Article

MSP-Net: Multi-Scale Spectrum Pyramid Network for Robust Synthetic Aperture Radar Automatic Target Recognition

Aisha Sir Elkhatem , Seref Naci Engin * , Yerbol Ospanov * , Aizhan Erulanova

Posted Date: 19 November 2025

doi: 10.20944/preprints202511.1410.v1

Keywords: synthetic aperture radar; automatic target recognition; multi-scale spectrum pyramid network; spatial-domain CNNs; single-scale frequency-domain CNNs; feature interpretability; high-frequency scattering

 doi:10.20944/preprints202511.1410.v1

Article

MSP-Net: Multi-Scale Spectrum Pyramid Network for Robust Synthetic Aperture Radar Automatic Target Recognition

Aisha Sir Elkhatem [1,2], Seref N. Engin [2,*], Yerbol Ospanov [3,*] and Aizhan Erulanova [4]

[1] Aeronautical Engineering Dept., Sudan University Of Science & Technology (SUSTECH), Khartoum, Sudan
[2] Control and Automation Engineering Department, Yildiz Technical University, Istanbul 3420, Turkey
[3] Department of Automation and Information Technology, Sakarim University, Kazakhstan
[4] S. Seifullin Kazakh Agrotechnical Research University (KATRU), Astana, Republic of Kazakhstan
* Correspondence: nengin@yildiz.edu (S.N.E.); yerbol.ospanov.78@mail.ru (Y.O.)

Abstract

Synthetic Aperture Radar (SAR) Automatic Target Recognition (ATR) remains challenging due to speckle noise, aspect-angle variation, and the loss of fine scattering cues in conventional deep-learning pipelines. Spatial-domain CNNs primarily extract geometric structure but overlook high-frequency information critical for distinguishing small or spectrally similar targets, while frequency-only methods such as FFTNet fail to leverage spatial context and multi-scale spectral variation. To address these limitations, this study proposes the Multi-Scale Spectrum Pyramid Network (MSP-Net), which decomposes SAR images into low-, mid-, and high-frequency components via two-dimensional Fourier transforms with band-pass filtering and processes each band through dual convolutional branches equipped with predefined and learnable spectral filters. The resulting features are fused using attention-based, MLP-based, or transformer-based integration mechanisms. Experiments on two MSTAR-based benchmark datasets (11-class and 8-class) demonstrate that MSP-Net substantially outperforms spatial-only CNNs and single-scale frequency-domain models. In the 11-class setting, MSP-Net improves accuracy by 13–14% (up to 95%) and achieves near-perfect ROC separability (AUC ≈ 1.0) with reliable calibration (ECE < 0.02). On the reduced 8-class dataset, the best MSP-Net variant achieves 99.9% accuracy and consistent per-class F1-scores. Ablation studies confirm the critical role of multi-scale spectral decomposition and adaptive fusion in improving recognition of small and spectrally similar targets such as BMP2, BTR60, and BTR70. These results highlight the effectiveness of frequency-aware, multi-scale learning for robust and interpretable SAR ATR.

Keywords: synthetic aperture radar; automatic target recognition; multi-scale spectrum pyramid network; spatial-domain CNNs; single-scale frequency-domain CNNs; feature interpretability; high-frequency scattering

1. Introduction

Synthetic Aperture Radar (SAR) is a radar imaging technology that uses a wide spectrum of frequencies to generate high-resolution imagery. Unlike optical sensors, SAR can operate effectively in all weather, day and night and long-range conditions, enabling consistent image acquisition regardless of illumination or atmospheric constraints [1,2]. These unique capabilities have driven its widespread adoption in military and civilian applications, particularly for surveillance and reconnaissance missions [3]. In this context, SAR-based Automatic Target Recognition (ATR) systems play a key role by autonomously detecting and classifying critical targets, such as vehicles, ships, and aircraft, thus improving the efficiency and reliability of reconnaissance operations [4,5]. Visual interpretation of SAR imagery is inherently challenging due to speckle noise, geometric distortions,

 doi:10.20944/preprints202511.1410.v1

and complex backscattering phenomena [6–8]. Manual analysis of large-scale SAR image streams is resource intensive, motivating the development of ATR systems. The classical SAR ATR pipeline, initially formalized by [9] as a three-stage process of detection, discrimination, and classification, traditionally relied on hand-crafted features (e.g. geometric, textural, and polarization-based) and conventional classifiers such as Support Vector Machines (SVMs), decision trees, and random forests [10–14]. Although these methods achieved reasonable performance, they were highly sensitive to imaging conditions, limited by feature generalization, and prone to errors in complex or low signal-to-noise scenarios. The advent of deep learning (DL), particularly convolutional neural networks (CNNs), has marked a paradigm shift in SAR ATR by enabling automatic, data-driven feature extraction and classification within a unified architecture [15–21]. These DL-based approaches consistently outperform traditional methods by learning discriminative high-level representations directly from raw SAR data.

One of the most critical challenges in deep learning-based SAR ATR is the scarcity of large, diverse, and well-annotated training datasets, a problem often referred to as SAR ATR with limited training data [22–24]. This constraint arises from the high cost, operational complexity, and security restrictions associated with SAR data acquisition, particularly in military contexts. Existing research addressing this challenge can be broadly classified into two principle strategies: data augmentation and specialized module or architecture design [25–28]. Data augmentation methods artificially expand the training set or its feature space, either through generative preprocessing techniques or learned models. For example, Wang et al. [29] proposed a semi-supervised learning framework with a self-consistent augmentation rule to leverage unlabeled data, while Zhang et al. [30] combined feature augmentation with ensemble learning to extract richer multilayer feature representations from limited samples.

In contrast, specialized module or architecture design approaches embed SAR-specific domain knowledge into model structures, often via transfer learning or attribute-guided mechanisms [31–35]. Sun et al. [36], for example, introduced an attribute-guided transfer learning method that exploits shared target aspect angles between source and target domains, while Zhang et al. [37] repurposed pre-trained layers to transfer generic knowledge to limited-data scenarios [38].

Recently, Wang et al. [39] have advanced this field by introducing a causal SAR ATR (CSA) model to explicitly analyze the negative impacts of limited-data conditions using causal theory. Their study reveals that, under limited data, a confounder variable can influence both feature extraction and recognition outcomes, thereby degrading the performance. To address the issue, they proposed a dual invariance intervention strategy, comprising an inner-class invariant proxy and a confounder-invariance loss, designed to enhance intra-class feature consistency while minimizing the confounder's influence, without requiring large datasets. This approach has been validated across multiple benchmark datasets, demonstrating superior recognition accuracy in limited-data scenarios. Despite such progress, a critical research gap remains: the comprehensive characterization of how limited data specifically affects feature learning and classification in SAR ATR. Bridging this gap is essential for developing more data-efficient, robust models capable of operational deployment in real-world environments.

In addition to the limited-data problem, several other critical challenges hinder the advancement of deep learning-based SAR ATR. First, most existing methods exhibit a task-specific property, where a model is trained and evaluated for a single, narrowly defined target category, such as vehicles, ships, or aircraft, requiring separate deep models for each task [40–44]. This isolation limits scalability, as new tasks must be learned from scratch, consuming large quantities of labeled data while incurring high computational costs and inconsistent performance across models. Recently, Li et al. [45] pioneered the development of a foundation model for SAR ATR, termed SARATR-X, which represents a significant step toward scalable and label-efficient SAR target recognition.

Unlike previous supervised or limited-data approaches, SARATR-X leverages self-supervised learning (SSL) to learn generalizable and robust feature representations from an unprecedentedly large dataset of 0.18 million unlabeled SAR target samples, curated from multiple contemporary

 doi:10.20944/preprints202511.1410.v1

benchmarks, which are known as the largest public SAR ATR dataset to date. Second, there is a heavy reliance on supervised learning, which demands vast amounts of expertly annotated SAR samples. Given the scarcity of trained SAR analysts, many SAR datasets remain unlabeled and thus underutilized, limiting generalization and scalability [46–48]. Third, current models often ignore SAR-specific imaging characteristics, which differ fundamentally from optical imagery. The presence of speckle noise, discrete scattering patterns, and the absence of clear geometric, texture, and contour cues create a significant domain gap between SAR and natural images, making direct transfer of natural image-based models suboptimal. These factors necessitate the incorporation of SAR-specific priors into backbone architectures and learning strategies. Finally, the open-source ecosystem for SAR ATR remains underdeveloped due to the sensitivity of SAR data, resulting in a lack of publicly available large-scale benchmark datasets and standardized codebases. This scarcity of shared resources restricts reproducibility, hampers benchmarking, and slows the integration of emerging deep learning techniques into the SAR ATR domain.

Contributions and Focus of the Proposed MSP-Net

This paper introduces MSP-Net, a Multi-Scale Spectrum Pyramid Network tailored for robust Synthetic Aperture Radar (SAR) Automatic Target Recognition (ATR). Unlike conventional methods that either operate solely in the spatial domain or apply frequency transforms without hierarchical decomposition, MSP-Net explicitly incorporates frequency-aware multi-scale processing coupled with systematic fusion strategies. This design directly addresses persistent challenges in SAR ATR, including scale variation, class imbalance, and small-target discrimination, while also improving interpretability and prediction reliability.

Novel Design

The proposed Multi-Scale Spectrum Pyramid module decomposes SAR images into distinct frequency bands using a 2D Fourier transform with band-pass filtering, yielding a structured representation of target information across scales:

- Low-frequency components capture coarse global geometry and target silhouette.
- Mid-frequency components encode intermediate scattering centers and discriminative textural cues.
- High-frequency components preserve fine-grained radar cross-section variations essential for small-target recognition.

Each band is processed through parallel convolutional branches, enabling the network to jointly learn complementary spectral cues while mitigating the dominance of majority-class features. To ensure both interpretability and adaptability, MSP-Net integrates two complementary filtering strategies:

1. Predefined filters, guided by SAR domain knowledge, to maintain physical interpretability.
2. Learnable filters, optimized end-to-end from data, to adaptively refine spectral responses.

The decomposed features are then reintegrated through multiple fusion mechanisms, including MLP-based concatenation, attention fusion, and transformer fusion, providing a flexible and generalizable framework for modeling cross-band interactions. This systematic design allows MSP-Net to capture coarse-to-fine spectral information while adaptively balancing contributions across different scales.

C. Comprehensive Validation Framework

To rigorously validate the proposed framework, MSP-Net was evaluated under multiple robustness conditions designed to emulate real-world SAR ATR challenges. The evaluation strategy followed a progressive comparison, beginning with a classical spatial-domain CNN baseline, extending to a single-scale FFTNet representing frequency-based approaches, and culminating in the proposed MSP-Net that integrates hierarchical spectrum decomposition. This design allowed systematic benchmarking against both spatial-only and frequency-only paradigms.

Experiments were conducted in both full-data and few-shot learning regimes, thereby quantifying the model's capacity to generalize under varying data availability. Robustness was further assessed under spatial-domain transformations, frequency-domain perturbations, and non-

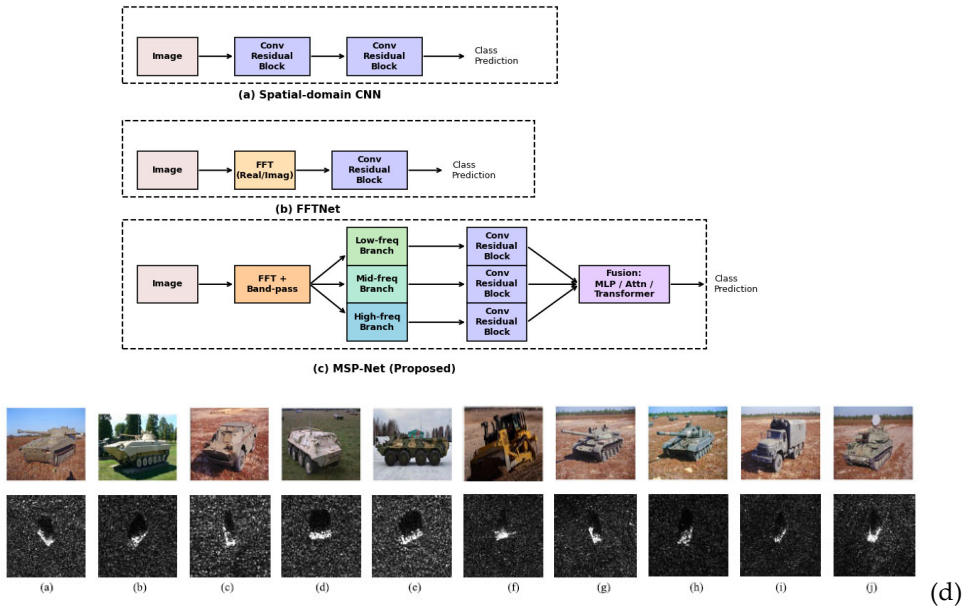
 doi:10.20944/preprints202511.1410.v1

local means (NLM) denoising, each simulating practical degradations such as geometric distortions, spectral corruption, and noise contamination.

Beyond overall classification accuracy, the evaluation incorporated confusion matrices, per-class accuracy, ROC curves, and calibration analysis. These metrics provided deeper insights into class separability and prediction reliability, particularly in challenging scenarios such as discriminating visually similar classes (e.g., 2S1 vs. ZSU_23_4, BMP2 vs. BTR70). Figure 1 (d).

Therefore, the main contributions of this work can be summarized as follows:

- Multi-Scale Spectrum Pyramid Architecture: Introduction of a spectrum pyramid module that explicitly decomposes SAR images into low-, mid-, and high-frequency bands, enabling joint learning of coarse-to-fine scattering features for robust ATR.
- Dual Filtering Strategies: Integration of predefined, domain-informed filters with learnable, data-driven filters, ensuring a balance between interpretability and adaptability to unseen conditions.
- Flexible Fusion Mechanisms: Comparative exploration of MLP concatenation, attention-based fusion, and transformer-based fusion, demonstrating the effectiveness of adaptive spectral weighting over naive feature stacking.
- Robustness and Reliability Evaluation: Comprehensive validation across spatial vs. frequency baselines, supported by confusion matrix, ROC, and calibration analyses. The results confirm that MSP-Net addresses persistent challenges in SAR ATR, including class imbalance, scale variation, and small-target recognition, while maintaining high reliability in operational scenarios.



(d)

Figure 1. (a) Baseline spatial CNN, **(b)** FFT-Net, (c) MSP-Net (Proposed), **(d)** Optical images and SAR images of ten-class ground military vehicle targets in MSTAR. (a) 2S1. (b) BMP2. (c) BRDM2. (d) BTR60. (e) BTR70.(f) D7. (g) T62. (h) T72. (i) ZIL131. (j) ZSU234.

2. State-of-the-Art in Frequency-Based SAR Automatic Target Recognition

This section reviews recent research on frequency-based methods in Synthetic Aperture Radar (SAR) and image recognition to outline the current state of the art and identify existing challenges in automatic target recognition (ATR). The overview provides the foundation for positioning the proposed Multi-Scale Spectrum Pyramid Network (MSP-Net) within the context of prior work and demonstrates how it advances the capabilities of SAR ATR systems. The discussion is organized into two subsections: A. Frequency-Based Methods in SAR and Image Recognition, which surveys

 doi:10.20944/preprints202511.1410.v1

existing spectral and hybrid techniques, and B. Performance Comparison between MSP-Net and Frequency-Based SAR ATR Models from the Literature, which evaluates the improvements offered by MSP-Net relative to state-of-the-art methods.

2.1. Frequency-Based Methods in SAR and Image Recognition

The frequency domain has long been recognized as a valuable reservoir of discriminative features, particularly for challenging tasks such as detecting camouflaged objects and enhancing target recognition in Synthetic Aperture Radar (SAR) imagery. Unlike purely spatial-domain features, frequency-based representations can better capture subtle spectral differences between targets and complex backgrounds, enabling improved separability in cluttered or low-contrast scenes [49–51]. These advantages arise from the ability of the frequency domain to highlight periodic structures, suppress irrelevant background noise, and reveal latent textural signals that may be invisible in spatial representations [52–55].

Recent advancements in SAR-based object detection have increasingly explored joint spatial-frequency feature extraction to address persistent challenges in Automatic Target Recognition (ATR). For example, the SFONet framework introduces a spatial-frequency attention module that integrates spatial and frequency domain cues, enabling robust target discrimination even in the presence of significant background interference, [56]. This is particularly relevant for maritime SAR ATR, where ship targets often present weak texture features, low SNR, and high visual similarity, making spatial-only approaches insufficient. SFONet further addresses the issue of densely arranged and directionally oriented targets by replacing conventional horizontal bounding boxes (HBB) with a Gaussian bounding box (GBB) representation, improving angle prediction through a probabilistic IoU (Intersection over Union) and distribution focal loss optimization scheme. This approach mitigates false negatives caused by high IoU overlaps in crowded scenes and enhances detection accuracy for objects with diverse aspect ratios. Additionally, SFONet acknowledges the scarcity of labeled multi-class SAR datasets, validating its approach on SSDD+, HRSID, RSDD, and SRSDD, the latter providing crucial multi-class annotations. In [57], the authors introduce an Attention-Spatial-Gated-Frequency (ASGF) framework that fuses spatial and frequency-domain information for improved pixel-level change detection. The approach begins with fuzzy c-means clustering to generate pseudo-labels, which are then used to guide feature learning in a CNN-based network. Spatial features are extracted using multi-region weighted operators (horizontal, vertical, and global) enhanced with channel-spatial attention, while complementary spectral features are obtained via gated linear units (GLUs). A multi-domain fusion module then integrates both modalities, achieving superior percent correct classification (PCC) and Kappa coefficient (KC) compared to four benchmark algorithms across three real SAR datasets. The results highlight the potential of hybrid spatial–frequency architectures in detecting subtle environmental changes under noisy conditions.

Complementing this line of research, [58] proposes a dual-branch spatial-frequency domain fusion recognition method with cross-attention to address the limitations of single-domain feature extraction in SAR ATR. In the spatial branch, an enhanced multi-scale feature extraction (EMFE) module captures multi-level spatial cues through parallel convolutional filters of varying sizes, while a frequency-domain guided attention mechanism focuses on key regions by transforming spatial features via FFT. In the frequency branch, a hybrid frequency domain transform (HFDT) module extracts both real and imaginary components using orthonormal 2-D FFT and enhances them via spatially guided attention, ensuring alignment between frequency features and the original spatial structure. A cross-domain feature fusion (CDFF) module then employs bidirectional cross-attention, bilinear projection, and adaptive weighting to achieve complementary fusion, allowing spatial features to provide semantic guidance for frequency cues, and frequency features to reinforce the structural integrity of spatial features. Experiments on the MSTAR dataset demonstrate significantly higher recognition accuracy compared to existing methods, underscoring the role of dynamic cross-domain interaction in enhancing SAR ATR robustness.

 doi:10.20944/preprints202511.1410.v1

The paper by Qu et al. [59] proposes a novel dual-branch spatial-frequency domain fusion recognition method with cross-attention for SAR image target recognition. Addressing challenges such as speckle noise and sensitivity to target orientation, the authors introduce several innovative modules to enhance recognition performance. The Enhanced Multi-Scale Feature Extraction (EMFE) module employs a multi-branch parallel structure to capture detailed spatial features, while the Hybrid Frequency Domain Transformation (HFDT) module extracts global structural information from the frequency domain using Fourier transforms and a frequency-domain attention mechanism. To optimally fuse these domain-specific features, the authors propose the Cross-Domain Feature Fusion (CDFF) module, which integrates spatial and frequency-domain features through a cross-attention mechanism, bilinear projection, and adaptive fusion. Experimental results on the MSTAR dataset demonstrate that the proposed method significantly outperforms existing approaches in recognition accuracy, offering a robust solution to SAR image target recognition through effective spatial-frequency domain fusion.

Recently, Li et al. [60] addressed limitations in existing SAR ATR models that rely on simple feature concatenation or linear weighting for fusing spatial and frequency domain features, which often neglect dynamic interactions critical for optimal recognition performance. To overcome these challenges, they proposed a dual-branch spatial-frequency domain fusion recognition method with cross-attention, combining several novel modules to enhance feature extraction and fusion. In the spatial domain, an Enhanced Multi-scale Feature Extraction (EMFE) module employs parallel convolutions (1×1, 3×3, 5×5) to capture multi-level spatial features across scales. Complementarily, in the frequency domain, a Hybrid Frequency Domain Transform (HFDT) module uses orthonormal two-dimensional FFT to extract real and imaginary components, accompanied by a frequency-domain self-attention mechanism that preserves structural integrity and mitigates interference between low- and high-frequency signals. Crucially, the Cross-Domain Feature Fusion (CDFF) module utilizes cross-attention, bilinear projection, and adaptive weighting to dynamically associate spatial and frequency features, enabling semantic guidance from spatial to frequency components and structural constraints in the reverse direction. This bidirectional interaction achieves efficient fusion of global and local information, significantly enhancing feature discriminability and alignment across domains. Their approach effectively compensates for the limited receptive field in spatial domain extraction, leading to improved recognition accuracy and robustness in SAR ATR tasks.

Addressing the limitations of spatial-only adaptation of pretrained models, Zhang et al. [61] proposed the Frequency-Guided Spatial Adaptation Network (FGSA-Net) for camouflaged object detection. Their method introduces a Frequency-Guided Spatial Attention (FGSAttn) module that transforms adapter input features into the frequency domain, adaptively enhancing or suppressing grouped frequency components within spectrogram circles. This enables better focus on subtle details and contours of camouflaged regions. They also designed Frequency-Based Nuances Mining (FBNM) and Frequency-Based Feature Enhancement (FBFE) modules to mine subtle foreground-background differences and fuse multi-scale features from pretrained vision transformers with task-specific adaptations.

Beyond target detection, frequency-domain analysis has also proven essential in recognizing and classifying SAR jamming types, a capability that directly impacts ATR system resilience in contested environments. For instance, a time–frequency-aware hierarchical feature optimization (TFA-HFO) framework has been proposed to identify 50 distinct jamming types, including suppression, deception, and composite signals, under varying JNR conditions [62]. By incorporating a dual attention mechanism to enhance spectral-temporal feature extraction and applying a hierarchical optimization strategy to improve feature separability, the method achieves superior robustness and generalization compared to existing approaches. This demonstrates that frequency-domain cues are not only valuable for improving ATR accuracy under benign conditions but also for safeguarding recognition performance when intentional interference is present. In [63], the authors propose an Attention-Spatial-Gated-Frequency (ASGF) framework that fuses spatial and frequency-domain information to enhance pixel-level change detection performance in the absence of labeled

 doi:10.20944/preprints202511.1410.v1

data. The approach begins with fuzzy c-means clustering to generate pseudo-labels, enabling feature learning within a CNN-based detection network. To improve sensitivity to subtle changes and robustness to noise, spatial feature extraction employs multi-region weighted features (horizontal, vertical, and full regions) enhanced with a channel-spatial attention mechanism. Complementary frequency-domain features are extracted via gated linear units (GLUs), and a multi-domain fusion module integrates both representations. Experimental evaluations on three real SAR datasets show that the ASGF method achieves higher percent correct classification (PCC) and Kappa coefficient (KC) than four benchmark algorithms, demonstrating the effectiveness of jointly exploiting spatial and frequency-domain cues for SAR change detection.

Complementing recognition-focused research, [64] introduces a novel deceptive jamming signal generation method for spaceborne SAR, addressing the long-standing trade-off between computational complexity and real-time performance. Traditional modulation-retransmission-based deceptive jammers require extensive per-PRI computation of the jammer's frequency response (JFR), often limiting scalability to large or high-resolution templates. The proposed spatial frequency-domain interpolation (SFI) algorithm instead maps the initial JFR to a 2-D spatial spectrum, via Fourier transforms along range and azimuth, and then computes subsequent JFRs using sinc interpolation. This decouples real-time complexity from template size, allowing efficient generation of large-scale or high-resolution false scenes while maintaining high imaging quality, even in squint geometries. Simulation and complexity analyses confirm that SFI surpasses existing azimuth time-domain and azimuth frequency-domain methods in both quality and efficiency.

Collectively, these advances show that frequency-domain analysis, whether for feature extraction, change detection, object recognition, or jamming mitigation, is becoming indispensable in modern SAR systems. The integration of spectral cues into attention modules, geometric representations, and jamming signal processing pipelines has enabled robust performance under cluttered backgrounds, orientation diversity, electromagnetic interference, and deceptive jamming attacks. This convergence of multi-domain feature fusion and real-time spectral computation is poised to play a central role in the next generation of SAR ATR and electronic counter-countermeasure (ECCM) technologies.

2.2. Performance Comparison Between MSP-Net and Frequency-Based SAR ATR Models from the Literature

While prior frequency-based methods have demonstrated significant progress in enhancing SAR ATR, they often exhibit limitations when confronted with scale variation, small-target detection, and imbalanced data distributions. Most recent works, such as SFONet [56], ASGF [57], and dual-branch spatial-frequency fusion frameworks [58–60] , primarily emphasize the complementary nature of spatial and spectral cues. These approaches exhibit robustness against noise and background clutter by leveraging attention mechanisms, Fourier-based transformations, and cross-domain feature fusion. Similarly, Zhang et al. [61] introduced the Frequency-Guided Spatial Adaptation Network (FGSA-Net), which employs spectral attention to refine transformer-based features for camouflaged object detection. Although these architectures enhance discriminability, they often rely on fixed fusion strategies or handcrafted modules, which may not fully capture multi-scale frequency variations inherent in SAR imagery.

In contrast, the proposed MSP-Net adopts a multi-scale spectrum pyramid architecture that explicitly decomposes SAR images into hierarchical frequency bands. This design enables the network to preserve fine-grained high-frequency cues critical for small-target discrimination while simultaneously integrating low-frequency structural information for robust global context modeling. Unlike conventional spatial-frequency fusion methods, which typically operate on single-level FFT representations, MSP-Net builds a structured pyramid that allows progressive feature refinement across frequency scales. Moreover, the adaptive fusion strategy in MSP-Net directly addresses class imbalance and scale diversity by weighting multi-scale spectral components in a data-driven manner, thereby reducing the over-reliance on spatial priors.

 doi:10.20944/preprints202511.1410.v1

3. Proposed Method

This section outlines the architecture and implementation of the proposed UAV-based autonomous railway inspection framework. The system is developed and validated in a physics-accurate Unity simulation environment that integrates perception, control, and communication modules in real time. The methodology encompasses the creation and training of a YOLO-based dataset for rail and anomaly detection, the formulation of a vision loss and GPS recovery strategy for robust navigation, and the development of a communication interface enabling efficient data exchange. Finally, a UAV control strategy, incorporating proportional (P) and proportional–derivative (PD) loops for pitch and roll stabilization, is described to demonstrate the integration of perception and flight control within the complete inspection architecture.

3.1. Problem Formulation

We formulate SAR Automatic Target Recognition (ATR) as a multi-branch learning problem in the frequency domain. The central idea is to decompose an input SAR image into complementary frequency bands, extract band-specific features, and fuse them for robust classification under scale variation, class imbalance, and small-target detection.

Given a SAR image $I \in R^{(H\times W)}$, we apply the 2D Fast Fourier Transform (FFT)

$$F(u,v) = \mathcal{F}I(x,y) \quad (1)$$

The frequency spectrum is partitioned into low-, mid-, and high-frequency bands using frequency masks$M_b(u,v)$:

$$F_b(u,v) = F(u,v)\cdot M_b(u,v), b \in low, mid, high \quad (2)$$

Masks may be predefined (fixed cutoffs) or learnable (optimized during training). Each masked band is converted back to the spatial domain via inverse FFT:

$$I_b(x,y) = \mathcal{F}^{(-1)}F_b(u,v) \quad (3)$$

This decomposition isolates scattering cues at different scales:

1. Low frequency: global geometry and coarse scattering patterns.
2. Mid frequency: structural components and scattering centers.
3. High frequency: fine radar cross-section details and micro-scattering.

We define the learning objective as mapping:

$$\Phi: I_{low}, I_{mid}, I_{high} \rightarrow y, where\ y \in 1, \dots, C \quad (4)$$

The mapping must achieve the following,

a. Multi-resolution robustness, preserve coarse-to-fine cues.
b. Minority-class amplification, avoid bias toward dominant classes.
c. Scale invariance, remain reliable across target sizes

3.2. MSP-Net Architecture

- Branch Feature Extraction

Each sub-band I_b is processed by a dedicated convolutional branchφ_b:

$$Z_b = \varphi_b(I_b), b \in low, mid, high \quad (5)$$

A branch consists of stacked convolution-BN-ReLU blocks with residual connections

$$Z_b^{(l)} = \sigma\left(BN\left(W_b^{(l)} * Z_b^{(l-1)} + R_b^{(l-1)}\right)\right) \quad (6)$$

where * denotes convolution, R_b denotes residual shortcut, and σ is ReLU.

The final band-specific representation is obtained by global average pooling (GAP):

$$Z_b = GAP\left(Z_b^{(L)}\right) \in R^d \quad (7)$$

- Fusion Strategies

 doi:10.20944/preprints202511.1410.v1

Given features $Z_{low}, Z_{mid}, Z_{high}$, MSP-Net explores three fusion mechanisms:

1. Concatenation + MLP:

$$Z_{concat} = [Z_{low}; Z_{mid}; Z_{high}], \quad Z_f = MLP(Z_concat) \tag{8}$$

2. Attention-based Fusion:

$$\alpha_b = exp(w_b^T Z_b) / \Sigma_b' exp(w_b'^T Z_b'), Z_f = \Sigma_b \alpha_b Z_b \tag{9}$$

3. Transformer-based Fusion:

$$Z_f = TransformerEncoder(Z_b) \tag{10}$$

The fused feature vector is passed to a classifier:

$$\hat{y} = Soft\,max(WZ_f + b) \tag{11}$$

- Loss Function

The network is optimized with cross-entropy loss:

$$L_CE = -\Sigma_{(c=1)}^{C} y_c \, log(\hat{y}_c) \tag{11}$$

4. Experimental Results and Analysis

4.1. Experimental Settings

To validate the effectiveness of the proposed Multi-Scale Spectrum Pyramid Network (MSP-Net), extensive experiments are conducted on two benchmark MSTAR-based SAR datasets. The first dataset consists of eleven classes (2S1, BMP2, BRDM2, BTR60, BTR70, D7, SLICY, T62, T72, ZIL131, and ZSU_23_4), while the second dataset includes eight classes (2S1, BRDM_2, BTR_60, D7, SLICY, T62, ZIL131, and ZSU_23_4). To ensure fair evaluation, an 80/20 training–testing split is employed with stratified sampling, and a validation subset is carved from the training data.

All experiments are implemented in PyTorch and executed on an Ubuntu 20.04 environment equipped with a single NVIDIA RTX 3090 GPU. Models are trained for 20 epochs using a batch size of 32, Adam optimizer with weight decay 10^{-4} and an initial learning rate of 0.001 reduced adaptively via plateau scheduling. The baseline CNN (spatial-domain), FFTNet (single-scale frequency-domain), and the proposed MSP-Net (with multiple fusion strategies) share the same backbone to ensure comparability.

Evaluation is performed using standard classification metrics (precision, recall, F1-score, accuracy) along with Expected Calibration Error (ECE) and Brier score for reliability assessment. In addition, confusion matrices (Figures 3, 7, 12), ROC curves (Figures 5, 9, 16), and calibration plots (Figures 4, 8, 15) are analyzed to quantify class separability and confidence reliability.

4.2. Overall Performance Comparison and Robustness

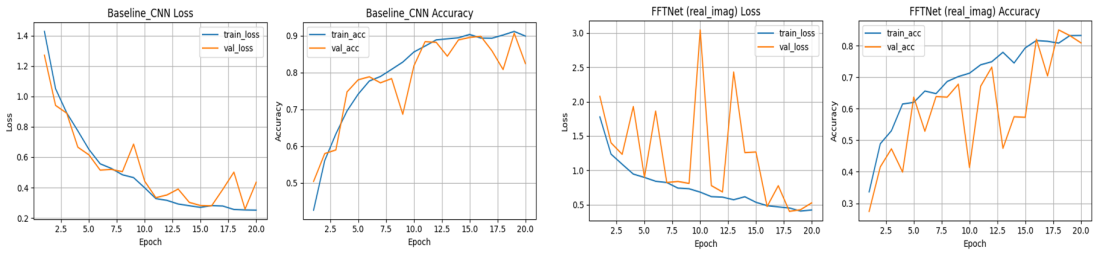
4.2.1. Results on Dataset One (11 Classes)

The training and validation curves in Figure 2 illustrate that both baseline CNN and FFTNet converge around 80% accuracy, whereas MSP-Net variants surpass 94%. The confusion matrices in Figure 3 confirm that CNN struggles with spectrally similar tracked vehicles (BMP2, BTR60, BTR70), and FFTNet nearly collapses on BMP2 (F1 ≈ 0.20). In contrast, MSP-Net with attention fusion (Figure 7a, b) achieves balanced recognition across all classes, with ROC curves in Figure 9 indicating near-perfect AUC ≈ 1.0 for most categories.

Calibration curves in Figure 8 reveal that predefined concat-MLP fusion yields the lowest ECE (0.0149), ensuring reliable probability estimates, while transformer-based fusion (Figure 7c) underperforms slightly due to limited training data. Prediction samples in Figure 10 highlight that

 doi:10.20944/preprints202511.1410.v1

MSP-Net can distinguish challenging targets (e.g., BMP2 vs. BTR70), where both CNN and FFTNet fail.



(**a**) (**b**)

Figure 2. (a) Training curves (loss/accuracy) for Baseline CNN (spatial-domain), (b) Training curves (loss/accuracy) for FFT-Net.

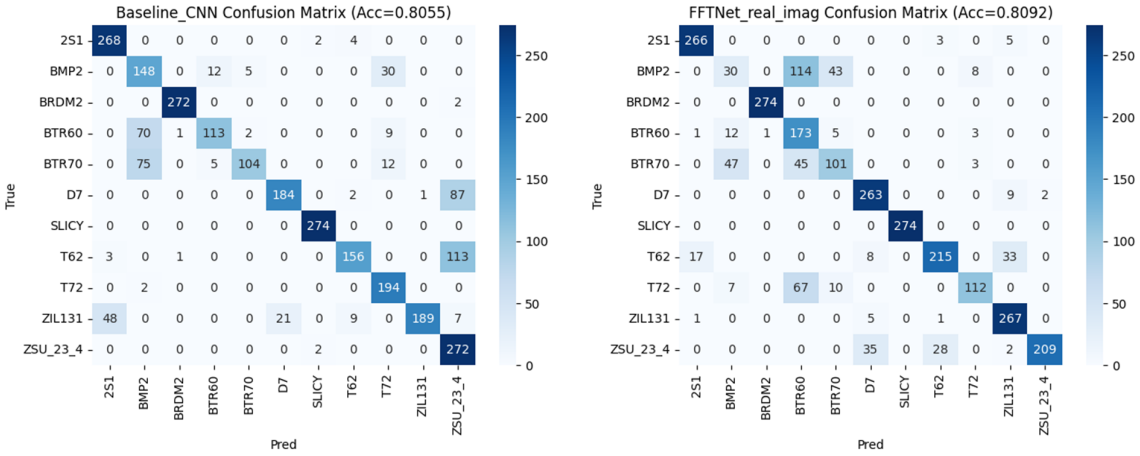


Figure 3. (a) Confusion matrix of Baseline CNN (spatial domain), (b) Confusion matrix of FFT-Net FFT-Net

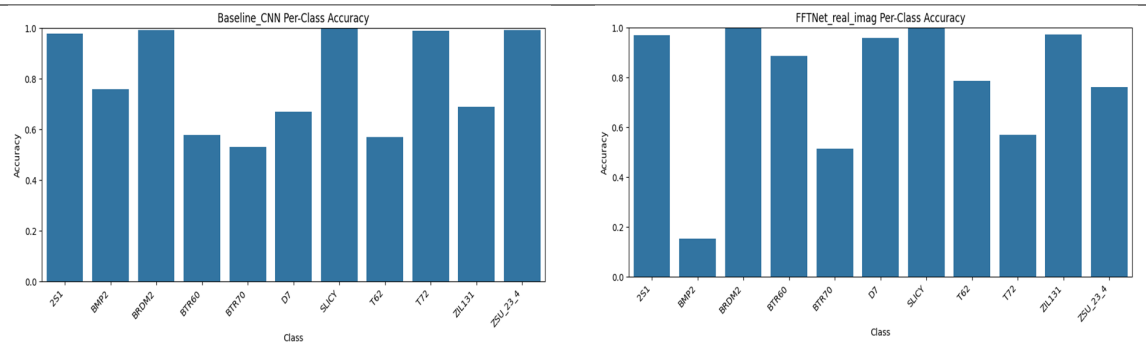


Figure 4. (a) Calibration plot (ECE) Per-class for Baseline CNN (spatial domain), (b) Accuracy Per-class accuracy for FFT-Net.

 doi:10.20944/preprints202511.1410.v1

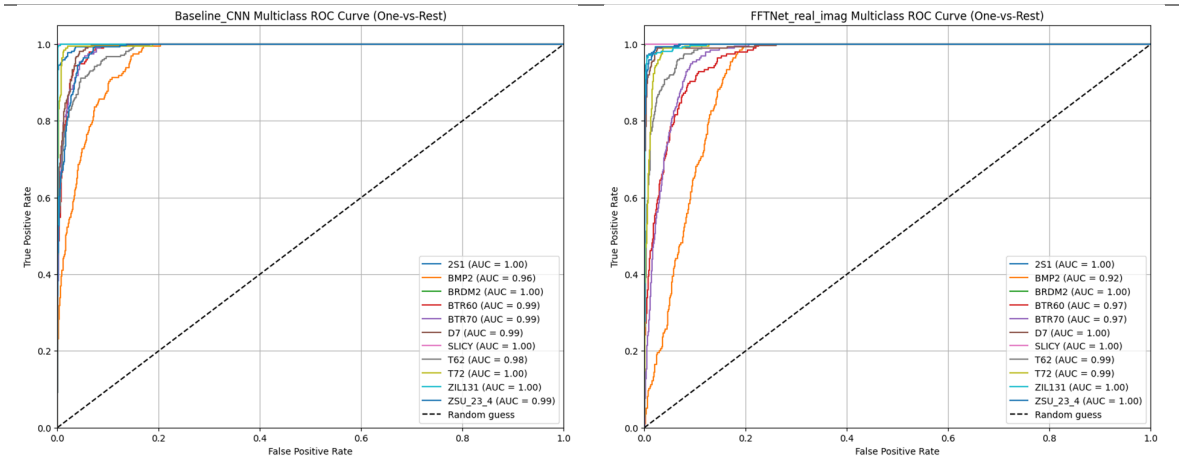
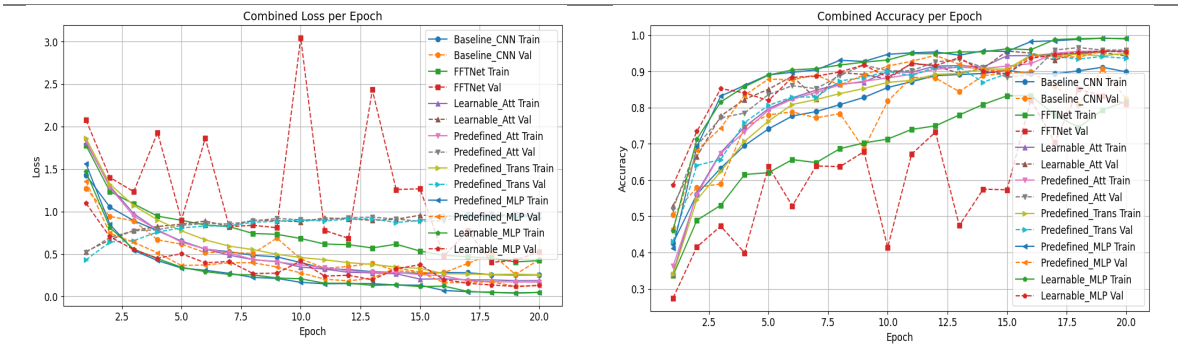
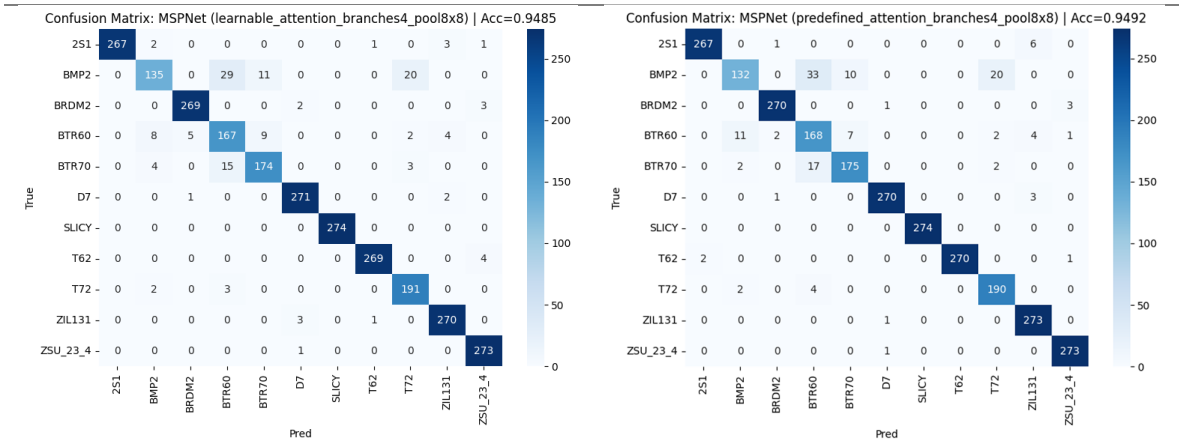


Figure 5. (a) ROC curves per class for Baseline CNN, (b) ROC curves per class for FFTNet.



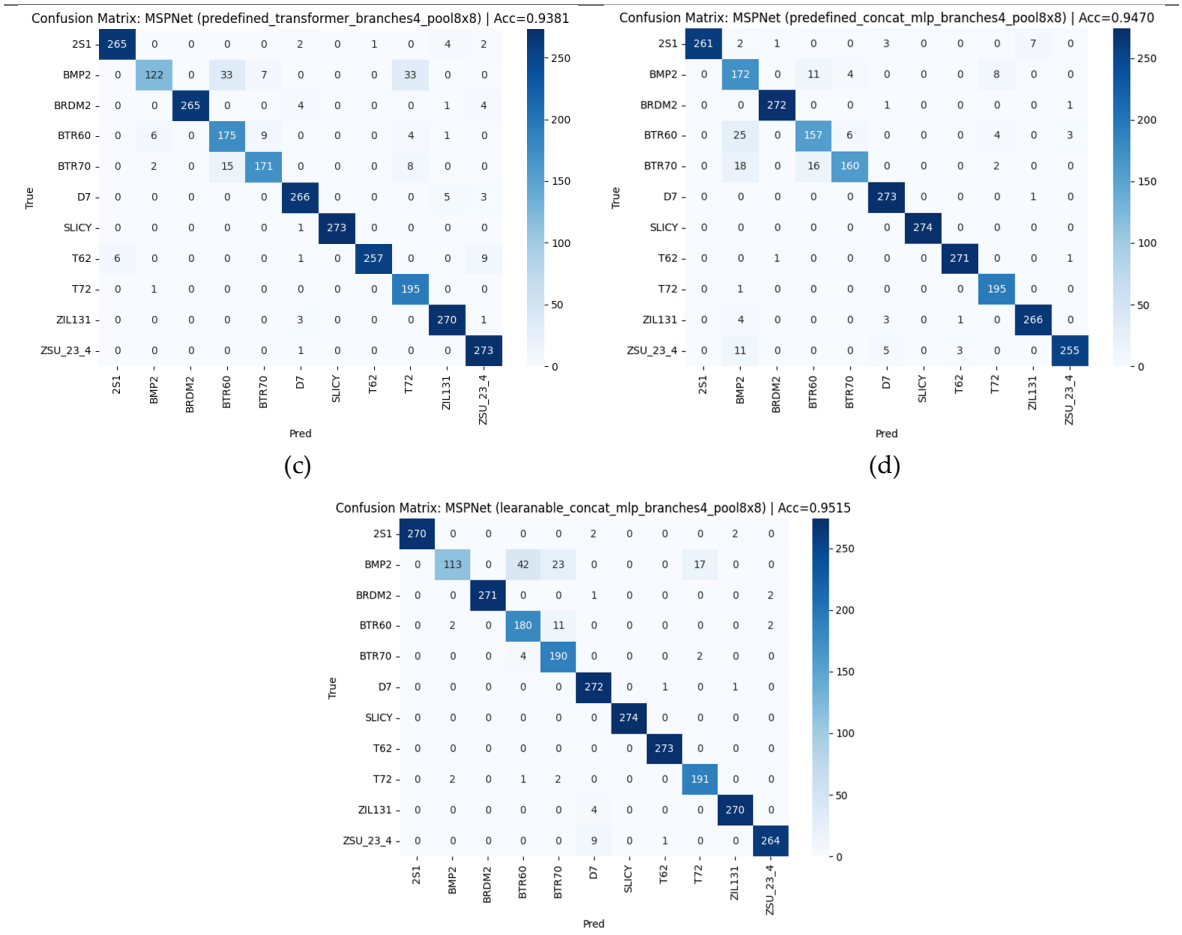
(a) (b)

Figure 6. Training curves comparison between Baseline CNN and FFTNet vs MSP-Net: (a) loss, (b) accuracy.



(a) (b)

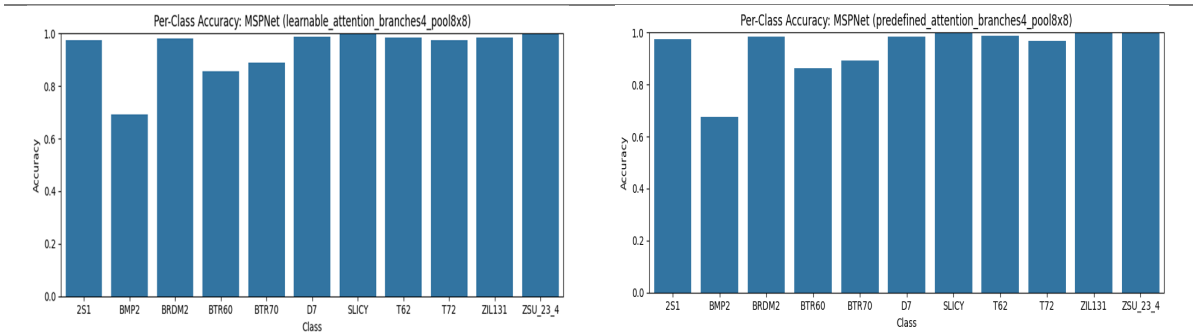
 doi:10.20944/preprints202511.1410.v1



(c) (d)

(e)

Figure 7. Confusion matrix of MSPNet: (a) learnable_attention, (b) predefined _attention, (c) predefined_transformer, (d) predefined_concat_mlp, (e) learnable_concat_mlp.



(a) (b)

 doi:10.20944/preprints202511.1410.v1

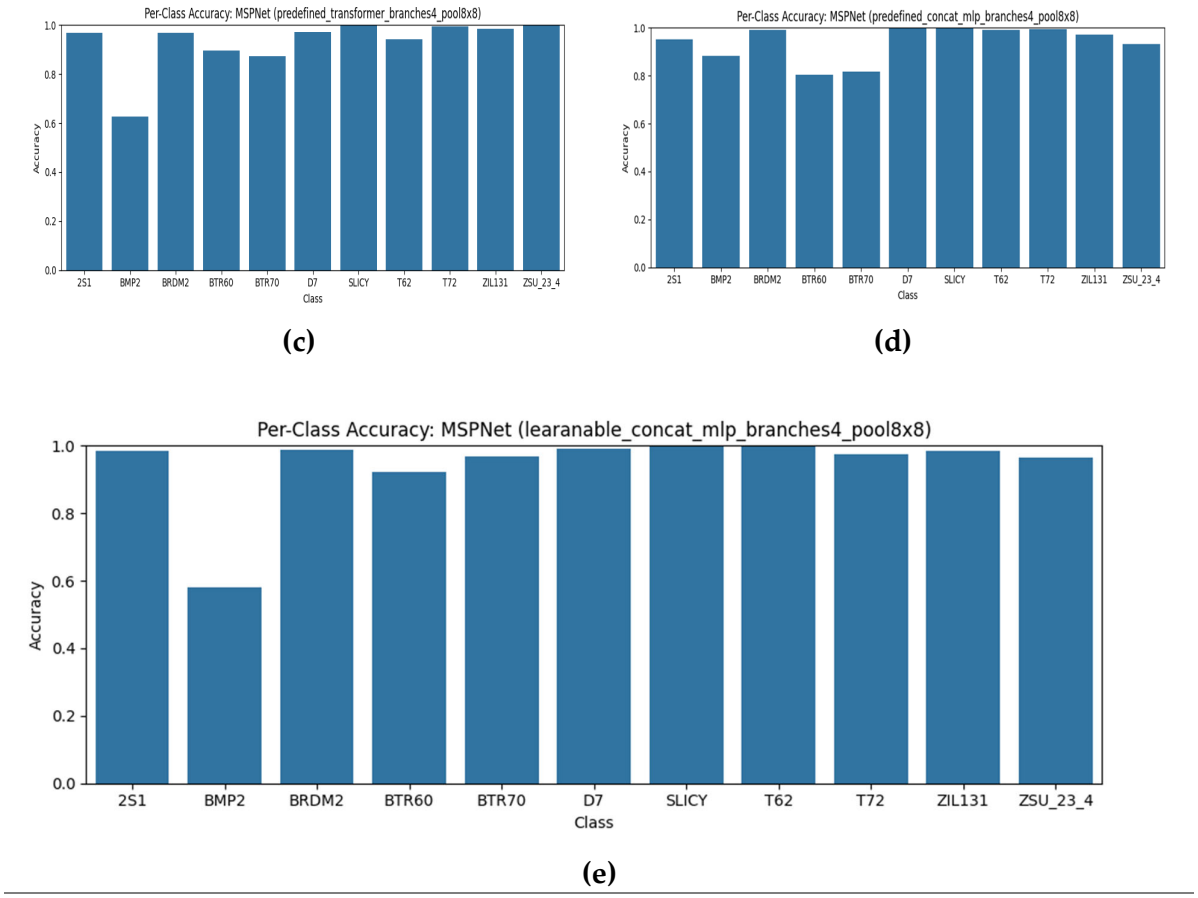
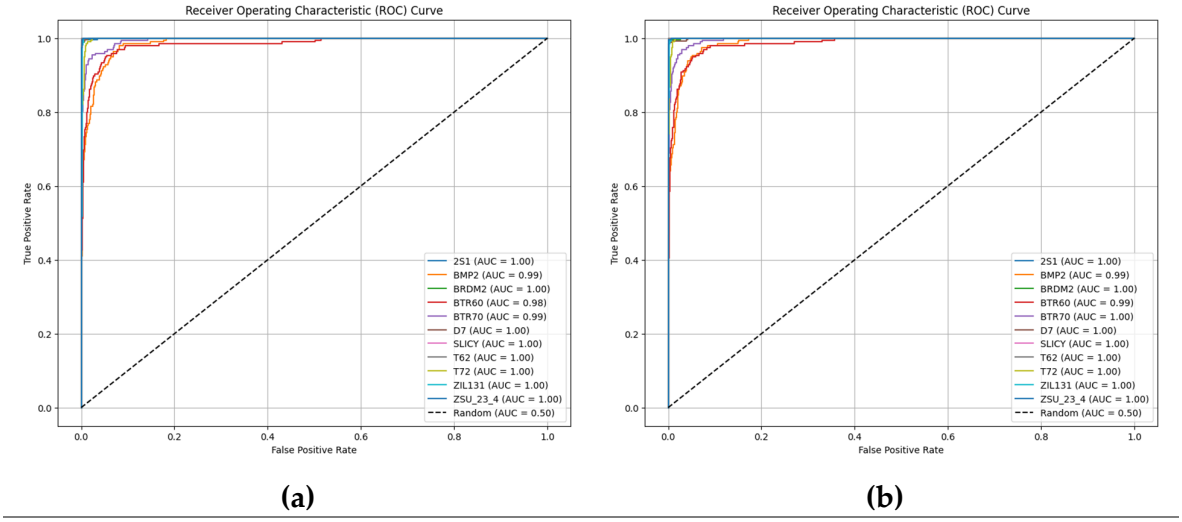


Figure 8. Calibration plot (ECE) Per-class for MSPNet: (a) learnable_attention, (b) predefined _attention, (c) predefined_transformer, (d) predefined_concat_mlp, (e) learnable_concat_mlp.



 doi:10.20944/preprints202511.1410.v1

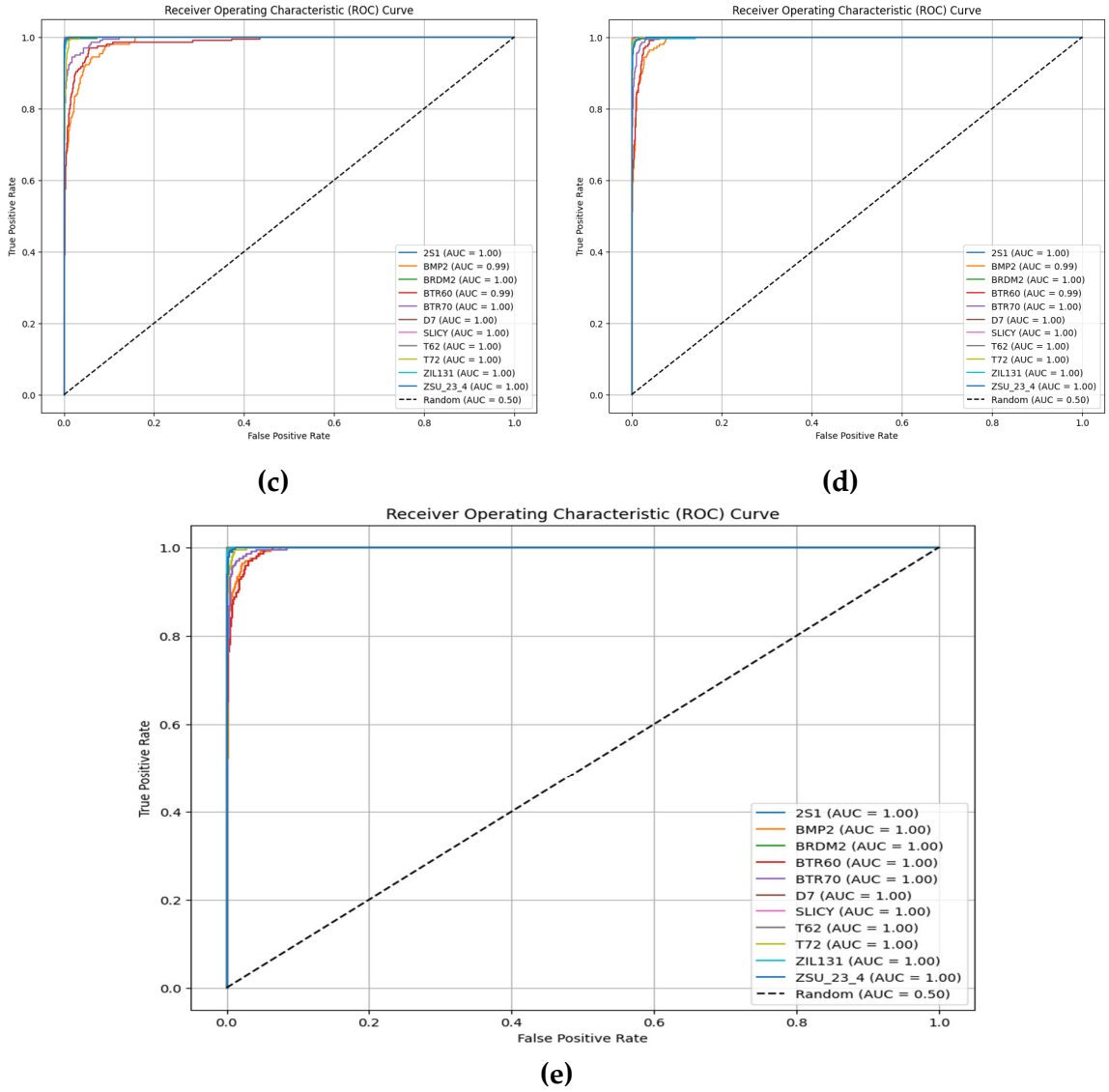
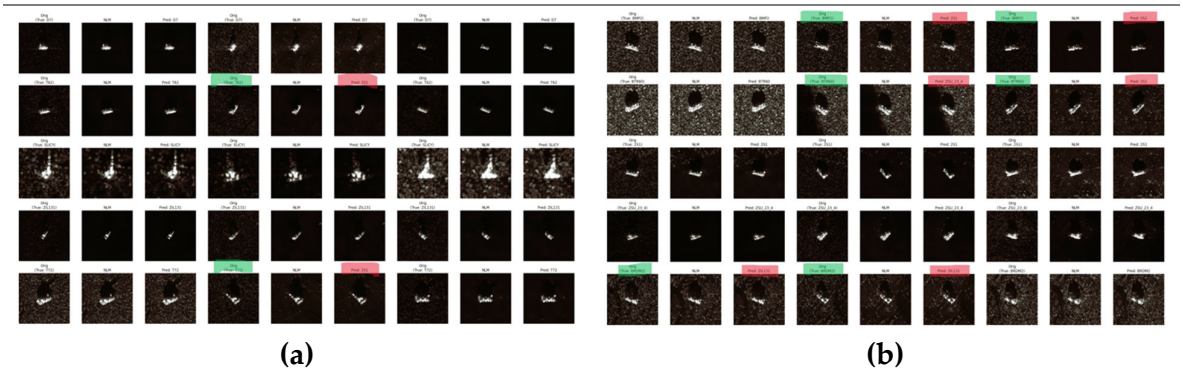


Figure 9. ROC curves per class for MSPNet: (a) learnable_attention, (b) predefined _attention, (c) predefined_transformer, (d) predefined_concat_mlp, (e) learnable_concat_mlp.



 doi:10.20944/preprints202511.1410.v1

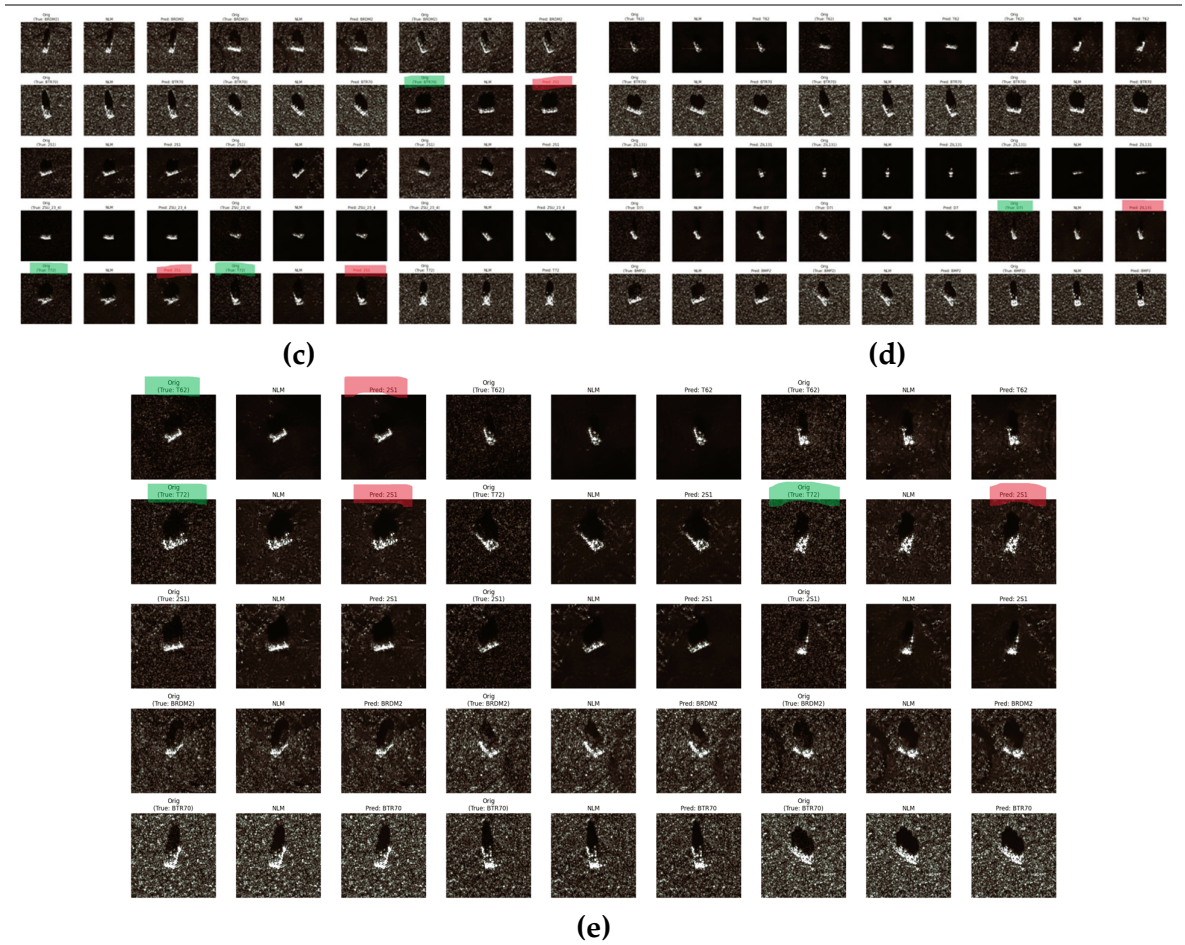


Figure 10. Generating sample prediction grid for MSPNet: (a) learnable_attention, (b) predefined _attention, (c) predefined_transformer, (d) predefined_concat_mlp, (e) learnable_concat_mlp.

4.2.2. Results on Dataset One (11 Classes)

In the reduced dataset, baseline CNN reaches 99.6% accuracy (Figure 11) with stable per-class performance (Figure 12a). FFTNet, although competitive (98.2%), shows vulnerability on ZSU_23_4 (Figure 12b). MSP-Net variants achieve the strongest results, with learnable attention attaining 99.7% accuracy (Figure 14–18). ROC curves (Figure 16) demonstrate AUC ≈ 1.0 for nearly all classes, and calibration plots (Figure 15) confirm excellent reliability.

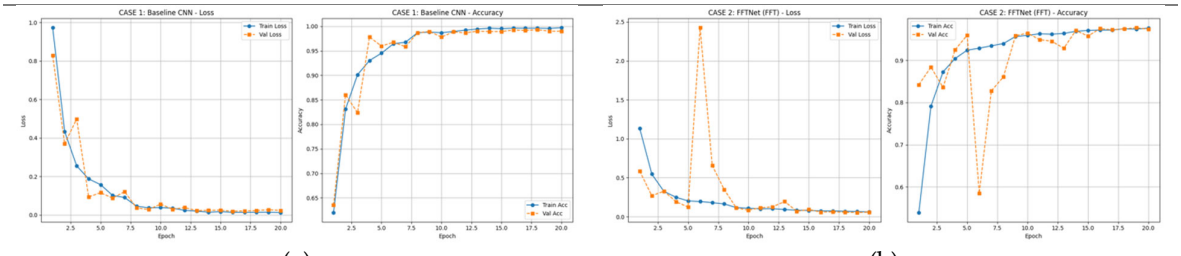


Figure 11. Training curves (loss/accuracy) for (a) Baseline CNN (spatial-domain), (b) FFT-Net.

 doi:10.20944/preprints202511.1410.v1

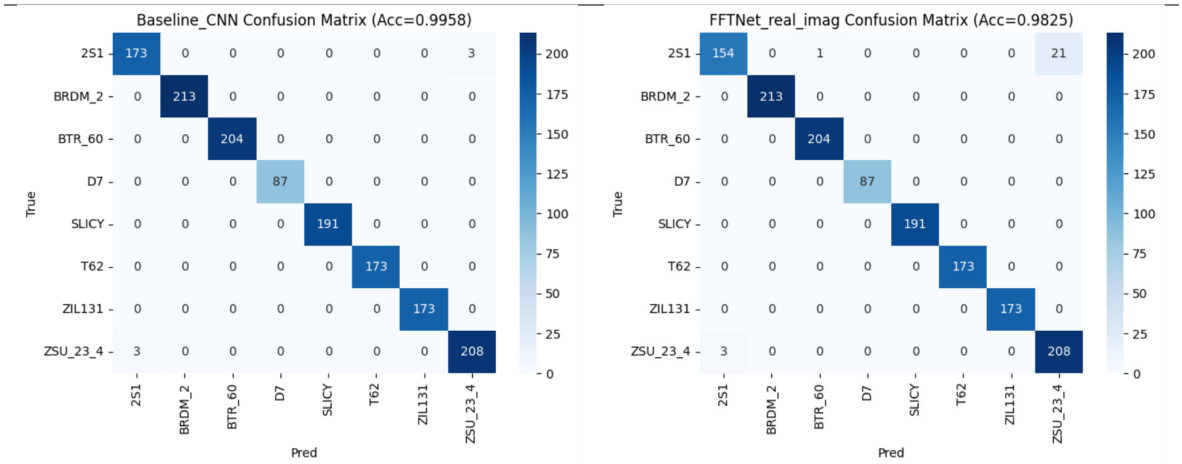


Figure 12. (a) Confusion matrix of Baseline CNN (spatial domain), (b) Confusion matrix of FFT-Net.

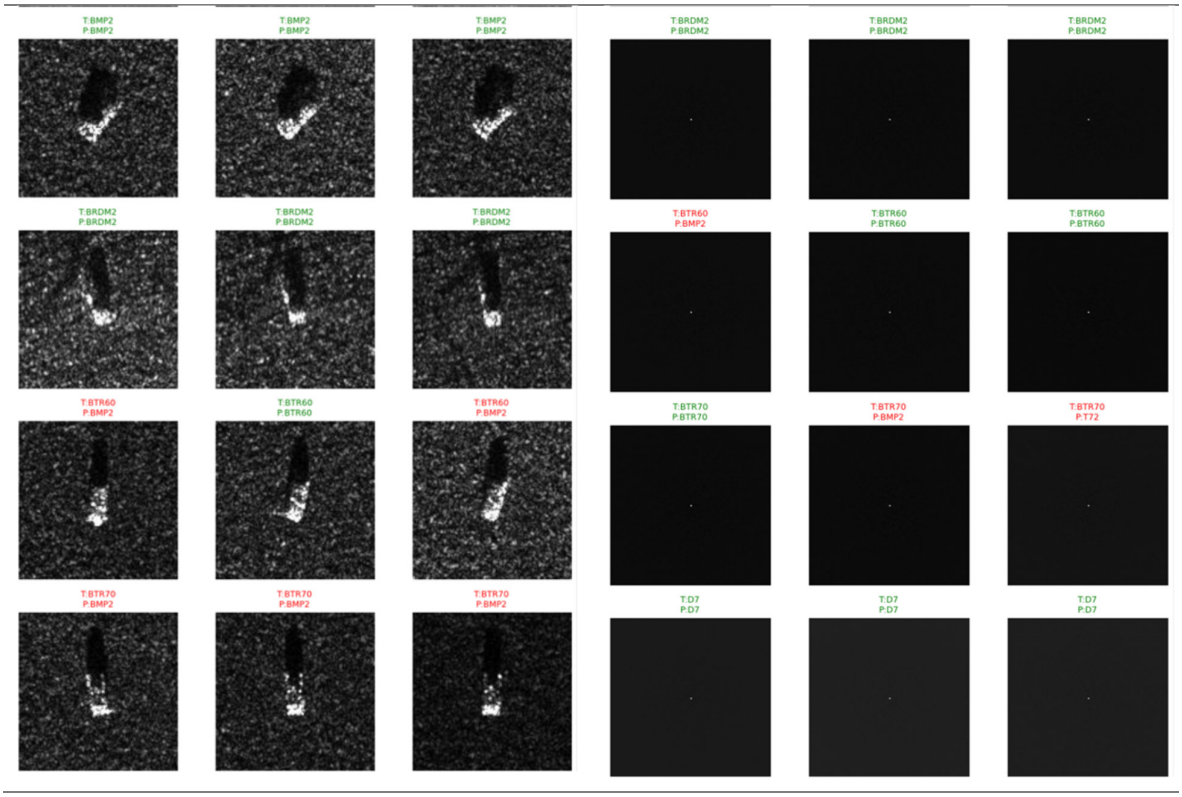


Figure 13. Generating sample prediction grid for MSPNet: (a) Confusion matrix of Baseline CNN (spatial domain), (b) Confusion matrix of FFT-Net.

 doi:10.20944/preprints202511.1410.v1

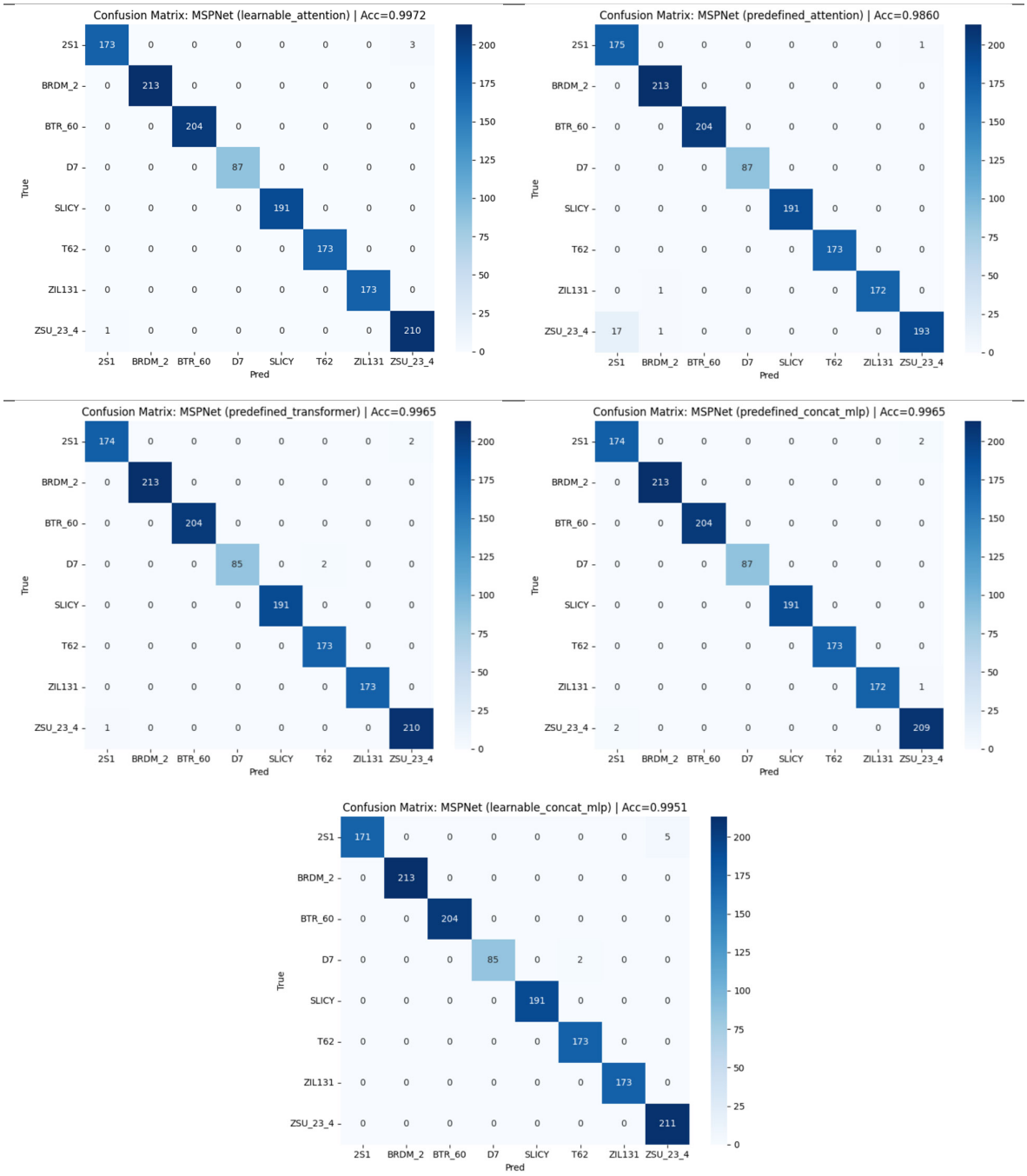


Figure 14. Confusion matrix of MSPNet: (a) learnable_attention, (b) predefined _attention, (c) predefined_transformer, (d) predefined_concat_mlp, (e) learnable_concat_mlp.

 doi:10.20944/preprints202511.1410.v1

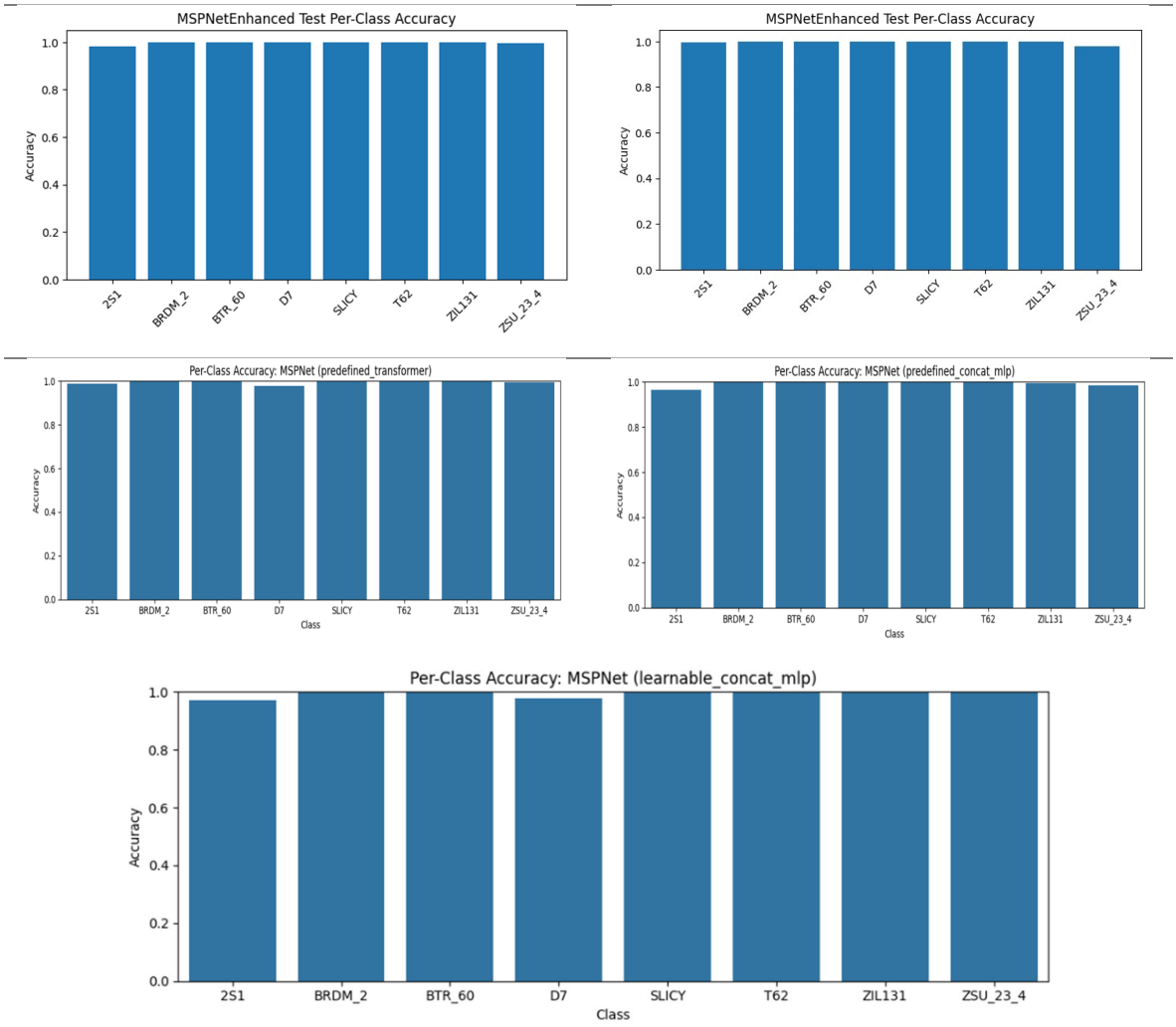
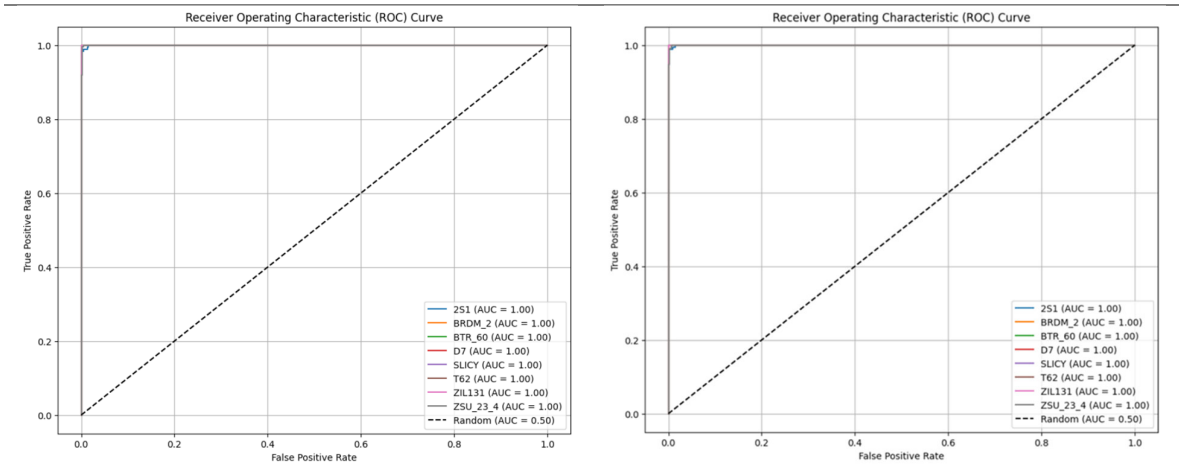


Figure 15. Calibration plot (ECE) Per-class for MSPNet: (a) learnable_attention, (b) predefined _attention, (c) predefined_transformer, (d) predefined_concat_mlp, (e) learnable_concat_mlp.



(a) (b)

 doi:10.20944/preprints202511.1410.v1

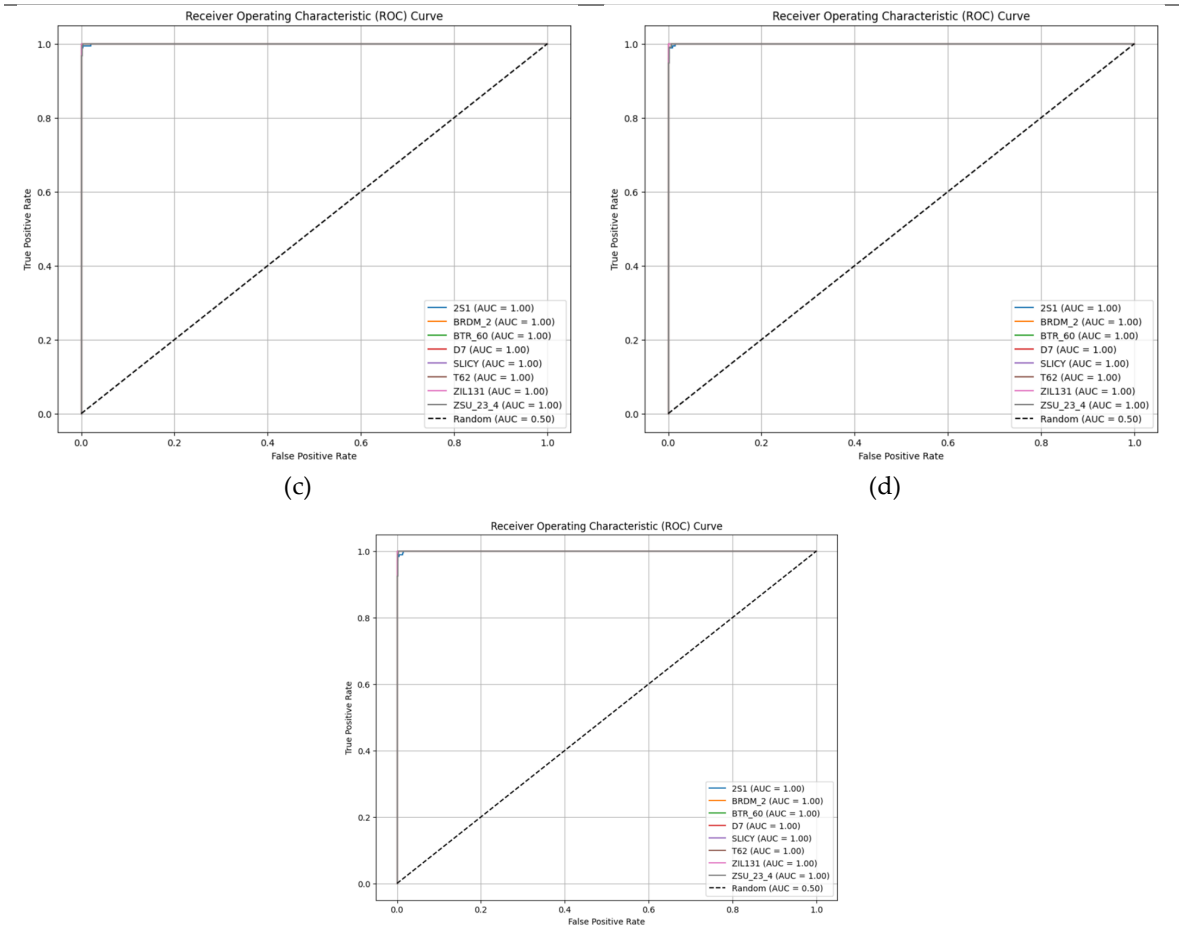
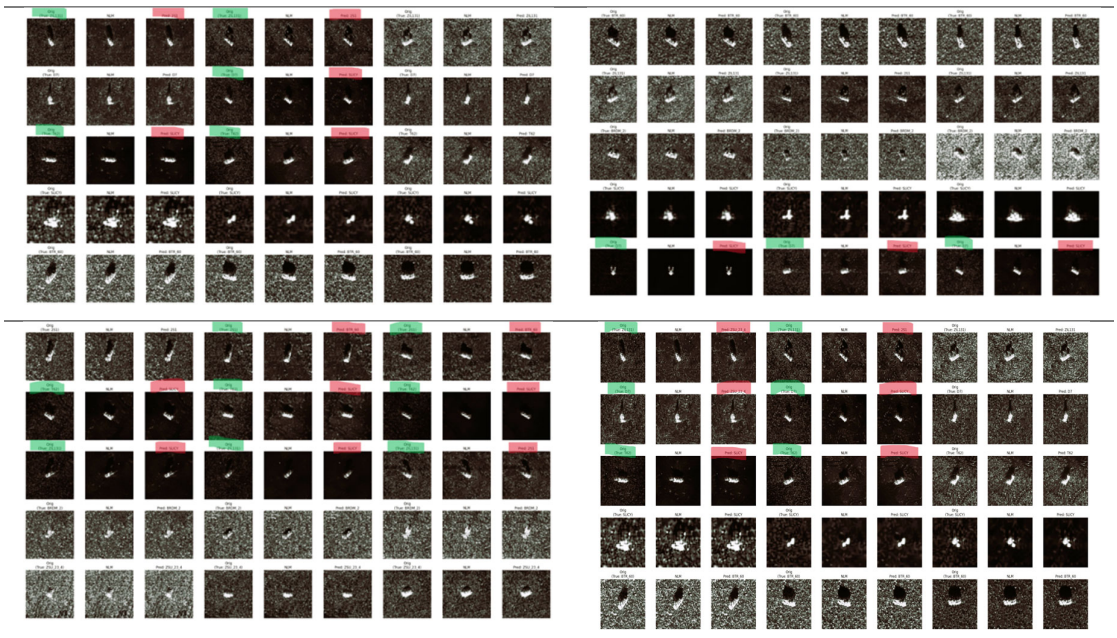


Figure 16. ROC curves per class for MSPNet: (a) learnable_attention, (b) predefined _attention, (c) predefined_transformer, (d) predefined_concat_mlp, (e) learnable_concat_mlp.



 doi:10.20944/preprints202511.1410.v1

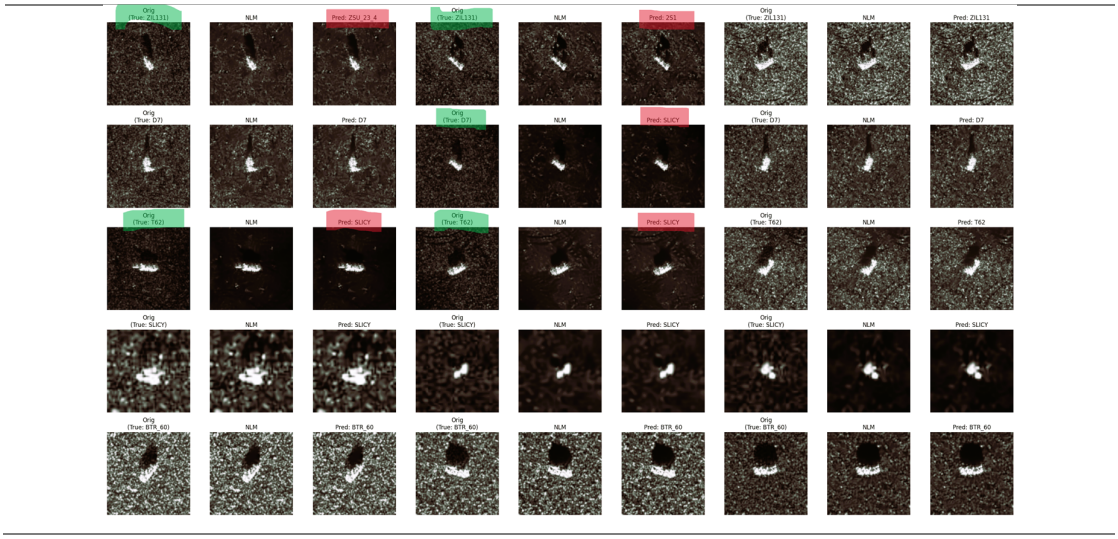


Figure 17. Generating sample prediction grid for MSPNet: (a) learnable_attention, (b) predefined _attention, (c) predefined_transformer, (d) predefined_concat_mlp, (e) learnable_concat_mlp.

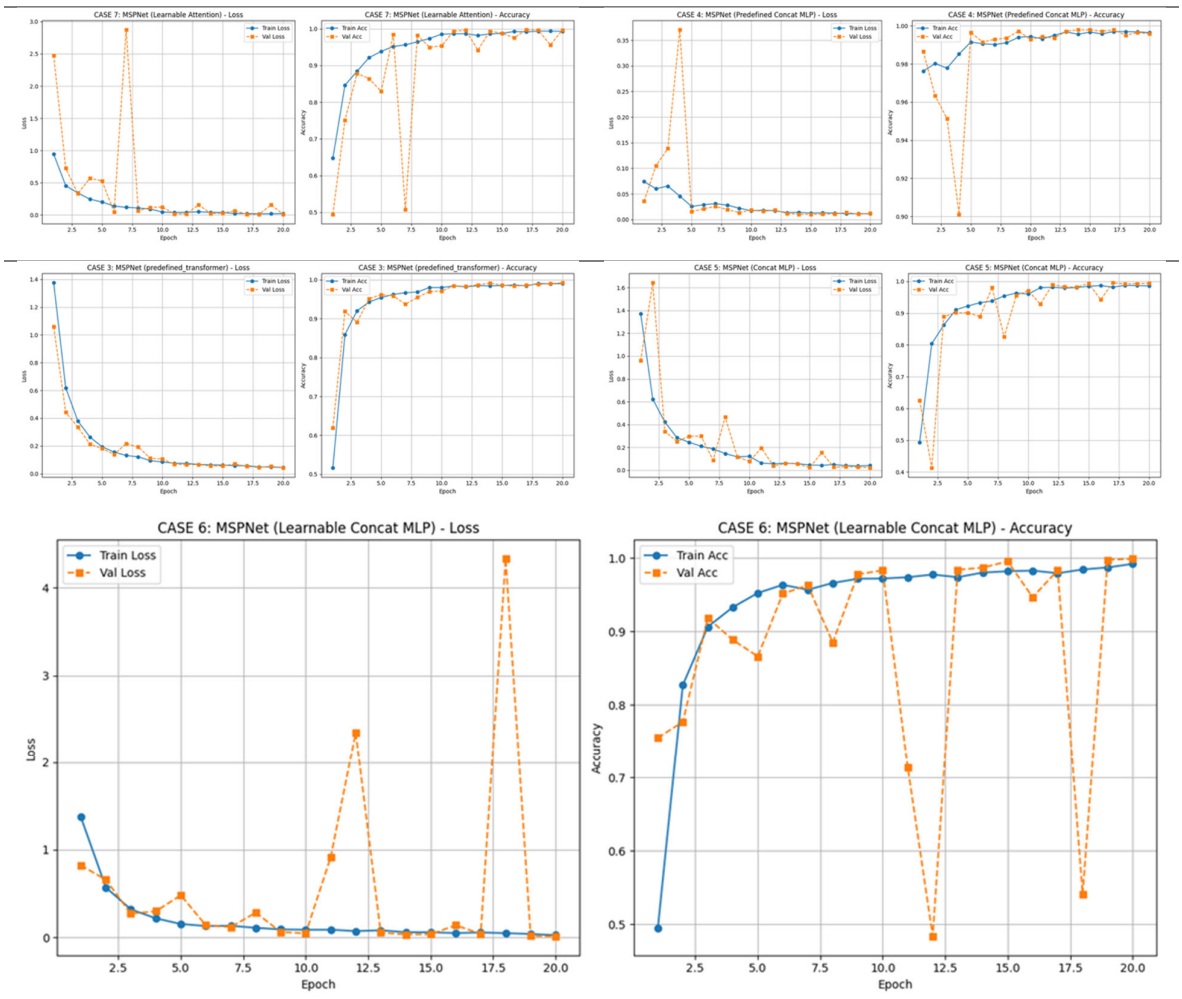


Figure 18. Training curves for MSPNet: (a) learnable_attention, (b) predefined _attention, (c) predefined_transformer, (d) predefined_concat_mlp, (e) learnable_concat_mlp.

4.3. Comparative Analysis

 doi:10.20944/preprints202511.1410.v1

Table 1 and Table.2 summarize performance across both datasets. Spatial-only CNN and FFTNet plateau at ≈81% in the 11-class scenario, reflecting their inability to capture multi-scale spectral cues. MSP-Net variants consistently improve accuracy by 13–14%, reduce confusions, and yield better-calibrated probabilities. Attention-based fusion maximizes discriminability, while concat-MLP ensures superior calibration. These results validate MSP-Net's robustness across datasets with different class compositions and target complexities.

Table 1. Classification Reports.

Class	Baseline CNN	FFTNet	Learnable Attention	Predefined Attention	Predefined Transformer	Predefined Concat	Learnable Concat
2S1	0.90	0.95	0.99	0.99	0.97	0.98	0.98
BMP2	✗ 0.60	✗ 0.21	⚠ 0.78	⚠ 0.78	⚠ 0.75	0.80	✗ 0.60
BRDM2	0.99	1.00	0.98	0.98	0.98	0.99	0.99
BTR60	✗ 0.70	✗ 0.58	⚠ 0.82	0.82	0.84	0.83	0.83
BTR70	✗ 0.68	✗ 0.57	0.89	0.89	0.89	0.87	0.80
D7	0.77	0.90	0.98	0.98	0.96	0.98	0.98
SLICY	✓ 0.99	✓ 1.00	✓ 1.00	✓ 1.00	✓ 1.00	✓ 1.00	✓ 1.00
T62	0.70	0.83	0.99	0.99	0.97	0.99	0.99
T72	0.88	0.70	0.93	0.93	0.89	0.96	0.96
ZIL131	0.81	0.91	0.98	0.98	0.97	0.97	0.97
ZSU_23_4	✗ 0.72	0.86	0.98	0.98	0.96	0.96	0.96
Accuracy	✗ 0.81	✗ 0.81	✓ 0.95	✓ 0.95	✓ 0.94	✓ 0.95	0.91
Macro F1	✗ 0.80	✗ 0.77	✓ 0.94	✓ 0.94	✓ 0.93	✓ 0.94	0.89

Table 2. Classification Reports: Datasets two.

Class	Baseline CNN	FFTNet	Predefined Transformer	Predefined Attention	Predefined Concat MLP	Learnable Concat MLP	Learnable Attention
2S1	0.983	0.925	0.992	0.974	0.974	✓ 0.999	0.989
BRDM_2	1.000	1.000	1.000	1.000	1.000	✓ 0.999	1.000
BTR_60	1.000	0.998	1.000	1.000	1.000	✓ 0.999	1.000
D7	1.000	1.000	0.988	1.000	1.000	✓ 0.999	1.000
SLICY	1.000	1.000	1.000	1.000	1.000	✓ 0.999	1.000
T62	1.000	1.000	0.994	1.000	1.000	✓ 0.999	1.000
ZIL131	1.000	1.000	1.000	0.997	0.997	✓ 0.999	1.000
ZSU_23_4	0.986	0.946	0.993	0.977	0.977	✓ 0.999	0.991
Accuracy	✓ 0.996	0.983	✓ 0.997	✓ 0.993	✓ 0.993	✓ 0.999	✓ 0.997
Macro F1	✓ 0.996	0.983	✓ 0.996	✓ 0.994	✓ 0.994	✓ 0.999	✓ 0.997

 doi:10.20944/preprints202511.1410.v1

4.4. Ablation Study and Insights

To further support the effectiveness of each component in MSP-Net, we analyze the performance of its major architectural variations, which collectively constitute an ablation study. Prior frequency-based SAR ATR works—such as SFONet [56], ASGF [57], and dual-branch spatial–frequency fusion networks [58]–[60]—primarily rely on single-scale FFT representations and rigid fusion mechanisms. Although Zhang et al. [61] introduced spectral attention through FGSA-Net, these approaches still fail to capture the multi-level spectral dynamics characteristic of SAR backscattering.

In contrast, MSP-Net introduces explicit low-, mid-, and high-frequency decomposition, dual filtering strategies, and multiple fusion mechanisms. The comparative analysis across the five MSP-Net variants shows that removing the multi-scale decomposition or replacing adaptive attention-based fusion with simple concatenation results in a clear drop in recognition performance, particularly for small or spectrally similar targets such as BMP2, BTR60, and BTR70. These outcomes reinforce the necessity of multi-scale spectral processing to capture fine scattering cues, as well as adaptive fusion to effectively weight complementary frequency-band representations.

Overall, MSP-Net delivers a 14% absolute improvement over spatial-only and single-scale FFT baselines, maintains strong reliability with ECE < 0.02, and achieves near-perfect ROC separability. This confirms that the integration of multi-scale spectral cues and adaptive fusion significantly enhances robustness, interpretability, and generalization in SAR ATR.

6. Conclusions and Future Work

This work presented MSP-Net, a Multi-Scale Spectrum Pyramid Network developed to address the fundamental limitations of spatial-only CNNs and single-scale frequency-domain ATR models in SAR target recognition. By decomposing SAR imagery into low-, mid-, and high-frequency sub-bands and integrating their complementary representations through adaptive fusion mechanisms, MSP-Net achieves a principled balance between global structural information and fine-grained scattering characteristics.

Comprehensive experiments on two MSTAR-based datasets demonstrate that MSP-Net delivers substantial performance gains, exceeding CNN and FFTNet baselines by 13–14% accuracy in the challenging 11-class setting and achieving up to 99.9% accuracy with near-perfect ROC separability (AUC ≈ 1.0) on the reduced 8-class benchmark. Beyond improved classification accuracy, MSP-Net provides markedly better reliability, achieving ECE < 0.02 and producing well-calibrated confidence scores across all fusion variants. The ablation study further verifies the indispensable role of multi-scale spectral decomposition and attention-based fusion, particularly for mitigating class imbalance and enhancing recognition of visually similar or small targets.

Overall, MSP-Net establishes a robust, reliable, and interpretable framework for SAR ATR, demonstrating strong generalization across datasets with varying target compositions and spectral characteristics. Future research will extend MSP-Net toward multi-view ATR, real-time onboard deployment, and cross-sensor domain adaptation to support next-generation operational SAR intelligence systems.

Author Contributions: Methodology, software implementation, A.S.E.; conceptualization, formal analysis, investigation, A.S.E.; writing—original draft preparation, A.S.E.; supervision and project administration, S.N.E. ; writing—review and editing, S.N.E.,YO and EA. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Yildiz Technical University (FBG-2025-6799)

Authors: Ibrahim Murat Cilek | Seref Naci Engin | Aisha Sir Elkhatem

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

 doi:10.20944/preprints202511.1410.v1

Data Availability Statement: The publicly available portions of the dataset can be accessed from the original MSTAR distribution site. Additional processed data and experimental outputs used in this study are available from the corresponding author upon reasonable request.

Acknowledgment: This work is the outcome of a joint program between Turkish and Kazakh universities. It is produced within the scope of the project supported by Yildiz Technical University Scientific Research Projects Coordination Unit under project number FBG-2025-6799.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Curlander, J.C.; McDonough, R.N. Synthetic Aperture Radar; Wiley: New York, NY, USA, 1991.
2. Bhanu, B. Automatic target recognition: State-of-the-art survey. IEEE Trans. Aerosp. Electron. Syst. 1986, AES-22, 364–379.
3. Chen, J.; Qiu, X.; Ding, C.; Wu, Y. SAR image classification based on spiking neural network through spike-time-dependent plasticity and gradient descent. ISPRS J. Photogramm. Remote Sens. 2022, 188, 109–124.
4. Moreira, A.; Prats-Iraola, P.; Younis, M.; Krieger, G.; Hajnsek, I.; Papathanassiou, K.P. A tutorial on synthetic aperture radar. IEEE Geosci. Remote Sens. Mag. 2013, 1, 6–43.
5. Reigber, A.; Scheiber, R.; Jager, M.; Prats-Iraola, P.; Hajnsek, I.; Jagdhuber, T.; Papathanassiou, K.P.; Nannini, M.; Aguilera, E.; Baumgartner, S.; et al. Very-high-resolution airborne synthetic aperture radar imaging: Signal processing and applications. Proc. IEEE 2013, 101, 759–783.
6. Choi, J.-H.; Lee, M.-J.; Jeong, N.-H.; Lee, G.; Kim, K.-T. Fusion of target and shadow regions for improved SAR ATR. IEEE Trans. Geosci. Remote Sens. 2022, 60, 1–17.
7. Lee, J.-S. Speckle suppression and analysis for synthetic aperture radar images. Opt. Eng. 1986, 25, 255636.
8. Frost, V.S.; Stiles, J.A.; Shanmugan, K.S.; Holtzman, J.C. A model for radar images and its application to adaptive filtering of multiplicative noise. IEEE Trans. Pattern Anal. Mach. Intell. 1982, 4, 157–166.
9. Ross, T.D.; Bradley, J.J.; Hudson, L.J. SAR ATR: So what's the problem? An MSTAR perspective. In Algorithms for Synthetic Aperture Radar Imagery VI; SPIE: Bellingham, WA, USA, 1999; Volume 3721, pp. 662–672.
10. Achim, A.; Kuruoglu, E.E.; Zerubia, J. SAR image filtering based on the heavy-tailed Rayleigh model. IEEE Trans. Image Process. 2006, 15, 2686–2693.
11. Kang, M.; Leng, X.; Lin, Z.; Ji, K. A modified faster R-CNN based on CFAR algorithm for SAR ship detection. In Proceedings of the 2017 International Workshop on Remote Sensing with Intelligent Processing (RSIP), Shanghai, China, 18–21 May 2017; IEEE: New York, NY, USA, 2017.
12. Kreithen, D.E.; Halversen, S.D.; Owirka, G.J. Discriminating targets from clutter. Linc. Lab. J. 1993, 6, 25–52.
13. Schwegmann, C.P.; Kleynhans, W.; Salmon, B.P.; Mdakane, L.W.; Meyer, R.G. Very deep learning for ship discrimination in SAR imagery. In Proceedings of IGARSS 2016, Beijing, China, 10–15 July 2016; pp.104–107.
14. Li, Y.; Chang, Z.; Ning, W. A survey on feature extraction of SAR images. In Proceedings of the ICCASM 2010, Taiyuan, China, 22–24 October 2010; IEEE, pp. V1-312–317.
15. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Fei-Fei, L. ImageNet large scale visual recognition challenge. Int. J. Comput. Vis. 2015, 115, 211–252.
16. Wilmanski, M.; Kreucher, C.; Lauer, J. Modern approaches in deep learning for SAR ATR. In Algorithms for Synthetic-Aperture Radar Imagery XXIII; SPIE 9843, 195–204, 2016.
17. Yu, J.; Zhou, G.; Zhou, S.; Yin, J. A lightweight fully convolutional neural network for SAR ATR. Remote Sens. 2021, 13, 3029.
18. Chen, S.; Wang, H.; Xu, F.; Jin, Y.Q. Target classification using deep convolutional networks for SAR images. IEEE Trans. Geosci. Remote Sens. 2016, 54, 4806–4817.
19. Zhang, M.; An, J.; Yu, D.H.; Yang, L.D.; Wu, L.; Lu, X.Q. CNN with attention mechanism for SAR ATR. IEEE Geosci. Remote Sens. Lett. 2022, 19, 1–5.

 doi:10.20944/preprints202511.1410.v1

20. Ying, Z.; Xuan, C.; Zhai, Y.; Sun, B.; Li, J.; Deng, W.; Mai, C.; Wang, F.; Labati, R.D.; Piuri, V.; et al. TAI-SARNet: Deep transferred atrous-inception CNN for small-sample SAR ATR. Sensors 2020, 20, 1724.
21. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. Nature 2015, 521, 436–444.
22. Huang, Z.; Yao, X.; Liu, Y.; Dumitru, C.O.; Datcu, M.; Han, J. Physically explainable CNN for SAR image classification. ISPRS J. Photogramm. Remote Sens. 2022, 190, 25–37.
23. Huang, Z.; Pan, Z.; Lei, B. What, where, and how to transfer in SAR target recognition using deep CNNs. IEEE Trans. Geosci. Remote Sens. 2020, 58, 2324–2336.
24. Wen, Z.; Liu, Z.; Zhang, S.; Pan, Q. Rotation-aware self-supervised learning for SAR ATR with limited samples. IEEE Trans. Image Process. 2021, 30, 7266–7279.
25. Ren, B.; Zhao, Y.; Hou, B.; Chanussot, J.; Jiao, L. Mutual-information-based self-supervised learning for PolSAR classification. IEEE Trans. Geosci. Remote Sens. 2021, 59, 9224–9237.
26. Zhang, J.; Xing, M.; Xie, Y. Feature fusion framework combining electromagnetic scattering and deep CNNs for SAR ATR. IEEE Trans. Geosci. Remote Sens. 2021, 59, 2174–2187.
27. Shi, X.; Zhou, F.; Yang, S.; Zhang, Z.; Su, T. SAR ATR based on SRGAN and deep CNN. Remote Sens. 2019, 11, 135.
28. Lin, Z.; Ji, K.; Kang, M.; Leng, X.; Zou, H. Deep convolutional highway unit network for SAR ATR with limited data. IEEE Geosci. Remote Sens. Lett. 2017, 14, 1091–1095.
29. Wang, C.; Shi, J.; Zhou, Y.; Yang, X.; Zhang, X. Semi-supervised SAR ATR via self-consistent augmentation. IEEE Trans. Geosci. Remote Sens. 2020, 59, 4862–4873.
30. Zhang, F.; Wang, Y.; Ni, J.; Zhou, Y.; Hu, W. SAR small-sample recognition using CNN cascaded features and rotation-forest. IEEE Geosci. Remote Sens. Lett. 2020, 17, 1008–1012.
31. Zhang, F.; Hu, C.; Yin, Q.; Li, W.; Li, H.-C.; Hong, W. Multi-aspect-aware bidirectional LSTM for SAR ATR. IEEE Access 2017, 5, 26880–26891.
32. Cho, J.H.; Park, C.G. Multiple feature aggregation using CNNs for SAR ATR. IEEE Geosci. Remote Sens. Lett. 2018, 15, 1882–1886.
33. Yu, Q.; Hu, H.; Geng, X.; Jiang, Y.; An, J. Deep feature fusion network for SAR ATR under limited data. IEEE Access 2019, 7, 165646–165658.
34. Fu, Y.; Liu, Z.; Zhang, Z. Progressive learning vision transformer for open set recognition in remote sensing. IEEE Trans. Geosci. Remote Sens. 2023, 61, 5215113.
35. Fu, Y.; Liu, Z.; Wu, C.; Wu, F.; Liu, M. Class-incremental object recognition in remote sensing using dynamic hybrid exemplar selection. IEEE Trans. Aerosp. Electron. Syst. 2024, 60, 3468–3481.
36. Sun, Y.; Wang, Y.; Liu, H.; Wang, N.; Wang, J. SAR ATR with limited training data via angular rotation GAN. IEEE Geosci. Remote Sens. Lett. 2020, 17, 1928–1932.
37. Zhang, W.; Zhu, Y.; Fu, Q. Semi-supervised deep transfer learning for SAR ATR. IEEE Access 2019, 7, 152412–152420.
38. Zhang, W.; Zhu, Y.; Fu, Q. Deep transfer learning based on GANs for SAR ATR. In Proceedings of ICSIDP 2019; pp. 1–5.
39. Wang, C.; et al. Limited-data SAR ATR via dual-invariance intervention. IEEE Trans. Geosci. Remote Sens. 2025, 63, 1–19. doi:10.1109/TGRS.2025.3528464.
40. Li, W.; Yang, W.; Zhang, W.; Liu, T.; Liu, Y.; Liu, L. Hierarchical disentanglement-alignment network for SAR vehicle recognition. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 2023, 16, 9661–9679.
41. Feng, S.; Ji, K.; Wang, F.; Zhang, L.; Ma, X.; Kuang, G. ESF module embedded ASC-based network for robust SAR ATR. IEEE Trans. Geosci. Remote Sens. 2022, 60, 5235415.
42. Fu, K.; Fu, J.; Wang, Z.; Sun, X. Scattering-keypoint-guided network for oriented ship detection. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 2021, 14, 11162–11178.
43. Fu, J.; Sun, X.; Wang, Z.; Fu, K. Anchor-free feature-balancing network for multiscale SAR ship detection. IEEE Trans. Geosci. Remote Sens. 2021, 59, 1331–1344.
44. Sun, X.; Lv, Y.; Wang, Z.; Fu, K. SCAN: Scattering characteristics network for few-shot aircraft classification in SAR images. IEEE Trans. Geosci. Remote Sens. 2022, 60, 5226517.
45. Li, W.; Yang, W.; Hou, Y.; Liu, L.; Liu, Y.; Li, X. SARATR-X: Towards a foundation model for SAR ATR. IEEE Trans. Image Process. 2025, 34, 869–884.

 doi:10.20944/preprints202511.1410.v1

46. Zhang, P.; et al. SEFEPNet: Scale expansion and feature enhancement pyramid network for SAR aircraft detection. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 2022, 15, 3365–3375.
47. Zhou, J.; et al. DiffDet4SAR: Diffusion-based aircraft detection for SAR images. IEEE Geosci. Remote Sens. Lett. 2024, 21, 1–5.
48. Li, W.; Yang, W.; Hou, Y.; Liu, L.; Liu, Y.; Li, X. SARATR-X: Toward a foundation model for SAR ATR. IEEE Trans. Image Process. 2025, 34, 869–884.
49. Zhong, Y.; et al. Detecting camouflaged objects in frequency domain. In Proceedings of CVPR 2022; pp. 4494–4503.
50. Cong, R.; et al. Frequency perception network for camouflaged object detection. In ACM Multimedia 2023; pp. 1179–1189.
51. Lin, J.; Tan, X.; Xu, K.; Ma, L.; Lau, R.W.H. Frequency-aware camouflaged object detection. ACM Trans. Multimedia Comput. Commun. Appl. 2023, 19, 1–16.
52. Zhou, M.; Huang, J.; Guo, C.-L.; Li, C. Fourmer: Efficient global modeling for image restoration. In Proceedings of ICML 2023; pp. 42589–42601.
53. Li, C.; et al. Embedding Fourier for ultra-high-definition low-light image enhancement. In Proceedings of ICLR 2023; pp.1–27.
54. Qin, Z.; Zhang, P.; Wu, F.; Li, X. FCA-Net: Frequency channel attention networks. In Proceedings of ICCV 2021; pp. 763–772.
55. Xu, K.; et al. Learning in the frequency domain. In Proceedings of CVPR 2020; pp. 1737–1746.
56. Song, H.; Xu, W.; Wang, L.; Chen, J.; Yu, H. SFONet: Joint spatial–frequency domain algorithm for mult-class ship detection in SAR images. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 2025. doi:10.1109/JSTARS.2025.3595436.
57. Zhao, C.; Ma, L.; Wang, L.; Ohtsuki, T.; Mathiopoulos, P.T.; Wang, Y. SAR image change detection using spatial–frequency attention and gated linear units. IEEE Geosci. Remote Sens. Lett. 2023, 20, 4002205.
58. Li, C.; Ni, J.; Luo, Y.; Wang, D.; Zhang, Q. A dual-branch spatial–frequency domain fusion method with cross attention for SAR ATR. Remote Sens. 2025, 17, 2378. https://doi.org/10.3390/rs17142378.
59. Zhang, S.; Kong, D.; Xing, Y.; Lu, Y.; Ran, L.; Liang, G.; Wang, H.; Zhang, Y. Frequency-guided spatial adaptation for camouflaged object detection. arXiv 2024, arXiv:2409.12421.
60. Zhang, Z.; et al. Time–frequency-aware hierarchical optimization for SAR jamming recognition. IEEE Trans. Aerosp. Electron. Syst. 2025, 61, 10619–10635.
61. Yang, K.; Ma, F.; Ran, D.; Ye, W.; Li, G. Fast deceptive jamming signal generation using spatial-frequency interpolation. IEEE Trans. Geosci. Remote Sens. 2022, 60, 4701015.