

# Article

# **Effective Feature Selection Method for Deep** Learning-Based Automatic Modulation Classification Scheme Using Higher-Order Statistics <sup>+</sup>

# Sang Hoon Lee, Kwang-Yul Kim<sup>D</sup> and Yoan Shin \*<sup>D</sup>

School of Electronic Engineering, Soongsil University, Seoul 06978, Korea; leesanghoon@ssu.ac.kr (S.H.L.); kky1117@ssu.ac.kr (K.-Y.K.)

- \* Correspondence: yashin@ssu.ac.kr; Tel.: +82-10-4234-0632
- + This paper is an extended version of the conference paper presented in the 1st International Conference on Artificial Intelligence in Information and Communication (ICAIIC 2019), Okinawa, Japan, 11-13 February 2019.

Received: 30 June 2019; Accepted: 8 January 2020; Published: 13 January 2020



Abstract: Recently, in order to satisfy the requirements of commercial communication systems and military communication systems, automatic modulation classification (AMC) schemes have been considered. As a result, various artificial intelligence algorithms such as a deep neural network (DNN), a convolutional neural network (CNN), and a recurrent neural network (RNN) have been studied to improve the AMC performance. However, since the AMC process should be operated in real time, the computational complexity must be considered low enough. Furthermore, there is a lack of research to consider the complexity of the AMC process using the data-mining method. In this paper, we propose a correlation coefficient-based effective feature selection method that can maintain the classification performance while reducing the computational complexity of the AMC process. The proposed method calculates the correlation coefficients of second, fourth, and sixth-order cumulants with the proposed formula and selects an effective feature according to the calculated values. In the proposed method, the deep learning-based AMC method is used to measure and compare the classification performance. From the simulation results, it is indicated that the AMC performance of the proposed method is superior to the conventional methods even though it uses a small number of features.

Keywords: automatic modulation classification; cumulant; correlation; effective feature; deep neural network

# 1. Introduction

In an effort to improve the transmission efficiency of satellite communication and mobile communication systems, the systems should consider adaptive changing parameters such as a modulation scheme, a transmission rate and a carrier frequency according to a channel state [1,2]. As part of this study, in order to effectively classify the modulation scheme, an automatic modulation classification (AMC) method has been widely studied [3,4]. Generally, the receiver can estimate the modulation scheme of the transmitted signal in the commercial system. However, since the communication parameters of the enemy cannot be accurately estimated in military communications, the research has been oriented to estimate the communication parameters by using only the received signals [5]. Thus, in order to improve the jamming performance against the enemy communication system, various research works have been undertaken to classify the modulation scheme by the AMC method [6]. The techniques for the AMC can be roughly classified into two types. The first type maximizes the likelihood function based on the statistical model from the received signal. However,



this method has poor performance due to the error that occurs when there is a change in the channel characteristics in the real environment. Moreover, considering different models, the computation of the algorithm becomes very complicated and the calculation effort becomes large [7]. The second type uses the machine-learning technique. This method uses training data to train machine learning models to classify modulation type. Assuming that the training data is similar to the actual data, it can demonstrate good performance even though the computational complexity is lower than the likelihood method. Therefore, in order to classify the modulation type quickly and accurately, the machine-learning algorithm is mainly used. The AMC scheme based on machine learning consists of a feature extraction step that extracts features from the received signal and a signal classification step that classifies the modulation type.

There are various techniques such as a deep neural network (DNN) [8], a convolutional neural network (CNN) [9] and a recurrent neural network (RNN) [10] for the AMC that have been studied. The CNN algorithm is a method that shows excellent performance in image processing. The research has been carried out to classify the signals by using the constellation images of the received signals as the features and to classify the signals by imaging the statistical characteristics [9]. The RNN algorithm is an excellent method for analyzing time-series data but requires algorithmic complexity and high calculation effort compared to performance [10]. On the other hand, the DNN algorithm can learn complex structures from various data and shows good performance for various machine-learning problems in recent years. The features frequently used in the AMC technique based on machine learning use the higher-order statistic cumulant and signal size, frequency, phase dispersion, and wavelet coefficient [11,12]. Therefore, in this paper, the cumulant is used as a feature for the AMC and as input data of the DNN algorithm. Various research works have focused on machine-learning methods, rather than analyzing features used as input data. Therefore, in this paper, we use only the features that greatly affect the classification performance through the proposed algorithm to reduce the computational complexity and to identify the received signal quickly while using the basic DNN structure algorithm. In reference [13], we have confirmed the difference of signal classification performance according to the features used as input data in the DNN algorithm and confirmed the features with high and low importance. Based on this, an effective feature selection method using a correlation coefficient is exploited to obtain the representative values and to verify the classification performance [14]. The proposed method is more effective than the conventional method, which uses mutual information and correlation coefficients in selecting five features [15,16]. In this paper, we compare the performance of the proposed method using only the correlation coefficient in various environments, the conventional method using mutual information and correlation coefficients, and three methods using only mutual information. In addition, we confirmed the performance of the proposed method with the addition of four kinds of sixth-order cumulants with large variability in the low signal-to-noise ratio (SNR) environment besides the second and fourth-order cumulant. In order to evaluate the proposed method, the representative value was selected from various cumulants by using each method and two kinds of the simulation were conducted. In the first simulation, in order to find the effective features values, we ranked the cumulants based on the calculation from each method. Then, we sequentially measured the classification performance by excluding the feature values one by one. In the second set of simulations, in order to measure the classification performance according to the group, the cumulants were divided into three groups (top, middle, and bottom) based on the ranking obtained from the efficient features extraction method. The following is a summary of how each group is divided.

- 1. Top group: the five highest important representative values of each method.
- 2. Middle group: the five medium important representative values of each method.
- 3. Bottom group: the five lowest important representative values of each method.

The cumulants in each group were used as the input data of the DNN algorithm to measure the classification performance. The three AMC environments that use the features of each group as input

values were implemented and the superiority of the proposed method was confirmed according to the group performance.

The rest of the paper is organized as follows. In Section 2, we explain the features and data analysis method. In Section 3, we introduce the proposed method and the conventional method. In Section 4 we describe the DNN structure used in this paper and present the simulation results. Finally, Section 5 provides the conclusions of the paper.

# 2. Data Analysis Techniques

# 2.1. Cumulant

The cumulant is one of the typical statistical features used in the hierarchical AMC scheme [17]. In this paper, the higher-order cumulants for baseband received signal samples r[n] generated in the additive white Gaussian noise (AWGN) channel are extracted as representative features and used as the inputs to the DNN algorithm. Since the proposed method exploits the correlation characteristics, we consider the high-order cumulants as the feature values. Table 1 summarizes the expressions for the second-, fourth- and sixth-order cumulants [16,18] used in this paper.

Higher-Order Cumulant		Expression
Second-Order Cumulant	$ C_{20} $ $ C_{21} $	M <sub>20</sub>    M <sub>21</sub>
Fourth-Order Cumulant	$\begin{array}{c}  C_{40}  \\  C_{41}  \\  C_{42}  \end{array}$	$ \begin{aligned} &  M_{40} - 3M_{20}^2  \\ &  M_{41} - 3M_{20}M_{21}  \\ &  M_{42} - M_{20}^2 - 2M_{21}^2  \end{aligned} $
Sixth-Order Cumulants	$\begin{array}{c}  C_{60}  \\  C_{61}  \\  C_{62}  \\  C_{63}  \end{array}$	$ \begin{array}{c} \left  M_{60} - 15M_{20}M_{40} + 30M_{20}^3 \right  \\ M_{61} - 10M_{20}M_{41} - 5M_{21}M_{40} + 30M_{21}M_{20}^2 \\ M_{62} - 6M_{20}M_{42} - 8M_{21}M_{41} - M_{22}M_{40} + 6M_{20}^2M_{22} + 24M_{21}^2M_{20} \right  \\ M_{63} - 9M_{21}M_{42} + 12M_{21}^3 - 3M_{20}M_{42} - 3M_{22}M_{41} + 18M_{20}M_{21}M_{22} \end{array} $

Table 1. The expressions for various cumulants.

Here,  $C_{xy}$  is the (x + y)-th order cumulant and  $M_{xy} \triangleq E[(r[n])^{x-y}(r^*[n])^y]$ , is the (x + y)-th order moment of the received signal r[n]. Table 2 summarizes the theoretical absolute value according to the modulation types; BPSK (binary phase shift keying), QPSK (quadrature phase shift keying), 8-PSK, 16-QAM (quadrature amplitude modulation).

	BPSK	QPSK	8-PSK	16-QAM	64-QAM
C <sub>20</sub>	1	0	0	0	0
$ C_{21} $	1	1	1	1	1
$ C_{40} $	2	1	0	0.68	0.62
$ C_{41} $	2	0	0	0	0
$ C_{42} $	2	1	1	0.68	0.62
$ C_{60} $	16	0	0	0	0
$ C_{61} $	16	4	0	2.08	1.79
$ C_{62} $	16	0	0	0	0
$ C_{63} $	16	4	4	2.08	1.79

Table 2. The theoretical absolute values of cumulants.

### 2.2. Correlation

In this paper, we use the correlation method which is one of data analysis methods to select the effective feature. Correlation refers to the similarity between data, so features with a high correlation

coefficient between feature values are relatively inefficient in the AMC processes. The Pearson correlation for the variables X, Y is [19]:

$$\operatorname{cor}(X,Y) = \frac{C(X,Y)}{\sigma_X \sigma_Y},\tag{1}$$

where C(X, Y) is the covariance of the variables X and Y. Thus, (1) can be expressed as:

$$\operatorname{cor}(X,Y) = \frac{\sum_{i=1}^{k} (x_i - \overline{x})(y_i - \overline{y})}{\sigma_X \sigma_Y},$$
(2)

where  $\sigma_X$  and  $\sigma_Y$  are the standard deviations of *X* and *Y*. From this correlation coefficient, information on other data can be obtained through one data. The proposed method uses correlation as a data analysis method to select effective features [19].

## 2.3. Mutual Information Quantity

When classifying a signal using the DNN algorithm, the input data should be selected to include as much information as possible. The mutual information quantity is one of the methods used in measuring the information of arbitrary variables used for this purpose [15]. When the modulation scheme information used in the transmitter is represented by c, mutual information values for the *i*-th feature and the module c are defined as:

$$I(x_i;c) = \iint P(x_i,c) \log \frac{P(x_i,c)}{P(x_i)P(c)} dx_i dc,$$
(3)

where  $P(x_i, c)$  is the joint probability distribution of  $x_i$  and c. The high mutual information value can be useful for the AMC because the features contains a lot of information about the module c [15].

# 3. Proposed Effective Feature Selection Method

#### 3.1. Conventional Effective Feature Selection Based on Mutual Information and Correlation

The conventional method based on mutual information and correlation performs preprocessing before using it for input data. This is to reduce the computational complexity of the algorithm and maintain the identification performance. The conventional mutual information and correlation method for extracting an efficient feature is expressed as [15]:

$$max_{x_{j}\in X-S_{m01}}\left[I(x_{j};c) - \frac{1}{m-1}\sum_{x_{i}\in S_{m-1}}I(x_{j};x_{i})\right],$$
(4)

where  $I(x_j; c)$  denotes the mutual information value between the feature value and the corresponding module,  $S_m$  denotes a set of selected feature values up to m runs, and X denotes a set of all feature values. A representative value according to each feature can be obtained from (4) and a feature having a large representative value is the most efficient feature. Table 3 shows the representative values of the second-, fourth-, and sixth-order cumulants of the conventional method [15]. As can be observed from the table, the conventional method indicates that the most effective feature is  $C_{60}$  and the most ineffective feature is  $C_{21}$  in the 10 dB SNR environment.

# 3.2. Conventional Effective Feature Selection Based on Mutual Information

The conventional mutual information method is used among the data analysis methods and the effective feature is selected from the information between the digital signals and the feature [20]. The mutual information quantities are expressed as

$$ECVI_{i} = \sum_{i}^{M} \left| I(r_{j}, t_{ij}) \right|, \tag{5}$$

where  $r_j$  denotes the *j*-th received signal and  $t_{ij}$  denotes the *i*-th characteristic value of the *j*-th received signal. If the amount of mutual information between the received signal and a specific feature is high, the features is valuable on the AMC process because it contains more information regarding the received signal. Therefore, the features having the largest representative value obtained from (5) can be considered an effective feature that greatly affects the AMC performance. Table 4 shows the representative values obtained from (5) of the mutual information method [20]. As shown in the table, mutual information method identifies  $C_{62}$  as the most effective feature and  $C_{21}$  as the most ineffective feature in the 10 dB SNR environment.

**Table 3.** Effective correlation values of each cumulant in various signal-to-noise ratio (SNR) environments for the conventional method. High values indicate strong influence on the classification performance, meaning that the associated cumulants are more effective features for AMC systems.

SNR	-10 dB	-5 dB	0 dB	5 dB	10 dB
C <sub>20</sub>	8.92	8.86	8.59	8.29	8.08
$C_{21}$	6.08	7.23	6.35	5.56	4.98
$C_{40}$	9.11	8.98	8.61	8.11	7.68
$C_{41}$	9.05	8.93	8.63	8.31	8.09
$C_{42}$	9.13	8.92	8.18	7.44	6.94
$C_{60}$	9.04	9.02	8.85	8.49	8.20
$C_{61}$	8.87	8.88	8.38	7.75	7.39
$C_{62}$	9.02	8.91	8.63	8.31	8.15
C <sub>63</sub>	9.07	8.93	8.29	7.51	7.01

**Table 4.** Effective correlation values of each cumulant in various SNR environments for the mutual information. High values indicate strong influence on the classification performance, meaning that the associated cumulants are more effective features for automatic modulation classification (AMC) systems.

SNR	-10 dB	-5 dB	0 dB	5 dB	10 dB
C <sub>20</sub>	8.93	8.91	8.75	8.60	8.50
$C_{21}$	6.13	7.34	6.58	5.95	5.43
$C_{40}$	9.13	9.06	8.72	8.30	7.98
$C_{41}$	9.08	9.01	8.84	8.61	8.55
$C_{42}$	9.12	8.94	8.41	7.82	7.39
$C_{60}$	9.09	8.89	8.90	8.70	8.51
$C_{61}$	9.04	8.99	8.69	8.16	7.83
$C_{62}$	9.11	9.04	8.85	8.68	8.60
$C_{63}$	9.14	9.07	8.54	7.90	7.46

#### 3.3. Proposed Effective Feature Selection Based on Correlation Coefficient

The optimal selection of the input data determines the optimal group of feature by comparing all combinations of feature. However, it is difficult to perform because it requires a large amount of computation. Therefore, in this paper, in order to reduce the computational complexity of the AMC and to maintain the classification performance, we proposed an effective feature method with a large influence on the classification performance based on the analysis of the correlation coefficient.

Thus, the effect of each feature on the classification performance should be numerically expressed as a representative value. The proposed method is expressed as:

$$ECV_{i} = \sum_{j=1}^{M} \sum_{\substack{k=1\\k\neq j}}^{M} \left| cor(x_{ik}, x_{ij}) \right|,$$
(6)

where *M* is the number of features,  $x_{ij}$  is the *j*-th feature of the *i*-th modulation type, and  $cor(x_{ik}, x_{ij})$  is the correlation coefficient between the two features. In this manner, one representative value can be obtained according to each characteristic. A characteristic with a large representative value has a little influence on the AMC performance. On the other hand, a feature with a small representative value has a strong influence on the classification performance and becomes the effective feature required for the AMC. As shown in Table 5, the proposed method indicates that the most effective feature is  $C_{40}$  and the most ineffective feature is  $C_{21}$  in the SNR 10 dB environment. As shown in Tables 3–5 the effective feature for each method is different, and the performance of each method is verified through two sets of simulations.

**Table 5.** Effective correlation values of each cumulant in various SNR environments for the proposed method. Unlike the other methods, small values indicate strong influence on the classification performance, meaning that the associated cumulants are more effective features for AMC systems.

SNR	-10 dB	-5 dB	0 dB	5 dB	10 dB
	0.37	1.73	6.07	12.45	16.97
$C_{21}$	1.80	3.07	9.07	15.63	18.34
$C_{40}$	0.56	1.35	3.52	7.69	12.19
$C_{41}$	1.19	3.1	8.21	14.62	18.23
$C_{42}$	0.73	0.95	9.28	15.31	18.07
$C_{60}$	1.82	0.65	1.77	8.08	14.02
$C_{61}$	2.96	3.55	8.47	13.46	15.24
$C_{62}$	3.34	4.77	8.5	14.64	18.07
C <sub>63</sub>	2.81	5.36	10.27	15.56	18.15

In the AMC structure of this paper, the modulated signals to be classified are generated in the AWGN channel, and the cumulants are extracted for each signal. The extracted cumulants are represented by one representative value through the proposed method as shown in Equation (6). In order to reduce the computational complexity of the algorithm and to classify the modulation type quickly, it extracts the top feature and classifies the modulation type after learning by using it as input data to the DNN algorithm. The proposed AMC structure is shown in Figure 1.



Figure 1. The AMC structure of the proposed scheme.

# 4. Simulation Results

## 4.1. Deep Neural Network (DNN) Structure and Simulation Environments

In this paper, five types of digital communication signals BPSK, QPSK, 8-PSK, 16-QAM, and 64-QAM are considered. Additionally, the nine characteristic values consisting of the second, fourth and sixth-order cumulant are used. The structure of the DNN algorithm consists of an input layer with nine feature in a fully connected layer structure, a hidden layer consisting of three layers, 40 nodes, 20 nodes, and 10 nodes, and finally an output layer for classifying signals. In the hidden layer, the Rectified Linear Unit (ReLU) function [21] is used and in the last output layer, each modulation type is classified by Softmax [22]. Table 6 shows the nonlinear activation functions considered in this paper. Since the Softmax function can produce the output in terms of probability, we can calculate the accuracy for each classified signal. The DNN structure for the first set of simulations is shown in Figure 2.

**Table 6.** Definitions of the non-linear activation functions used in the deep neural network (DNN). The ReLU was used for all hidden layers, while the Softmax was used for the output layer.



Figure 2. DNN structure for the first set of simulations.

In both sets of simulations, we trained the DNN algorithm several times for hyperparameter optimization. Since the DNN is a very complex structure, it is difficult to find the optimal weighting coefficients in one calculation. Therefore, in this paper we set up the hyperparameters and trained the DNN algorithms through the following standard procedures. In the first step, we adjusted the hyperparameters and trained the DNN using the backpropagation algorithm based on the gradient descent, and applied the batch normalization to prevent the overfitting during the training. Next, the validation errors were counted and the training was stopped to prevent overfitting when the validation error started to increase. Also, when the validation errors did not decrease anymore, we continued to train by cutting the learning rate in half. We utilized 20% of the input data for the validation.

In order to train the above DNN structure, the epoch was set to 200, batch size to 64, and a total of 50,000 units of data (10,000 digital modulation symbols by each of 5 modulation schemes considered) were generated in various SNR environments. Then, 9 features ( $C_{20}$ ,  $C_{21}$ ,  $C_{40}$ ,  $C_{41}$ ,  $C_{42}$ ,  $C_{60}$ ,  $C_{61}$ ,  $C_{62}$ ,  $C_{63}$ ) were considered for each digital modulation symbol, yielding 450,000 features used as the input data. In other words, the number of training data units is 450,000, and that of the test and the validation data is 90,000 each, which is 20% of the training data. The parameters of the first DNN obtained through the above process are summarized in Table 7.

Parameters	Value
Number of input nodes	9
Number of hidden layer	3
Number of nodes of 1st hidden layer	40
Number of nodes of 2st hidden layer	20
Number of nodes of 3st hidden layer	10
Number of output nodes	5
Activation function of hidden layer	ReLU
Activation function of output layer	Softmax
Number of training data	450,000
Number of test data	90,000
Number of validation data	90,000
Epochs	200
Bacth Size	64

**Table 7.** The DNN parameters used in the first simulation for optimal feature extraction and performance verification.

#### 4.2. Simulation Result

In this paper, we propose an efficient feature extraction method to reduce training time while maintaining AMC performance. In order to evaluate the proposed method, the representative value was selected from various cumulants by using each method and two sets of simulations were conducted. In the first simulation, in order to find the effective feature values, we ranked the cumulants based on the calculation from each method. Then, we measured the classification performance sequentially by excluding the feature values one by one. The structure of DNN is the same environment except for the input layer. Table 8 summaries the classification performance according to elimination of each feature.

SNR	-10 dB	-5 dB	0 dB	5 dB	10 dB
ALL	86.37	89.67	95.70	98.61	99.91
$C_{20}$	77.44	83.99	92.05	93.03	98.38
C <sub>21</sub>	83.10	84.60	91.85	97.56	98.92
$C_{40}$	79.16	83.91	91.03	96.73	98.12
$C_{41}$	83.19	84.93	91.92	97.52	98.40
C <sub>42</sub>	82.90	83.67	91.47	97.47	98.69
$C_{60}$	83.09	83.14	91.14	96.91	98.35
$C_{61}$	83.17	84.53	91.52	97.30	98.27
C <sub>62</sub>	83.28	84.87	92.14	97.42	98.55
C <sub>63</sub>	83.11	85.46	91.40	97.34	98.73

**Table 8.** Classification performance according to elimination of each feature [%]. The feature with the lowest value is the most essential for the classification.

In the 10 dB SNR environment, the most essential or effective feature is  $C_{40}$  and the most unnecessary or ineffective feature is  $C_{21}$ . In the case of the proposed method, these features  $C_{40}$  and  $C_{21}$  are identified precisely. On the other hand, the mutual information method identified  $C_{62}$  as the most essential feature and  $C_{21}$  as the most unnecessary feature. In the case of the conventional method, the most effective feature was extracted as  $C_{60}$  and the most unnecessary feature was extracted as  $C_{21}$ . In the proposed method, the most effective feature and the most unnecessary feature were accurately identified in the 10 dB SNR environment, while the other two methods accurately identified the unnecessary feature but failed to extract the most effective feature. In other words, the proposed method shows superior performance in extracting effective features compared to the conventional methods. Table 9 shows the difference in the classification performance when all the features are used and when the effective features are excluded by each method. If a method shows the highest value for a given SNR value in the table, that method is the best in correctly identifying the effective features.

SNR	-10 dB	-5 dB	0 dB	5 dB	10 dB
Proposed	8.93	6.53	4.56	1.88	1.79
Mutual information	3.26	4.21	4.56	1.70	1.36
Conventional	3.47	6.53	4.56	1.70	1.56

**Table 9.** Difference in classification performance when all features are used and when the effective features are excluded by each method.

In [14], only the second and the fourth-order cumulants are considered, and the variation of characteristic values is small even in low SNR environments. Therefore, there was little variation in efficiency ranking in a low SNR environment. However, in this paper, it can be seen that the order of efficiency fluctuates significantly in a low SNR environment due to the sixth-order cumulant with high variability. In this manner, when the feature with a large variability is used, the ranking of the efficiency value of each feature can be changed according to each SNR. However, since the performance is also changed to a similar trend, it becomes suitable even in environments using the feature with high volatility. The proposed method shows higher performance not only when using the second- and fourth-order cumulants but also when using the sixth-order cumulant.

In the second set of simulations, in order to measure the classification performance according to the group, the cumulants were divided into three groups (top, middle, and bottom) based on the ranking obtained from the efficient feature-extraction method. The cumulants in each group were used as the inputs to the DNN algorithm to measure the classification performance. The parameters of the second DNN are summarized in Table 10.

Parameters	Value
Number of input nodes	5
Number of hidden layer	2
Number of nodes of 1st hidden layer	30
Number of nodes of 2st hidden layer	10
Number of output nodes	5
Activation function of hidden layer	ReLU
Activation function of output layer	Softmax
Number of training data	250,000
Number of test data	50,000
Number of validation data	50,000
Epochs	200
Batch Size	64

**Table 10.** The DNN parameters used in the second simulation for optimal feature group extraction and performance verification.

Figure 3 and Table 11 shows the simulation results of the proposed method and Figures 4 and 5 show the results of the conventional method. Tables 12–14 represent the features used for each group. In Figures 3–5, the desirable result is that the best classification performance is achieved when the top group is used as the input data, while the worst classification performance is achieved when the bottom group is used as the input data. In this respect, the conventional methods are unsatisfactory since the top group cannot always obtain the best performance in all SNR ranges. However, the proposed method shows a stable and the best performance over a wide SNR range of –2 dB to 10 dB. Even if the same amount of data is exploited, there is a large difference in performance depending on the features used. Also, even in the low SNR environment, the performance of the top group is better than that of the bottom group in the high SNR environment. Therefore, we conclude that the proposed method is very effective at extracting the useful feature group. Figure 6 shows the classification performance when only the features of the top group of each method is used as the input data. This figure also highlights that the proposed method shows superior performance in all SNR environments. Therefore,

we conclude that the proposed method is very effective at extracting the input data group. Figure 6 shows the classification performance when only the features of the top group of each method are used as the input data. where the proposed method shows superior performance in all SNR environments.



**Figure 3.** Classification performance of each group obtained by the proposed method. The top group achieves the best classification performance while the bottom group achieves the worse performance, which shows the validity of the proposed method.

**Table 11.** Classification performance of each group obtained through the proposed method.

SNR	-10 dB	-5 dB	0 dB	5 dB	10 dB
Top group	80.67	81.02	90.65	96.17	97.45
Middle group	80	81.76	86.69	95.05	96.46
Bottom group	80	82.06	85.73	88.86	91.59



**Figure 4.** Classification performance of each group obtained by the mutual information method. The bottom group achieves the best classification performance while the top group achieves the worse performance, which.



Figure 5. Classification performance of each group obtained by the conventional method.

SNR	Top Group	Middle Group	Bottom Group
-10 dB	$C_{62}, C_{61}, C_{63}, C_{60}, C_{21}$	$C_{63}, C_{60}, C_{21}, C_{41}, C_{42}$	$C_{20}, C_{40}, C_{42}, C_{41}, C_{21}$
-5 dB	$C_{63}, C_{62}, C_{61}, C_{41}, C_{21}$	$C_{61}, C_{41}, C_{21}, C_{20}, C_{40}$	$C_{60}, C_{42}, C_{40}, C_{20}, C_{21}$
0 dB	$C_{63}, C_{42}, C_{21}, C_{62}, C_{61}$	$C_{21}, C_{62}, C_{61}, C_{41}, C_{20}$	$C_{60}, C_{40}, C_{20}, C_{41}, C_{61}$
5 dB	$C_{21}, C_{63}, C_{42}, C_{62}, C_{41}$	$C_{42}, C_{62}, C_{41}, C_{61}, C_{20}$	$C_{40}, C_{60}, C_{20}, C_{61}, C_{41}$
10 dB	$C_{21}, C_{41}, C_{63}, C_{42}, C_{62}$	$C_{63}, C_{42}, C_{62}, C_{20}, C_{61}$	$C_{40}, C_{60}, C_{61}, C_{20}, C_{62}$

**Table 12.** Feature of each group used in the proposed method.

 Table 13. Feature of each group used in the mutual information method.

SNR	Top Group	Middle Group	Bottom Group
-10 dB	$C_{63}, C_{40}, C_{42}, C_{62}, C_{60}$	$C_{42}, C_{62}, C_{60}, C_{41}, C_{61}$	$C_{21}, C_{20}, C_{61}, C_{41}, C_{60}$
-5 dB	$C_{63}, C_{40}, C_{62}, C_{41}, C_{61}$	$C_{62}, C_{41}, C_{61}, C_{42}, C_{20}$	$C_{21}, C_{60}, C_{20}, C_{42}, C_{61}$
0 dB	$C_{60}, C_{62}, C_{41}, C_{20}, C_{40}$	$C_{41}, C_{20}, C_{40}, C_{61}, C_{63}$	$C_{21}, C_{42}, C_{63}, C_{61}, C_{40}$
5 dB	$C_{60}, C_{62}, C_{41}, C_{20}, C_{40}$	$C_{41}, C_{20}, C_{40}, C_{61}, C_{63}$	$C_{21}, C_{42}, C_{63}, C_{61}, C_{40}$
10 dB	$C_{62}, C_{41}, C_{60}, C_{20}, C_{40}$	$C_{60}, C_{20}, C_{40}, C_{61}, C_{63}$	$C_{21}, C_{42}, C_{63}, C_{61}, C_{40}$

 Table 14. Feature of each group used in the conventional method.

SNR	Top Group	Middle Group	Bottom Group
-10 dB	$C_{42}, C_{40}, C_{63}, C_{41}, C_{60}$	$C_{63}, C_{41}, C_{60}, C_{62}, C_{20}$	$C_{21}, C_{61}, C_{20}, C_{62}, C_{60}$
-5 dB	$C_{60}, C_{40}, C_{41}, C_{63}, C_{42}$	$C_{41}, C_{63}, C_{42}, C_{62}, C_{61}$	$C_{21}, C_{20}, C_{61}, C_{62}, C_{42}$
0 dB	$C_{60}, C_{62}, C_{41}, C_{40}, C_{20}$	$C_{41}, C_{40}, C_{20}, C_{61}, C_{63}$	$C_{21}, C_{42}, C_{63}, C_{61}, C_{20}$
5 dB	$C_{60}, C_{62}, C_{41}, C_{20}, C_{40}$	$C_{41}, C_{20}, C_{40}, C_{61}, C_{63}$	$C_{21}, C_{42}, C_{63}, C_{61}, C_{40}$
10 dB	$C_{60}, C_{62}, C_{41}, C_{20}, C_{40}$	$C_{41}, C_{20}, C_{40}, C_{61}, C_{63}$	$C_{21}, C_{42}, C_{63}, C_{61}, C_{40}$



**Figure 6.** Classification performance when only the features of the top group of each method are used as the input data.

## 5. Conclusions

Recently, the DNN-based AMC scheme has been studied as a method to improve jamming performance. However, research on the features used as the input data is insufficient and most studies aim at improving the calculation and the performance of the algorithm. In this paper, we propose an efficient feature-extraction method for the DNN-based AMC, we analyze feature used as input data, and we select an effective feature through the proposed method. From the results, it can be established that even if the same amount of data is used, the difference in classification rate performance according to each feature is large and the task of extracting efficient features is important. The optimal activity of selecting input data will be to find the optimal feature group by comparing the performance according to the combination of all the features. However, this is a difficult method to actually perform because it requires a large amount of calculation. Therefore, it is necessary to analyze features such as conventional techniques and proposed methods that use mutual information and correlation between data. It is expected that the AMC with high classification performance can be realized with a small computation effort by extracting the efficient feature values using the proposed method. Thus, we conclude that the proposed method can be considered a method to improve the performance of the AMC for military communication systems, AMC-based jamming systems, and the automatic coding and modulation for commercial wireless communication systems.

**Author Contributions:** S.H.L. contributed to this work in experiment planning, experiment measurements, data analysis and manuscript preparation. K.-Y.K. manuscript preparation, data analysis. Y.S. contributed in experiment planning, data analysis, manuscript preparation. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2019-2018-0-01424) supervised by the IITP (Institute for Information and communications Technology Promotion).

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

# References

- Karrs, K.; Kuzdeba, S.; Petersen, J. Modulation Recognition Using Hierarchical Deep Neural Networks. In Proceedings of the 2017 IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN), Piscataway, NJ, USA, 6–9 March 2017.
- Dobre, O.A.; Abdi, A.; Bar-Ness, Y.; Su, W. Blind Modulation Classification: A Concept Whose Time Has Come. In Proceedings of the Sarnoff Symposium on Advances in Wired and Wireless Communication, Princeton, NJ, USA, 18–19 April 2005; pp. 223–228.
- Swami, A.; Sadler, B.M. Hierarchical Digital Modulation Classification Using Cumulants. *IEEE Trans. Commun.* 2000, 48, 416–429. [CrossRef]
- 4. Hazza, A.; Shoaib, M.; Alshebeili, S.A.; Fahad, A. An Overview of Feature-Based Methods for Digital Modulation Classification. In Proceedings of the 1st International Conference on Communications, Signal Processing, and Their Applications ICCSPA 2013, Sharjah, UAE, 12–14 February 2013; pp. 1–6.
- 5. Lee, J.H.; Kim, J.K.; Kim, B.D.; Yoon, D.W.; Choi, J.W. Robust Automatic Modulation Classification Technique for Fading Channels Via Deep Neural Network. *Entropy* **2017**, *19*, 454.
- 6. Amuru, S.; Buehrer, M. Optimal Jamming Against Digital Modulation. *IEEE Trans. Inf. Forensics Secur.* 2015, 10, 2212–2224. [CrossRef]
- 7. Wei, W.; Mendel, J.M. Maximum-Likelihood Classification for Digital Amplitude-Phase Modulations. *IEEE Trans. Commun.* **2010**, *48*, 189–193. [CrossRef]
- 8. Ali, A.; Yangyu, F. Automatic Modulation Classification Using Deep Learning Based on Sparse Autoencoders with Nonnegativity Constraints. *IEEE Signal Process. Lett.* **2017**, *24*, 1626–1630. [CrossRef]
- 9. Zhang, M.; Diao, M.; Guo, L. Convolutional Neural Networks for Automatic Cognitive Radio Waveform Recognition. *IEEE Access* 2017, *5*, 11074–11082. [CrossRef]
- Hong, D.; Zhang, Z.; Xu, X. Automatic Modulation Classification Using Recurrent Neural Networks. In Proceedings of the 3rd IEEE International Conference on Computer and Communications (ICCC), Chengdu, China, 13–16 December 2017; pp. 695–700.
- An, N.; Li, B.; Huang, M. Modulation Classification of Higher Order MQAM Signals Using Mixed-Order Moments and Fisher Criterion. In Proceedings of the 2nd International Conference on Computer and Automation Engineering (ICCAE), Singapore, 26–28 February 2010; pp. 150–153.
- Huang, F.Q.; Zhong, Z.M.; Xu, Y.T.; Ren, G.C. Modulation Recognition of Symbol Shaped Digital Signals. In Proceedings of the 2008 International Conference on Communications, Circuits and Systems, Xiamen, China, 25–27 May 2008; pp. 328–332.
- Lee, S.H.; Kim, K.Y.; Kim, T.H.; Shin, Y. Analysis on Classification Accuracy by Data Filtering in DNN-Based Automatic Signal Classification. In Proceedings of the KICS Summer Conference 2018, Jeju, Korean, 20–22 June 2018; pp. 692–693.
- 14. Lee, S.H.; Kim, K.Y.; Kim, J.H.; Shin, Y. Effective Feature-Based Automatic Modulation Classification Method Using DNN Algorithm. In Proceedings of the International Conference on Artificial Intelligence in Information and Communication (ICAIIC), Okinawa, Japan, 11–12 February 2019; pp. 557–559.
- Peng, H.; Long, F.; Ding, C. Feature Election Based on Mutual Information: Criteria of Maxdependency, Max-Relevance, and Min-Redundancy. *IEEE Trans. Pattern. Anal. Mach. Intel.* 2006, 27, 1226–1238. [CrossRef] [PubMed]
- Peng, S.L.; Jiang, H.; Wang, H.; Alwageed, H.; Yao, Y.D. Modulation Classification Using Convolutional Neural Network Based Deep Learning Model. In Proceedings of the 26th Wireless and Optical Communication Conference (WOCC), Newark, NJ, USA, 7–8 April 2017; pp. 1–5.
- Orlic, V.D.; Dukic, M.L. Automatic Modulation Classification: Sixth-Order Cumulant Features as a Solution for Real-World Challenges. In Proceedings of the 20th Telecommunications Forum (TELFOR), Belgrade, Serbia, 20–22 November 2012; pp. 392–399.
- Chang, D.C.; Shin, P.K. Cumulants-Based Modulation Classification Technique in Multipath Fading Channels. IET Commun. 2015, 9, 828–835. [CrossRef]
- 19. Sheugh, L.; Alizadeh, S.H. A Note on Pearson Correlation Coefficient as a Metric of Similarity in Recommender System. In Proceedings of the 2015 AI & Robotics (IRANOPEN), Qazvin, Iran, 12 April 2015.

- Ebihara, T.; Taoka, H.; Miki, N.; Sawahashi, M. Performance of Outer-Loop Control for AMC Based on Mutual Information in MIMO-OFDM Downlink. In Proceedings of the 2012 IEEE 75th Vehicular Technology Conference (VTC Spring), Yokohama, Japan, 6–9 May 2012; pp. 990–995.
- 21. Zhan, C.; Woodland, P.C. Parameterised Sigmoid and ReLU Hidden Activation Functions for DNN Acoustic Modelling. In Proceedings of the INTERSPEECH 2015, Dresden, Germany, 6–10 September 2015; pp. 3224–3228.
- 22. Loffe, S.; Szegedy, C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *arXiv* 2015, arXiv:1502.03167v3.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).