ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940

www.ijasem.org

Vol 19, Issue 2, 2025

Dendritic Mixture-of-Experts (D-MoE): A Robust Architecture for Low-SNR Radio Signal Classification

Dr. P. Dhana Lakshmi[1], K. Hara Naga Sai Satyanarayana[2], Y. Adarsh[3], Ch. Mahesh[4]

Assistant Professor[1], UG Students[2,3,4]
Department of Electronics and Communication Engineering,
Acharya Nagarjuna University College of Engineering and Technology,
Nagarjuna Nagar, Guntur, Andhra Pradesh, INDIA

ABSTRACT

Automatic Modulation Classification (AMC) is vital for cognitive radio and spectrum monitoring, but standard deep learning models often fail at low Signal-to-Noise Ratios (SNR < 0 dB). This paper introduces D-MoE, a novel Mixture-of-Experts architecture designed for enhanced robustness across diverse SNR conditions. D-MoE employs a pre-trained Denoising Autoencoder (DAE) to extract noise-resilient features, feeding them to a specialized low-SNR expert (DendriticGRU), while a separate high-SNR expert (ResNet-SE) processes raw I/Q data. An SNR-based gate routes signals appropriately. Trained on the RadioML 2016.10a dataset, D-MoE achieves a strong overall test accuracy of 62.54%. Significantly, it demonstrates improved low-SNR performance, reaching approximately 50% accuracy at -6 dB and 73% at -2 dB, while maintaining excellent accuracy (>93%) at high SNRs (+10 dB and above). This performance surpasses baseline CNN (34%) and non-denoising MoE (57%) models, validating the D-MoE approach of combining feature denoising with expert specialization for robust wide-range AMC.

Keywords: Automatic Modulation Classification, Deep Learning, Mixture-of-Experts, Denoising Autoencoder, Low SNR, RadioML, Cognitive Radio, Robust Signal Processing.

1. INTRODUCTION

The escalating demand for wireless connectivity across diverse applications, from mobile broadband and Internet of Things (IoT) to satellite communications and radar systems, exerts immense pressure on the finite radio frequency (RF) spectrum. Efficient and intelligent utilization of this resource is paramount. Automatic Modulation Classification (AMC), the task of identifying the modulation scheme (e.g., BPSK, QAM, FM) of an unknown intercepted signal without prior knowledge of its parameters, serves as a fundamental enabling technology for enhancing spectral awareness and adaptability [1]. By determining how information is encoded onto a carrier wave, AMC enables enhanced spectrum situational awareness. This awareness is fundamental for advanced wireless applications, including cognitive radio (where radios intelligently adapt to their environment), dynamic spectrum access (DSA), spectrum monitoring and enforcement (identifying signals and detecting interference or unauthorized use), electronic warfare (EW) signal intelligence (SIGINT), and ensuring compatibility in heterogeneous wireless networks [2, 3, 4].

Historically, AMC methodologies were rooted in classical signal processing and statistical decision theory. These approaches involved meticulous extraction of features designed to capture specific characteristics of different modulation types. Examples include analyzing higher-order statistics like cumulants and moments, or exploiting cyclostationary properties inherent in modulated signals due to underlying periodicities (e.g., symbol rate, carrier frequency) [5]. These features, while theoretically grounded, often require substantial domain expertise for feature design and selection, can be sensitive to inaccuracies in parameter estimation (like carrier frequency or symbol timing offsets), and critically, their performance typically degrades substantially in the presence of noise, fading, and interference – conditions ubiquitous in real-world wireless channels, especially at low Signal-to-Noise Ratios (SNRs).

The application of deep learning (DL) marked a paradigm shift in AMC research. Inspired by successes in computer vision and other domains, researchers began applying deep neural network architectures, particularly Convolutional Neural Networks (CNNs), directly to the raw complex baseband (In-phase and Quadrature, I/Q) representation of radio signals [2, 6]. This end-to-end learning approach demonstrated the powerful ability of DL models to automatically learn relevant hierarchical features from the data, often outperforming classical feature-based methods, especially when sufficient signal power (moderate-to-high SNR) is available. The release of benchmark datasets like RadioML [2] further spurred development in this area. However, despite rapid progress in DL architectures for AMC [7, 8], a persistent and significant challenge remains: the vulnerability of these models to noise. Most standard deep learning models exhibit a sharp decline in classification accuracy as the SNR drops below approximately 0 dB [1], severely limiting their utility in many practical scenarios involving weak signals, long distances, or noisy/interfered channels.

Addressing this crucial low-SNR performance gap is the primary motivation for this work. We propose a novel hybrid

deep learning architecture, the Dendritic Mixture-of-Experts (D-MoE), specifically conceived to enhance AMC robustness across a broad SNR spectrum (–20 dB to +18 dB), with a particular emphasis on the challenging sub-0 dB regime. D-MoE employs a multi-faceted strategy inspired by principles of noise reduction and specialized processing. It incorporates: 1) An explicit denoising stage using a pre-trained Denoising Autoencoder (DAE) [9] to learn noise-invariant feature representations. 2) A Mixture-of-Experts framework where distinct neural network "experts" are specialized for different SNR conditions. 3) A deterministic gating mechanism based on the input SNR to route signals to the most appropriate expert. Expert A utilizes a ResNet-SE architecture [10, 11] for high-SNR signals, while Expert B employs a unique DendriticGRU structure fed by DAE features for low-SNR signals [12]. This strategic fusion of denoising and expert specialization aims to provide resilient classification accuracy where standard approaches fail.

2. RELATED WORK

The pursuit of accurate and robust AMC has led to diverse research directions, from classical signal processing to advanced deep learning paradigms.

1) Classical Feature-Based Methods: Initial AMC research focused on identifying features robust to noise and channel variations. Techniques included statistical analysis of instantaneous amplitude, phase, and frequency; utilization of higher-order statistics like cumulants and moments which possess theoretical invariance to Gaussian noise [5]; and methods based on cyclostationarity, exploiting the inherent periodicities in communication signals. These features were typically classified using traditional machine learning algorithms like SVMs or decision trees. While foundational, these methods often require precise parameter estimation and struggle significantly at low SNRs where feature extraction becomes unreliable.

2) Foundational Deep Learning Approaches: The success of deep learning prompted its application to AMC, treating I/Q sequences as time-series data. O'Shea et al. [2, 6] demonstrated the power of end-to-end learning using relatively shallow CNNs applied directly to raw I/Q samples from their RadioML datasets. This approach eliminated the need for manual feature engineering and achieved superior performance at moderate-to-high SNRs. Subsequent studies explored deeper CNN architectures [13], investigated methods for faster inference [14], and developed complex-valued CNNs [7] that naturally handle complex I/Q inputs and achieved very high accuracy, primarily driven by high-SNR performance. However, the performance gap at low SNRs persisted across most of these architectures [1].

3) Hybrid CNN-RNN Architectures: To better model the temporal nature of modulated signals, hybrid architectures combining CNNs for spatial/local feature extraction with Recurrent Neural Networks (RNNs) for sequential modeling were proposed. CLDNNs, integrating CNNs, LSTMs, and DNNs [15, 8], and models using GRUs [12, 16], showed promise in capturing temporal dependencies, potentially improving performance for certain modulation types.

4) Attention Mechanisms and Transformers: Leveraging breakthroughs in natural language processing, attention mechanisms and Transformer architectures were explored for AMC. Attention allows models to dynamically focus on the most salient parts of the input sequence [17]. Transformers, based entirely on self-attention mechanisms, can capture long-range dependencies effectively. Various adaptations, including combinations with CNNs and Graph Neural Networks (GNNs) like CTGNet [18], have been proposed. Multi-modal approaches like IQFormer [19], which fuse information from raw I/Q, spectrograms, and constellations via Transformers, currently represent the state-of-the-art in terms of peak accuracy on benchmarks, though often at the cost of significant computational complexity.

5) Mixture-of-Experts and Dendritic Models: Recognizing that a single monolithic network might struggle across diverse conditions, modular approaches have been explored. The MoE concept, using specialized experts, was introduced to AMC by Gao et al. [20], who gated between Transformer and ResNet experts based on estimated SNR. Biologically inspired dendritic structures, mimicking neural signal integration, were incorporated into hybrid models by Yin et al. [21], showing performance benefits. Our D-MoE architecture builds upon the MoE principle but differs significantly by incorporating an explicit DAE denoising stage specifically to provide noise-robust features to the low-SNR expert (DendriticGRU).

6) Alternative Input Representations: A distinct line of research avoids direct I/Q processing by first transforming the signal into a 2D representation. Spectrograms (time-frequency images) processed by standard or adapted computer vision models (CNNs [22], Vision Transformers [23]) have proven effective. Similarly, constellation diagrams visualized as images have been used with image classification techniques [24]. While successful, these methods rely on the quality of the transformation and differ fundamentally from approaches processing the 1D time-domain signal.

7) Advanced Training Strategies and Architectures: Other research directions include improving performance through advanced training techniques or novel architectures. Data augmentation specific to RF signals, particularly using generative models like GANs [25] or recent diffusion models [26, 27], aims to improve model generalization. Self-supervised learning methods offer ways to pre-train models on unlabeled data. Adversarial training seeks to improve robustness against intentional interference [28, 32]. Other novel architectures like Capsule Networks and Graph Convolutional Networks continue to be explored for AMC tasks.

8) Surveys and Context: Several recent survey papers [3, 1, 4] provide essential context, benchmark various approaches, and consistently highlight robust low-SNR performance as a key remaining challenge, motivating the development of architectures like the proposed D-MoE focused specifically on this problem. Additionally, work on spectrum sensing techniques [29, 30, 31] provides broader context on related cognitive radio functionalities.

3. DATASET AND PREPROCESSING

3.1 RadioML 2016.10a Dataset

This study utilizes the widely adopted RadioML 2016.10a dataset [2]. This dataset serves as a standard benchmark, fa-

cilitating comparisons across different studies. It comprises 220,000 synthetically generated signal examples. Each example represents a short segment of a radio signal captured as 128 complex time samples (I/Q data), resulting in an input shape of (2, 128) per example. The dataset features 11 distinct modulation formats: analog modulations AM-DSB, AM-SSB, WBFM; and digital modulations BPSK, QPSK, 8PSK, PAM4, GFSK, CPFSK, QAM16, QAM64. A key attribute is its coverage of a broad range of signal quality conditions simulated through varying Signal-to-Noise Ratios (SNRs). Samples are provided at 20 discrete SNR levels, ranging from -20 dB to +18 dB in uniform 2 dB increments. The dataset is balanced, containing 1000 distinct examples for each modulation type at each SNR level.

3.2 Data Partitioning Strategy

To ensure rigorous model training and unbiased evaluation, the dataset was partitioned into three distinct subsets: training, validation, and testing. A stratified splitting approach based on the modulation type label was employed during splitting to maintain the proportional representation of each class across all three subsets. The specific split ratios used were 70% for the training set (resulting in 154,000 samples), 15% for the validation set (33,000 samples), and 15% for the test set (33,000 samples).

3.3 Preprocessing Steps

Standard preprocessing techniques were applied to the raw I/Q data before feeding it into the neural networks. Firstly, each I/Q sequence (2x128) was standardized channel-wise using statistics derived solely from the training set. Secondly, categorical modulation labels were converted into integer representations [0-10]. Finally, during training only, dynamic data augmentations were applied to batches, including AWGN addition (20 dB SNR, p=0.5) and random Time Shifts (±10 samples, p=0.5).

4. PROPOSED D-MOE ARCHITECTURE

The D-MoE architecture is predicated on the hypothesis that optimal AMC performance across a wide SNR range necessitates both noise mitigation and specialized processing. To this end, it integrates a Denoising Autoencoder (DAE) front-end with a Mixture-of-Experts (MoE) framework containing distinct high-SNR and low-SNR processing pathways, coordinated by an SNR-based gate. The overall structure is visualized in Figure 1.

4.1 Denoising Autoencoder (DAE)

The DAE forms the foundation of the low-SNR processing path, tasked with learning noise-resilient signal representations through unsupervised pre-training. Its primary function extends beyond simple data compression; it is specifically pre-trained to learn a transformation that maps noisy input signals to a latent representation emphasizing robust, underlying signal characteristics while suppressing noise artifacts [9].

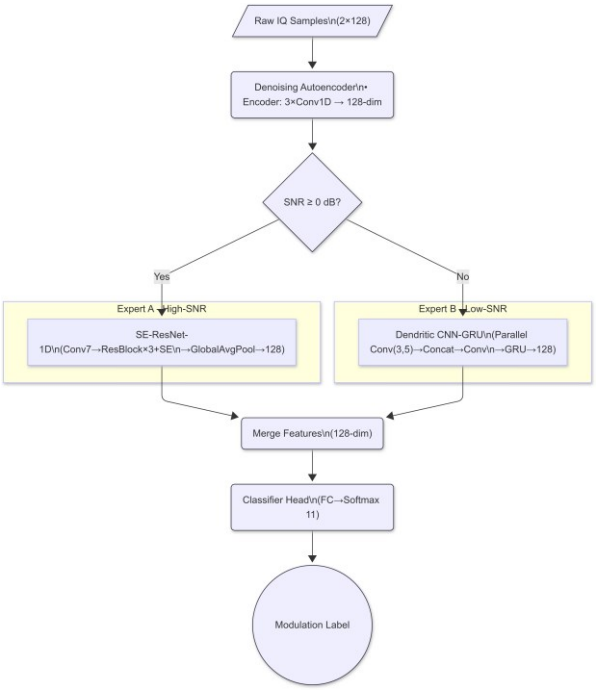

Fig 1: Block diagram of the proposed D-MoE architecture, illustrating data flow through the DAE, SNR Gate, Expert A (ResNetSE), Expert B (DendriticGRU), and Classification Head.

4.1.1 Rationale

Operating directly on heavily noise-corrupted signals (low SNR) presents significant challenges for classifiers. Noise can obscure discriminative features and lead to poor generalization. By pre-training a DAE on a denoising task (reconstructing clean signals from noisy versions), we force the encoder component to learn features that are inherently less sensitive to noise variations. This effectively provides a "cleaned" or "denoised" feature set for subsequent processing steps, specifically for the low-SNR expert pathway.

4.1.2 Encoder Architecture Details

The encoder maps the input (Batch x 2 x 128) to a latent feature map (Batch x 128 x 32). This is achieved through three sequential convolutional blocks. The first block uses a 1D convolution (in=2, out=32, kernel=7, padding=3), followed by Batch Normalization, ReLU activation, and Max Pooling (kernel=2), reducing sequence length to 64. The larger kernel size captures broader initial patterns. The second block applies a similar sequence (Conv1D in=32, out=64, k=5, p=2; BN; ReLU; MaxPool k=2), reducing length to 32 and refining features with a medium kernel size. The final block performs feature extraction with a smaller kernel (Conv1D in=64, out=128, k=3, p=1; BN; ReLU), resulting in an output feature map of 128 channels and length 32. Batch Normalization stabilizes training, ReLU provides non-linearity, and Max Pooling reduces dimensionality and offers minor translation invariance.

4.1.3 Decoder Architecture Details

The decoder symmetrically reconstructs the input shape using transposed convolutions and upsampling. The first decoder block applies a transposed convolution (in=128, out=64, k=3, p=1), followed by Batch Normalization, ReLU, and Upsampling (factor=2), increasing length to 64. The second block

continues this pattern (ConvTranspose1d in=64, out=32, k=5, p=2; BN; ReLU; Upsample factor=2), increasing length back to 128. The final layer uses a transposed convolution (in=32, out=2, k=7, p=3) to output the reconstructed 2x128 signal.

4.1.4 Pre-training and Freezing

The DAE is pre-trained for 15 epochs minimizing the MSE loss between the reconstructed output and the original (cleaner) input signal, using the Adam optimizer with a learning rate schedule (1e-3 reduced to 1e-4). After this phase, the trained encoder weights are frozen. This crucial step preserves the learned denoising capability, ensures the low-SNR expert receives consistent features, improves computational efficiency during classifier training, and acts as a regularizer. The frozen encoder's output (128x32 feature map) becomes the input for Expert B.

4.2 Expert A (High-SNR ResNetSE)

Expert A is tailored for classifying signals with relatively high quality (SNR ≥ 0 dB), operating directly on raw I/Q data. It employs a robust ResNet-style architecture [10] with integrated Squeeze-and-Excitation (SE) blocks [11].

4.2.1 Rationale

For relatively clean signals, a powerful feature extractor capable of learning complex hierarchies is beneficial. ResNets enable training very deep networks via skip connections, preventing vanishing gradients. SE blocks add channel attention, allowing the network to adaptively focus on the most informative feature channels for the specific input, potentially improving discrimination between similar modulation types.

4.2.2 Architecture Details

Expert A follows a 1D ResNet pattern, taking the standardized raw I/Q (Batch x 2 x 128) as input. An initial block (Conv1D k=7, s=2, out=64; BN; ReLU; MaxPool1D k=3, s=2) rapidly reduces sequence length (128→64→32) and increases channel depth. This is followed by three sequential residual stages using `ResidualBlock1D_SE` modules, with 2 blocks per stage. Channels progress from 64 to 64 (Stage 1, length 32), then 64 to 128 (Stage 2, length 16, with downsampling via stride=2 at the start), and finally 128 to 128 (Stage 3, length 16). Each `ResidualBlock1D_SE` module implements the standard Conv-BN-ReLU-Conv-BN structure (using kernel size 3), followed by an SE module (Global Avg Pool → FC-ReLU → FC-Sigmoid → Rescale), before adding the skip connection and applying the final ReLU. Finally, Global Average Pooling (`AdaptiveAvgPool1d(1)`) aggregates the final stage's feature maps into a 128-dimensional vector.

4.2.3 Output

The 128-dimensional feature vector from Global Average Pooling serves as the high-SNR signal representation for the gating mechanism.

4.3 Expert B (Low-SNR DendriticGRU)

Expert B specializes in low-SNR (SNR < 0 dB) classification, utilizing features from the frozen DAE encoder. Its architecture uniquely combines multi-scale CNN processing with GRU-based temporal modeling.

4.3.1 Rationale

Operating on denoised features allows Expert B to focus on residual discriminative patterns. The "dendritic" parallel CNN paths with different kernel sizes (k=3, k=5) allow the network to simultaneously analyze features at different temporal scales, potentially capturing noise-obscured information. The subsequent GRU [12] models temporal dependencies within this enhanced feature sequence.

4.3.2 Architecture Details

The input to Expert B is the frozen DAE encoder output (Batch x 128 Channels x 32 Length). The Dendritic CNN component consists of two parallel paths: Path 1 applies a 1D convolution (128→32, k=3, p=1) followed by BN and ReLU, while Path 2 uses a larger kernel (Conv1D 128→32, k=5, p=2) followed by BN and ReLU. The outputs of these paths are concatenated along the channel dimension (resulting in 64 channels) and passed through an integration layer (Conv1D 64→64, k=3, p=1; BN; ReLU), producing a feature map of shape (Batch x 64 x 32). For the GRU layer, this feature map is permuted to (Batch x 32 Length x 64 Features) to match the `batch_first=True` expectation. A single GRU layer with an input size of 64 and a hidden size of 128 processes this sequence.

4.3.3 Output

The final output of Expert B is the 128-dimensional final hidden state (`h_n`) of the GRU layer.

4.4 Gating and Classification Head

This component selects the appropriate expert features and makes the final prediction.

4.4.1 Gate

A deterministic hard gate is used, based on the input signal's SNR label. It selects features from Expert A if the SNR is greater than or equal to 0 dB, otherwise it selects features from Expert B.

4.4.2 Head

The classification head consists of a single Linear layer that maps the selected 128-dimensional expert features to 11 output classes, followed by a Softmax activation function to produce probabilities.

5. IMPLEMENTATION DETAILS

5.1 Platform and Software

The D-MoE model was implemented using PyTorch (v1.8+). All training and evaluation were performed within the Google Colaboratory environment, utilizing NVIDIA Tesla T4 GPUs provided by the platform. Standard Python libraries including NumPy, Scikit-learn, and Matplotlib were used for data manipulation, evaluation metrics, and visualization.

5.2 Data Handling

Custom PyTorch 'Dataset' classes were created for the training, validation, and test splits. 'DataLoader' instances were configured with a 'batch_size' of 96. Due to potential resource limitations and stability issues observed with multi-processing data loading in the Colab environment, 'num_workers' was set to 0 (disabling parallel loading) and 'pin_memory' was set to 'False'. While ensuring stability, this configuration limits data loading throughput compared to optimal settings. Data standardization statistics (mean, std dev) were computed only on the training set and applied consistently. Augmentations (AWGN, Time Shift) were applied randomly during training only.

5.3 Training Procedures

The training strategy involved two distinct phases. First, the DAE was pre-trained (Phase 1) solely for signal reconstruction over 15 epochs using MSE loss. The Adam optimizer ($\beta_1 = 0.9, \beta_2 = 0.999$) was used with an initial learning rate of 1e-3. The 'ReduceLROnPlateau' scheduler monitored the validation MSE loss with a patience of 3 epochs and a reduction factor of 0.1, decreasing the LR towards the end of pre-training. Second, the classifier was trained (Phase 2). The DAE encoder weights were frozen, and the parameters of Expert A, Expert B, and the Classification Head were trained jointly for 50 epochs. The objective function was Cross-Entropy loss with label smoothing ($\epsilon = 0.1$). The Adam optimizer was used again with an initial learning rate of 1e-4 and weight decay (L2 penalty) of 1e-5. The 'ReduceLROnPlateau' scheduler monitored validation accuracy, reducing the learning rate (factor=0.1, patience=5, min_lr=1e-6) when improvement stalled. The model checkpoint achieving the highest validation accuracy (Epoch 42, 62.28%) was saved as the final model for evaluation.

6. EVALUATION PROTOCOL

Performance was evaluated using a comprehensive suite of metrics on the unseen test set. This included overall accuracy (the fraction of correctly classified samples across all classes and SNRs), per-SNR accuracy (calculated individually for each SNR level from -20 dB to +18 dB to assess robustness), and per-class metrics (Precision, Recall/Sensitivity, F1-Score, Specificity calculated for each modulation type over the entire test set, along with Macro and Weighted averages). Additionally, an overall confusion matrix was generated to visualize misclassification patterns between modulation types across all SNRs. Finally, DAE feature visualization using t-SNE and PCA was performed on the frozen DAE encoder's output features (for a subset of test data), colored by class and SNR, to qualitatively assess the learned representation's structure and noise robustness.

7. RESULTS AND ANALYSIS

7.1 Overall and Per-SNR Performance

Evaluation of the best D-MoE model on the held-out test set yielded an overall classification accuracy of 62.54%. This represents a strong performance level on this challenging benchmark dataset. The detailed per-SNR accuracy is presented in Table 1 and visualized in Figure 2. The plot highlights the model's effectiveness across the SNR range. Accuracy degrades gracefully below 0 dB, achieving 50% at -6 dB and 73% at -2 dB, significantly better than the rapid collapse often seen in standard models. Above the 0 dB gating threshold, performance rapidly increases, surpassing 92% accuracy for SNRs ≥ +4 dB, confirming that the high-SNR expert path functions effectively. This demonstrates the success of the MoE strategy in combining robustness at low SNRs with high performance at high SNRs.

Table 1: PER-SNR TEST ACCURACY (%) ON RADIOML 2016.10A

SNR (dB)	Acc. (%)	SNR (dB)	Acc. (%)	SNR (dB)	Acc. (%)
-20	10.66	-6	50.28	8	92.85
-18	9.47	-4	70.50	10	93.61
-16	11.14	-2	72.84	12	93.62
-14	15.96	0	88.72	14	92.90
-12	17.17	2	91.95	16	93.56
-10	27.05	4	92.35	18	92.59
		6	92.39		

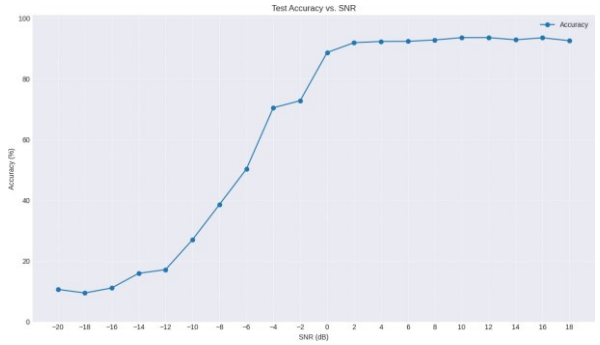

Fig 2: Test Accuracy vs. SNR for the D-MoE model.

7.2 Baseline Comparisons

The efficacy of the D-MoE design choices is highlighted by comparisons against baseline models (Table 2). D-MoE's 62.54% accuracy represents a substantial improvement over a standard CNN baseline (34%) implemented under similar conditions. More significantly, it outperforms an ablated version of the architecture lacking the DAE component (MoE w/o DAE: 57%) by over 5 absolute percentage points, directly quantifying the benefit derived from the explicit denoising feature extraction for the low-SNR expert branch. Compared to published results, D-MoE surpasses the overall accuracy reported for the foundational VT-CNN2 model [2] and offers a more balanced performance profile across SNRs compared to models like OPResNet-18 [1], which prioritize peak high-SNR accuracy at the cost of severe low-SNR degradation.

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT
ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

Table 2: COMPARISON WITH BASELINE MODELS

Model	Overall Acc (%)	Notes
Simple CNN (Baseline)	34%	Our implementation
MoE w/o DAE (Baseline)	57%	Our implementation
VT-CNN2 [2]	≈61%	Literature value
OPResNet-18 [1]	94% (High SNR)	Literature value
	5-18% (< -10dB)	
D-MoE (Proposed)	**62.54%**	This work

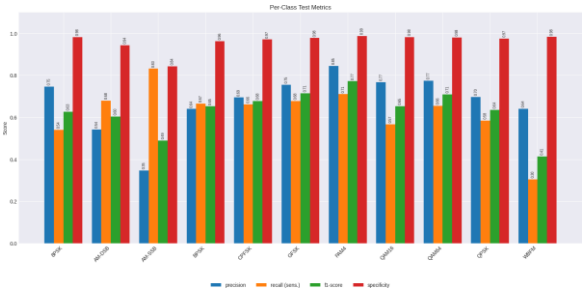

Fig 4: Per-class test metrics (Precision, Recall, F1-Score, Specificity).

7.3 DAE Effectiveness

The quantitative improvement shown in Table 2 underscores the DAE's importance. Qualitatively, the t-SNE and PCA visualizations of the DAE encoder's latent features (Fig. 3) provide further evidence of its utility. The visualizations demonstrate that the encoder learns a structured representation that captures both modulation-specific information (evident from partial class clustering in t-SNE) and SNR-related characteristics (evident from the clear gradient in PCA). This confirms that the DAE transforms the noisy input into a feature space that is more amenable to classification, particularly for the low-SNR expert.

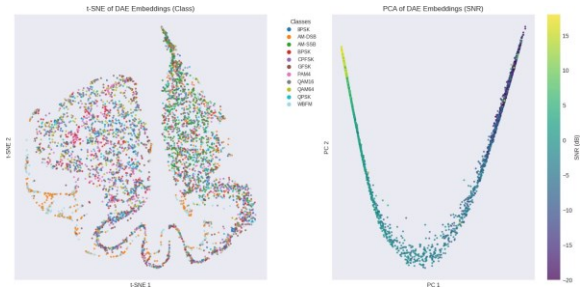

Fig 3: t-SNE (left, colored by class) and PCA (right, colored by SNR) visualizations of DAE latent features.

7.4 Per-Class Analysis

A detailed breakdown of performance for each modulation type is provided in Table 3 and Fig. 4. D-MoE achieves high F1-scores (above 0.65) for several digital modulations, including PAM4, GFSK, CPFSK, BPSK, QAM16, and QAM64, indicating effective classification for these types. Analog modulations, particularly AM-SSB (F1=0.490) and WBFM (F1=0.413), present greater difficulty, a common finding in AMC literature. The overall confusion matrix (Fig. 5) visually details these classification tendencies, highlighting significant confusion between WBFM and AM signals, and also between QAM16 and QAM64, especially at lower SNRs where their constellation differences become less distinct.

Table 3: PER-CLASS PERFORMANCE METRICS (TEST SET)

Class	Precision	Recall	F1-Score	Specificity
8PSK	0.746	0.540	0.627	0.982
AM-DSB	0.543	0.679	0.603	0.943
AM-SSB	0.347	0.832	0.490	0.844
BPSK	0.640	0.665	0.653	0.963
CPFSK	0.695	0.662	0.678	0.971
GFSK	0.755	0.678	0.714	0.978
PAM4	0.845	0.710	0.772	0.987
QAM16	0.767	0.568	0.653	0.983
QAM64	0.774	0.655	0.710	0.981
QPSK	0.697	0.585	0.636	0.975
WBFM	0.641	0.305	0.413	0.983
Accuracy		0.6254		
Macro Avg	0.677	0.625	0.632	N/A
Weighted Avg	0.677	0.625	0.632	N/A

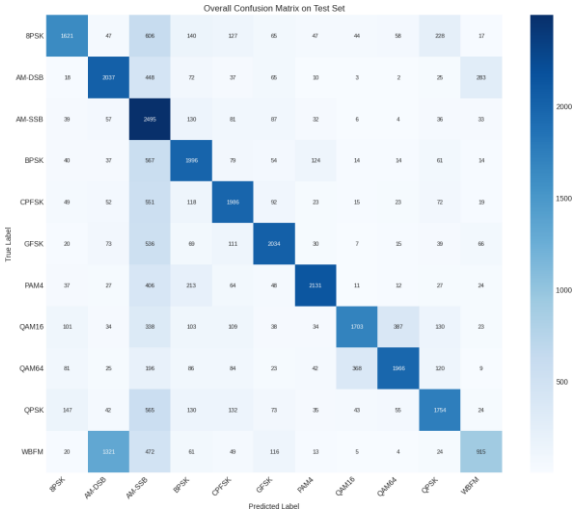

Fig 5: Overall normalized confusion matrix on the test set.

8. DISCUSSION

The comprehensive evaluation confirms the effectiveness of the D-MoE architecture and its underlying design principles. The core achievement is the demonstrably improved robustness to noise compared to baseline approaches. The significant accuracy gain over the MoE variant without the DAE (+5.5% absolute accuracy) provides strong quantitative evidence for the value of explicit noise-robust feature extraction via the pre-trained DAE before low-SNR classification. This allows

Expert B (DendriticGRU) to operate on a cleaner representation, focusing its capacity on discerning modulation patterns rather than combating noise simultaneously. The DendriticGRU structure within Expert B, combining multi-scale convolutional analysis with GRU-based temporal modeling, appears well-suited for this task.

Simultaneously, the ResNetSE architecture of Expert A ensures that performance is not compromised at high SNRs; operating directly on raw I/Q, it achieves accuracy levels (>92-93%) consistent with strong deep learning models in favorable conditions. The smooth transition and sustained high performance seen in the per-SNR accuracy curve (Fig. 2) validate the overall MoE strategy facilitated by the deterministic SNR gate. The overall accuracy of 62.54% positions D-MoE favorably against foundational models like VT-CNN2 and offers a more balanced performance profile across SNRs than models heavily optimized only for high-SNR regimes.

However, the analysis also reveals limitations and areas for future investigation. While significantly improved, performance in the very low SNR range (below -10 dB) remains modest, suggesting that extreme noise levels still pose a considerable challenge even with the DAE preprocessing. The overall accuracy, while respectable, does not match the highest reported figures achieved by significantly more complex Transformer-based architectures or multi-modal fusion models [19], representing a potential trade-off between the D-MoE's architectural complexity and maximum achievable performance. The deterministic hard gate at 0 dB, while simple and effective here, represents a point of fragility; its performance relies on accurate SNR knowledge (available in the dataset but needing estimation in practice) and may induce abrupt transitions for signals near the threshold. A learned soft gate could potentially offer more graceful adaptation and robustness to SNR estimation errors. Furthermore, the persistent difficulty in distinguishing specific modulation groups (WBFM/AM, QAM16/QAM64) suggests that the feature representations, even after denoising and expert processing, may not fully capture the most subtle discriminating characteristics under all conditions. Addressing these specific confusions might require future architectural refinements or feature engineering more tailored to these difficult classes. Finally, the practical limitation imposed by the `num_workers=0` configuration highlights challenges in optimizing training efficiency, particularly in resource-constrained environments like Google Colab.

9. CONCLUSION

This paper presented D-MoE, a novel Dendritic Mixture-of-Experts architecture designed to enhance the robustness of Automatic Modulation Classification, particularly against the deleterious effects of low Signal-to-Noise Ratios. By strategically integrating a pre-trained Denoising Autoencoder for noise-resilient feature extraction with specialized, SNR-gated expert networks—a ResNetSE for high SNRs and a DendriticGRU for low SNRs—the proposed architecture achieves a strong balance between low-SNR robustness and high-SNR accuracy. Evaluated rigorously on the standard RadioML 2016.10a benchmark dataset, D-MoE achieved a competitive overall test accuracy of 62.54%. It significantly outperformed relevant baseline models, including a comparable MoE structure without the DAE component, quantitatively validating the benefit of the integrated denoising strategy. The per-SNR analysis confirmed substantially improved resilience in the challenging sub-0 dB regime compared to typical models, while high-SNR performance remained excellent. This work validates the hypothesis that combining explicit denoising with expert network specialization is a highly effective strategy for building more reliable and practical AMC systems capable of operating effectively across diverse and noisy wireless environments. The D-MoE architecture represents a valuable contribution towards robust signal classification for cognitive radio and related applications.

ACKNOWLEDGMENT

The authors wish to express their sincere gratitude to their project supervisor, Dr. P. Dhana Lakshmi, for the invalu- able guidance and support provided throughout this research project.

REFERENCES

[1] F. Zhang, *et al.*, "Deep Learning Based AMC: Models, Datasets, and Challenges," *arXiv:2207.09647*, 2022.

[2] T. J. O'Shea, T. Roy, and T. C. Clancy, "Over-the-Air Deep Learning Based Radio Signal Classification," *IEEE JSTSP*, vol. 12, no. 1, pp. 168–179, 2018.

[3] E. Jafarigol, *et al.*, "AI/ML-Based AMC: Recent Trends & Future Possibilities," *arXiv:2502.05315*, 2025.

[4] T. J. O'Shea and J. Hoydis, "An Introduction to Deep Learning for the Physical Layer," *IEEE JSAC*, vol. 39, no. 6, pp. 1344–1367, 2021.

[5] L. Qian and C. Zhu, "Modulation Classification Based on Cyclic Spectral Features and Neural Network," in *3rd Int. Congress Image & Signal Proc.*, 2010.

[6] T. C. Clancy, T. J. O'Shea, and N. West, "Convolutional Radio Modulation Recognition Networks," in *Proc. Int. Conf. Eng. Appl. Neural Netw. (EANN)*, 2016, pp. 215-226.

[7] J. Krzyston, R. Bhattacharjea, and A. Stark, "High-Capacity Complex Convolutional Neural Networks for I/Q Modulation Classification," *arXiv:2010.10717*, 2020.

[8] 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, Sept. 2018.

[9] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, "Extracting and Composing Robust Features with Denoising Autoencoders," in *Proc. 25th Int. Conf. Machine Learning (ICML)*, 2008, pp. 1096-1103.

[10] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 770-778.

[11] J. Hu, L. Shen, and G. Sun, "Squeeze-and-Excitation Networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2018, pp. 7132–7141.

[12] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation," in *Proc. Conf. Empir. Methods in Natural Lang. Proc. (EMNLP)*, 2014, pp. 1724–1734.

[13] X. Liu, D. Yang, and A. El Gamal, "Deep Neural Network Architectures for Modulation Classification," *arXiv:1712.00443*, 2017.

[14] S. Ramjee, S. Ju, D. Yang, X. Liu, A. E. El Gamal, and Y. C. Eldar, "Fast Deep Learning for Automatic Modulation Classification," *arXiv:1901.05850*, 2019.

[15] T. N. Sainath, A. Mohamed, B. Kingsbury, and B. Ramabhadran, "Convolutional, Long Short-Term Memory, and DNN (CLDNN) Combinations for LVCSR," in *ICASSP*, 2015.

[16] M. Alazab and J. Hu, "Fusion GRU Deep Learning Neural Network (FGDNN) for AMC." *EURASIP J. Wireless Commun. Netw.*, vol. 2023, p. 16, 2023.

[17] X. Wu, S. Wei, and Y. Zhou, "Deep Multi-Scale Representation Learning with Attention for AMC," *arXiv:2209.03764*, 2022.

[18] Z. Zou, *et al.*, "CTGNet: CNN-Transformer Graph Neural Network for Modulation Classification," *IEEE Trans. Circuits Syst. I*, vol. 70, no. 11, pp. 4546-4558, Nov. 2023.

[19] M. Shao *et al.*, "IQFormer: A Novel Transformer-Based Model With Multi-Modality Fusion for AMC," *arXiv:2503.00123*, 2025.

[20] J. Gao, Q. Cao, and Y. Chen, "MoE-AMC: Enhancing AMC Performance Using Mixture-of-Experts," *arXiv:2312.02298*, 2023.

[21] P. Yin, *et al.*, "Analysis on Dendritic Deep Learning Model for AMR Task," *Cybersecurity*, vol. 7, p. 77, 2024.

[22] M. Waqas, M. Ashraf, and M. Zakwan, "Modulation Classification Through Deep Learning Using Resolution-Transformed Spectrograms," *arXiv:2306.04655*, 2023.

[23] X. Zhang, *et al.*, "FE-SKVIT: Vision Transformer on IQ Spectrograms for AMC," *IEEE Trans. Multimedia*, vol. 25, pp. 9174-9186, 2023.

[24] Y. Sun and E. Ball, "AMC Using Techniques from Image Classification," *IET Communications*, vol. 16, no. 11, pp. 1303–1314, 2022.

[25] I. Goodfellow, *et al.*, "Generative Adversarial Nets," in *NIPS*, 2014, pp. 2672–2680.

[26] Y. Xu, L. Huang, L. Zhang, L. Qian, and X. Yang, "Diffusion-Based Radio Signal Augmentation for AMC," *Electronics*, vol. 13, no. 11, p. 2063, 2024.

[27] M. Mirmozafari and X. Li, "DiRSA: Diffusion-Based Radio Signal Augmentation for AMC," *arXiv:2401.13992*, 2024.

[28] L. Yang, *et al.*, "Adversarial Training for AMC Robustness Against Interference," *IEEE Trans. Wireless Commun.*, vol. 21, no. 9, pp. 7183-7195, Sept. 2022.

[29] K. Murali, P. Dhana Lakshmi, and V. Navyasree, "Spectrum sensing based on Forward Active Channel Allocation in wireless 5G communications," *International Journal for Modern Trends in Science and Technology*, vol. 02, no. 02, Feb. 2016.

[30] P. Dhana Lakshmi and N. Venkateswara Rao, "SPECTRUM SENSING FOR GREEN COGNITIVE RADIO COMMUNICATIONS: A SURVEY," *Int. J. Elec. Elecn. Eng. Telcomm.*, vol. 6, no. 2, Apr. 2017.

[31] D. Lakshmi Potteti and N. Venkateswara Rao, "On the performance of single ring law based sensing approaches for opportunistic spectrum access," *AEU - International Journal of Electronics and Communications*, vol. 88, pp. 76-83, May 2018.

[32] M. Sadeghi and E. G. Larsson, "Adversarial attacks on deep learning based radio signal classification," *IEEE Wireless Communications Letters*, vol. 8, no. 1, pp. 213-216, Feb. 2019.