A Novel Combination of PCA and LSTM for Multi-carrier Modulation Signals Classification

KM Sejal
Department of Computer Engineering
National Institute of Technology
Kurukshetra, India
sejal2000bui@gmail.com

Mohit Dua
Department of Computer Engineering
National Institute of Technology
Kurukshetra, India
er.mohitdua@nitkkr.ac.in

Abstract: Fast communication is essential in today's world, and signal demodulation is a key factor affecting latency. Traditional demodulation methods are complex and time-consuming, making Automatic Modulation Classification (AMC) crucial for modern communication systems, particularly in wireless communication and signal processing. This paper presents a novel approach to classify five types of Multicarrier Modulation (MCM) signals. It also analyzes two subcarrier waveforms for each signal type, resulting in 10 unique MCM signals. This paper is divided into 2 parts – feature extraction from signals and classification of signals using those extracted features. To extract relevant features, Principal Component Analysis (PCA) and Recursive Feature Elimination (RFE) are applied. A Long Short-Term Memory (LSTM) network is then used to classify the MCM signals. Our results show that the combination of PCA and RFE. LSTM achieves superior performance in categorizing the MCM signals, making it a promising approach for efficient and accurate AMC. This paper is classifying MCM signals of range -16dB to 16dB with the accuracy of 91% to 97%.

Keywords—Automatic Modulation Classification, Principal Component Analysis, Recursive Feature Elimination, Long Short Term Memory

I. INTRODUCTION

The 5G generation has raised expectations for better dependability, reduced latency, and quicker data speeds. Ensuring the fulfillment of this requirement calls for a communication infrastructure that is robust, safe and effective. The fundamental method that enables flawless radio wave exchange via cell phones and quick data transmission via fiber optic cables is called modulation. Modulation serves as a versatile foundation of communication by reducing latency for near real-time responsiveness. It employs high-frequency carriers to ensure stable long-distance transmission and enhances data rates through signal compression. Additionally, by adding international noise to signals, preventing unlawful interception and eavesdropping efforts, and safe guarding data privacy, several modulation techniques improve data security. Wireless communication networks are among the most well-known businesses and domains where Automatic Modulation Classification, or AMC, is utilized. In this sector, AMC is essential for determining the incoming signal's modulation schemes, which helps to optimize decoding, processing of signals and allocating resources. This optimization raises quality of service and improves system performance across Advancing communication frameworks, such as 4G, 5G, and beyond. AMC is also applied in key military and defense scenarios such as electronic warfare (EW) and signal intelligence (SIGINT). In recent years, automatic modulation categorization, or AMC, has drawn a lot of attention from researchers. Numerous conventional techniques, including instantaneous parameters [1], cyclostationary features [2], [3], and high-order cumulant (HOC) features [4], [5], and judgment based on likelihood for classification within single carrier signals blended with multicarrier signals, [6] – [7] have been proposed. O'Shea [8] achieves encouraging results using convolution neural network (CNN) and raw I/Q data, avoiding the need to generate expert features and reducing complexity. However, various caused effects brought about by harsh and packed settings, like dilation, multipath interference, I/Q balance, and phase offset, may influence the accuracy of AMC utilizing raw I/Q data, as discussed and evaluated in [8] and [9]. The use of multi-carrier modulation (MCM) has emerged as a key strategy for meeting growing data needs. By partitioning data across numerous subcarriers, each transmitted on a unique carrier frequency; Multicarrier Modulation (MCM) addresses bandwidth requirements and bolsters robustness against noise and interference. This shift marks a substantial advancement in achieving efficient, resilient communication amid growing data rates. In signal analysis, PCA and Recursive Feature Elimination (RFE) are applied to raw data for feature selection, followed by the application of Long Short-Term Memory (LSTM) for classification, providing a robust method to classify the modulating signals.

This paper's remaining sections are arranged as follows: In Section 2, previous research on the subject is reviewed. The proposed work is described in Section 3. The description about dataset is shown in Section 4 and experimental results are shown in this section's A part. A discussion is given in Section 4's B, and the study is concluded in Section 5.

II. LITERATURE REVIEW

Much work has gone into determining which feature from a modulated signal is most suited for classification. Ho et al.'s study [10] largely examined BPSK and BASK modulation types at -5dB SNR, and they employed wavelet transform to extract signal features with 54.0% accuracy. Moser et al. classified modulated signals using the Instantaneous feature. They analyzed 9 distinct types of modulated signals with Signal-to-Noise Ratios (SNR) of 20 dB and 10 dB at two different threshold levels, 8192 and 2048, and achieved a 100% accuracy rate at the threshold level of 8192. When working with 6 different types of modulated signals, Cumulants is used to extract the primary characteristics for classification by Xie et al. [11], achieving 99% accuracy at -5dB and 100% accuracy at -2dB. Constellation diagrams are a crucial tool for classifying modulated signals, as demonstrated by Peng et al., who classified BPSK, 4ASK, OQPSK QPSK, and 8PSK with 100% accuracy at 4dB SNR. The effectiveness of the Short

Time Fourier Transform (STFT) in feature extraction of various signals within their respective domains is shown by three studies: Zhang and others. In [12], Wu et al. used the STFT as feature extraction technique of modulated signal and convert all into a grayscale spectrogram. This resulted in an accuracy of 99.0%. Various studies have demonstrated the effectiveness of PCA in reducing the dimensionality of feature sets while retaining the most significant information, thereby enhancing the performance of classification algorithms. Additionally, the application of PCA in [13] further highlights its robustness in handling complex modulated signals, making it a valuable tool in AMC research.

To classify the signals after most valuable feature extraction for classification, many classification techniques are used by the authors in their papers. Because LSTM networks can handle sequential data, time-dependent modulation pattern analysis is a perfect fit for them. The benefits of LSTMs in capturing temporal dependencies and long-range correlations in signal sequences were demonstrated by research conducted by Rajendran et al. [15] (2018). In Wu et al.'s work, CNN is also utilized to categorize 16 different types of modulation signals. The authors compare DL based and ML based approach 99% accuracy while using SVM yields 78.0% accuracy. The contribution of this study is summarized as follows:

1. This paper contains 2 parts- feature extraction and classification.

2. In this paper, 5 multi carrier modulation signals are used which are further classified by 2 types of sub carrier modulation, i.e., QAM (Quadrature Amplitude Modulation) 16 and QAM64.

3. PCA and RFE are used for feature extraction of signals.

4. The classification of 10 MCM signals is performed using LSTM.

III. METHODOLOGY

The proposed work of this paper is divided into 2 stages- feature extraction and classification of MCM signals. The steps of proposed work are given in Fig. 1. In the first stage PCA is used on CSV files to reduce the dimension of large dataset by taking useful features and further recursive feature elimination technique is applied for choosing most dominant features. In second stage a filtered set of features are trained by LSTM to classify MCM signals.

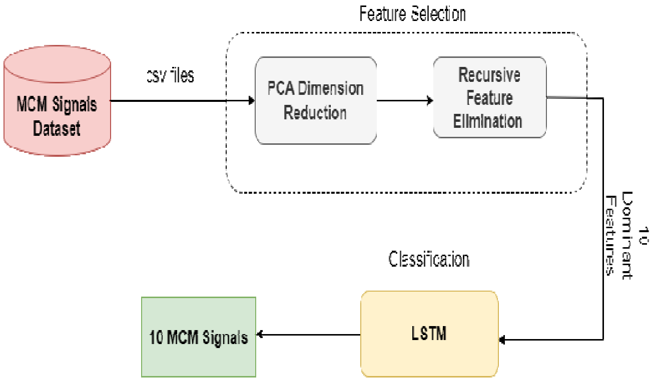


Fig. 1. Proposed Methodology

Step1- In this paper, we propose a novel approach to AMC of MCM signals. The proposed method begins with feature extraction using PCA, which reduces the dimensionality of the signal while retaining the most informative features. However, PCA may not always select the most relevant features, which can lead to suboptimal classification performance. To address this limitation, we incorporate RFE as a post-processing step to identify and eliminate irrelevant features. RFE is a powerful technique that recursively eliminates the least important features until a specified number of features is reached, thereby improving the robustness and accuracy of the classification model. By combining PCA and RFE, we create a hybrid feature set that leverages the strengths of both techniques. The resulting feature set, which is containing 10 dominant features, is shown to be highly effective in capturing the underlying patterns and characteristics of MCM signals, leading to improved AMC performance. The result PCA stage is given in figure 2.

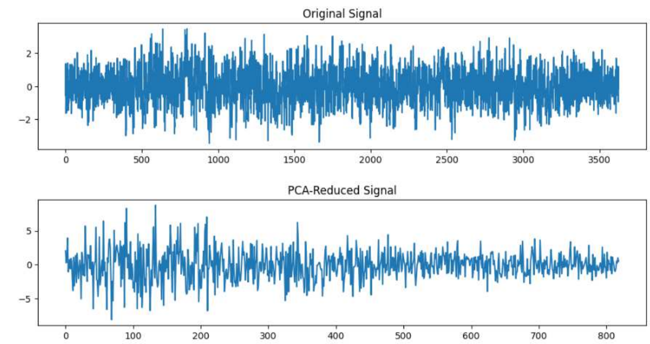


Fig. 2. Original Signal vs. PCA Reduced Signal

Step2- The selected features are then fed into a LSTM network for classification. The LSTM architecture is particularly well-suited for AMC of MCM signals due to its ability to learn long-term dependencies and patterns in sequential data. The LSTM network can effectively capture the complex relationships between the features and the modulation types, enabling accurate classification of the 10 MCM signals. Furthermore, the LSTM's ability to handle noisy and variable-length input sequences makes it an ideal choice for AMC applications, where the signals may be subject to channel impairments and interference. By leveraging the strengths of PCA, RFE, and LSTM, our proposed approach achieves superior performance in classifying MCM signals, demonstrating its potential for real-world AMC applications.

IV. RESULTS AND DISCUSSION

A simulated dataset is used for this work, featuring 5 multi-carrier modulation containing sub-carrier modulations QAM16 and QAM64 with varying SNRs from -20 dB to 30 dB. Those 5 multi-carrier modulations are FBMC (Filter Bank Multi-carrier), FOFDM (Filtered Orthogonal Frequency Division Multiplexing), OFDM (Orthogonal Frequency Division Multiplexing), UFMC (Universal Filtered Multi-carrier), WOLA (Weighted Overlap and Add based OFDM). The dataset is structured to mimic signal fading and time selectivity using a realistic free-space wireless channel model, and comprises 10,000 samples per SNR, each labeled with the modulation type and SNR value. White Gaussian noise is added to the dataset to simulate real-world noise conditions. This large dataset serves as an important training and testing ground for machine learning techniques in spectrum sensing and multi-carrier modulation signal detection. TABLE I. describes the features and no. samples present of each modulated signal present in dataset.

TABLE I. DATASET OF MCM SIGNALS

Modulated Signal	No. of samples	No. of features
FBMC_QAM16	1000	3626
FBMC_QAM64	1000	3626
FOFDM_QAM16	1000	3627
FOFDM_QAM64	1000	3627
OFDM_QAM16	1000	3626
OFDM_QAM64	1000	3626
UFMC_QAM16	1000	3626
UFMC_QAM64	1000	3626
WOLA_QAM16	1000	3627
WOLA_QAM64	1000	3627

A. Results and Performance Analysis:

The result of PCA and RFE is 10 dominant features, which are column numbers given in dataset. Those are ‘15’, ‘16’, ‘36’, ‘255’, ‘523’, ‘561’, ‘694’, ‘708’, ‘794’.

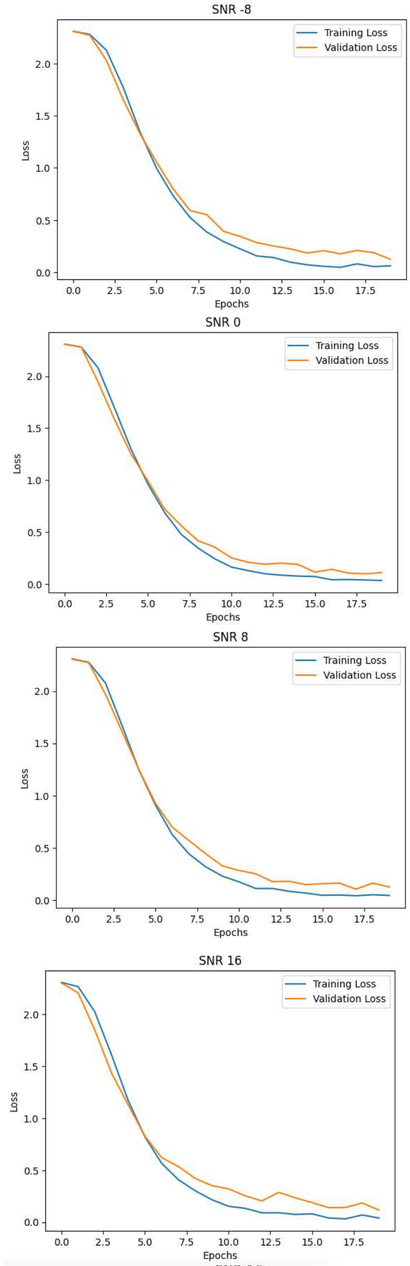
The proposed method demonstrates superior performance across a wide range of SNR values, specifically from -16dB to 16dB. This extensive range includes scenarios with very low to high signal-to-noise ratios, indicating the robustness of our approach. This model is giving high accuracy on 16dB and low accuracy on -16dB. This proposed model is giving monotonically increasing accuracy. TABLE II. represents the evaluation metrics of proposed method on different SNR values -16dB to 16dB calculated on test data.

TABLE II. RESULT METRICS OF TEST DATA

SNR Value	F1 Score	Precision	Recall	Accuracy
-16	0.90	0.91	0.90	0.91
-8	0.94	0.95	0.94	0.94
0	0.96	0.97	0.96	0.96
8	0.97	0.98	0.97	0.97
16	0.97	0.98	0.97	0.97

TABLE III. presents the loss graph across various SNR values, illustrating the model's performance in MCM classification. The depicted loss trends demonstrate that the proposed method maintains consistent and suitable classification accuracy over a broad range of SNR conditions, confirming its robustness and effectiveness in handling diverse signal-to-noise environments for multi-carrier modulation classification. Fig. 3 presents accuracy graph of SNR 16dB.

TABLE III. LOSS GRAPHS -16 DB TO 16DB



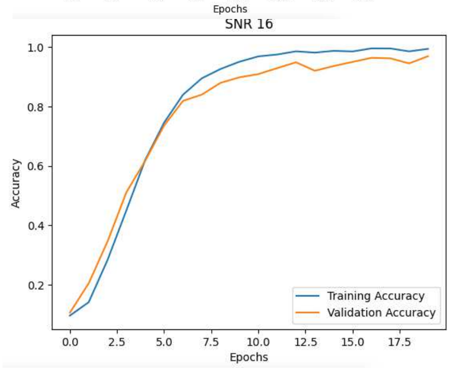


Fig. 3. Accuracy Graph of 16dB

B. Discussion:

The main goal of our proposed model is to classify the MCM Signals so that information retrieval can be easy. Our approach employs PCA and RFE for feature extraction and utilizes LSTM for classification. The results, presented in the results section, clearly indicate the efficacy of our proposed method, particularly when combining 2 feature extraction and dimension reduction methods with a neural network. It makes whole classification process very lightweight. TABLE IV. offers a comparison of our method with existing approaches, showcasing its superior performance.

TABLE IV. COMPARATIVE ANALYSIS

Techniques	Dataset	SNR	Accuracy
Recurrent Neural Network [8]	WB-FM, AMDSB, AM-SSB, BPSK, 8PSK, 16QAM, QPSK, 64QAM, BFSK, CPFSK and PAM4	-20dB – 18dB	91%
High-order cumulants and Deep Learning [11]	2FSK, 2PSK, 4FSK, 2ASK, 4PSK and 4ASK	-5dB	99%
STFT Spectrogram and Xception Model [17]	FBMC, FOFDM, OFDM, UFMC, WOLA with sub-carrier QAM16 and QAM64	-10dB 0dB 4dB 10dB	91% 95% 95% 99%
Proposed	FBMC, FOFDM, OFDM, UFMC, WOLA with sub-carrier QAM16 and QAM64	-16dB -8dB 0dB 8dB 16dB	91% 94% 96% 97% 97%

V. CONCLUSION

This paper presents an innovative approach for efficient and accurate Automatic Modulation Classification (AMC) of Multicarrier Modulation (MCM) signals, addressing the complexities of traditional demodulation methods. By combining Principal Component Analysis (PCA) and Recursive Feature Elimination (RFE) for feature extraction with a Long Short-Term Memory (LSTM) network for classification, our method successfully classifies five MCM types and their subcarrier waveforms (QAM16 and QAM64), achieving 91% to 97% accuracy within a SNR range of -16 dB to 16 dB. This technique enhances AMC accuracy and efficiency, promising significant improvements for real-time communication systems. Future research directions include expanding classification to additional modulation schemes and subcarrier formats, enhancing algorithm adaptability to varying channel conditions, and optimizing computational efficiency for real-time deployment. Additionally, integrating deep learning or reinforcement learning can further refine classification accuracy, while investigating FPGA/ASIC acceleration, scalability, and adversarial attack resilience will be critical. Exploring practical applications in mobile and satellite networks can demonstrate the method's effectiveness and broaden its real-world utility.

REFERENCES

[1] G. M. Bouchou, H. Wang, and M. E. H. Lakhdari, "Automatic digital modulation recognition based on stacked sparse autoencoder," in Proceedings of the 2017 IEEE 17th International Conference on Communication Technology (ICCT), pp. 28-32, Oct. 2017.

[2] H. Wang and L. Guo, "A new method of automatic modulation recognition based on dimension reduction," in Proceedings of the 2017 Forum on Cooperative Positioning and Service (CPGPS), pp. 316-320, May 2017.

[3] Q. Zhang, O. A. Dobre, Y. A. Eldemerdash, S. Rajan, and R. Inkol, ‘‘Second-order cyclostationarity of BT-SCLD signals: Theoretical developments and applications to signal classification and blind parameter estimation,’’ IEEE Trans. Wireless Commun., vol. 12, no. 4, pp. 1501–1511, Apr. 2013.

[4] D.-C. Chang and P.-K. Shih, ‘‘Cumulants-based modulation classification technique in multipath fading channels,’’ IET Commun., vol. 9, no. 6, pp. 828–835, 2015.

[5] W. Su, ‘‘Feature space analysis of modulation classification using very high-order statistics,’’ IEEE Commun. Lett., vol. 17, no. 9, pp. 1688–1691, Sep. 2013.

[6] B. Dulek, O. Ozdemir, P. K. Varshney, and W. Su, ‘‘Distributed maximum likelihood classification of linear modulations over nonidentical flat blockfading Gaussian channels,’’ IEEE Trans. Wireless Commun., vol. 14, no. 2, pp. 724–737, Feb. 2015.

[7] Y. Yuan, P. Zhao, B. Wang, and B. Wu, ‘‘Hybrid maximum likelihood modulation classification for continuous phase modulations,’’ IEEE Commun. Lett. , vol. 20, no. 3, pp. 450–453, Mar. 2016.

[8] T. J. O'Shea, J. Corgan, and T. C. Clancy, "Convolutional radio modulation recognition networks," in Engineering Applications of Neural Networks, Springer, pp. 213–226, 2016.

[9] S. C. Hauser, W. C. Headley, and A. J. Michaels, "Signal detection effects on deep neural networks utilizing raw IQ for modulation classification," in Proceedings of the IEEE Military Communications Conference (MILCOM), Baltimore, MD, USA, pp. 121–127, Oct. 2017.

[10] K. M. Ho, C. Vaz, and D. G. Daut, "A wavelet-based method for classification of binary digitally modulated signals," in Proceedings of the 2009 IEEE Sarnoff Symposium, pp. 1-5, 2009.

[11] W. Xie, S. Hu, C. Yu, P. Zhu, X. Peng, and J. Ouyang, "Deep learning in digital modulation recognition using high order cumulants," *IEEE Access*, vol. 7, pp. 63760-63766, 2019.

[12] J. Wu, Y. Zhong, and A. Chen, "Radio modulation classification using STFT spectrogram and CNN," in Proceedings of the 7th International Conference on Computer and Communications (ICCC), Chengdu, China, pp. 178-182, Dec. 10-13, 2021.

[13] K. Zhang, E. L. Xu, and Z. Feng, "PFS: A novel modulation classification scheme for mixed signals," in Proceedings of the IEEE Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), Montreal, QC, Canada, pp. 1-5, Oct. 2017.

[14] N. Chakravarty, M. Dua, and S. Dua, "Automatic modulation classification using amalgam CNN-LSTM," in Proc. 2023 IEEE Radio and Antenna Days of the Indian Ocean (RADIO), pp. 1-2, May 2023.

[15] S. Sharma, S. Maiti, S. Mythirayee, S. Rajendran, and B. Raju, "A Deep Dive into Sleep: Single-Channel EEG-Based Sleep Stage Classification with Model Interpretability," arXiv preprint arXiv:2309.07156, 2023.

[16] A. Tailor, M. Dua, and P. Verma, "Automatic classification of multi-carrier modulation signal using STFT spectrogram and deep CNN," Physica Scripta, vol. 99, no. 7, p. 076009, 2024.