Deep Ensemble Learning for Automatic Modulation Classification

Jiali Nie (✉ 20185371@csuft.edu.cn)
  Central South University of Forestry and Technology
Wenke Tan
  Central South University of Forestry and Technology
Houmei Zhang
  Central South University of Forestry and Technology

RESEARCH

Deep Ensemble Learning for Automatic Modulation Classification

Jiali Nie*, Wenke Tan and Houmei Zhang

*Correspondence:
20185371@csuft.edu.cn
School of Computer and
Information Engineering, Central
South University Of Forestry And
Technology, Changsha, Hunan
Full list of author information is
available at the end of the article

Abstract

Automatic modulation classification (AMC) plays an increasingly vital role in cognitive radio (CR), cognitive electronic warfare, and other areas. It aims at classifying the modulated modes of the received signals accurately and provides a guarantee for the subsequent detailed parameter identification. Deep learning (DL) methods allow the computer to automatically learn the pattern features and integrate features into the process of building the model, thereby reducing the incompleteness caused by artificial design features. At the same time, the DL methods have been applied in the AMC field as its powerful ability to process complex data and have achieved excellent performance in recent years. In this paper, we propose a deep ensemble learning AMC network, which uses a multi-model ensemble method to fuse multiple DL features. Specifically, different DL models are integrated by ensemble learning, which enhances the learning ability of the single model. With the proposed ensemble model trained on a measured wireless signal dataset, we conclude that the ensemble structure of Inception and CLDNN can fuse spatial features and temporal features, and achieve state-of-the-art performance in AMC tasks. Besides, the impact of the inphase/quadrature (I/Q) sample-length on wireless signals is further investigated, and find that the classification accuracy of the deep ensemble model is improved by 0.7% to 10% compared to the single model under various sample-length. Simultaneously, we visualize convergence clustering with t-distributed stochastic neighbor embedding (t-SNE), and the visualization results prove that the deep ensemble model has a stronger clustering ability than a single model.

Keywords: automatic modulation classification; deep learning; cognitive radio; ensemble learning; LSTM; inception

1 Introduction

As the 6th generation mobile networks (6G) satisfies the growing demand for communication applications, and the new generation of communication technology represented by 6G will bring new impetus to industrial development and innovation. Wireless communication architectures are gradually becoming intelligent and automated. Among them, some researchers proposed intelligent solutions for automatic modulation classification (AMC), which promotes wireless signal recognition, spectrum monitoring [1, 2]. In wireless signal application scenarios, cognitive radio (CR) mainly addresses spectrum allocation and management, and AMC is a critical task to be solved urgently [3, 4, 5, 6].

For the specific signal processing, we first use an appropriate classifier to classify the modulated modes, then select the corresponding demodulator to recover the signal data according to the modulated modes. The traditional classification methods

can be summarized as likelihood-based (LB) and feature-based (FB) [7]. The former is based on the theory of probability and hypothesis testing, and the latter selects the statistical feature that reflects the difference over modulated modes manually. There are primarily the following methods: parameter statistical feature method, high-order analytical feature method, constellation diagram method, wavelet transform feature method, cyclic spectrum feature method, etc. [8, 9]. However, since the implementation of manual operations, it is challenging to design an end-to-end classification system. Furthermore, in the environment with a low signal-to-noise ratio (SNR), these traditional methods are arduous to play a useful role.

With the emergence of big data and the increase in computing capacity, DL has become an essential tool of artificial intelligence technology and has been widely used in face recognition, speech recognition, natural language processing, search recommendation, game confrontation, etc. [10, 11]. For AMC, DL-based methods utilize the self-organization and self-learning mechanism of neural networks to extract the wireless signal feature and fit feature space using nonlinear DL methods. Finally, the back-end classifier decides to complete the classification and recognition of the modulated signal.

With the help of dominant computing power, the deep network learns complex feature representations from the complex data, and for wireless signal data similarly. The DL-based modulated modes classification can divide into two categories: statistical features based that use the DL model to classify, and the classification process is complicated. The other one extracts features by the convolutional neural networks (CNNs) and uses the DL network to classify the signal data automatically. The classification of modulated modes has some differences compared to image classification. Digital modulated classification locates at the receiving end of a digital communication system, which is different from natural image classification, and the wireless modulated signal may face many uncertain channel interferences during transmissions, such as frequency offset and thermal noise. Eliminating the influence of these noises and improving the accuracy of AMC is a complicated process, and many researchers are trying to solve these issues in AMC.

Our method utilizes a fusion layer to measure the importance and compatibility of the primary classifier. In backpropagation training, the weights associated with the classifier are learned to achieve the best combination of basic classifiers in the set. Besides, ensemble learning can integrate heterogeneous models through backpropagation, thereby taking advantage of multiple features obtained through assign weights to classifiers successfully. Ensemble learning tries to combine various weak supervision models, and achieves more comprehensive and powerful supervision models. The basic idea of ensemble learning is that even if a weak classifier gets a wrong prediction, other weak classifiers can correct them. In this paper, we integrate the DL model to enhance the feature extraction ability of the model and finally achieve the purpose of improving the classification accuracy. Our contributions are as follows:

(1) We applied ensemble learning to the modulated modes classification issues. Through the combination of multiple integration methods, an end-to-end ensemble learning classifier is obtained. This method can provide abundant features and improve the problem of poor adaptability under low SNR and low sampling rates.

(2) Specifically, for the AMC task, we combined two by two based on the baseline model and exploring the classification ability under various conditions. The comparative experiment found that the deep ensemble method achieves high classification recognition accuracy and stability.

(3) Through comparative experiments, analyze the performance of the model by comparing the experimental indicators such as accuracy, recall, and F1 score. The experimental results show that the ensemble learning method can effectively improve the classification accuracy and stability of the model. By comparing the classification capabilities under different signal sampling lengths, we found that when the signal length is shorter, it is about 10% higher than the single baseline model. Through the t-SNE graph, the intermediate output process of the classifier is obtained, and it is found that the integrated model has a strong clustering ability.

2 Related Work

In this section, we will review some recent relevant literature on specific DL-based signal classification methods and ensemble learning algorithms.

2.1 DL-Based Methods

With the widespread application of CR, DL-based AMC has attracted much attention. In [12], the author proposed a modulation recognition algorithm based on the constellation and subtractive clustering. Xing Z. proposed a modulated modes recognition algorithm to classify MPSK based on cyclic spectral peak positions and compressed sensing [13]. In [14], the author offered a modulated recognition algorithm based on wavelet transform features, compressed sensing, and support vector machines (SVM). After compressed sensing processing, the wavelet transform features have a lower dimension while retaining all the feature information. In [15], the author proposed an algorithm based on entropy features and random forests, which brought in a higher recognition rate and lower temporal complexity. Timothy J. O' shea directly applied the method of the deep neural network to modulation signal recognition, designed an end-to-end recognition network, and used the powerful fitting ability of the neural network to improve the classification accuracy and generalization ability significantly [16].

Yun Lin used the neural network, Softmax, KNN, and deep self-classifier to classify various modulated modes such as MASK, MPSK, MFSK, MSK, MQAM, and made the promising results [17]. Krishna Karra experimented on the signal dataset from [17], and used a hierarchical deep neural network (H-DNN) to complete the recognition tasks of signal data type, modulated type, and modulated order, and achieved good recognition effect [18]. Yu Wang found that the recognition rate of DL models is limited by high-order QAM and designed a front-back neural network [19]. Former CNN focuses on the inter-class classification of modulated modes, and latter CNN focuses on the intra-class recognition to solve the high-order QAM classification issue. To further eliminate the effect of low SNRs, Keerthi et al. developed a CNN network based on the research of [16], and added the dropout layer, auxiliary information, and adopted hybrid learning to classify 11 modulated modes. The simulation results showed that the recognition rate could reach more than 90% under the SNR is -2dB [20].

2.2 Ensemble Learning Methods

Ensemble learning methods are widely used for dealing with multiclass classification problems, and improve overall classification accuracy by making features work better and promoting each model [21]. A host of algorithms are built based on the idea of fusion, including Boosting, Random Forest, XGBoost [22], GBDT, etc., which can be divided into data layer fusion, feature layer fusion, and decision layer fusion [23, 24]. The construction of the base model and the selection of the fusion method is a critical step, which ensure the participating models achieve better performance. Suitable fusion methods improve the effect of integration effectively. Decision fusion methods include label output-based fusion methods, probability output-based fusion methods, and bayesian fusion methods. Yue Zhao made the model combination easier by designing the combo toolkit [25].

Deep feature-level fusion schemes have an excellent capability of data representations for improving the performance of AMC, and a host of researchers have done work in this area. For instance, Karpathy proposed various fusion schemes to process 117 types of features [26]. Sun used a restricted Boltzmann machine to perform inference from complementary high-level features [27]. Ng designed an end-to-end framework using the long short-term memory (LSTM) to process the features extracted from video frames by CNN [28]. Those feature-level fusion methods for the following processing achieved promising classification results. For ensemble learning in AMC tasks, one has to consider not only the information in the spatial domain but also the complex associations in the temporal domain. However, due to the fusion of spatial and temporal features, it would unavoidably induce high-dimensional representations, which raise significant challenges for the efficiency of AMC.

3 Methods

3.1 Ensemble model

In general, different DL models have distinct structures under different application scenarios, and their recognition capabilities are varied [29]. So far, there is no universal artificial intelligence model that has superior performance in various contexts. To enhance the stability of the model and combine the predictive capabilities of various features, the idea of the ensemble method appeared naturally. The idea is to make full use of the advantages of different algorithms, fuse multiple features, learn from each other, and make a comprehensive judgment to form a robust algorithm framework with strong adaptability. Simultaneously, these features have complementary advantages, which improve the generalization ability, and have superior stability.

We consider combining the feature extraction layer of the Inception module with CLDNN to extract powerful feature expressions. Overall, we design an end-to-end structure from signal data input to feature extraction processing, feature fusion, and final decision output. In Fig. 1, we showed the overall structure of the ensemble learning model. At the input layer, the signal data enters the two models separately. After processing by the convolutional layer, LSTM layer, and fully connected (FC) layer of the two branch models, obtain the features lastly. Then cascading the output of these two branch models together, and the concatenation function can be expressed as $Z = C(f_{m1}(z_1), f_{m2}(z_2))$, where z_1 is the output of one of the models

in the upper layer, z_2 is the other one. f_{m1} and f_{m2} are branch models, C is the cascading stitching operation, and Z is the result of the cascade. C is the input of the last FC layer, and the output is calculated and determined by the activation function $Softmax$. The following sections will introduce the details of the single model under the deep ensemble model.

3.2 Inception module

Convolution kernels of different shapes and sizes in the CNN structure can extract various features. According to this principle, Google first proposed GoogLeNet, which is composed of the Inception module. Its main idea is to use different sizes of convolution kernels to perform convolution calculations and obtain various scale-space features. It uses a 1*1 convolution kernel to reduce the dimension and computational complexity. The Inception module utilizes the principle of sparse matrix decomposition into dense matrix calculations to speed up the convergence speed. Then the feature vectors obtained by the convolution are fused to improve the feature perception ability. These fused features can significantly improve classification performance.

Similarly, in the automatic classification task of wireless signals, multi-scale receptive fields are required to obtain attribute information such as phase and frequency characteristics of signal data. We introduce the main idea of the Inception module and design a feature extraction module for signal data. According to the two-dimensional image convolution model, different convolution kernel scales are improved and develop the one-dimensional convolution kernels for signal data. As shown in Fig. 2, for the Inception module details, it contains four different convolution kernels and a maximum pooling layer. The sizes of the convolution kernels are 1*1, 1*3, 1*5, 1*8, and the kernel size of the Maxpooling layer is 1*2, which can reduce the deviation of the estimated mean caused by the parameter error of the convolution layer. Finally, the three-way feature stack is accumulated and fused as the output of the Inception module.

3.3 CLDNN Module

The CLDNN model performs prominently on automatic speech recognition (ASR). Through the unique module design, the compelling features are related to improving the model classification ability. Its structure contains the following functions: CNN can reduce frequency offset changes, LSTM can provide long-term memory, and DNN can non-linearly map features to an abstract space for effective separation. Recurrent Neural Network (RNN) is heavily used for learning persistent features from temporal series data. LSTM is a particular type of RNN which is efficient in learning long-term dependencies. The time-domain convolution of input features is added to CNN to reduce variance changes further, and then use associations in the temporal domain. In ASR, the speaker ID can be output in combination with the speaker depth feature. Whether DNN or LSTM, it does perform very well in specific effects, but it cannot perform well in all tasks of speech processing. Generally, each neural network has its advantages, and CLDNN is a combination of benefits.

For data with sequence attributes, the relationship between the current time and the time before-after is a powerful feature, and making full use of this feature can

enhance the recognition ability of the model. Similarly, communication signals have temporal attributes, and we introduce a feature extraction layer of the CLDNN network structure. For the CLDNN structure we used, the main modules are four convolutional layers, two LSTM layers [30]. LSTM cells have an internal state (ct) along with three gates, and the block diagram of the LSTM cell is presented in Fig. 3, along with the corresponding equations.

Gates.

$$
\begin{aligned}
i_t &= \sigma(W_i[x_t, h_{t-1}] + b_i) \\
f_t &= \sigma(W_f[x_t, h_{t-1}] + b_f) \\
o_t &= \sigma(W_o[x_t, h_{t-1}] + b_o)
\end{aligned}
\tag{1}
$$

Input transform.

$$
\tilde{c}_t = tanh(W_c[x_t, h_{t-1}] + b_c
\tag{2}
$$

State update.

$$
\begin{aligned}
c_t &= f_t \cdot c_{t-1} + i_t \cdot \tilde{c}_t \\
h_t &= o_t \cdot tanh(c_t)
\end{aligned}
\tag{3}
$$

where t is the current moment, $t-1$ is the previous moment. f_t is forget gate, i_t is input gate, o_t is output gate, c_t is the cell state, $\tilde{c}_t$ is the cell state candidate value, h_t is the hidden layer state value. Activation function σ is defined as $\sigma(x) = 1/(1+e^{-x})$, and $tanh(x) = (e^x - e^{-x})/(e^x + e^{-x})$. W and b are the weight and bias, and the different subscripts are used to distinguish. LSTM controls information through these structures. Firstly, the forget gate combines the previously hidden layer state value h_{t-1} and the current input x_t, and decides to discard specific old information through the sigmoid function σ. Secondly, the input gate and $tanh$ decide what new information is saved from the hidden layer activation value h_{t-1} and the current input value x_t at the previous moment, and obtain the candidate value $\tilde{c}_t$. Then, combining the forget gate and the input gate to discard and saving the information to obtain the cell state c_t at the current moment. Finally, the output gate combine with $tanh$ decides which information in h_{t-1} and x_t, c_t is output as the hidden layer states h_t at this moment. LSTM discard, maintain and update information through a series of gating structures. As the final result of h_t is obtained by multiple functions, and combined with the sum operation, it is not easy to produce gradient disappearance during backpropagation.

To solve the issue of gradient disappearance during model training, and increase the sparsity to facilitate calculation. The scaled exponential linear units (SELU) is selected as the activation function for all available layers except the last fully connected layer.

$$
f_{selu}(z)_j = \lambda \begin{cases} z, & z > 0 \\ \alpha e^z - \alpha, & z \le 0 \end{cases}
\tag{4}
$$

where z is the calculated result of the previous layer, α is the hyperparameter. We use $Softmax$ as the final decision output of the fully connected layer, which is defined as follow,

$$f_{Softmax}(z_j) = \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}} \tag{5}$$

where z_j represents the j-th input value of the $Softmax$ layer. Then we can take it as a probability, and the classification decision result can be judged according to the probability. We use cross-entropy as the loss function, which can speed up model convergence, and it is defined as follows,

$$H(p,q) = -\sum_{i=1}^{N} p(x_i)log(q(x_i)) \tag{6}$$

where $p(x_i)$ is the true probability distribution and $q(x_i)$ is the predict probability distribution.

4 Experimental

4.1 Datasets and Setting

To verify the effectiveness of the proposed model, we use the RadioML2018.10a as experimental data to complete the classification task and evaluate the DL model performance. RadioML2018.10a is a measured modulated signals dataset and released by the O'Shea team in [16]. Compared with the previous data set, it expended to 24 categories. The authenticity is closer to the signal data applied in the actual scene, which can be expressed as:

$$r(k) = s(k) + n(k) = he^{j(2\pi fk+\theta)}x(k) + n(k) \tag{7}$$

where $s(k)$ is the discrete wireless signal samples, $n(k)$ is the noise included in the measured dataset and it includes carrier frequency offset (CFO), symbol rate offset (SRO), delay spread, and thermal noise. These noises are common interferences of measured wireless signal data in channel propagation. k is the sample index for a sampling frequency f_s. h is the signal amplitude, $x(k)$ denotes the k-th symbol generated from constellation with unit average power. The SNR of the signal is defined as $SNR = 10log_{10}h^2/2\sigma^2$ (from -20dB to + 30dB), where h^2 is the signal power and σ^2 is the noise power. The 24 categories included in this data set are 32PSK, 16APSK, 32QAM, FM, GMSK, 32APSK, OQPSK, 8ASK, BPSK, 8PSK, AM-SSB-SC, 4ASK, 16PSK, 64APSK, 128QAM, 128APSK, AM-DSB-SC, AM-SSB-WC, 64QAM, QPSK, 256QAM, AM-DSB-WC, OOK, 16QAM, which represented by IQ channels, and the length of each channel is 1024. The data of each signal is stored in a [2, 1024] matrix form. All implementations of the network are based on Keras with Tensorflow as the backend. GPU composed of one NVIDIA GeForce RTX 2080Ti is applied for training and testing networks.

The appropriate performance metric is indispensable for evaluating the performance of wireless signal classification. The overall accuracy cannot meticulously

evaluate the performance of classification. In addition, we introduce some performance metrics for comprehensively evaluating the performance of classifying tasks. False positives (FP), false negatives (FN), and true positives (TP) are often utilized for assessing the performance of the machine learning algorithms. Positive and negative represent the judgment results of the model, and true and false evaluate whether the judgment result of the model is correct. On the basis, precision, recall, and F1-score are calculated for the task. Precision means the proportion of actual cases (TP) for all positive cases (TP + FP) determined by the model. For all positive cases (TP + FN) in the dataset, the proportion of positive instances (TP) correctly judged by the model to all positive cases is the recall. F1–score is the harmonic mean of precision and recall. They are defined as follows.

$$Precision = \frac{TP}{TP + FP} \tag{8}$$

$$Recall = \frac{TP}{TP + FN} \tag{9}$$

$$F1 - score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{10}$$

4.2 Baselines

In some recent studies, a host of AMC methods based on DL structures have been proposed, and these deep models have excellent performance. In the paper [16], the author offers two AMC methods based on DL structure: VGGNet and ResNet. The author conducted experiments on the measured data set RadioML2018a. From the experimental results, ResNet achieved the best results. In this experiment, we take the four single model of e.t., VGGNet, ResNet, CLDNN, and Inception, as the baseline as the baseline and compare them with the deep ensemble learning model.

The VGGNet has shown in Fig. 4(a). The model is composed of seven convolutional blocks and two fully connected layers. Each convolution block contains a convolutional layer and a maximum pooling layer. Since the signal data is high-dimensional, and directly using the fully connected layer network structure to fit such high-dimensional data has a massive redundancy, it can be solved by the convolution layer and the maximum pooling layer. The convolutional layer learns the relationship between several adjacent values. The maximum pooling can downsample, reduce dimensionality, remove redundant information, compress features, and simplify network complexity. Finally, the fully connected layer learns the deep spatial characteristics of the low-dimensional modulated signal data through the powerful fitting ability. The $Softmax$ function determines the decision result by the calculated probability value, and output the final category.

With the deepening of the neural network layer, various problems in the network training cause the model not to converge ideally. Generally, the parameter initialization of each convolutional layer is close to 0. The training process is to update the weight parameters of each layer continuously. However, the gradient of the model disappears as the network layer deepens so that the superficial parameters cannot be updated. Redundant layers in the deep network make backpropagation arduous,

Table 1 Comparison of classification accuracy of algorithms under different SNRs

Methods	-12dB	-6dB	0dB	6dB	12dB	18dB	24dB	30dB
CLDNN	0.082	0.25	0.478	0.803	0.93	0.935	0.937	0.934
VGGNet	0.098	0.268	0.538	0.834	0.932	0.936	0.934	0.935
ResNet	0.068	0.2	0.475	0.835	0.963	0.965	0.965	0.966
Inception	0.082	0.239	0.53	0.839	0.943	0.951	0.95	0.95
ResNet+CLDNN	0.084	0.25	0.523	0.847	0.97	0.97	0.97	0.97
VGGNet+CLDNN	**0.1**	**0.274**	**0.551**	0.886	0.96	0.962	0.963	0.962
ResNet+VGGNet	0.094	0.233	0.541	0.887	0.96	0.962	0.962	0.962
ResNet+Inception	0.083	0.209	0.52	0.853	0.969	0.97	0.97	0.969
VGGNet+Inception	0.095	0.249	0.545	0.85	0.939	0.943	0.942	0.944
Our model	0.096	0.261	0.542	**0.896**	**0.971**	**0.976**	**0.976**	**0.975**

and the network is expected to reduce the impact of these unnecessary layers. By adding identity mapping to ensure that the input and output of this layer are the same, and reduce the impact of the redundant layer. The main idea of ResNet is to add a direct connection channel in the network. The former network structure is a nonlinear transformation of the performance input, while the directly connected channel allows retaining a certain proportion of the output of the previous network layer. The first one undergoes convolution processing through two weight layers to obtain $f(x)$. The second one maps directly without any processing. Finally, we add the output values of the two channels, and then calculate the activation function to get the result $relu(f(x) + x)$, where $relu$ is defined as $relu(x) = max(0, x)$. $relu$ makes the network sparse and improves computing efficiency. Fig. 4(b) shown the details of ResNet. The residual stack is composed of a linear convolution layer and a Maxpooling layer. Overall, the network consists of six residual blocks and two fully connected layers. The activation function in the fully connected layer is $selu$, which can help maintain a larger gradient optimization.

Table 2 Comparison of multi-model with baseline CNNs

Methods	Accuracy	Precision	Recall	F1
CLDNN	0.585	0.65	0.6	0.6
VGGNet	0.6	0.65	0.6	0.6
ResNet	0.594	0.68	0.6	0.62
Inception	0.6	0.65	0.62	0.62
ResNet+CLDNN	0.611	0.68	0.62	0.64
VGGNet+CLDNN	0.606	0.67	0.6	0.62
ResNet+VGGNet	0.613	0.69	0.61	0.63
ResNet+Inception	0.605	0.68	0.62	0.64
VGGNet+Inception	0.606	0.69	0.63	0.64
Our model	**0.62**	**0.7**	**0.62**	**0.64**

5 Results and Discussion

We compare the deep ensemble model with other networks are shown in Table 1. The baseline models used in the experiments are CLDNN, VGGNet, ResNet, and Inception. Besides, we integrated these four baseline models and compared them to each other to obtain conclusions. We can learn that the performance of these proposed models is better than single models and other ensemble models, especially in the low SNR part, from 0dB to 10dB. When it compares to the Resnet, the best promotion is back up to 8.1%. The experimental results demonstrate that the deep ensemble model is able to apply to different types of communication data and has a stable and compelling performance.

In Table 1, the experimental results of the deep ensemble learning model and the single model under different SNRs are presented. The results demonstrate that compared with the classification accuracy of a single model, the deep ensemble learning model has an excellent performance improvement, and the classification accuracy under each SNR is improved by 0.2% to 8%. In detail, the experimental results of different fusion ways are different. Under the SNR of -12dB and -6dB, the classification result of the VGGNet and CLDNN integrated model is the optimal performance. In the SNR exceeds 10dB, The ensemble method of Inception and CLDNN achieves the best classification accuracy. The existence of these differences depends on the diversity and collaboration of the features of the fusion model combination. The presence of these differences depends on the variety and cooperation of the fusion model combination. Overall, the deep ensemble model we proposed achieved the highest accuracy of the measured modulated modes classification. The reason for our model improvement is that the model integrates the powerful spatial feature extraction function Inception and the better temporal representation of CLDNN. By combining these models, rich features are merged. In the face of various channel noises and physical differences between the transmitter and receiver, the model still has a robust performance.

In Fig. 5, the confusion matrix of our model from 0db to 30dB is shown partly. It can be seen that the classification accuracy of 24 different measured signals increases with the increase of SNR. In an environment with the SNR exceeds 10dB, our model classification accuracy can reach 97.6%. These matrices demonstrate that the high-order QAM is more confusing than other categories. Since high-order QAM are more susceptible to interference from noise signals in IQ data representation, making high-dimensional data indistinguishable in low-dimensional projections. Secondly, the error mainly comes from the confusion between AM-SSB-WC and AM-SSB-SC. After a discrete sampling of the analog signal, the DL model has a specific error when classifying the analog signal in this scenario, which indicates more effective analog signal characteristics need to be obtained.

As shown in Fig. 6, we compare the performance of deep ensemble learning networks under different signal sampling lengths (1024, 512, and 256). The results of the experiment demonstrate our model works best under various conditions, and the accuracy of the deep ensemble learning model at different sampling lengths increase from 0.2% to 12%. Among them, when the sampling length is 256, the deep ensemble learning model generally improves the accuracy by about 10%, which benefits from the deep ensemble model can extract more abundant signal features. Besides, with the increase of signal length, the classification accuracy is also improved. It proves that only with ample features can help the model perform better. Thus , when the signal adopts a shorter length, the rich features extracted in our model can help the classifier keep the classification ability.

In Table 2, the average accuracy, precision, recall, and F1-score of the signal data at all SNRs are compared by the four single models with the deep ensemble learning model. Experimental results verified by different classification criteria prove that the method of integrating multiple deep features has better classification results than a single DL model. The overall classification accuracy is improved by 4%. F1-score is the overall performance of the model's stability and accuracy. The comparison of

F1-score demonstrates that the overall performance of the deep ensemble model is compelling improved.

t-SNE is a method of data reduction and visualization by transforming the similarity between data points into probability. It reduces the high-dimensional data to 2-3 dimensions, and finally displays the distribution in low-dimensional coordinates. Although other data dimensionality reduction and visualization methods, like Isomap, LLE, variants, are more suitable for developing a single continuous low-dimensional manifold. By using t-SNE, the similarity between samples can be displayed accurately. In this paper, we utilize t-SNE for assessing the separability of data and act on the output of the last fully connected layer. Fig. 7(a) and Fig. 7(b) show the t-SNE distribution of Inception and CLDNN, which is the component of our model. Fig. 7(c) shows the t-SNE distribution of Resnet with the best performance in a single model, while Fig. 7(d) represents the distribution of the deep ensemble learning network, both of them are in the case of SNR is 30dB. We can see that the model clusters the same kind of signals into a region, while the deep ensemble learning network mainly converts the same category data to a line. The multi-model fusion network divides features more precisely and classifies them more accurately.

6 Conclusion

A single DL model is proposed with a single feature extraction function in this paper. Based on the single DL model, a novel end-to-end deep ensemble learning model is designed to improving AMC stability and classification accuracy. Experimental results validate the effectiveness of the proposed schemes in this paper.

Author details
School of Computer and Information Engineering, Central South University Of Forestry And Technology, Changsha, Hunan.

References
 1. Ettefagh, Y., Moghaddam, M.H., Eghbalian, S.: An adaptive neural network approach for automatic modulation recognition, 1–5 (2017)
 2. Chen, S., Shen, B., Wang, X., Yoo, S.: A strong machine learning classifier and decision stumps based hybrid adaboost classification algorithm for cognitive radios. Sensors **19**, 5077 (2019)
 3. Bouraoui, R., Besbes, H.: Cooperative spectrum sensing for cognitive radio networks: Fusion rules performance analysis, 493–498 (2016)
 4. Dobre, O.A., Abdi, A., Bar-Ness, Y., Su, W.: Survey of automatic modulation classification techniques: classical approaches and new trends. IET Communications **1**(2), 137–156 (2007)
 5. Taira, S., Murakami, E.: Automatic classification of analogue modulation signals by statistical parameters, 202–207 (1999)
 6. Sun, X., Su, S., Zuo, Z., Guo, X., Tan, X.: Modulation classification using compressed sensing and decision tree–support vector machine in cognitive radio system. Sensors **20**, 1438 (2020)
 7. Huang, S., Yao, Y., Wei, Z., Feng, Z., Zhang, P.: Automatic modulation classification of overlapped sources using multiple cumulants. IEEE Transactions on Vehicular Technology **66**(7), 6089–6101 (2017)
 8. Gardner, W.A., Spooner, C.M.: Signal interception: performance advantages of cyclic-feature detectors. IEEE Transactions on Communications **40**(1), 149–159 (1992)
 9. Spooner, C., Gardner, W.: Robust feature detection for signal interception. Communications, IEEE Transactions on **42**, 2165–2173 (1994)
10. Yang, Y., Hou , C., Lang, Y., Guan, D., Huang, D., Xu, J.: Open-set human activity recognition based on micro-doppler signatures. Pattern Recognition **85**, 60–69 (2018)
11. Yang, Y., Hou, C., Lang, Y., Sakamoto, T., He, Y., Xiang, W.: Omnidirectional motion classification with monostatic radar system using micro-doppler signatures. IEEE Transactions on Geoscience and Remote Sensing **58**(5), 3574–3587 (2020)
12. Wang, F., Wang, Y., Chen, X.: Graphic constellations and dbn based automatic modulation classification, 1–5 (2017)
13. Xing, Z., Zhou, J., Ye, J., Yan, J., Zou, L., Wan, Q.: Automatic modulation recognition of psk signals using nonuniform compressive samples based on high order statistics, 611–614 (2014)
14. Ya, T., Lin, Y., Wang, H.: Modulation recognition of digital signal based on deep auto-ancoder network, 256–260 (2017)
15. Zhang, Z., Li, Y., Zhu, X., Lin, Y.: A method for modulation recognition based on entropy features and random forest, 243–246 (2017)
16. O'Shea, T.J., Roy, T., Clancy, T.C.: Over the air deep learning based radio signal classification. IEEE Journal of Selected Topics in Signal Processing **12**(1), 168–179 (2018)
17. Zhu, X., Lin, Y., Dou, Z.: Automatic recognition of communication signal modulation based on neural network, 223–226 (2016)
18. Karra, K., Kuzdeba, S., Petersen, J.: Modulation recognition using hierarchical deep neural networks, 1–3 (2017)
19. Wang, Y., Liu, M., Yang, J., Gui, G.: Data-driven deep learning for automatic modulation recognition in cognitive radios. IEEE Transactions on Vehicular Technology **68**(4), 4074–4077 (2019)
20. Kumar Arumugam, K.S., Kadampot, I.A., Tahmasbi, M., Shah, S., Bloch, M., Pokutta, S.: Modulation recognition using side information and hybrid learning, 1–2 (2017)
21. Polikar, R.: Ensemble machine learning (2009)
22. Chen, T., Guestrin, C.: Xgboost: A scalable tree boosting system (2016)
23. Kittler, J., Hatef, M., Duin, R.P.W., Matas, J.: On combining classifiers. IEEE Transactions on Pattern Analysis and Machine Intelligence **20**(3), 226–239 (1998)
24. Sanderson, C., Paliwal, K.: Identity verification using speech and face information. Digital Signal Processing **14**, 449–480 (2004)
25. Zhao, Y., Wang, X., Cheng, C., Ding, X.: Combining machine learning models using combo library (2019)
26. Karpathy, A., Toderici, G., Shetty, S., Leung, T., Li, F.F.: Large-scale video classification with convolutional neural networks, 1725–1732 (2014)
27. Montavon, G., Orr, B., Müller, K.: Neural networks: Tricks of the trade: Second edition. Lecture Notes in Computer ence, 599–619 (2012)
28. Y H, N., Hausknecht, M., Vijayanarasimhan, S.: Beyond short snippets: Deep networks for video classification[c]. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE (2015)
29. Zhang, D., Ding, W., Zhang, B., Xie, C., Li, H., Liu, C., J.Han: Automatic modulation classification based on deep learning for unmanned aerial vehicles. Sensors **18**, 924 (2018)
30. Rajendran, S., Meert, W., Giustiniano, D., Lenders, V., Pollin, S.: Distributed deep learning models for wireless signal classification with low-cost spectrum sensors. IEEE Transactions on Cognitive Communications and Networking **4**(3), 433–445 (2017)

Figure legends
Figure 1. Deep ensemble learning architecture for AMC.
Figure 2. Inception block in ensemble model.
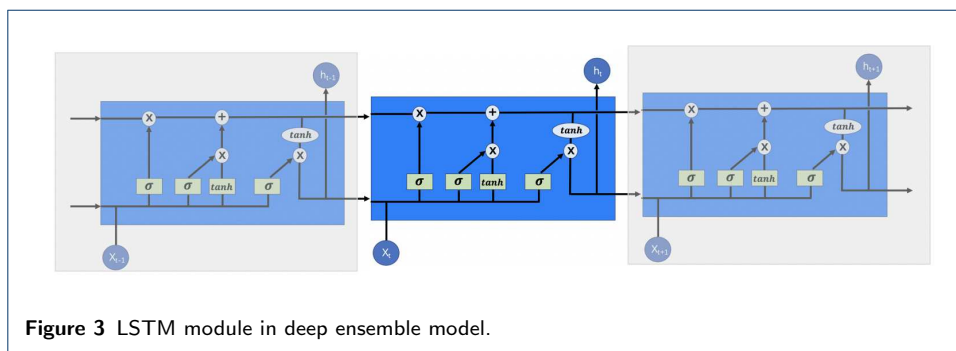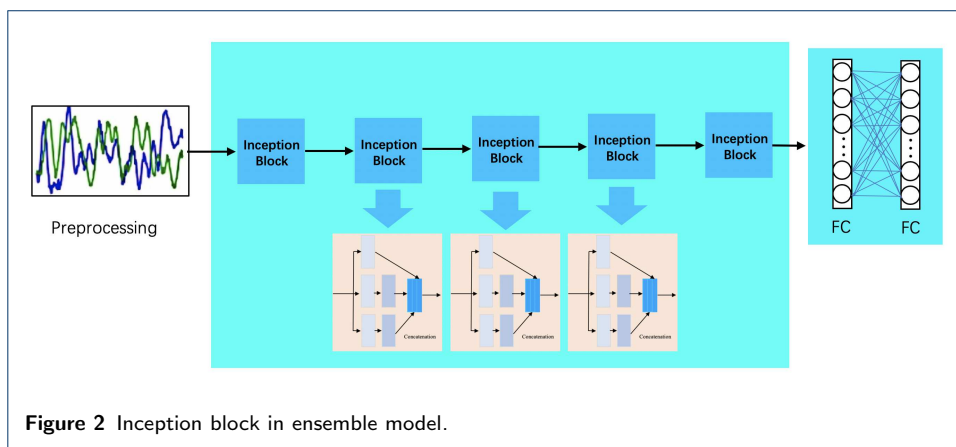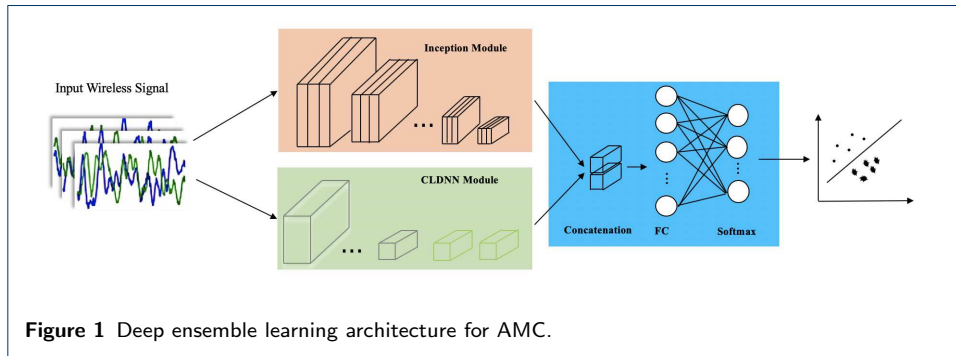Figure 3. LSTM module in deep ensemble model.
Figure 4. (a) is the structure of VGGNet, (b) is the structure of the residual network model.
Figure 5. Confusion matrix of our model in various SNRs. With the sequences:(a) SNR = 0 dB, (b) SNR = 10 dB, (c) SNR = 20 dB, (d) SNR = 30 dB.
Figure 6. The performance of different model versus sample length: (a) 256, (b) 512, (c) 1024.
Figure 7. t-SNE plots of different models with a SNR of 30dB:(a) Inception, (b) CLDNN, (c) ResNet, (d) Our Model.
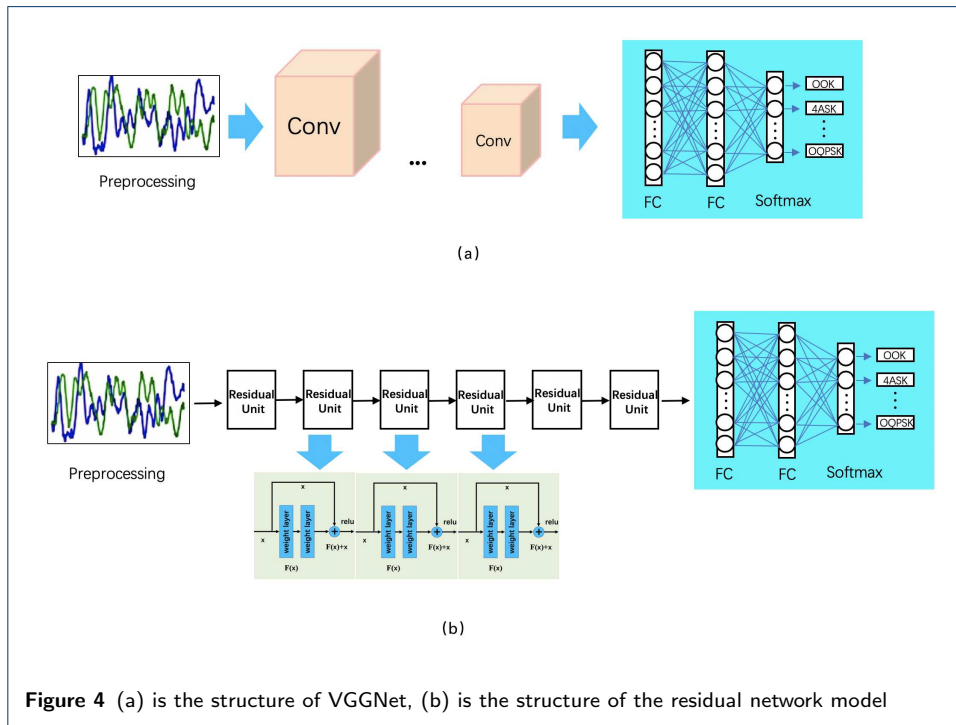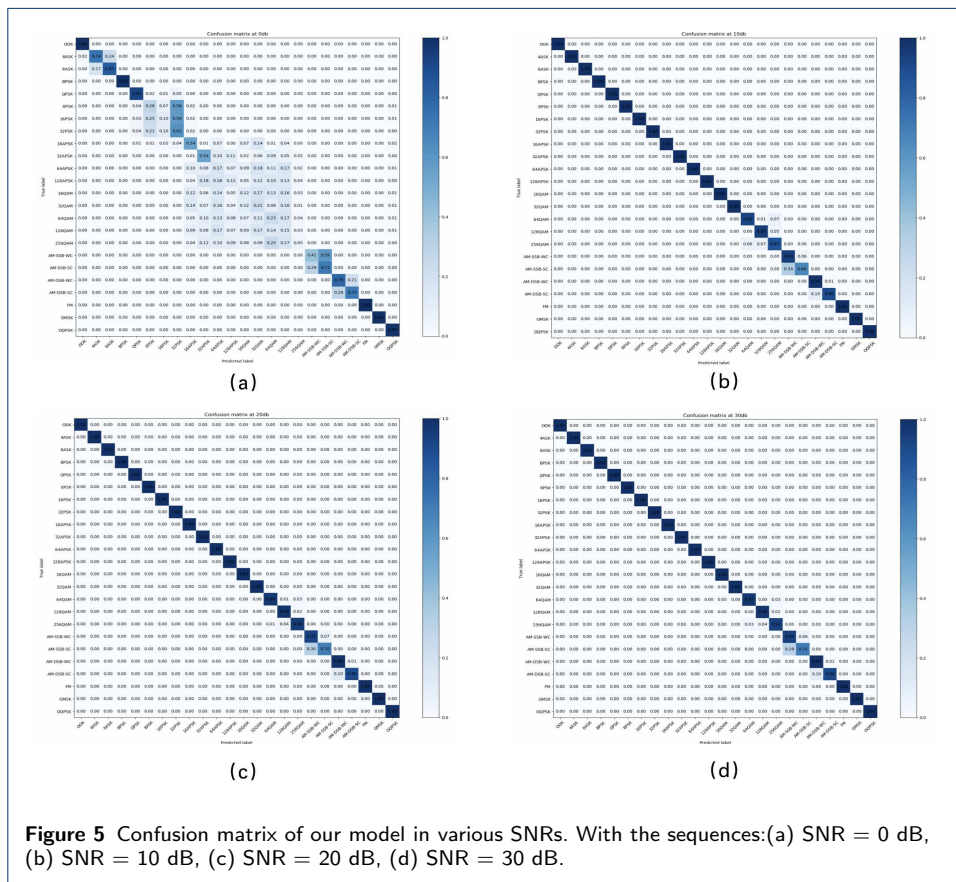
Figures

Figure 1 Deep ensemble learning architecture for AMC.

Figure 2 Inception block in ensemble model.

Figure 3 LSTM module in deep ensemble model.

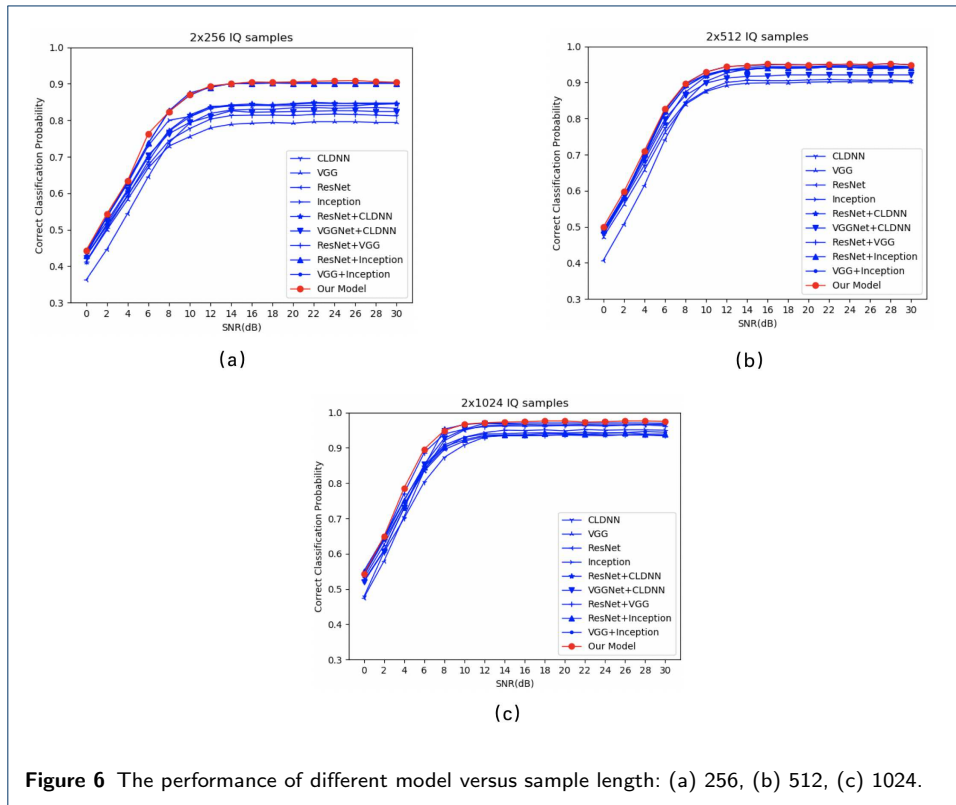**Figure 4** (a) is the structure of VGGNet, (b) is the structure of the residual network model
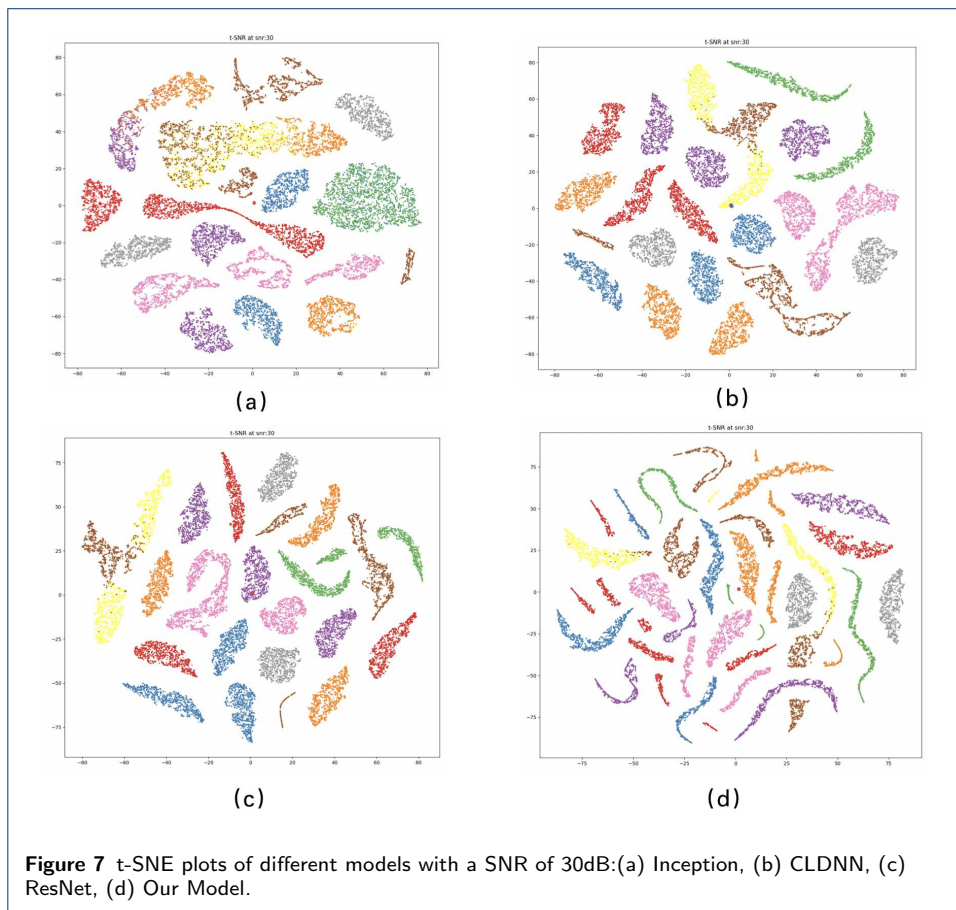

Figure 5 Confusion matrix of our model in various SNRs. With the sequences:(a) SNR = 0 dB, (b) SNR = 10 dB, (c) SNR = 20 dB, (d) SNR = 30 dB.

Figure 6 The performance of different model versus sample length: (a) 256, (b) 512, (c) 1024.

Figure 7 t-SNE plots of different models with a SNR of 30dB:(a) Inception, (b) CLDNN, (c) ResNet, (d) Our Model.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- bmcmathphys.bst
- bmcarticle.bib
- bmcartbiblio.sty
- bmcart.cls
- spbasic.bst
- vancouver.bst