A Kind of Wireless Modulation Recognition Method Based on DenseNet and BLSTM

XIAOSONG XIE¹, GUANGSONG YANG², MENGXI JIANG¹,², QIUBO YE¹, AND CHEN-FU YANG³,⁴ (Member, IEEE)
¹School of Information Engineering, Jimei University, Xiamen 361021, China
²School of Information Science and Technology, Xiamen University, Xiamen 361005, China
³Department of Chemical and Materials Engineering, National University of Kaohsiung, Kaohsiung 811, Taiwan
⁴Department of Aeronautical Engineering, Chaoyang University of Technology, Taichung 413, Taiwan
Corresponding authors: Qiubo Ye (qbye@jmu.edu.cn) and Cheng-Fu Yang (cfyang@nuk.edu.tw)

This work was supported by the Natural Science Foundation of Fujian Province under Grant 2021J01133207.

ABSTRACT Deep learning has achieved remarkable results in various fields, such as image recognition and classification. However, in the recognition of radio modulation methods, deep learning for different modulation methods of radio signal recognition results are not satisfactory. In this paper, we propose to use densely connected convolutional networks combined with bidirectional recurrent neural networks to identify the radios of 11 different modulation methods. The final results show that our method is more accurate than the traditional convolution neural network in modulation recognition.

INDEX TERMS Radio modulation recognition, deep learning, densely connected convolutional networks.

I. INTRODUCTION

With the development of communication technology, wireless communication environment is becoming more and more complex. There are many communication signals with different modulation methods in the air, and how to identify and monitor the modulation of these unknown signals arouses much interest. Modulation recognition technology has been widely used in military and civil fields. For example, we can monitor enemy’s communication radio stations by recognizing modulated signals in battle. We also can use modulation recognition for interference identification and monitoring spectrum in our daily life.

Signal modulation recognition can be divided into two categories: one is based on likelihood ratio decision theory [1], and the other is based on statistical pattern recognition [2]. The former has higher classification accuracy but higher computational complexity [3]. The latter has lower computational complexity and is easy to calculate and implement, but has lower classification accuracy [4].

Deep learning technology has been widely used in image processing [5] and natural language processing [6], [7]. In recent years, it has also been applied to radio modulation recognition. In [8], one-dimensional radio signal was transformed into spectrum image by STFT (Short Time Fourier Transform) and input into convolutional neural networks (CNN). In [9], the authors used phase and quadrature (IQ) samples to train CNN, and obtained good results in low SNR QAM signal classification. The long-term dependent features of LSTM are helpful to the feature extraction of wireless signals. However, with the increase of neural network layers, the number of parameters will increase and the recognition time will be prolonged.

To extract more time series features, the methods of broadening the width of neural network or increasing the depth of neural network are used in the above works, but the scale of the neural network is often limited by hardware conditions. In order to extract the deeper features of radio signal after modulation without over fitting, we propose a new network structure for wireless signal recognition, combined with DenseNet (Densely Connected Convolutional Networks) [10] and BLSTM (Binary LSTM). The former can deepen the depth of neural network with the property of anti-over fitting, and the latter can improve the classification accuracy due to its robustness. Finally, we test and prove the effectiveness of our strategy algorithm by using the RadioML2016.10a dataset. [11]

II. RELATED WORKS

A. DENSENET NEURAL NETWORK

DenseNet is a kind of neural network structure proposed in the IEEE Conference on Computer Vision and Pattern Recognition in 2017. It has been widely used in many fields. For example, Zhang et al. have applied it in remote sensing
classification scenes with high classification accuracy [12]. Liang et al. used this technique for predicting glioma genotypes [13], achieving end-to-end training with high accuracy. Yang et al. used this technique for JPEG steganography [14], which reduced the training parameters. Song et al. used this technology to finger vein recognition [15], which had better anti-noise ability. Hou and Zheng used DenseNet neural network for modulation classification [16], but the network depth was still small. This paper expands the depth of DenseNet neural network to take full advantages of the DenseNet neural network.

B. BLSTM

BLSTM [17] is a bi-directional neural network proposed in 2015, and also plays a very important role in many fields. For example, Chen et al. applied this technology to sentence type classification to improve emotional analysis [18]. Experiments have proved that it has good performance. Luo et al. used it for document-level chemical named entity recognition [19], which makes the neural network model more robust. Liang et al. used this technology for facial expression recognition to better learn space-time information [20]. Stymne et al. used it in cross-language pronoun prediction with very high accuracy [21]. Reference [22] applied it to positive blood culture detection, which can better capture space-time information for detection.

C. CNN BASED MODULATION RECOGNITION

Fan et al. used CNN for radio modulation recognition [23], which has good robustness. Rajendran et al. used a two-layer LSTM structure to process and extract time information [24]. West and Tim O’Shea used CLDNN (Compute Library for Deep Neural Networks) [25], which can not only extract features from data of multiple time points and reduce invalid data, but also deal with time-related data.

Zhu et al. used an improved AlexNet neural network [26], which achieved similar performance as the likelihood ratio judgment theory recognition method with strong robustness. Zhang et al. used HetNet neural network [27], which achieved a higher transmission rate without increasing the overhead of the system. Shah and Dang combined DNN and RBFN [28], which improved accuracy compared with K-NearestNeighbor. Peng et al. used GoogleNet neural network [29] proposed by Google, which had a certain improvement in classification accuracy compared with Support Vector Machine neural network. Based on CNN, Guo used the maximum pooled layer to increase the number of convolution layers, which had good robustness and accuracy [30]. Reference [31] combined semi-supervised and CNN network to identify radio modulation modes, which was also robust. Li, et al. used a deep learning network consisting of stacked autocoders and softmax regression for recognition [32], which still had good performances at low signal-to-noise ratio. However, none of the above methods used DenseNet, and the parameters in these training processes will be larger.

In this paper, we propose a new structure combined with DenseNet and BLSTM, and use it for radio modulation recognition, which solves the problem of lower classification accuracy.

III. NEURAL NETWORK MODEL

A. CNN MODEL

CNN has the advantages of local perception, weight sharing and shift invariance, which can spatially share the local correlation of each adjacent layer. A basic hypothesis of CNN is that the input data are localized and shifted unchanged, while the wireless signal sampling data accord with it. The general architecture of the CNN is shown in Figure 1. Each convolution layer will extract different data features from the input data. Finally, the classification function outputs a vector of 1 * p, which represents the probability of P modulation methods for the input data.

![CNN neural network](image1)

**FIGURE 1.** CNN neural network.

Radio modulation recognition can be regarded as a classification problem. We set the input signal as x and the output as y, x is a vector of 1 * m * n, y is a vector of 1 * p, representing p kinds of different modulation methods. A neural network is similar to a huge non-linear function, which can be infinitely near the real model by constantly adjusting the constant parameters of this function.

Figure 2 is an example of the neural network with 4 layers in CNN. $W^l_{jk}$ is the weight of the k-th neuron of layer l-1 connected to the j-th neuron of layer l, $b^l_j$ is the bias of the j-th neuron of layer l, $z^l_j$ is the input of the activation function of the j-th neuron of layer l, and $a^l_j$ is the output of the activation function of the j-th neuron of layer l.

![Neural network layer structure in CNN](image2)

**FIGURE 2.** Neural network layer structure in CNN.

We can express the network structure as follows:

$$z^l_j = \sum_k W^l_{jk} a^{l-1}_k + b^l_j$$  \hspace{1cm} (1)
\[
\alpha_j^l = \sigma \left( \sum_k w_{jk} \alpha_{k}^{l-1} + b_j^l \right)
\]  

(2)

Equation 2 is for calculation layer by layer, where \( \sigma \) is activation function, assuming the output of the last layer is \( y_1 \) to \( y_k \). Then we use the softMax function as the output layer to calculate the probability of each category, and output \( y_1 \) to \( y_p \), which represents the probability of \( p \) kinds of modulation methods. Let \( s_i \) represent the \( i \)th output of the softMax function, which can be expressed as,

\[
s_i = \frac{e^{y_i}}{\sum_k e^{y_k}}
\]

(3)

\[\sum_{i=1}^{k} s_i = 1\]

(4)

where \( e \) is a natural constant.

Cross entropy loss function is used to process the output of softMax, and we can get the loss \( c \) of cross entropy as,

\[
c = \text{crossEntropy}(y, s) = -\sum_{i=1}^{k} y_i \log(s_i)
\]

(5)

The cross entropy is used to measure the difference between two probability distributions. The smaller its value is, the smaller the difference between the real value and the predicted value is, and the closer the probability distribution is. After the loss is obtained, the parameters can be updated.

We first get,

\[
w' = \frac{\partial c}{\partial w} = \frac{1}{n} \sum x(\sigma(z) - y)
\]

(6)

\[
b' = \frac{\partial c}{\partial b} = \frac{1}{n} \sum x(\sigma(z) - y)
\]

(7)

Secondly the parameters \( w \) and \( b \) are updated with \( \alpha \) as the learning rate:

\[
w = w - \alpha \ast w'
\]

(8)

\[
b = b - \alpha \ast b'
\]

(9)

Then the weight bias of the upper layer is updated with the same method, which is the process of CNN back propagation.

### B. DENSENET NEURAL NETWORK MODEL

To solve the problem of gradient vanishing caused by too deep network layers, we introduce the DenseNet neural network, which reduces the parameters by adding bypass multiplexing, so that the layers of the whole neural network are deep enough to extract enough features. Figure 3 shows the neural network structure of DenseNet. As can be seen in Figure 3, the input of each neural network layer comes from the output of all previous layers. The output of \( X_0 \) will be the input of \( H_0 \), the output of \( X_0 \) and \( X_1 \) will be the input of \( H_1 \), and so on. Each neural network layer will extract features from the input data, and the features will become more obvious with the increase of the depth of the layer.

The traditional neural network improves the network performance by deepening the network hierarchy and widening the network structure. However, DenseNet greatly reduces the number of network parameters through feature reuse and bypass setting, and alleviates the problem of gradient vanishing to a certain extent. In the DenseNet neural network, if the output of layer \( i \) is represented by \( X_i \) and the activation function is represented by \( H \), \( X_i \) can be represented by equation (10) below.

\[
X_i = H([X_0, X_1, X_2, \cdots, X_{i-1}])
\]

(10)

DenseNet makes the extracted data more complete through such information reuse, which is also the reason why DenseNet has a better effect in data feature extraction.

### C. BLSTM

The traditional LSTM neural network is widely used to learn persistent features from time series data, which benefits from the special gate structure of LSTM showed in Figure 4.

The core of LSTM is the neuron state, which is the segment \( C_{t-1}C_t \) in Figure 4. LSTM consists of a forgetting gate, input gate and output gate. The forgetting gate \( f_t \) determines which information is discarded or retained in the neuron state according to the information of \( h_{t-1} \) and \( x_t \). The input gate decides which information is updated through \( h_{t-1} \) and \( x_t \). \( h_{t-1} \) and \( x_t \) will get new information through a tanh layer, which may be updated into the new neuron. The output gate decides which information is output to the neuron state according to \( h_{t-1} \) and \( x_t \). These three gates can be expressed from equations (11) to (16), where \( x_t \) represents the input at the \( i \)-th time and \( h_t \) represents the output at the \( i \)-th time.
Forgetting gate:
\[ f_t = \sigma(w_f \cdot [h_{t-1}, x_t] + b_f) \]  
(11)

Input gate:
\[ i_t = \sigma(w_i \cdot [h_{t-1}, x_t] + b_i) \]  
(12)
\[ \tilde{C}_t = \tanh(w_C \cdot [h_{t-1}, x_t] + b_C) \]  
(13)

Output gate:
\[ o_t = \sigma(w_o \cdot [h_{t-1}, x_t] + b_o) \]  
(14)
\[ h_t = o_t \cdot \tanh(C_t) \]  
(15)

Neuron state update:
\[ C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \]  
(16)

The traditional recurrent neural network (RNN) and LSTM neural network can only predict the output of the next time according to the time sequence information of the previous time. However, in some cases, the output of the current time is not only related to the previous state, but also to the future state. The basic idea of bi-directional RNN is that each training sequence has two recurrent neural networks (i.e. forward and backward) which are connected with an output layer. This structure provides some information to the output layer, including past and future context information for each point in the input sequence. Figure 5 shows a bidirectional RNN, in which the six weights (w1-w6) are reused at each time step. In bidirectional RNN, there are two hidden states: one is a positive sequence, as shown in the left to the right solid line in Figure 5; the other is a reverse sequence, as shown in the right to the left solid line in the figure. When calculating the state variables at time t, for positive sequence, it is easy to understand that the input is the hidden state of the previous time and the system input of the current time; for reverse sequence, the input is the hidden state of the next time and the system input of the current time. BLSTM is an optimized LSTM, which can train forward and backward time series respectively, and the output data can obtain context information.

**IV. MODULATION RECOGNITION MODEL BASED ON DENSENET + BLSTM + DNN**

In this section, we propose a neural network structure of DenseNet + BLSTM + DNN shown in Figure 6. The signal data first go through batch regularization and ReLu layer, then go through Dense layer in DenseNet, splice two BLSTM layers, and then use softmax function for classification.

The format of input data is batch, 1, 2, 128, and the parameter value of nb_filter is 10. Then, the data are input to a 1*5 convolution layer, and the output of the convolution layer will be processed by batch normalization. The convolution layer is followed by four DenseBlocks, and a Transition-Layer is used in the middle of each two DenseBlocks. There are 10 DenseLayers in DenseBlock, and the DenseLayers in DenseBlock are connected according to Equation (10). A DenseLayer consists of two convolution layers. Before each convolution layer, batch normalization is used to process the input data. The data flow direction of DenseLayer is shown in Figure 7. In DenseLayer, the input data first go through batch regularization, then the LeakyReLU activation function is used, then a convolution neural network is used, then the batch regularization and LeakyReLU activation functions are used, and then the output is obtained from a convolution neural network.

**V. EXPERIMENTS AND EVALUATION**

**A. CLASSIFICATION ACCURACY**

The experiments were done by using RML2016.10a_dict dataset, to evaluate the classification accuracy of different models such as CLDNN, inception, ResNet in Reference [33]. From the experimental results shown in Figure 10, it can be seen that the proposed scheme (DenseNet + BLSTM + DNN) has a high classification accuracy, approximately 5% more than CLDNN, 10% more than CNN, 15% more than inception, 35% more than ResNet, and slightly higher accuracy at low SNR than other models.

Here, we will analyze the classification accuracy of the proposed scheme and CLDNN for different modulation
modes under different SNR. The confusion matrices when SNR = 18dB and SNR = 0dB are shown in Figure 8 and Figure 9 respectively. It can be seen in the figures that the classification accuracy of AM-SSB, BPSK, CPFSK and GFSK is above 90%, and the accuracy of the proposed scheme is slightly higher than that of CLDNN. The classification accuracy of other modulation modes is shown in Table 1.

The confusion matrices of SNR = −14dB are shown in Figure 11. We can see that neither method can recognize the modulation mode correctly, because it is difficult to recognize the features of noise and signal at low SNR.

**B. THE INFLUENCE OF DENSENET CONVOLUTION KERNEL SIZE**

In the network structure shown in Figure 6, DenseNet can reduce the parameters and increase the network depth level. In this section, we will analyze the influence of different convolution kernel sizes on the classification accuracy of DenseNet, and design four kinds of DenseNet neural network models with different convolution kernel sizes with fixed network structures. The classification accuracy of neural networks is evaluated by using RML2016.10a_diet dataset under different SNR. In this experiment, six convolution kernels of different sizes are designed. Under the same network structure, the convolution kernels of DenseBlock in DenseNet are modified to 1*3, 1*5, 1*7, 1*9, 1*11 and 1*13, respectively.

The number of experiments is 200, epochs are set to 15, and Adam optimizer is used.

Figure 12 shows the classification accuracy after 15 epoch networks convergence. It can be seen that at high SNR, the classification accuracy of 1*5 and 1*9 DenseNet convolution kernels is almost the same, reaching more than 70%, while the classification accuracy of other DenseNet convolution kernels is about 65%. When the SNR decreases from 5dB to −10dB, the difference of classification accuracy of different convolution kernel sizes gradually decreases, and the classification accuracy of 1*5 and 1*9 sizes is still slightly higher.

At low SNR, the classification accuracy of 1*3 convolution
TABLE 1. Classification accuracy of different modulation modes.

| modulation mode | SNR=18dB | SNR=0dB |
|-----------------|----------|---------|
|                 | Proposed Model | CLDNN | Proposed Model | CLDNN |
| AM-SSB          | 0.99      | 0.61    | 0.76          | 0.91   |
| QAM16           | 0.69      | 0.41    | 0.64          | 0.48   |
| QAM64           | 0.74      | 0.53    | 0.69          | 0.46   |
| QPSK            | 0.97      | 0.80    | 0.49          | 0.52   |

kernel is about 20%, while that of 1*5 and 1*9 convolution kernels is about 15%, and that of other convolution kernels is between 10% and 15%.

Figure 13 shows the confusion matrices with SNR = 18dB. We further display the classification accuracy of different convolution kernel sizes in Table 2. From the table, we can see that different convolution kernel sizes lead to different classification accuracy for different modulation modes. Kernel size 1*3 has high accuracy for AM-DSB, kernel size 1*5 has high accuracy for AM-DSB, BPSK, GFSK and PAM4, and kernel size 1*9 has high accuracy for BPSK and CPFSK. Combining the results of Figure 13 with those of Figure 12, it can also be observed that the larger convolution kernel size does not lead to better accuracy. When the convolution kernels are 1*11 and 1*13, the classification accuracy decreases.

C. INFLUENCE OF BLOCK SIZE

The network structure of DenseNet is composed of multiple blocks. In this section, we will discuss the influence of different numbers of blocks on classification accuracy. Four kinds of DenseNet blocks with different sizes are used in the experiment, and the classification accuracy is calculated under different SNR. Epoch is set to 15, the batch size is 200, and convolution kernel size of the dense block is 1 * 9.

Figure 14 shows the comparison of the results after convergence of 15 epoch networks. It can be seen that with the increase of the number of blocks, the classification accuracy
TABLE 2. Classification accuracy of different convolution kernel sizes (SNR = 18dB).

| Kernel size | 1*3 | 1*5 | 1*7 | 1*9 | 1*11 | 1*13 |
|-------------|-----|-----|-----|-----|------|------|
| 8PSK        | 0.67| 0.62| 0.66| 0.67| 0.65 | 0.49 |
| AM-DSB      | 0.67| 0.37| 0.08| 0.32| 0.48 | 0.22 |
| AM-SSB      | 0.65| 0.75| 0.65| 0.65| 0.58 | 0.78 |
| BPSK        | 0.79| 0.98| 0.89| 0.98| 0.97 | 0.83 |
| QPSK        | 0.97| 0.96| 0.98| 1.00| 0.83 | 0.97 |
| GFSK        | 0.94| 0.99| 0.93| 0.91| 0.99 | 0.94 |
| PAM4        | 0.78| 0.93| 0.93| 0.88| 0.93 | 0.89 |
| QAM16       | 0.44| 0.58| 0.26| 0.49| 0.3  | 0.25 |
| QAM64       | 0.55| 0.31| 0.45| 0.45| 0.48 | 0.51 |
| WPBF        | 0.73| 0.76| 0.68| 0.68| 0.51 | 0.86 |

VI. CONCLUSION

In order to solve the problem that the classification accuracy of traditional radio modulation recognition methods is not high, the deep learning method is used to recognize radio modulation modes, and the neural network structure of DenseNet + BLSTM + DNN is proposed. In this combined structure, the signals first pass through a convolution layer and then access to a batch regularization layer and LeakyRelu layer, followed by a DenseNet neural network to deepen the neural network level through the bypass multiplexing of DenseNet. Then, the signals are input to two BLSTM layers to extract the time sequence information, and finally input to softMax for classification. By adding bypass multiplexing to reduce parameters, the depth of neural network is increased, and the deeper information is extracted.

Compared with the existing neural network models, the overall classification accuracy of our proposed scheme is improved by 8%. In addition, through the experimental we can observe that, by changing the convolution kernel size and the number of blocks, different modulation modes have different classification accuracy. In our future work, we will consider changing the convolution kernel size adaptively to recognize different modulation modes.

ACKNOWLEDGMENT

(Xiaosong Xie and Guangsong Yang are co-first authors.)

REFERENCES

[1] T. Toda, A. W. Black, and K. Tokuda, “Voice conversion based on maximum-likelihood estimation of spectral parameter trajectory,” IEEE Trans. Audio, Speech, Language Process., vol. 15, no. 8, pp. 2222–2235, Nov. 2007.
[2] A. K. Jain, R. P. W. Duin, and J. C. Mao, “Statistical pattern recognition: A review,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 22, no. 1, pp. 4–37, Jan. 2000.
[3] T. Nakatani, B. H. Huang, T. Yoshioka, K. Kinoshita, M. Delcroix, and M. Miyoshi, “Speech dereverberation based on maximum-likelihood estimation with time-varying Gaussian source model,” IEEE Trans. Audio, Speech, Language Process., vol. 16, no. 8, pp. 1512–1527, Nov. 2008.
[4] T. Lambrou, P. Kudumakis, R. Speller, M. Sandler, and A. Linney, “Classification of audio signals using statistical features on time and wavelet transform domains,” in Proc. IEEE Int. Conf. Acoust., Speech Signal Process., (ICASSP), vol. 6, May 1998, pp. 3621–3624.
[5] J. Deng, W. Dong, R. Socher, L. J. Li, K. Li, and L. Fei-Fei, “ImageNet: A large-scale hierarchical image database,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2009, pp. 248–255.
[6] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, “Inception-v4, Inception-ResNet and the impact of residual connections on learning,” in Proc. 31st AAAI Conf. Artif. Intell., Feb. 2017, pp. 1–12.
[7] G. E. Dahl, D. Yu, L. Deng, and A. Acero, “Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition,” IEEE Trans. Audio, Speech, Language Process., vol. 20, no. 1, pp. 30–42, Jan. 2012.
[8] Y. Zeng, M. Zhang, F. Han, Y. Gong, and J. Zhang, “Speech recognition and convolutional neural network for automatic modulation recognition,” IEEE Wireless Commun. Lett., vol. 8, no. 3, pp. 929–932, Jun. 2019.
[9] Y. Wang, M. Liu, J. Yang, and G. Gui, “Data-driven deep learning for automatic modulation recognition in cognitive radios,” IEEE Trans. Veh. Technol., vol. 68, no. 4, pp. 4074–4077, Apr. 2019.
[10] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 4700–4708.
[11] T. J. O’Shea and N. West, “Radio machine learning dataset generation with GNU radio,” in Proc. GNU Radio Conf., vol. 1, no. 1, 2016, pp. 1–6.
[12] J. Zhang, C. Lu, X. Li, H.-J. Kim, and J. Wang, “A full convolutional network based on DenseNet for remote sensing scene classification,” *Math. Biosci. Eng.*, vol. 16, no. 5, pp. 3345–3367, 2019.

[13] S. Liang, R. Zhang, D. Liang, T. Song, T. Ai, C. Xia, L. Xia, and Y. Wang, “Multimodal 3D DenseNet for IDH genotype prediction in gliomas,” *Genes*, vol. 9, no. 8, p. 382, Jul. 2018.

[14] J. Yang, Y.-Q. Shi, E. K. Wong, and X. Kang, “JPEG steganalysis based on DenseNet,” 2017, arXiv:1711.09335. [Online]. Available: http://arxiv.org/abs/1711.09335

[15] J. M. Song, W. Kim, and K. R. Park, “Finger-vein recognition based on deep DenseNet using composite image,” *IEEE Access*, vol. 7, pp. 66845–66863, 2019.

[16] T. Hou and Y. Zheng, “Modulation recognition of communication signal based on deep learning,” *Wireless Electr.*, vol. 49, no. 9, pp. 796–800, 2019.

[17] Z. Huang, W. Xu, and K. Yu, “Bidirectional LSTM-CRF models for sequence tagging,” 2015, arXiv:1508.01991. [Online]. Available: http://arxiv.org/abs/1508.01991

[18] T. Chen, R. Xu, Y. He, and X. Wang, “Improving sentiment analysis via sentence type classification using BiLSTM-CRF and CNN,” *Expert Syst. Appl.*, vol. 72, pp. 221–230, Apr. 2017.

[19] L. Luo, Z. Yang, P. Yang, Z. Zhang, L. Wang, H. Lin, and J. Wang, “An attention-based BiLSTM-CRF approach to document-level chemical named entity recognition,” *Bioinformatics*, vol. 34, no. 8, pp. 1381–1388, 2017.

[20] D. Liang, H. Liang, Z. Yu, and Y. Zhang, “Deep convolutional BiLSTM fusion network for facial expression recognition,” *Vis. Comput.*, vol. 36, no. 3, pp. 499–508, Mar. 2020.

[21] S. Stymne, S. Loáiciga, and F. Cap, “A BiLSTM-based system for cross-lingual pronoun prediction,” in Proc. 3rd Workshop Discourse Mach. Transl., 2017, pp. 47–53.

[22] L. De Baets, J. Ryusinick, T. Peiffer, J. Decruyenaere, F. De Turck, F. Ongenae, and T. Dhaene, “Positive blood culture detection in time series data using a BiLSTM network,” 2016, arXiv:1612.00962. [Online]. Available: http://arxiv.org/abs/1612.00962

[23] F. Meng, P. Chen, L. Wu, and X. Wang, “Automatic modulation classification: A deep learning enabled approach,” *IEEE Trans. Veh. Technol.*, vol. 67, no. 11, pp. 10760–10772, Nov. 2018.

[24] S. Rajendran, W. Meert, D. Giustiniano, V. Lenders, and S. Pollin, “Distributed deep learning models for wireless signal classification with distributed low-cost spectrum sensors,” *IEEE Trans. Cogn. Commun. Netw.*, vol. 4, no. 3, pp. 433–445, 2018.

[25] T. N. Sainath, O. Vinyals, A. Senior, and H. Sak, “Convolutional, long short-term memory, fully connected deep neural networks,” in Proc. *IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2015, pp. 4580–4584.

[26] L. Zha, Z. Gao, and Z. Zhu, “Blind modulation classification via accelerated deep learning,” in Proc. *IEEE 5th Int. Conf. Comput. Commun. (ICCOC)*, Dec. 2019, pp. 2102–2107.

[27] L. Zhang, J. Tan, Y. Liang, G. Feng, and D. Niyato, “Deep reinforcement learning-based modulation and coding scheme selection in cognitive heterogeneous networks,” *IEEE Trans. Wireless Commun.*, vol. 18, no. 6, pp. 3281–3294, Jun. 2019.

[28] M. H. Shah and X. Dang, “Low-complexity deep learning and RBFN architectures for modulation classification of space-time block-code (STBC)-MIMO systems,” *Digital Signal Process.*, vol. 99, Apr. 2020, Art. no. 102656.

[29] S. Peng, H. Jiang, H. Wang, A. Hathal, Z. Yu, S. M. Mazrouei, and Y. D. Yao, “Modulation classification based on signal constellation diagrams and deep learning,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 3, pp. 718–727, Mar. 2019.

[30] J. Guo, H. Zhang, J. Xu, and Z. Chen, “Pattern recognition of wireless modulation signals based on deep learning,” in Proc. *IEEE 6th Int. Symp. Electromagn. Comput. (ISEMC)*, Nov. 2019, pp. 1–5.

[31] X. He, L. Lin, and J. Xie, “Radio signals modulation mode recognition based on semisupervised deep learning,” in Proc. *IEEE 2nd Int. Conf. Automat., Electron. Elect. Eng. (AUTEEE)*, Nov. 2019, pp. 330–333.

[32] J. Li, L. Qi, and Y. Lin, “Research on modulation identification of digital signals based on deep learning,” in Proc. *IEEE Int. Conf. Electron. Inf. Commun. Technol. (ICIECT)*, Aug. 2016, pp. 402–405.

[33] N. E. West and T. O’Shea, “Deep architectures for modulation recognition,” in Proc. *IEEE Int. Symp. Dyn. Spec. Access Netw. (DySPAN)*, Mar. 2017, pp. 1–6.

**XIAOSONG XIE** is currently pursuing the degree with the School of Computer Engineering, Jimei University. His research interests include artificial intelligence and modern communication technology.

**GUANGSONG YANG** received the Ph.D. degree from the School of Information Technology, Xiamen University. He was a Visiting Scholar at UC Davis, USA, from 2009 to 2010, and at Griffith University, Australia, from 2017 to 2018. He is currently a Professor with the School of Information Engineering, Jimei University, Xiamen, China. His current research interests include wireless sensor networks and intelligent information processing.

**MENGXI JIANG** received the B.S. degree in computer science and technology from Fuzhou University, China, in 2013, and the M.S. degree in computer technology from Xi’an Polytechnic University, China, in 2016. She is currently pursuing the Ph.D. degree in computer science and technology with Xiamen University. Her research interests include machine learning and computer vision.

**QUIBO YE** received the bachelor’s degree from Hefei University of Technology, China, the master’s degree from the North China Electric Power University, and the Ph.D. degree from the University of Manitoba, Canada, all in electrical engineering. He was a Research Scientist/Project Leader at Communications Research Centre, Canada, a Visiting Assistant Professor with Rose–Hulman Institute of Technology, USA, and a Research and Development Engineer at Zeland Software, Inc., USA. He was an Adjunct Research Professor with the Electronics Department, Carleton University, Canada. He is currently a Chair Professor with the School of Information Engineering, Jimei University, Fujian, China. He has published books, chapters of books, and more than 70 research papers. His research interests include computational electromagnetics, antennas, EMC/EMI, wave propagation, wireless communication, and applications of AI in communications. He has received the Outstanding Engineer Award from IEEE Ottawa and other awards for his outstanding performance as the Chair of the IEEE Ottawa AP/MTT Chapter. He was the Chair of the 2011 IASTED International Conference on Wireless Communication and the 2016 IEEE International Symposium on EMC.

**CHENG-FU YANG** (Member, IEEE) received the B.S., Master’s, and Ph.D. degrees from the Department of Electrical Engineering, Cheng Kung University, in 1986, 1988, and 1993, respectively. In 2004, he became a Professor of chemical and materials engineering with the National University of Kaohsiung (NUK). He obtained “The Outstanding Contribution Award of the Chinese Ceramic Society,” in 2009. In 2010, he was the first (and only) person to become a Distinguished Professor of NUK. He became a fellow of the Taiwanese Institute of Knowledge Innovation (TIKI), in 2014, and the Institution of Engineering and Technology (IET), in 2015. He became a Mingjiang Scholar and an Invited Chair Professor of Jimei University, Xiamen, Fujian, China, and an Honorary Chair Professor of Chaoyang University of Technology, in 2020.