Multimodal Machine Learning Algorithm for Enhanced Signal Modulation Recognition in Wireless Communication Systems

Bojun Zhang
School of Intelligence and Computing, Tianjin University
Tianjin, China
1020244023@tju.edu.cn

2024 IEEE 49th Conference on Local Computer Networks (LCN) | 979-8-3503-8800-8/24/$31.00 ©2024 IEEE | DOI: 10.1109/LCN60385.2024.10639739

***Abstract*—Automatic modulation identification plays an important role in wireless communication systems. With the development of deep learning technology, more and more researches have begun to adopt deep learning methods for modulation identification. However, traditional deep learning methods face interference from nonlinearity, noise, and time-varying nature of the signal when processing the signal, resulting in degradation of classification performance. Meanwhile, single modal features are difficult to fully capture the time and spatial domain information of the signal, which is important for accurately identifying the signal modulation type. To overcome these limitations, this study proposes a multi-modal deep learning based signal modulation recognition scheme. The scheme improves the modulation recognition performance by fusing features from different modalities, including time, frequency and spatial domain features. Experimental results show that the proposed method achieves significant performance improvement in the signal modulation recognition task. This provides a more robust and accurate modulation identification capability for wireless communication systems.**

***Index Terms*—modulation recognition, multimodal, deep learning**

I. Introduction

With the rapid development and wide application of wireless communications, automatic modulation identification has become an important task. As shown in Figure 1,automatic modulation identification is the process of determining the type of modulation it employs based on the received signal. Automatic modulation identification has a wide range of applications in wireless communication systems. First, accurate modulation identification can help the receiving end to correctly demodulate the signal, thus realizing high-quality data transmission. Second, modulation recognition can be used in radio spectrum monitoring and management to help monitor the type and distribution of radio signals to optimize the use of spectrum resources. In addition, in the field of radio communication security, modulation identification can be used to detect and identify potentially interfering signals and illegal signals. Conventional signal modulation recognition methods usually use manually designed features and classifiers based [1]–[3], but these methods have some challenges when dealing with complex wireless signals.

In recent years, deep learning has made significant progress in various fields, including computer vision and natural language processing. However, in the field of signal modulation recognition, traditional deep learning methods face some problems in processing signals. The method [4] of directly feeding the signal into the neural network for classification may be disturbed by the nonlinearity, noise, and time-varying nature of the signal, which leads to a degradation of the classification performance. In addition, there are some limitations in converting signals into spectrograms [5] and using convolutional neural networks for classification because spectrograms do not fully capture the time and spatial domain information of the signal, which is important for accurately identifying the type of signal modulation.

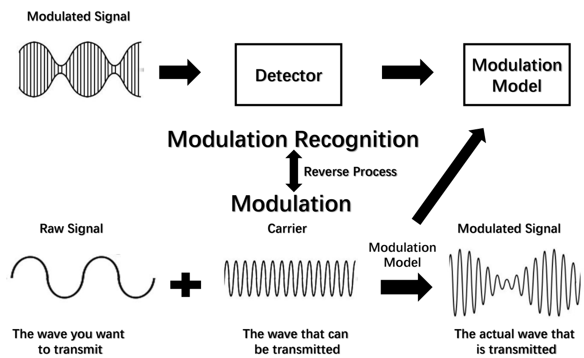


Fig. 1. Signal Modulation Recognition Schematic

In order to overcome the limitations of traditional deep learning methods and to improve the performance of signal modulation recognition, the introduction of multimodal deep learning becomes an effective approach. Multimodal deep learning utilizes the features of multiple data modalities (e.g., time, frequency, and spatial domains, etc.) for signal modulation recognition, thus capturing the spatio-temporal features of signals more comprehensively and improving recognition accuracy.

In this study, we propose a signal modulation recognition scheme based on multimodal deep learning, aiming to address the limitations of traditional deep learning methods in signal modulation recognition. We obtain more comprehensive signal information by employing features from multiple data modalities, such as Gramian Angular Field (GAF) [6] and Gramian

Angular Difference Field (GADF) [6], Markov Transition Field (MTF) [7], Recursive Plot (RP) [8], Motif Difference Field (MDF) [9], Relative Position Matrix (RPM) [10], and Short-Time Fourier Transform (STFT) [5]. By combining these features of different modes, we are able to capture the time, frequency, and spatial domain features of the signal more accurately, thus improving the performance of signal modulation identification.

Next, we feed the extracted multimodal features into a well-designed deep learning model for signal modulation recognition. We employ residual networks and LSTM structures to efficiently learn the abstract representations and temporal features of the signals. The main contributions of this study include:

1) A multi-modal deep learning based signal modulation recognition scheme is proposed, which can fully utilize the feature information of different modes.
2) Novel feature extraction methods and deep learning model structures are introduced to improve the accuracy and robustness of signal modulation recognition.
3) The effectiveness of the proposed scheme is verified through large-scale experiments and compared and analyzed with traditional methods.

By combining multimodal features and deep learning models, we aim to achieve highly accurate and efficient signal modulation identification, providing more reliable and efficient modulation identification capabilities for wireless communication systems.

II. RELATED WORK

A. deep Learning model for Modulation Recognition

Lee et al [11]. construct a four-layer fully connected network to perform signal modulation recognition. liu et al [12]. used CNN to extract signal features. Zhang et al [13]. use the DenseNet structure and use residual connections to construct a deep neural network for signal modulation recognition. Liu et al [14]. use CNN networks to extract the features of the signals, followed by similarity to construct the adjacency matrix between the signal samples and use graph convolution for signal modulation recognition. SSRCNN [15]: propose a semi-supervised learning (SSL) framework that can efficiently extract knowledge from unlabeled data by designing loss functions and neural network structures. SQRNN [16]: propose an automatic constraint classifier architecture that exploits the low time slot feature of the transformation threshold to enhance the learning capability of the model. However, none of the above algorithms take into account the existence of low signal-to-noise ratio situations and the inconsistency between the distribution of test and training data. As a result, some key information is lost and a full representation of the target problem is not possible.

B. Multimodal features of time series

Recurrence Plot (RP) [8] is a method for converting a time series into a two-dimensional image. It constructs a binary matrix by comparing the similarities between sample points in a time series. In a recurrence plot, the rows and columns of the matrix correspond to the sample points in the time series, and the elements of the matrix indicate the similarity between the corresponding sample points. Recurrence plots can capture repeating patterns and periodicity in a time series. Gramian Angular Summation/Difference Field (GASF/GADF) [6] is a method for converting a time series into a two-dimensional image. They obtain a two-dimensional image by converting the time series to polar coordinate representation and then computing it using sine and cosine functions.GASF and GADF capture the periodicity and trend of the time series, respectively. Markov Transition Field (MTF) [7] is a method for converting a time series into a two-dimensional image. It constructs a two-dimensional matrix by analyzing the state transfer relationships in a time series, where the elements of the matrix represent the transfer probabilities from one state to another. The Markov transfer field captures the state transfer patterns and sequence properties in the time series. Motif Difference Field (MDF) [9] extracts features by calculating the difference between signal samples. This method highlights the variation and trend information in the signal and helps to capture the dynamic features of the modulated signal. Relative Position Matrix (RPM) [10] extracts features by analyzing the relative order and relationship between signal samples. This method captures the sequential orderliness and correlation in the signal and helps to differentiate between different modulation classes. The Short-Time Fourier Transform (STFT) [5] provides spectral information. Phase Space Reconstruction (PSR) [17] is a method for converting a time series into a point cloud representation in phase space. It uses delayed embedding techniques in the time series to convert a univariate time series into a set of points in a multidimensional phase space. The dynamic characteristics and structure of the time series can then be analyzed by visualizing the set of points in phase space, e.g., by plotting a scatterplot of the phase space or generating an image representation of the phase space.

III. FRAMEWORK OVERVIEW

This section aims to provide a detailed description of the multimodal feature extraction component and model architecture in our system. The primary objective is to transform the signal modulation recognition problem into a classification task. To achieve this, we have constructed a multimodal feature extractor designed to capture diverse aspects of the signal's characteristics. Subsequently, these features are fed into a deep learning model for the ultimate classification prediction. This section will delve into the construction process of the model and the design principles behind each component.

A. Multimodal Feature Extractor

As shown in Figure 2, No longer limited to signal processing, we draw on a variety of advanced time series feature extraction techniques [18], including Gramian Angular Field (GAF) [6] and Gramian Angular Difference Field (GADF) [6], Markov Transition Field (MTF) [7], Recursive Plot (RP) [8], Motif Difference Field (MDF) [9], Relative Position Matrix

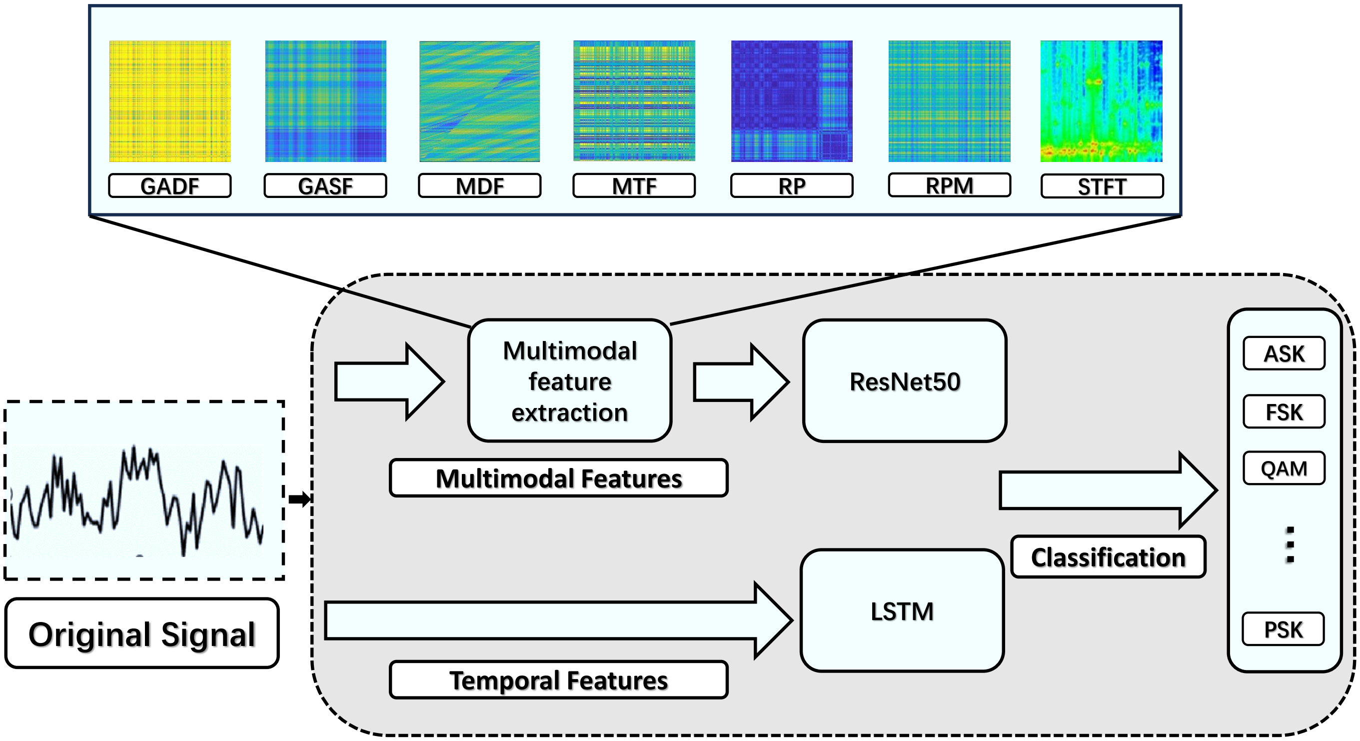


Fig. 2. Signal Modulation Recognition System Architecture

(RPM) [10], and Short-Time Fourier Transform (STFT) [5]. to construct the multimodal features of the original signal.

Compared to the traditional approach of representing the original signal as a spectrogram using Short-Time Fourier Transform (STFT) [5], our chosen algorithms offer several advantages.

Firstly, Gramian Angular Field (GAF) [6] and Gramian Angular Difference Field (GADF) [6] excel in capturing the temporal dependencies and non-linearity in signal evolution, providing a more comprehensive feature representation, including spectral information, periodicity, phase relationships, and trends. This multidimensional feature representation better captures the modulation patterns and dynamic characteristics of signals, supporting a more comprehensive understanding of modulation modes.

Meanwhile, Markov Transition Field (MTF) [7] extracts features by analyzing the state transition relationships between signal samples. This method captures state transition patterns during signal modulation, aiding in the identification of differences between various modulation categories.

Additionally, Recursive Plot (RP) [8] technique utilizes higher-order statistical information by modeling recursive relationships between signal samples to extract features. This approach captures recursive structures and dependencies within the signal, revealing inherent patterns that conventional methods might overlook, enabling the model to recognize complex features that could be neglected by traditional approaches.

The Motif Difference Field (MDF) [9] extracts features by computing the differences between signal samples. This method emphasizes changes and trends in the signal, aiding in capturing the dynamic characteristics of modulated signals.

The Relative Position Matrix (RPM) [10] extracts features by analyzing the relative order and relationships between signal samples. By incorporating spatial relations and structural information into the feature set, it can capture the sequential order and correlation within the signal, facilitating the differentiation between different modulation classes. This enables the model to capture spatial dependencies within the signal, particularly beneficial when dealing with complex spatial arrangements in signal modulation patterns.

It can be seen that the multimodal features we constructed have significant advantages over traditional algorithms, and are able to capture the multidimensional features such as temporal evolution, nonlinear properties, spectral information, periodicity, phase relationship, and change trend of the signal in a more comprehensive and integrated way. This comprehensive feature representation makes our model perform better in the comprehensive understanding of different modulation modes and provides strong support for the performance improvement of the signal modulation identification task.

B. Network Architecture

As the primary objective of this paper is to validate the effectiveness of the proposed multimodal feature extraction, we used a widely used neural network architecture to construct the model. Multi-modal features previously extracted for the real and imaginary parts of the original signal, respectively, were uniformly resized to a size of 128×128. Subsequently, these features are concatenated into a 3D vector of dimensions $14 \times 128 \times 128$. Following this, standard normalization is applied along each channel of the vector. Finally, the normalized features are fed into the ResNet50 [19] backbone network for feature learning.

We understand that the time series of the original signal contains rich temporal features. Therefore, our multimodal features not only encompass those constructed through two-dimensional images but also include the temporal characteristics of the original signal [20]. To comprehensively utilize this information, As shown in Figure 2, we designed a multi-input deep learning model. In this model, we feed the constructed multimodal image features into the ResNet50 [19] backbone for feature learning [21]. Simultaneously, we input the real and imaginary parts of the original signal into two separate LSTM [22] layers (with no shared parameters) forming the backbone network, extracting temporal features of the signal's real and imaginary components, respectively.

After extracting high-dimensional features from the two backbone networks, we concatenate the features from both channels. These concatenated features are then fed into two fully connected layers. In the first fully connected layer, we apply leakyrelu and BN operations to create a non-linear mapping for the network layer. For the second fully connected layer, we directly apply softmax for the final prediction of the signal modulation category.

IV. IMPLEMENTATION AND EVALUATION

This section will cover the experimental part of our study. We will begin by introducing our experimental setup, followed by comparative experiments and ablation experiments to validate the effectiveness of the proposed algorithm.

A. Experimental Setup

We employed GNU-radio and Python tools for signal simulation and dataset construction. The dataset comprises 11 modulation signals, including 8 digital modulation types (8PSK, BPSK, CPFSK, GFSK, PAM4, 16QAM, 64QAM, QPSK) and 3 analog modulation types (AM-DSB, AM-SSB, WBFM). Each modulation type covers 20 Signal-to-Noise Ratio (SNR) levels, totaling 220,000 samples. In contrast to traditional signal simulations using random sequences as data sources, we utilized Shakespeare's Gutenberg works as the baseband signal for digital modulation and the series "Serial Episode" for analog modulation. In terms of noise environments, Additive White Gaussian Noise (AWGN) was introduced, considering diverse channel scenarios, including AWGN, selective fading (Rician + Rayleigh), Center Frequency Offset (CFO), and Sample Rate Offset (SRO). The sampling rate was set to 200 kHz, with delays ranging from $[0.0, 0.9, 1.7]$. The SNR ranges from -20dB to 20dB, with 2dB intervals, yielding 1,000 samples for each SNR. Each sample includes In-phase (I) and Quadrature-phase (Q) signals, with each signal comprising 128 points. Consequently, the dataset's overall size is $220,000 \times 2 \times 128$. Also to verify the generalizability and robustness of our model for unseen data, we use two public datasets RML2018.01a [20] and HisarMod2019.1 [23] for testing.

Our hardware components include a laptop and a high-performance server, Powerleader PR2730G, equipped with Nvidia Tesla P100 GPU. For programming, we use Python, PyTs [24] library for multimodal feature extraction, and PyTorch library for model construction. The dataset is divided into a training set, consisting of approximately 176,000 samples, and a test set with 44,000 samples, both in npy format. The categorical cross-entropy was set as the loss function and Adam's algorithm was used as the optimizer with a cosine annealing learning rate optimization algorithm. In all experiments, the initial learning rate starts at 0.001 and the batch size is set to 400. if the validation loss does not decrease within 5 periods, the learning rate is halved. If the validation loss remains stable for 50 periods, the training process stops. And ten-fold cross-validation is used to avoid the randomness of the parameters. We evaluate the model using the modulation type recognition accuracy (Accuracy) metric.

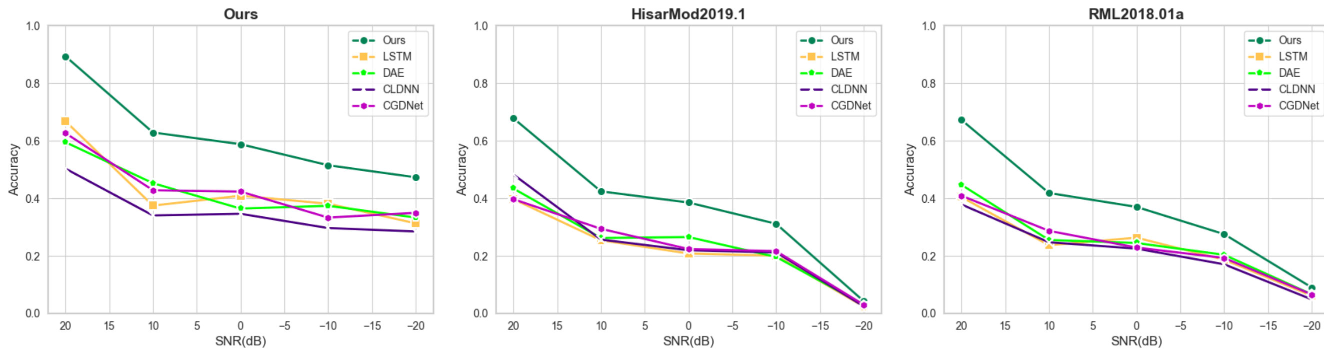
We use comparative experiments and ablation experiments to verify the effectiveness of the algorithm proposed in this paper. We have selected several classical algorithms as the baseline of this paper, the following is the baseline of this paper:

1) LSTM [4]: used the LSTM algorithm to extract time-domain features of the signal for modulation identification.
2) DAE [25]: employed a self-encoder network to compress the noise of the original signal thereby increasing the modulation recognition robustness.
3) CLDNN [26]: use Inception structure and LSTM to improve learning synchronization and equalization.
4) CGDNet [27]: improved LSTM to GRU based on CLDNN [26] and added Gaussian discard to ensure the modulation recognition rate on the basis of reducing the complexity of the algorithm to some extent.

The effectiveness of this paper's algorithm is verified by comparing it with the above baseline algorithm.

B. Comparative Experiments

As shown in Figure 3, We conducted a comprehensive validation experiment to thoroughly compare the accuracy of signal modulation. The experiment covered a wide range from -20SNR to 20SNR, aiming to simulate signal modulation scenarios in different signal-to-noise ratio (SNR) environments. The results indicate that our model outperforms other baseline algorithms significantly at various SNR levels. Particularly noteworthy is the fact that, under low SNR conditions, while other algorithms experience a substantial decrease in accuracy, our algorithm, although affected to some extent, maintains a relatively stable performance. This suggests that our approach demonstrates promising performance across different noise environments, especially in challenging low SNR conditions. The experimental findings clearly highlight the significant



Accuracy on Our Dataset Accuracy on HisarMod2019.1 Accuracy on RML2018.01a

Fig. 3. Recognition Accuracy at Different SNR on Three Datasets

advantages of our adopted multi-modal feature construction method and deep learning model in the task of signal modulation recognition. Compared to other baseline algorithms, our system exhibits greater robustness across various SNR levels, particularly in low SNR conditions. This strongly indicates that our approach can effectively capture signal modulation patterns in complex noise environments, thereby enhancing the overall system performance.

To validate the model's generalization and robustness on unseen data, we applied the pre-trained model to two additional publicly available datasets, distinct from our training data and without transfer learning. We compared the performance of various algorithms by directly testing the model. While the accuracy of other algorithms might significantly decline on these datasets due to different data distributions, our algorithm maintains high accuracy, surpassing other baseline algorithms. Observing that other algorithms experience noticeable accuracy drops on unknown data, likely due to their weak adaptability to different data distributions during training, our model, in contrast, performs remarkably well on unseen data. This superior performance may be attributed to our model's comprehensive consideration of the complexity and diversity of signal modulation tasks during training. The introduction of multimodal features enhances the model's adaptability to different datasets.

Compared to traditional algorithms, our multimodal features are more flexible and diverse, providing a more comprehensive expression of the multidimensional properties of signals. Traditional methods might be limited to spectral information, unable to capture deeper and multimodal features of signals. Thus, by introducing these diversified feature extraction methods, our model can more comprehensively and accurately represent signal characteristics, yielding significant performance advantages in various data distributions and noise environments. The integrated utilization of these multimodal features enhances the model's adaptability, thereby improving its generalization and robustness in practical applications.

C. Ablation Experiments

As shown in Table I, In order to validate the effectiveness of our algorithm, we conducted ablation experiments, systematically replacing the algorithm's backbone network and gradually removing components of the multimodal features to understand their impact on model performance. Specifically, we attempted to substitute the backbone network with ResNet32 and ResNet18, and replaced the temporal feature extraction backbone network with TCN [28] and GRU [29]. Simultaneously, we excluded various combinations of multimodal features, including Gramian Angular Field (GAF), Gramian Angular Difference Field (GADF), Markov Transition Field(MTF), Recursive Plot(RP), Motif Difference Field(MDF), Relative Position Matrix(RPM), and Short-Time Fourier Transform (STFT). The experimental results indicate that changing the backbone network has a relatively minor impact on the model's performance, with a limited decrease in accuracy. However, removing any part of the multimodal features significantly reduces the model's accuracy.

This experimental phenomenon underscores the crucial role of multimodal features in the model's performance. Further analysis reveals that each multimodal feature collaborates synergistically, complementing each other in extracting signal features. This synergy enables the model to more comprehensively and accurately capture different signal modulation patterns. The experiment confirms the importance of multimodal features, providing not only additional dimensions of information but also relationships between these dimensions that contribute to capturing the complex characteristics of signals. Therefore, the experimental results emphasize the critical role of multimodal features in our algorithm, signifying their essential significance in enhancing model performance and adaptability to diverse data distributions.

V. CONCLUSION

In this paper, we propose a new algorithm for signal modulation recognition in wireless communication systems. We adopt a multimodal deep learning algorithm to fuse features from different modalities in the time, frequency, and spatial domains to improve the performance of modulation identification. Compared with traditional algorithms, our algorithm achieves higher accuracy and reliability in signal modulation recognition. By taking full advantage of the different modal features, our model is able to better capture the information of

TABLE I
ACCURACY OF ABLATION EXPERIMENTS ON THREE DATASETS

Dataset	Our Dataset	RML2018.01a	HisarMod2019.1
w/o GAF	0.91	0.78	0.73
w/o GADF	0.92	0.77	0.74
w/o MTF	0.88	0.62	0.54
w/o RP	0.83	0.52	0.64
w/o MDF	0.84	0.75	0.56
w/o RPM	0.79	0.58	0.51
w/o STFT	0.89	0.65	0.71
$Original \quad Algorithm_{ResNet32}$	0.93	0.78	0.74
$Original \quad Algorithm_{ResNet18}$	0.91	0.61	0.72
$Original \quad Algorithm_{TCN}$	0.92	0.80	0.85
$Original \quad Algorithm_{GRU}$	0.94	0.79	0.83
$Original \quad Algorithm$	0.93	0.81	0.84

the signal in both the time and spatial domains, thus improving the accuracy of modulation identification. Experimental results show that our proposed algorithm achieves significant performance improvement in signal modulation identification tasks.

REFERENCES

[1] J. A. Sills, "Maximum-likelihood modulation classification for psk/qam," in *MILCOM 1999. IEEE Military Communications. Conference Proceedings (Cat. No. 99CH36341)*, vol. 1. IEEE, 1999, pp. 217–220.

[2] C.-S. Park, J.-H. Choi, S.-P. Nah, W. Jang, and D. Y. Kim, "Automatic modulation recognition of digital signals using wavelet features and svm," in *2008 10th International conference on advanced communication technology*, vol. 1. IEEE, 2008, pp. 387–390.

[3] R. Roy and T. Kailath, "Esprit-estimation of signal parameters via rotational invariance techniques," *IEEE Transactions on acoustics, speech, and signal processing*, vol. 37, no. 7, pp. 984–995, 1989.

[4] S. Rajendran, W. Meert, D. Giustiniano, V. Lenders, and S. Pollin, "Deep learning models for wireless signal classification with distributed low-cost spectrum sensors," *IEEE Transactions on Cognitive Communications and Networking*, vol. 4, no. 3, pp. 433–445, 2018.

[5] Y. Zeng, M. Zhang, F. Han, Y. Gong, and J. Zhang, "Spectrum analysis and convolutional neural network for automatic modulation recognition," *IEEE Wireless Communications Letters*, vol. 8, no. 3, pp. 929–932, 2019.

[6] Z. Wang and T. Oates, "Imaging time-series to improve classification and imputation," in *Proceedings of the 24th International Conference on Artificial Intelligence*, 2015, pp. 3939–3945.

[7] Z. Wang, T. Oates *et al.*, "Encoding time series as images for visual inspection and classification using tiled convolutional neural networks," in *Workshops at the twenty-ninth AAAI conference on artificial intelligence*, vol. 1. AAAI Menlo Park, CA, USA, 2015.

[8] J.-P. Eckmann, S. O. Kamphorst, D. Ruelle *et al.*, "Recurrence plots of dynamical systems," *World Scientific Series on Nonlinear Science Series A*, vol. 16, pp. 441–446, 1995.

[9] X. Li, Y. Kang, and F. Li, "Forecasting with time series imaging," *Expert Systems with Applications*, vol. 160, p. 113680, 2020.

[10] W. Chen and K. Shi, "A deep learning framework for time series classification using relative position matrix and convolutional neural network," *Neurocomputing*, vol. 359, pp. 384–394, 2019.

[11] J. Lee, B. Kim, J. Kim, D. Yoon, and J. W. Choi, "Deep neural network-based blind modulation classification for fading channels," in *Proc. of IEEE ICTC*, 2017, pp. 551–554.

[12] X. Liu, D. Yang, and A. El Gamal, "Deep neural network architectures for modulation classification," in *2017 51st Asilomar Conference on Signals, Systems, and Computers*. IEEE, 2017, pp. 915–919.

[13] L. Zhang, C. Lin, W. Yan, Q. Ling, and Y. Wang, "Real-time ofdm signal modulation classification based on deep learning and software-defined radio," *IEEE Communications Letters*, vol. 25, no. 9, pp. 2988–2992, 2021.

[14] Y. Liu, Y. Liu, and C. Yang, "Modulation recognition with graph convolutional network," *IEEE Wireless Communications Letters*, vol. 9, no. 5, pp. 624–627, 2020.

[15] Y. Dong, X. Jiang, L. Cheng, and Q. Shi, "Ssrcnn: A semi-supervised learning framework for signal recognition," *IEEE Transactions on Cognitive Communications and Networking*, vol. 7, no. 3, pp. 780–789, 2021.

[16] P. Ghasemzadeh, M. Hempel, and H. Sharif, "Gs-qrnn: A high-efficiency automatic modulation classifier for cognitive radio iot," *IEEE Internet of Things Journal*, vol. 9, no. 12, pp. 9467–9477, 2022.

[17] E. N. Lorenz, "Deterministic nonperiodic flow," *Journal of atmospheric sciences*, vol. 20, no. 2, pp. 130–141, 1963.

[18] H. Ismail Fawaz, G. Forestier, J. Weber, L. Idoumghar, and P.-A. Muller, "Deep learning for time series classification: a review," *Data mining and knowledge discovery*, vol. 33, no. 4, pp. 917–963, 2019.

[19] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[20] T. J. O'Shea, T. Roy, and T. C. Clancy, "Over-the-air deep learning based radio signal classification," *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 1, pp. 168–179, 2018.

[21] F. Zhang, C. Luo, J. Xu, Y. Luo, and F.-C. Zheng, "Deep learning based automatic modulation recognition: Models, datasets, and challenges," *Digital Signal Processing*, vol. 129, p. 103650, 2022.

[22] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[23] K. Tekbıyık, A. R. Ekti, A. Görçin, G. K. Kurt, and C. Keçeci, "Robust and fast automatic modulation classification with cnn under multipath fading channels," in *2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring)*. IEEE, 2020, pp. 1–6.

[24] J. Faouzi and H. Janati, "pyts: A python package for time series classification," *Journal of Machine Learning Research*, vol. 21, no. 46, pp. 1–6, 2020. [Online]. Available: http://jmlr.org/papers/v21/19-763.html

[25] A. Dai, H. Zhang, and H. Sun, "Automatic modulation classification using stacked sparse auto-encoders," in *2016 IEEE 13th international conference on signal processing (ICSP)*. IEEE, 2016, pp. 248–252.

[26] N. E. West and T. O'shea, "Deep architectures for modulation recognition," in *2017 IEEE international symposium on dynamic spectrum access networks (DySPAN)*. IEEE, 2017, pp. 1–6.

[27] J. N. Njoku, M. E. Morocho-Cayamcela, and W. Lim, "Cgdnet: Efficient hybrid deep learning model for robust automatic modulation recognition," *IEEE Networking Letters*, vol. 3, no. 2, pp. 47–51, 2021.

[28] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," *arXiv preprint arXiv:1803.01271*, 2018.

[29] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," in *NIPS 2014 Workshop on Deep Learning, December 2014*, 2014.