Dual-Stream CNN–LSTM Fusion for Pilot-Assisted Automatic Modulation Classification in OTFS Systems

Dhruvil Mali Kamal Captain Dayshaun Kakadiya Arvind Chandak
Department of Electronics Engineering,
Sardar Vallabhbhai National Institute of Technology, Surat, India
Email: {u23ec062, kamalcaptain, u23ec064, u23ec061}@eced.svnit.ac.in

***Abstract*—OFDM has become the foundation of modern wireless communications through the standardization of many of today's mobile communication systems. However, its performance degrades significantly in high-mobility environments due to Doppler-induced inter-carrier interference (ICI). OTFS modulation is proposed as a promising solution, providing improved robustness and reliability in time-varying wireless channels. By mapping data into Delay-Doppler domain instead of a single time or frequency domain allows the receiver to obtain the benefit of Time-Frequency diversity with providing an inherent sparse channel representation. In this work, a hybrid dual-stream CNN-LSTM fusion-based Automatic Modulation Classification (AMC) framework is proposed for OTFS systems. The integration of embedded pilot symbols further enhances classification performance, particularly at low signal-to-noise ratio (SNR) levels. Experimental results show that the proposed model achieves up to 99.22% classification accuracy at 25 dB SNR and greater than 90% accuracy above 5 dB. The proposed AMC system model outperforms with better results than existing models in terms of classification accuracy, including operation at very low SNR levels.**

***Index Terms*—Orthogonal Time Frequency Space (OTFS), Automatic Modulation Classification (AMC), Deep Learning, High-Mobility Communications, Delay-Doppler Domain.**

I. Introduction

In modern wireless communication systems, the ability to maintain reliable connectivity in high-mobility environments such as high-speed railways and vehicle-to-everything (V2X) scenarios is a major challenge. These environments are characterized by severe Doppler shifts and time-varying channel conditions that complicate signal reception [1]. To address these challenges, the development of an intelligent Automatic Modulation Classification (AMC) system has become more essential [2], [3] and it is important to examine behavior of existing multi-carrier modulation techniques in high speed environments.

For decades, Orthogonal Frequency Division Multiplexing (OFDM) has been the standard method of multi-carrier modulation by powering 4G and 5G networks due to its high spectral efficiency [4]. In high-speed environments, the fast movement between the transmitter and receiver causes Inter-Carrier Interference (ICI) and significant spectral overlapping, which creates a major problem for OFDM systems by leading to performance degradation [5]. While techniques have been developed to improve classification in OFDM, the waveform's natural sensitivity to frequency shifts caused by rapid movement remains a major performance bottleneck [5].

To overcome these limitations, OTFS has been introduced as a promising solution for future 6G systems [4]. Unlike traditional schemes, OTFS maps information symbols to the Delay-Doppler (DD) domain, which effectively transforms a fading, time-varying channel into a near-constant complex gain for all symbols [2]. This allows OTFS to achieve full channel diversity across both time and frequency. This makes OTFS more reliable and robust against extreme Doppler spreads than OFDM [1], [6].

OTFS provides significant advantages in high-mobility channels. But identifying modulation schemes within this framework remains a challenging task due to residual interference and the complex structure of Delay-Doppler domain signals [1], [4]. This identification is done by the Automatic Modulation Classification (AMC). For this, deep learning provides an effective approach for wireless communications [7] solving challenges in 6G technologies by improving performance. AMC serves as an intermediate step between signal detection and demodulation which allows a receiver to identify the modulation format of a received signal without prior knowledge of the transmission parameters [2]. Recently, deep learning-based Automatic Modulation Classification (AMC) methods, particularly those employing Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks, have demonstrated strong capability in extracting discriminative features directly from raw signal data compared to classical likelihood-based approaches [8], [6]. However, the performance of many existing models degrades as the number of modulation classification categories increases [9], [6]. For feature extraction, the hybrid architectures such as CNN-LSTM would be an effective solution for OTFS-based AMC. In these architectures, CNNs effectively capture spatial patterns and high-dimensional representations from in-phase and quadrature (IQ) samples, while LSTM networks model long-term temporal dependencies within the signal sequence [3]. By integrating these complementary capabilities, CNN-LSTM architectures can effectively learn complex signal dynamics, mitigating issues such as gradient vanishing under extreme Doppler conditions while avoiding the high compu-

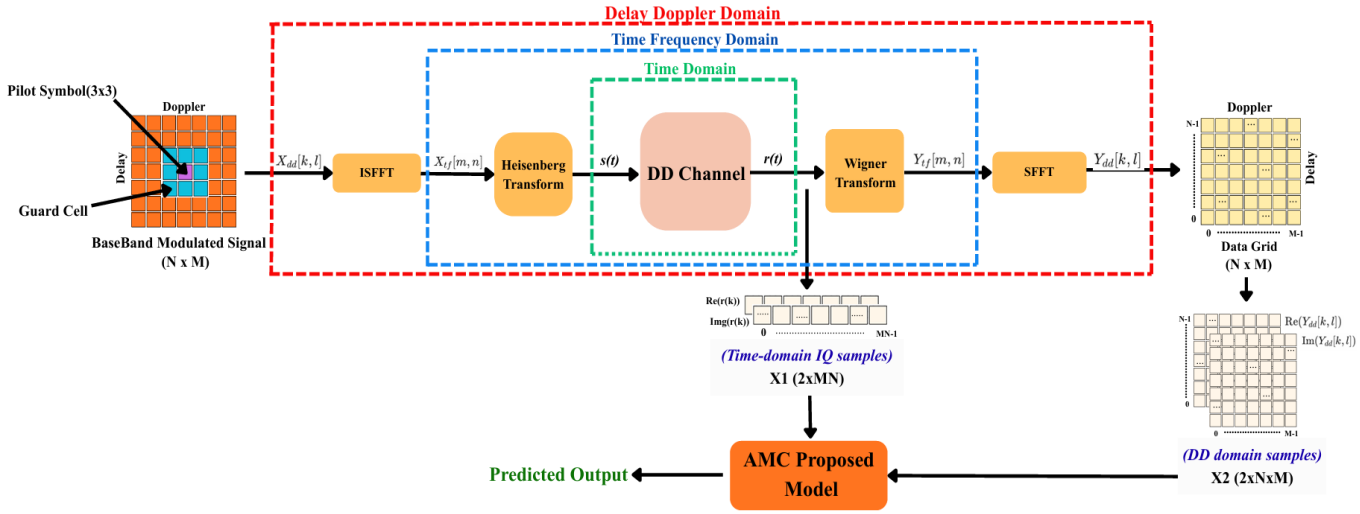


Fig. 1: Modulation, demodulation and AMC process flow for an OTFS system.

tational complexity associated with classical likelihood-based techniques when transmission parameters increase [2], [6].

In this paper, we are proposing a dual-stream flexible deep learning architecture specifically made for AMC in OTFS-based 6G networks. Our approach diverges from conventional single-input models by fusing convolutional features with temporal sequences to better capture the unique Delay-Doppler characteristics of the channel. In the model, for feature discriminability under severe channel impairments, the proposed framework extends AMC to nine modulation schemes like BPSK, QPSK, 8PSK, OOK, 4ASK, 8ASK and even higher-order schemes such as 16QAM, 64QAM, and 256QAM and can be used to classify with flexible number of modulation classification schemes.

II. System Model

The modulation, demodulation, and AMC process flow is shown in Fig. 1. An OTFS system first maps information symbols in Delay-Doppler (DD) domain, which are transformed into the Time-Frequency (TF) domain using ISFFT. Then converted into a Time-domain signal via the Heisenberg transform for transmission. At the receiver, this signal is processed using the Wigner transform and SFFT used to recover DD-domain symbols.

A. Modulation and Demodulation Procedure

We consider an OTFS system in which information symbols are mapped onto the Delay-Doppler (DD) domain [1], [4], providing inherent robustness to time-varying channels, where the transmitted symbols are denoted as $x[k,l]$ for $k = 0, \ldots, N-1$ and $l = 0, \ldots, M-1$, with N and M representing the number of Doppler and delay bins, respectively.

The DD grid is defined as:

$$\Gamma = \left\{ \left(\frac{k}{NT}, \frac{l}{M\Delta f} \right) \right\} \tag{1}$$

where T is the symbol duration and Δf is the subcarrier spacing. Each symbol occupies a unique delay and Doppler coordinate.

The DD-domain symbols are transformed into the Time-Frequency (TF) domain using the Inverse Symplectic Finite Fourier Transform (ISFFT) [1] and it is defined as:

$$X[n,m] = \frac{1}{\sqrt{NM}} \sum_{k=0}^{N-1} \sum_{l=0}^{M-1} x[k,l] e^{j2\pi\left(\frac{nk}{N} - \frac{ml}{M}\right)} \tag{2}$$

This operation spreads each DD symbol over the entire TF grid makes it robust against channel variations. The TF-domain signal is then converted into a continuous-time waveform $s(t)$ by applying Heisenberg transform [6] with the transmit pulse $g_{tx}(t)$ as:

$$s(t) = \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} X[n,m] g_{tx}(t - nT) e^{j2\pi m\Delta f(t-nT)} \tag{3}$$

The wireless channel in the DD domain is modeled as:

$$h(\tau,\nu) = \sum_{i=1}^{P} h_i \delta(\tau - \tau_i)\delta(\nu - \nu_i) \tag{4}$$

where P is the number of paths in the DD domain [1], [5]. The variables h_i, τ_i, and ν_i refer to the channel gain, delay, and Doppler shift, respectively.

The received signal becomes:

$$r(t) = \iint h(\tau,\nu) s(t-\tau) e^{j2\pi\nu(t-\tau)} d\tau d\nu + w(t) \tag{5}$$

where $w(t)$ is Additive White Gaussian Noise (AWGN).

At the receiver, the received symbols in the TF domain $Y_{tf}[n,m]$ can be obtained via the Wigner Transform [6] :

$$Y_{tf}[n,m] = \int r(t) g_{rx}^{*}(t - nT) e^{-j2\pi m\Delta f(t-nT)} dt \tag{6}$$

where $g_{rx}(t)$ denotes the pulse shaping filter at the receiver. By applying the Symplectic Finite Fourier transform (SFFT) to the sample gives us the symbols in the DD domain:

$$y_{dd}[k,l] = \frac{1}{\sqrt{NM}} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} Y_{tf}[n,m] e^{-j2\pi\left(\frac{nk}{N} - \frac{ml}{M}\right)} \quad (7)$$

This representation is further separated into its real and imaginary components and used for automatic modulation classification (AMC).

III. Proposed Methodology

In this section, we present a detailed explanation of the proposed hybrid deep learning architecture for AMC in the OTFS system. This work leverages distinct signal representations to achieve robust classification across all nine modulation schemes **BPSK, QPSK, 8PSK, OOK, 4ASK, 8ASK, 16QAM, 64QAM, and 256QAM**.

A. Proposed DL-Based AMC Architecture

Fig. 2 refers the proposed model employs a dual-stream CNN–LSTM Fusion deep learning architecture to extract complementary features from both DD-domain representations and Time-domain (Time Input) IQ samples of the received signal. The Delay-Doppler (DD) domain representation (DD Grid input) has a shape of $(M, N, 2)$, where $M = 64$ and $N = 32$ in our case.

The DD-domain branch processes a $64 \times 32 \times 2$ input using three successive Conv2D layers with 32, 64, and 128 filters, respectively. The first two layers are followed by Batch Normalization and (2×2) Max-Pooling, while the third convolution layer uses Average Pooling to preserve high-level feature information. The resulting features are passed through a 128-neuron Dense layer to generate M_1.

Simultaneously, the time-domain branch processes 2048×2 IQ samples through two Conv1D layers with 64 and 128 filters and kernel sizes of (1×7) and (1×5), respectively, each followed by Batch Normalization and Max-Pooling layer with pool size of 2. The extracted features are then fed into a 64-unit LSTM layer [8] to capture temporal dependencies, followed by a 128-neuron Dense layer to produce M_2.

These outputs M_1 and M_2 are concatenated to form a unified feature vector. This fused internal output is passed through a Dense layer with 256 neurons, followed by a Dropout layer with a rate of 0.3 to prevent overfitting. Subsequent Dense layer with 128 neurons for further feature refinement. The Softmax classification layer provides the probability distribution across the nine modulation classes and produces the final output. All convolution and dense layers utilize the ReLU activation function.

B. Dataset Generation

A diverse dataset was generated to simulate a realistic high-mobility environment across various SNR levels for all nine modulation schemes. The system parameters were set

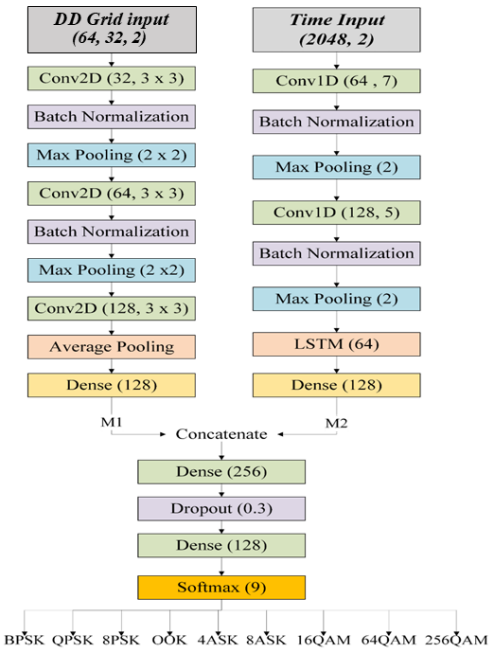


Fig. 2: Proposed dual-stream model architecture.

with $M = 64$ delay bins and $N = 32$ doppler bins with a subcarrier spacing of 15 kHz. The channel was modeled using the Extended Vehicular A (EVA) power delay with nine different propagation taps. To simulate high mobility conditions, a vehicle speed of 120 km/h was applied, and the Doppler shift for each tap was calculated using Jakes' formula.

A 3×3 pilot block (9 pilot grid elements) is inserted at the center of the delay–Doppler grid. The pilot is enclosed within a 7×7 protected region containing 49 grid elements in total, of which the remaining 40 grid elements are assigned as zero-power guard symbols to prevent interference between pilot and data symbols during channel spreading.

In addition, Additive White Gaussian Noise (AWGN) is added to the received signal to simulate thermal noise and interference effects to visualize practical wireless system. The dataset comprises 1,500 examples per modulation technique for each SNR level, ranging from -10 dB to 25 dB in steps of 5 dB. The resulting signals were normalized prior to training. This normalization is applied to both Time-domain IQ samples and delay–Doppler domain samples.

C. Training Process

The generated dataset was split into two subsets, where 80% of the dataset was used for training and the remaining 20% for validation testing. The weights were optimized using the Adam optimizer. During training, an exponential learning rate decay schedule was applied. The initial learning rate was set to 0.002 with a decay factor of 0.9. The decay was applied every 5 epochs to stabilize the convergence process. For this multi-class classification problem, the categorical cross-entropy loss

function was used. The training process was conducted for 30 epochs with a batch size of 128 samples.

IV. Result and Discussion

The performance of the proposed hybrid dual-stream CNN–LSTM fusion-based automatic modulation classification (AMC) scheme for OTFS systems over delay–Doppler channels is evaluated through simulations. Classification performance is evaluated in terms of classification accuracy (%). To improve classification capability, distinct pilot signals are embedded for different modulation schemes, enhancing overall accuracy. The key simulation parameters are listed in TABLE I. The multipath channel is modeled using the Extended Vehicular A (EVA) model, where path delays are defined accordingly and Doppler shifts for each delay tap are generated using Jakes' model.

TABLE I: Simulation Parameters and System Specifications

Parameters	Specifications
Carrier Frequency (f_c)	5 GHz
Subcarrier Spacing (Δf)	15 kHz
Sampling Frequency (f_s)	960 kHz
Delay-Doppler Grid ($M \times N$)	64 × 32
Channel Model	EVA
Maximum Doppler Spread ($f_{D,max}$)	555 Hz
Maximum Delay Spread ($l_{\max}$)	2
Number of Channel Taps	9

The performance evaluation of the proposed model is conducted on a test dataset comprising 28,800 samples.

The classification performance of the proposed dual-stream hybrid CNN–LSTM model for different modulation schemes across varying SNR levels is shown in Fig. 3. The overall classification accuracy increases significantly with SNR. Even at low SNRs (-10 dB and -5 dB), the model achieves robust performance due to the use of unique pilot signals for each modulation scheme, which enhance feature discriminability. Some confusion remains among higher-order modulation schemes, as indicated by the brighter off-diagonal elements in the confusion matrices. As the SNR increases to 0 dB and 5 dB, the classification performance further improves, with most modulation schemes achieving high accuracy.

The proposed model achieves near-perfect classification performance at SNR values of 10 dB and above. All modulation schemes attain accuracy exceeding 96% at 10 dB or higher SNR with several approaching 100%. This is confirmed by the confusion matrices (Fig. 5) where the matrices become strongly diagonal indicating highly accurate classification with less errors. Fig. 4(b) ensures that proposed model achieves over all test accuracy of greater than 90% for 5 dB. All these accuracy is mapped for completely unseen data.

The differences in classification accuracy across modulation schemes primarily come from the complexity of their constellations. OOK performs better than other schemes because its constellation points are well separated and less affected by noise. In the case of higher-order modulation schemes like 64QAM and 256QAM, they have closely packed constellation points, making them more vulnerable to noise and channel issues at low SNR values. As the SNR increases, the distinguishing features become clearer which leads to better classification accuracy for all modulation schemes.

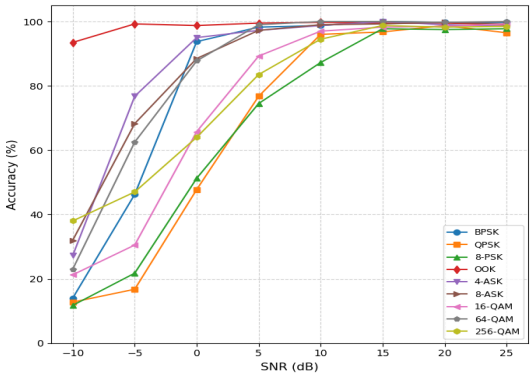
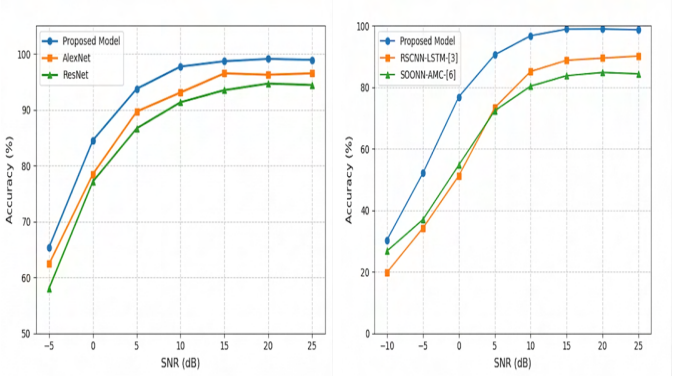


Fig. 3: Accuracy vs. SNR plot for various modulation schemes.

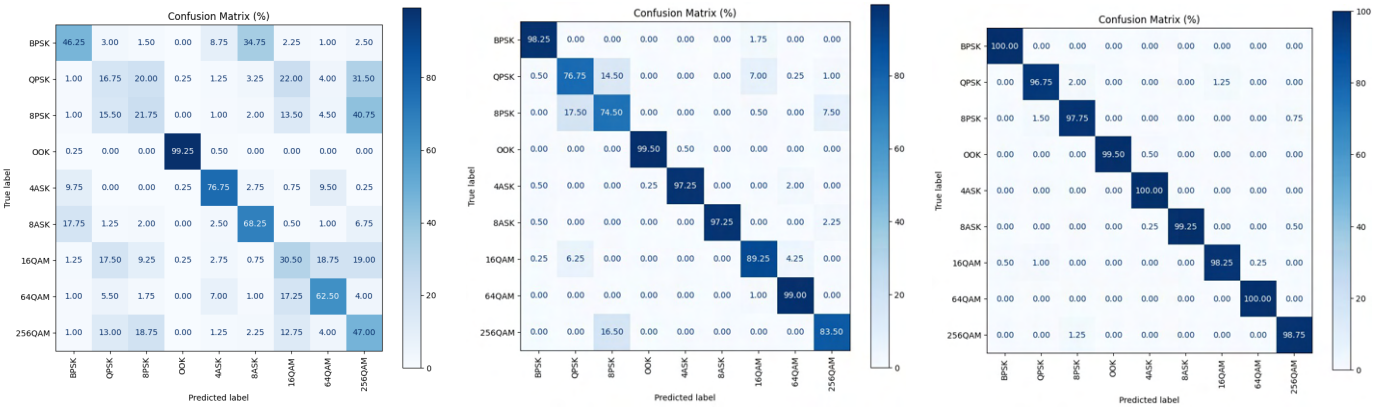
Resnet [10] and Alexnet [11] are two popular predefined models which can be applied in Automatic Modulation Classification (AMC). For the six-class modulation classification task consisting of BPSK, QPSK, 8PSK, OOK, 4ASK, and 8ASK, the proposed model achieves an overall classification accuracy of 91.73% over the testing dataset from SNR ranging -5 dB to 25 dB SNR. As the Fig. 4(a) demonstrate the same condition the proposed model outperforms both Resnet [10] and Alexnet [11] architectures across varying SNR conditions. The proposed model shows better robustness at low and moderate SNRs than the predefined models. This suggests that the proposed model is more accurate than Resnet and Alexnet models and adaptable with number of classification.



(a) (b)

Fig. 4: Performance comparison of proposed model with (a) predefined models and (b) existing research paper models.

Fig. 4(b) illustrates the classification performance comparison of the proposed model with two baseline models (RSCNN-LSTM-[3] and SOONN-AMC-[6]) across different SNR levels.



(a) (b) (c)

Fig. 5: Confusion matrices at (a) -5 dB SNR, (b) 5 dB SNR and (c) 15 dB SNR.

All models show improved performance as SNR increases. However, the proposed model maintains a high performance gap. For example at 0 dB SNR, the proposed model achieves approximately 77% accuracy while RSCNN-LSTM-[3] and SOONN-AMC-[6] achieve around 51% and 55%, respectively. At medium SNR levels (5–10 dB) the proposed model exceeds 90% accuracy while the baseline [3], [6] models stays behind by a large margin. The model proposed at higher SNR is capable of classifying modulation schemes in a near perfect way and exceeds the models of existing research articles [3], [6]. The proposed model is able to differentiate the higher order schemes such as 16QAM, 64QAM and 256QAM more accurately at SNR greater than 5 dB, which is challenging task for other pre-existing models. At low SNR values (-10 dB to -5 dB), the proposed model shows robustness in noisy environments due to the effective feature extraction capability of the hybrid architecture & the insertion of pilot signals. All these factors enhance the model discriminative capability. The proposed model approaches near-perfect classification accuracy ($\approx 98\%$) at high SNR values of 15 dB and above while RSCNN-LSTM-[3] and SOONN-AMC-[6] saturate at accuracy levels of approximately 90% and 85%. This shows that the proposed architecture learns more discriminative features and generalizes better at different channel conditions.

V. Conclusion

This work proposes a hybrid dual-stream CNN–LSTM-based AMC framework for OTFS systems over Delay–Doppler channels. Simulation results show that the proposed model achieves 96.78% accuracy at 10 dB and above, and more than 90% accuracy at SNR values greater than 5 dB on completely unseen data. It outperforms both existing research models and predefined architectures across all SNR levels. This performance gain is due to the dual-stream CNN–LSTM fusion and use of differentiated pilot signals, which enhances feature discriminability with reliable and efficient modulation classification in OTFS systems.

Acknowledgment

The authors would like to thank all individuals and institutions who contributed to this work through valuable discussions, technical support, and resources. Special appreciation is extended to the research community for their foundational contributions to OTFS and deep learning-based AMC.

References

[1] R. Hadani, S. Rakib, S. Kons, M. Tsatsanis, A. Monk, C. Ibars, J. Delfeld, Y. Hebron, A. J. Goldsmith, A. F. Molisch, and R. Calderbank, "Orthogonal Time Frequency Space Modulation," *Proc. IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 1–6, 2017.

[2] O. A. Dobre, A. Abdi, Y. Bar-Ness, and W. Su, "Survey of Automatic Modulation Classification Techniques: Classical Approaches and New Trends," *IET Communications*, vol. 1, pp. 137–156, 2007.

[3] A. Kumar, Manish, and U. Satija, "Residual Stack-Aided Hybrid CNN-LSTM-Based Automatic Modulation Classification for Orthogonal Time-Frequency Space System," *IEEE Communications Letters*, vol. 27, no. 12, pp. 3255–3259, 2023.

[4] R. Hadani, S. Rakib, A. F. Molisch, C. Ibars, A. Monk, M. Tsatsanis, J. Delfeld, A. Goldsmith, and R. Calderbank, "Orthogonal Time Frequency Space (OTFS) Modulation for Millimeter-Wave Communications Systems," *Proc. IEEE MTT-S International Microwave Symposium (IMS)*, pp. 681–683, 2017.

[5] B. A. Ozden, E. Aydin, E. Aslandogan, H. b. Ilhan, E. Basar, M. Wen, M. Di Renzo, and V. Poor, "A Survey of OTFS-Based Index Modulation Techniques: Challenges, Benefits, and Future Directions for 6G and Beyond," *arXiv*, 2025.

[6] S. Pratap Singh, M. Singh, and U. Satija, "Self-Organized Operational Neural Network-Based Modulation Classification for OTFS System," *IEEE Wireless Communications Letters*, vol. 15, pp. 320–324, 2026.

[7] M. Marchese, H. Wymeersch, P. Spallaccini, S. Chinnici, and P. Savazzi, "Reduced-Latency DL-Based Fractional Channel Estimation in OTFS Receivers," *IEEE Communications Letters*, pp. 1–6, 2023.

[8] J. Cheng, L. Dong, and M. Lapata, "Long Short-Term Memory-Networks for Machine Reading," *Proc. Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 551–561, 2016.

[9] P. Maurya, K. Captain, A. Parmar, and P. K. Shah, "Modulation Classification for OTFS-NOMA in Heterogeneous User Mobility Profile," *Proc. International Conference on Signal Processing and Communications (SPCOM)*, pp. 1–5, 2024.

[10] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016.

[11] M. Z. Alom, T. M. Taha, C. Yakopcic, S. Westberg, P. Sidike, M. S. Nasrin, B. C. Van Essen, A. A. S. Awwal, and V. K. Asari, "The History Began from AlexNet: A Comprehensive Survey on Deep Learning Approaches," *Electronics*, vol. 8, no. 3, p. 292, 2019.