang, W.; Glisic; S. Hierarchical hypothesis and feature-based blind modulation classification for linearly modulated signals. IEEE Trans. Veh. Technol. 2017, 66, 11057–11069. [CrossRef]
6. Dobre, O.A.; Bar-Ness, Y.; Su, W. Higher-order cyclic cumulants for high order modulation classification. IEEE Mil. Commun. Conf. 2003, 1, 112–117.
7. Su, W. Feature space analysis of modulation classification using very high-order statistics. IEEE Commun. Lett. 2003, 17, 1688–1691. [CrossRef]
8. Orlic, V.D.; Dukic, M.L. Automatic modulation classification algorithm using higher-order cumulants under real-world channel conditions. IEEE Commun. Lett. 2009, 3, 917–919. [CrossRef]
9. Ebrahimzadeh, A.; Ghazalian, R. Blind digital modulation classification in software radio using the optimized classifier and feature subset selection. Eng. Appl. Artif. Intell. 2011, 24, 50–59. [CrossRef]
10. Ghauri, S.A.; Qureshi, I.M.; Shah, I. Modulation classification using cyclostationary features on fading channels. Res. J. Appl. Sci. Eng. Technol. 2014, 24, 5331–5339. [CrossRef]
11. Kavalov, D.; Kalinin, V. Neural network surface acoustic wave RF signal processor for digital modulation recognition. IEEE Trans. Ultrason. Ferroelectr. Freq. Control 2002, 49, 1280–1290. [CrossRef]
12. Wu, Z.; Zhou, S.; Yin, Z.; Ma, B.; Yang, Z. Robust automatic modulation classification under varying noise conditions. IEEE Access 2017, 5, 19733–19741. [CrossRef]
13. Afan, A.; Yangyu, F.; Liu, S. Automatic modulation classification of digital modulation signals with stacked auto-encoders. Digit. Signal Process. 2017, 71, 108–116.
14. O’Shea, T.J.; Corgan, J.; Clancy, T.C. Convolutional radio modulation recognition networks. *Commun. Comput. Inf. Sci.* 2016, 629, 213–226.
15. Meng, F.; Chen, P.; Wu, L.; Wang, X. Automatic modulation classification: A deep learning enabled approach. *IEEE Trans. Veh. Technol.* 2018, 67, 10760–10771. [CrossRef]
16. Ahmadi, N. Using fuzzy clustering and TTSAS algorithm for modulation classification based on constellation diagram. *Eng. Appl. Artif. Intell.* 2010, 23, 357–370. [CrossRef]
17. Zheng, S.; Qi, P.; Chen, S.; Yang, X. Fusion methods for CNN-based automatic modulation classification. *IEEE Access* 2019, 7, 66496–66504. [CrossRef]
18. Gao, L.; Zhang, X.; Gao, J.; You, S. Fusion image based radar signal feature extraction and modulation recognition. *IEEE Access* 2019, 7, 13135–13148. [CrossRef]
19. Welch, P.D. A direct digital method of power spectrum estimation. *IBM J. Res. Dev.* 1961, 5, 141–156. [CrossRef]
20. Ye, J.; Deng, P.; Li, P.; Yan, L.; Pan, W.; Zou, X.; Hao, M. Photonic-assisted modulation classification for RF signals using probabilistic neural network. In Proceedings of the Optical Fiber Communications Conference and Exhibition, San Diego, CA, USA, 3–7 March 2019; pp; 1–3.
21. Kumar, Y.; Sheoran, M.; Jajoo, G.; Yadav, S.K. Automatic modulation classification based on constellation density using deep learning. *IEEE Commun. Lett.* 2020, 24, 1275–1278. [CrossRef]
22. Haldar, M.; Abdool, M.; Ramanathan, P.; Xu, T.; Yang, S.; Duan, H.; Zhang, Q.; Barrow-Williams, N.; Turnbull, B.C.; Collins, B.M.; et al. Applying deep learning to airbnb search. *arXiv* 2018, arXiv:1810.09591.
23. Ramachandran, P.; Zoph, B.; Le Quoc, V. Searching for activation functions. *arXiv* 2017, arXiv:1710.05941.
24. Klambauer, G.; Unterthiner, T.; Mayr, A.; Hochreiter, S. Self-normalizing neural networks. *arXiv* 2017, arXiv:1706.02515
25. Shi, Y.; Xu, H.; Jiang, L.; Liu, Y. Few-shot modulation classification method based on feature dimension reduction and pseudo-label training. *IEEE Access* 2020, 8, 140411–140425. [CrossRef]
26. Lopatka, J.; Macrej, P. Automatic modulation classification using statistical moments and a fuzzy classifier. In Proceedings of the 5th International Conference on Signal Processing, Beijing, China, 21–24 August 2000; pp. 1500–1506.
27. Pajic, M.S.; Veinovic, M.; Peric, M.; Orlic, V.D. Modulation order reduction method for improving the performance of AMC algorithm based on sixth–order cumulants. *IEEE Access* 2020, 8, 106386–106394. [CrossRef]
28. Barrera Alvarez, J.L.; Hernandez Montero, F.E. Classification of MPSK signals through eighth-order statistical signal processing. *IEEE Lat. Am. Trans.* 2017, 15, 1601–1607. [CrossRef]
29. Hasan, A.N.; Shongwe, T. The use of multiclass support vector machines and probabilistic neural networks for signal classification and noise detection in PLC/OFDM channels. In Proceedings of the International Conference Radioelektronika, Bratislava, Slovakia, 15–16 May 2020; pp. 1–6.