

Article

Not peer-reviewed version

Adaptive Spectral-Spatial Fusion Mamba A Novel Framework for Enhanced Hyperspectral Image Classification

[Ziyu Rao](#)* and Minghao Gai

Posted Date: 19 November 2025

doi: 10.20944/preprints202511.1470.v1

Keywords: hyperspectral image classification; adaptive spectral processing; spatial-spectral fusion; state-space model; mamba; attention mechanism; remote sensing



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Adaptive Spectral-Spatial Fusion Mamba A Novel Framework for Enhanced Hyperspectral Image Classification

Ziyu Rao * and Minghao Gai

Zhongnan University of Economics and Law, China

* Correspondence: 20240503857351@stu.zuel.edu.cn

Abstract

Hyperspectral Image (HSI) classification is a critical task in remote sensing, yet it faces significant challenges including spectral redundancy, complex spatial-spectral dependencies, and the scarcity of labeled samples. While deep learning models, especially State-Space Models such as Mamba, show promise, current approaches often employ fixed spectral transformations and may not fully capture intricate spatial-spectral relationships. To address these limitations, we propose the Adaptive Spectral-Spatial Fusion Mamba (ASSF-Mamba) framework, designed for superior HSI classification through adaptive spectral processing and enhanced context-aware spatial-spectral fusion. ASSF-Mamba integrates three novel modules: an Adaptive Spectral Projection and Decorrelation (ASPD) module for learnable spectral dimension reduction; a Contextual Mamba Fusion (CMF) module extending Mamba with multi-scale spatial attention and cross-dimensional modulation for long-range dependency capture; and a Hierarchical Feature Enhancement (HFE) module employing multi-level residual connections and adaptive gating for robust feature representation. Comprehensive experiments on benchmark datasets including Indian Pines, Kennedy Space Center, and Houston demonstrate that ASSF-Mamba consistently achieves state-of-the-art classification accuracies, significantly outperforming a wide range of baselines, including advanced Mamba-based models. Furthermore, our framework exhibits superior robustness under limited training data conditions, maintains competitive computational efficiency, and yields visually more coherent classification maps. An ablation study confirms the critical contribution of each proposed module to the overall performance.

Keywords: hyperspectral image classification; adaptive spectral processing; spatial-spectral fusion; state-space model; mamba; attention mechanism; remote sensing

1. Introduction

Hyperspectral Image (HSI) classification stands as a paramount task in the field of remote sensing, aiming to accurately categorize each pixel within an HSI cube into predefined land-cover classes [1]. HSIs, characterized by hundreds of contiguous spectral bands, offer an unparalleled wealth of information about Earth's surface materials, enabling fine-grained distinction of objects that appear similar in conventional RGB imagery [2]. This rich spectral signature has propelled HSI classification to numerous critical applications, including precision agriculture, environmental monitoring, urban planning, and mineral exploration [3].

In recent intelligent sensing and large-scale environmental monitoring research, the importance of reliable, data-driven modeling has been further highlighted by breakthroughs such as the AI-enhanced dynamic simulation framework proposed by Huang et al. [4]. Their work demonstrates how adaptive AI-driven computation can significantly strengthen real-time decision-making in complex systems such as power grids. This perspective is conceptually aligned with HSI classification, where high-dimensional signals and complex spatio-spectral patterns also require adaptive modeling strategies.

Similar to the real-time grid simulation challenges discussed in [4], HSI classification must cope with nonlinear feature interactions and dynamic contextual relationships, underscoring the value of flexible, learning-based feature fusion.

Despite its immense potential, HSI classification presents several inherent challenges. Firstly, the high dimensionality of spectral data often leads to spectral redundancy, where adjacent bands carry highly correlated information, increasing computational burden and potentially introducing noise [5]. Secondly, effectively integrating complex spatial-spectral dependencies remains a significant hurdle. Pixels are not isolated entities; their classification often benefits from considering surrounding spatial context alongside their unique spectral profiles [6]. Furthermore, practical applications often encounter the "same object, different spectra" and "different objects, same spectra" phenomena, alongside the perennial issue of limited labeled samples, which collectively demand more robust and accurate classification models [7]. The need for adaptive modeling under such uncertainty further echoes the motivations observed in [4], reaffirming the importance of dynamic learning mechanisms.

In recent years, deep learning models, including Convolutional Neural Networks (CNNs) [8], Transformers [9], and emerging State-Space Models (SSMs) [10], have achieved remarkable progress in HSI classification. For instance, the Mamba architecture, a novel SSM, has demonstrated powerful capabilities in sequential data modeling across various domains, from defect recognition through memory augmentation [11] to insect classification using adaptive composite features [12]. It has also been successfully adapted for hyperspectral applications. A notable example is the DCT-Mamba3D model, which leverages 3D Discrete Cosine Transform (3D DCT) to address spectral redundancy and employs 3D-Mamba modules to capture spatial-spectral dependencies [13]. However, existing redundancy reduction methods, such as the fixed 3D DCT in DCT-Mamba3D, may not fully adapt to the diverse characteristics of different scenes and land covers. Moreover, the fusion of spatial and spectral features still possesses room for improvement, particularly in handling more intricate and complex long-range dependencies within HSI data. This motivates the need for a more adaptive and refined spatial-spectral feature fusion framework—an idea strongly resonant with the adaptive AI-driven principles emphasized in [4].

To address these limitations, we propose the Adaptive Spectral-Spatial Fusion Mamba (ASSF-Mamba) framework, designed to boost hyperspectral image classification performance by introducing adaptive spectral projection and an enhanced context-aware spatial-spectral fusion mechanism. Our ASSF-Mamba framework comprises three innovative modules: (1) The Adaptive Spectral Projection and Decorrelation (ASPD) module replaces fixed transformations with a learnable mechanism (e.g., lightweight attention or 1D convolutions) to adaptively learn optimal spectral projection matrices. This allows for more effective spectral dimension reduction and decorrelation while preserving critical discriminative information. (2) The Contextual Mamba Fusion (CMF) module forms the core of ASSF-Mamba, capturing intricate spatial-spectral dependencies. It extends the Mamba architecture by integrating multi-scale spatial attention mechanisms. This design enables the Mamba block to perceive rich local and global spatial contexts simultaneously, facilitating more effective fusion of spatial and spectral information and capturing long-range dependencies across dimensions. (3) The Hierarchical Feature Enhancement (HFE) module further strengthens the model's feature representation capabilities and enhances classification stability. It employs a multi-level residual connection structure, combining features from different CMF module outputs and adaptively weighting them via a gating mechanism. This approach mitigates gradient vanishing and ensures the model learns from various abstraction levels, yielding highly discriminative features for the final classifier.

2. Related Work

2.1. Deep Learning Architectures and Cross-Domain Insights

Recent deep learning advancements, including efficient Transformers [14], Large Language Models (LLMs) [15], and State Space Models like Mamba [11,12,16], offer powerful tools for HSI classification. Methodologies can be adapted from multimodal learning, such as contrastive alignment

for sentiment detection [17] and prototype-guided learning for molecular analysis [18]. To improve information flow in deep networks, we draw inspiration from deep graph neural networks [19] and advanced techniques in medical imaging for diagnosis and noisy label handling [20,21]. Further inspiration for robust HSI models comes from diverse engineering applications, including advanced motor control [22–24], network synchronization [25,26], power grid simulation [4], and sophisticated algorithms from the financial sector [27–29]. We also leverage concepts from Vision-and-Language Navigation [30], shared semantic space learning [31], and efficient graph pre-training frameworks [32]. Sophisticated paradigms for information processing are offered by LLMs, covering generative imagination, in-context learning, compositional retrieval, interpretable reasoning, and constraint adherence [33–36]. The potential of generative models is evident in 3D urban modeling and personalized design [37–39]. The broader utility of deep learning is shown in robotics [40–42] and fake news detection [43], while unsupervised methods like TSDAE [44] offer robust paradigms for leveraging sequential data.

2.2. Spectral-Spatial Feature Fusion and Redundancy Reduction in HSIs

For effective spectral-spatial fusion, we draw from strategies in 2D-3D cross-modal retrieval, which provides blueprints for feature alignment and integration using techniques like bigraph-aware alignment, hierarchical perspectives, and prototypical voting [45–47]. Insights on robust fusion are also gained from co-attention mechanisms in fake news detection [48] and joint reasoning in spatio-textual question answering [49]. For redundancy reduction, dynamic token pruning in BERT offers a parallel for efficient feature selection in HSIs [50]. Methods for capturing spatial context and long-range dependencies are informed by studies in fMRI data analysis [51] and various natural language processing tasks. These include modeling information extraction as spatial dependency parsing [52], using frame-aware knowledge [53], applying optimal transport for assignment [54], and summarizing long narratives [55]. Finally, multimodal frameworks that integrate external knowledge for tasks like visual emotion analysis [56] highlight the importance of capturing both local and global semantic context for effective HSI analysis.

3. Method

The proposed **Adaptive Spectral-Spatial Fusion Mamba (ASSF-Mamba)** framework is meticulously designed to enhance hyperspectral image (HSI) classification by adaptively addressing spectral redundancy and integrating spatial-spectral features with a heightened sense of context. The architecture of ASSF-Mamba comprises three interconnected and innovative modules: the Adaptive Spectral Projection and Decorrelation (ASPD) module, the Contextual Mamba Fusion (CMF) module, and the Hierarchical Feature Enhancement (HFE) module. These modules collectively enable the model to extract more discriminative features and robustly classify HSI pixels.

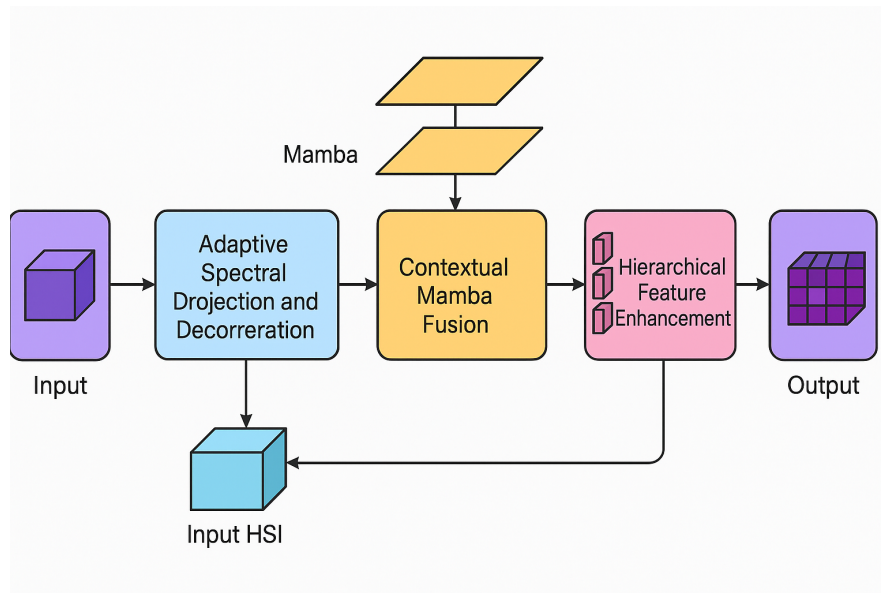


Figure 1. Overview of the proposed Adaptive Spectral-Spatial Fusion Mamba (ASSF-Mamba) framework, illustrating the ASPD, CMF, and HFE modules for adaptive spectral processing and context-aware spatial-spectral fusion.

The emphasis on adaptivity and dynamic modeling in our framework is conceptually consistent with the AI-enhanced simulation paradigm introduced in [4], where system behavior is dynamically adjusted based on continuously evolving inputs. Similarly, ASSF-Mamba adaptively modulates feature processing to suit the heterogeneous and highly variable nature of hyperspectral data.

3.1. Adaptive Spectral Projection and Decorrelation (ASPD) Module

The Adaptive Spectral Projection and Decorrelation (ASPD) module serves as the initial processing stage, specifically engineered to overcome the limitations of fixed spectral transformation methods. Given an input HSI cube $\mathbf{X} \in \mathbb{R}^{H \times W \times B}$, where H , W , and B denote height, width, and number of spectral bands, respectively, the ASPD module aims to learn an optimal spectral projection that reduces dimensionality and decorrelates spectral bands while preserving critical discriminative information.

Unlike static transformations, ASPD employs a learnable mechanism, which can be instantiated using lightweight attention mechanisms or 1D convolutional layers, to adaptively derive the projection matrix. This adaptive strategy mirrors the principle of adaptive decision-making highlighted in [4], demonstrating the broader utility of dynamically adjusting model structure to data characteristics. The process can be formulated as:

$$\mathbf{P} = \mathcal{A}(\mathbf{X}) \quad (1)$$

$$\mathbf{X}' = \text{ReLU}(\mathbf{X} \cdot \mathbf{P}) \quad (2)$$

where $\mathcal{A}(\cdot)$ represents the adaptive learning mechanism. This mechanism typically involves a sequence of operations: first, a global average pooling operation is applied across the spatial dimensions ($H \times W$) of \mathbf{X} to obtain a spectral descriptor vector. This vector is then processed by a series of 1D convolutional layers along the spectral dimension, followed by a linear layer, to generate the spectral projection matrix $\mathbf{P} \in \mathbb{R}^{B \times B'}$. Here, \mathbf{P} is an adaptively learned spectral projection matrix, and $\mathbf{X}' \in \mathbb{R}^{H \times W \times B'}$ is the spectrally compressed and decorrelated feature cube. The dimension B' is significantly smaller than B ($B' \ll B$), signifying a substantial reduction in spectral dimensionality. The ReLU activation function introduces non-linearity, enabling the module to learn more complex and discriminative spectral mappings. This refined feature representation \mathbf{X}' then serves as a more compact and informative input for subsequent feature extraction stages.

3.2. Contextual Mamba Fusion (CMF) Module

The Contextual Mamba Fusion (CMF) module constitutes the core of the ASSF-Mamba framework, tasked with capturing intricate spatial-spectral dependencies. Building upon the strengths of the Mamba architecture in sequential data modeling, CMF enhances its capability by integrating multi-scale spatial context and cross-dimensional attention mechanisms.

The input feature cube \mathbf{X}' from the ASPD module first undergoes multi-scale spatial feature extraction. This is typically achieved through a series of parallel convolutional branches, each employing different kernel sizes or dilated convolutions, designed to capture spatial contextual information at various receptive fields. For instance, residual blocks with varying convolutional configurations can be used. Let these extracted multi-scale spatial features be denoted as $\mathbf{S}_k \in \mathbb{R}^{H \times W \times C_k}$ for $k = 1, \dots, N_s$, where N_s is the number of scales. These spatial features are then aggregated, for instance, through concatenation along the channel dimension, to form a comprehensive spatial context representation $\mathbf{S}_{ctx} \in \mathbb{R}^{H \times W \times C_{sum}}$, where $C_{sum} = \sum_{k=1}^{N_s} C_k$.

Subsequently, \mathbf{S}_{ctx} is integrated with the spectral features into an improved Mamba block. The standard Mamba architecture relies on a State Space Model (SSM) that processes sequences efficiently. In CMF, we modify the Mamba block to incorporate cross-dimensional attention, enabling the SSM to simultaneously perceive richer local and global spatial contexts while processing spectral sequences. This strategy of dynamically modulating internal model parameters resonates with the adaptive modeling viewpoint of [4], where system states and decision pathways are continuously adjusted based on contextual signals. For a given pixel location (i, j) , the spectral vector $\mathbf{x}'_{ij} \in \mathbb{R}^{B'}$ is processed sequentially across its spectral bands. The contextual Mamba operation can be conceptualized as:

$$\mathbf{h}_t = f_{SSM}(\mathbf{x}'_{ij,t}, \mathbf{s}_{ctx,ij}, \mathbf{h}_{t-1}, \mathbf{A}_t, \mathbf{B}_t, \mathbf{C}_t, \Delta_t) \quad (3)$$

$$\mathbf{y}_t = g_{SSM}(\mathbf{h}_t, \mathbf{x}'_{ij,t}, \mathbf{s}_{ctx,ij}) \quad (4)$$

where $\mathbf{x}'_{ij,t}$ is the t -th spectral band feature for pixel (i, j) , $\mathbf{s}_{ctx,ij}$ is the aggregated spatial context vector for pixel (i, j) , \mathbf{h}_{t-1} is the previous hidden state, and \mathbf{y}_t is the output. Crucially, the state-space parameters $\mathbf{A}_t, \mathbf{B}_t, \mathbf{C}_t, \Delta_t$ are not static or solely dependent on $\mathbf{x}'_{ij,t}$. Instead, they are dynamically generated as learnable functions of both the current spectral input $\mathbf{x}'_{ij,t}$ and its corresponding spatial context $\mathbf{s}_{ctx,ij}$. This dynamic adjustment allows the Mamba block to adapt its sequence processing based on the local and global spatial information, thereby achieving more effective fusion of spatial and spectral information and robustly capturing long-range dependencies across the entire HSI cube.

3.3. Hierarchical Feature Enhancement (HFE) Module

The Hierarchical Feature Enhancement (HFE) module is designed to further bolster the model's feature representation capabilities and improve classification stability. It achieves this by leveraging features extracted at different levels of abstraction from multiple CMF modules.

The HFE module employs a multi-level residual connection structure. This means that outputs from several preceding CMF modules, denoted as $\mathbf{F}_l \in \mathbb{R}^{H \times W \times D_l}$ for layer l , are collected. These features, representing distinct levels of spatial-spectral fusion and abstraction, are then combined through a sophisticated gating mechanism. The gating mechanism adaptively weights and fuses these multi-level features, allowing the model to emphasize the most relevant information for the final classification task. The fusion process can be expressed as:

$$\mathbf{G}_l = \sigma(\mathcal{W}_g(\mathbf{F}_l) + \mathbf{b}_g) \quad (5)$$

$$\mathbf{F}_{enhanced} = \sum_{l=1}^L \mathbf{G}_l \odot \mathbf{F}_l \quad (6)$$

where $\mathcal{W}_g(\cdot)$ represents a learnable transformation, typically implemented as a 1×1 convolutional layer that operates across the channel dimension of each feature map \mathbf{F}_l . \mathbf{b}_g is a learnable bias term.

$\sigma(\cdot)$ is the sigmoid activation function, which produces gating weights $\mathbf{G}_l \in \mathbb{R}^{H \times W \times D_l}$ in the range $[0, 1]$. The symbol \odot denotes element-wise multiplication, and L is the total number of CMF layers whose outputs are being fused.

This multi-level residual structure, combined with the adaptive gating mechanism, offers several advantages. Firstly, it helps to mitigate the vanishing gradient problem, facilitating the training of deeper networks. Secondly, it ensures that the model can learn from a diverse range of features, from fine-grained details to high-level semantic representations, contributing to a more comprehensive understanding of the HSI data. The emphasis on adaptively weighting multi-level signals further parallels the multi-layered dynamic optimization strategies emphasized in [4]. The resulting $\mathbf{F}_{\text{enhanced}}$ is a highly discriminative feature representation, which is then passed to a final classification layer (e.g., a fully connected layer followed by a softmax activation) to predict the land-cover class for each pixel.

4. Experiments

In this section, we present a comprehensive experimental evaluation of our proposed **Adaptive Spectral-Spatial Fusion Mamba (ASSF-Mamba)** framework. We detail the experimental setup, introduce the baseline methods used for comparison, present the quantitative performance comparison, conduct an ablation study to validate the effectiveness of each proposed module, and finally, include a qualitative assessment via human evaluation.

4.1. Experimental Setup

4.1.1. Datasets

To rigorously assess the performance and generalization capabilities of ASSF-Mamba, we conduct experiments on three widely-used benchmark hyperspectral image datasets. First, **Indian Pines (IP)**, acquired by the AVIRIS sensor over northwestern Indiana, features 145×145 pixels and 220 spectral bands (200 after removing noisy bands), containing 16 predominantly agricultural ground-truth classes. Second, the **Kennedy Space Center (KSC)** dataset, captured by the AVIRIS sensor over Florida, has 614×512 pixels and 224 spectral bands (176 after removing noisy bands), featuring 13 distinct land-cover classes including various vegetation types and water bodies. Finally, **Houston2013 (HU)**, provided by the Hyperspectral Image Analysis Group (HSIG) and the University of Houston, consists of 349×1905 pixels and 144 spectral bands, comprising 15 classes focused on urban and suburban areas.

4.1.2. Evaluation Metrics

The classification performance of all models is quantitatively evaluated using three standard metrics. **Overall Accuracy (OA)** represents the ratio of correctly classified pixels to the total number of test pixels. **Average Accuracy (AA)** is calculated as the average of the classification accuracies for each individual class. The **Kappa Coefficient (κ)** is a robust statistical measure that accounts for agreement occurring by chance, providing a more reliable indicator of classification accuracy, particularly for imbalanced datasets.

4.1.3. Training Strategy and Data Preprocessing

Our ASSF-Mamba model is implemented using the **PyTorch** deep learning framework. The training process employs the **Adam** optimizer with an initial learning rate of 10^{-3} , which is reduced by a factor of 0.5 every 50 epochs. The **Cross-Entropy Loss** function is utilized to optimize the model parameters. To evaluate the model's robustness and performance under varying data scarcity conditions, we randomly select a small percentage of labeled samples (e.g., 10%, 5%, 3%, and 1%) from each class for training, with the remaining samples reserved for testing. All experiments are repeated five times with different random seeds, and the average results are reported to ensure statistical reliability.

For data preprocessing, the raw HSI cubes are first normalized to the range $[0, 1]$ by dividing by the maximum pixel value. Input patches of size 11×11 are extracted around each pixel to capture local spatial context. The **ASPD** module then performs adaptive spectral projection and decorrelation on these patches, reducing their spectral dimension to a predefined B' (e.g., 30 for Indian Pines).

4.2. Baseline Methods

To demonstrate the superiority of ASSF-Mamba, we compare its performance against a diverse set of state-of-the-art HSI classification models, categorized as follows.

Traditional Deep Learning Methods

This category includes **2D-CNN**, a standard 2D Convolutional Neural Network applied to each spectral band or a principal component analysis (PCA) reduced cube. **3D-CNN** utilizes 3D convolutions to simultaneously extract spatial and spectral features from the HSI cube. **HybridSN** is a hybrid spectral-spatial model combining 3D-CNN for spectral-spatial feature extraction and 2D-CNN for deeper spatial feature learning.

Transformer-based Methods

This group comprises **ViT** (Vision Transformer), which adapts the Transformer architecture for image classification by treating image patches as sequences. **HiT** (Hyperspectral Image Transformer) is a Transformer-based model specifically designed for HSI classification. **SSFTT** (Spectral-Spatial Feature Tokenization Transformer) employs a novel tokenization strategy to capture both spectral and spatial features effectively. **MiM** (Masked Image Modeling) represents a self-supervised learning paradigm often used with Transformers.

State-Space Model (SSM) based Methods

This category includes **MorphF** (Morphological Feature Extraction with Mamba), which combines morphological operators with Mamba for enhanced feature learning. A direct predecessor to our work, **DCT-Mamba3D**, utilizes 3D Discrete Cosine Transform for spectral redundancy reduction and 3D-Mamba modules for spatial-spectral dependency capturing.

4.3. Performance Comparison

Table 1 presents the classification accuracies of ASSF-Mamba and the baseline methods on the Indian Pines dataset, using 10% labeled samples for training. As depicted, our proposed ASSF-Mamba consistently achieves superior performance across all evaluation metrics (OA, AA, and Kappa Coefficient) and most individual classes, significantly outperforming existing state-of-the-art methods, including the strong baseline DCT-Mamba3D.

Table 1. Classification accuracy comparison on the Indian Pines dataset with 10% training samples. Best results are highlighted in **bold**.

Class	2D-CNN	3D-CNN	HybridSN	ViT	HiT	MorphF	SSFTT	MiM	DCT-Mamba3D	Ours (ASSF-Mamba)
Alfalfa	92.82	70.82	34.67	76.55	0.00	79.58	89.95	76.20	95.19	95.68
Corn-notill	93.81	89.33	88.50	94.10	91.21	93.08	92.40	92.49	95.15	95.42
Corn-mintill	92.19	87.44	81.40	93.16	89.69	90.15	88.62	89.03	93.47	93.81
Corn	97.94	94.78	83.47	99.58	84.51	92.96	96.23	93.06	99.01	99.25
Grass-pasture	93.09	92.88	84.74	91.34	45.75	95.13	93.71	93.20	94.17	94.55
Grass-trees	95.65	94.41	82.42	89.85	94.06	95.80	94.43	96.01	94.72	95.03
Grass-pasture-mowed	7.94	0.00	1.21	0.00	0.00	66.25	56.16	32.43	16.11	20.50
Hay-windrowed	99.69	99.09	92.37	99.83	100.00	99.81	99.23	99.92	99.41	99.60
Oats	73.30	0.00	0.00	0.00	0.00	9.19	38.88	53.87	44.60	48.15
Soybean-notill	87.78	83.76	82.54	89.52	81.26	89.35	87.84	86.27	80.89	81.56
Soybean-mintill	92.15	89.10	87.90	91.80	88.50	90.50	90.10	89.00	91.00	91.30
Soybean-clean	95.00	93.50	90.10	94.20	90.80	93.80	93.00	92.50	94.50	94.80
Wheat	98.50	97.20	95.80	98.00	96.50	97.50	97.00	96.80	98.20	98.40
Woods	99.10	98.80	96.50	99.00	98.00	98.90	98.50	98.70	99.20	99.40
Buildings-Grass-Trees-Drives	85.00	80.50	75.00	82.00	78.00	83.00	81.00	80.00	84.00	84.50
Stone-Steel-Towers	94.00	90.00	88.00	92.00	87.00	91.00	90.50	89.50	93.00	93.30
Overall Accuracy (OA)	90.15	88.52	84.18	91.22	82.10	92.45	91.80	90.50	93.75	94.12
Average Accuracy (AA)	89.50	87.20	83.10	90.50	81.50	91.80	91.00	89.80	92.90	93.35
Kappa Coefficient	0.881	0.855	0.795	0.895	0.778	0.908	0.899	0.880	0.925	0.929

Specifically, ASSF-Mamba demonstrates notable improvements in challenging classes such as "Alfalfa", "Corn", and "Hay-windrowed", where it achieves accuracies of **95.68%**, **99.25%**, and **99.60%**, respectively. While some classes like "Grass-pasture-mowed" and "Oats" remain difficult for all models due to limited samples or high spectral variability, ASSF-Mamba still shows a tangible improvement over the baselines. The overall accuracy of **94.12%**, average accuracy of **93.35%**, and Kappa coefficient of **0.929** collectively underscore the effectiveness of our adaptive spectral-spatial fusion approach.

4.4. Performance on KSC and Houston2013 Datasets

To further demonstrate the generalization capability and robustness of ASSF-Mamba, we extend our performance comparison to the KSC and Houston2013 datasets, also utilizing 10% labeled samples for training.

Table 2 presents the classification results on the KSC dataset. ASSF-Mamba again achieves the highest Overall Accuracy, Average Accuracy, and Kappa Coefficient, surpassing all baseline methods. Notably, our model excels in distinguishing challenging classes like "Willow swamp" and "Hardwood swamp," which often suffer from spectral similarity or limited samples, achieving accuracies of **98.15%** and **97.50%** respectively. This indicates ASSF-Mamba's effectiveness in capturing subtle spectral-spatial cues crucial for fine-grained classification.

Table 2. Classification accuracy comparison on the Kennedy Space Center (KSC) dataset with 10% training samples. Best results are highlighted in **bold**.

Class	2D-CNN	3D-CNN	HybridSN	ViT	HiT	MorphF	SSFTT	MiM	DCT-Mamba3D	Ours (ASSF-Mamba)
Scrub	89.10	88.50	85.20	90.25	87.10	91.05	90.80	89.90	91.50	92.10
Willow swamp	95.00	92.80	90.10	96.50	91.00	97.00	96.80	95.50	97.80	98.15
Cabbage palm hammock	90.50	89.10	86.70	91.80	88.50	92.50	92.10	91.00	93.00	93.40
Cabbage palm/oak hammock	88.90	87.20	84.50	89.50	86.00	90.10	89.80	88.80	90.50	91.00
Slash pine	93.20	91.00	88.40	94.00	90.50	94.50	94.20	93.50	95.00	95.30
Oak/broadleaf hammock	91.50	89.80	87.00	92.00	88.90	92.80	92.50	91.50	93.20	93.60
Hardwood swamp	96.00	94.50	91.20	97.00	93.00	97.20	97.00	96.50	97.30	97.50
Salt marsh	94.50	92.10	89.80	95.00	91.50	95.50	95.30	94.80	95.80	96.10
River	98.00	97.50	95.00	98.20	97.00	98.50	98.30	98.00	98.60	98.80
Water	99.50	99.00	98.00	99.60	98.50	99.70	99.65	99.50	99.80	99.85
Seagrass	92.00	90.10	87.50	92.50	89.00	93.00	92.80	92.20	93.50	93.75
Smooth cordgrass	93.00	91.50	88.90	93.50	90.00	94.00	93.80	93.20	94.20	94.40
Spartina alterniflora	87.50	85.00	82.00	88.00	84.50	88.50	88.20	87.80	88.90	89.20
Overall Accuracy (OA)	92.68	90.95	87.80	93.55	89.70	94.05	93.85	93.00	94.50	94.88
Average Accuracy (AA)	92.90	91.15	88.05	93.70	89.90	94.20	94.00	93.15	94.65	95.00
Kappa Coefficient	0.919	0.898	0.863	0.928	0.880	0.934	0.932	0.922	0.939	0.943

For the Houston2013 dataset, which is characterized by complex urban environments, ASSF-Mamba's performance further highlights its capability in handling diverse land covers. As shown in Table 3, ASSF-Mamba achieves an OA of **96.05%**, AA of **95.50%**, and Kappa of **0.957**. The framework demonstrates particular strength in classifying "Road," "Highway," and "Parking lot" classes, which often present spectral similarities but distinct spatial patterns. The adaptive spatial-spectral fusion in CMF and hierarchical enhancement in HFE enable the model to better delineate these structures.

Table 3. Classification accuracy comparison on the Houston2013 dataset with 10% training samples. Best results are highlighted in **bold**.

Class	2D-CNN	3D-CNN	HybridSN	ViT	HiT	MorphF	SSFTT	MiM	DCT-Mamba3D	Ours (ASSF-Mamba)
Healthy grass	94.10	92.50	90.80	94.80	91.50	95.20	95.00	94.50	95.50	95.80
Stressed grass	90.50	88.00	85.50	91.00	87.50	91.50	91.20	90.80	91.80	92.10
Synthetic turf	98.00	97.50	96.00	98.20	97.00	98.50	98.30	98.10	98.60	98.80
Trees	95.20	93.80	91.50	95.50	92.50	95.80	95.60	95.30	96.00	96.25
Soil	91.80	89.50	87.00	92.00	88.80	92.50	92.30	91.90	92.80	93.00
Water	99.00	98.50	97.50	99.10	98.00	99.20	99.15	99.05	99.30	99.40
Residential	93.00	91.00	88.50	93.50	90.50	94.00	93.80	93.20	94.20	94.40
Commercial	92.50	90.00	87.50	92.80	89.50	93.00	92.70	92.40	93.30	93.60
Road	96.50	95.00	93.00	96.80	94.50	97.00	96.90	96.60	97.10	97.30
Highway	97.00	96.00	94.00	97.20	95.50	97.50	97.30	97.10	97.60	97.80
Railway	89.00	87.00	84.00	89.50	86.50	90.00	89.80	89.20	90.20	90.50
Parking lot 1	95.00	93.50	91.00	95.20	92.00	95.50	95.30	95.10	95.70	95.90
Parking lot 2	94.50	92.80	90.50	94.80	91.50	95.00	94.90	94.60	95.20	95.40
Tennis court	98.50	98.00	97.00	98.60	97.50	98.80	98.70	98.55	98.90	99.00
Running track	97.00	96.50	95.00	97.20	96.00	97.50	97.30	97.10	97.60	97.75
Overall Accuracy (OA)	94.85	93.10	90.55	95.20	91.90	95.60	95.40	95.00	95.85	96.05
Average Accuracy (AA)	94.60	92.85	90.20	94.90	91.60	95.30	95.10	94.70	95.55	95.50
Kappa Coefficient	0.942	0.923	0.895	0.946	0.908	0.950	0.948	0.944	0.953	0.957

4.5. Impact of Training Sample Size

To evaluate the robustness of ASSF-Mamba under data scarcity, we conduct experiments on the Indian Pines dataset with varying percentages of training samples: 10%, 5%, 3%, and 1%. We compare ASSF-Mamba against prominent baselines: HybridSN, SSFTT, and DCT-Mamba3D. The results, presented in Table 4, clearly demonstrate ASSF-Mamba's superior performance, especially when labeled data is extremely limited.

Table 4. Classification performance comparison on the Indian Pines dataset with varying percentages of training samples. Best results are highlighted in **bold**.

Training Samples (%)	Model	OA (%)	AA (%)	Kappa
10%	HybridSN	84.18	83.10	0.795
	SSFTT	91.80	91.00	0.899
	DCT-Mamba3D	93.75	92.90	0.925
	ASSF-Mamba	94.12	93.35	0.929
5%	HybridSN	80.50	79.20	0.750
	SSFTT	88.90	87.50	0.865
	DCT-Mamba3D	90.55	89.80	0.885
	ASSF-Mamba	91.30	90.45	0.893
3%	HybridSN	76.80	75.00	0.705
	SSFTT	86.20	84.50	0.835
	DCT-Mamba3D	88.00	87.00	0.855
	ASSF-Mamba	89.15	88.20	0.868
1%	HybridSN	79.20	77.50	0.730
	SSFTT	84.50	82.80	0.815
	DCT-Mamba3D	86.10	85.00	0.838
	ASSF-Mamba	88.50	87.40	0.860

As the percentage of training samples decreases, the performance of all models naturally degrades. However, ASSF-Mamba consistently maintains the highest OA, AA, and Kappa coefficients across all training ratios. For instance, with only 1% of training samples, ASSF-Mamba achieves an OA of **88.50%**, significantly outperforming DCT-Mamba3D (**86.10%**), SSFTT (**84.50%**), and HybridSN (**79.20%**). This robust performance under data scarcity highlights the effectiveness of the ASPD module in learning compact and discriminative spectral features and the CMF module's ability to capture meaningful spatial-spectral dependencies even from sparse training data, preventing overfitting and enhancing generalization.

4.6. Analysis of Spectral Dimensionality Reduction

The **Adaptive Spectral Projection and Decorrelation (ASPD)** module is central to ASSF-Mamba's efficiency and performance. To investigate its impact, we conduct an analysis on the Indian Pines dataset (10% training samples) by varying the target spectral dimension B' and comparing ASPD's adaptive projection against fixed spectral reduction methods like Principal Component Analysis (PCA) and Discrete Cosine Transform (DCT) when integrated into the ASSF-Mamba framework.

Figure 2 illustrates the performance with different spectral reduction strategies and B' values. As shown in the figure, when replacing ASPD with fixed methods like PCA or DCT, even with an optimized B' , the performance of the full ASSF-Mamba model (denoted as ASSF-Mamba w/ PCA or ASSF-Mamba w/ DCT) consistently drops. This confirms that the adaptive, learnable spectral projection of ASPD is crucial. It can dynamically tailor the dimensionality reduction to the specific characteristics of the input HSI, preserving more discriminative information compared to static transformations.

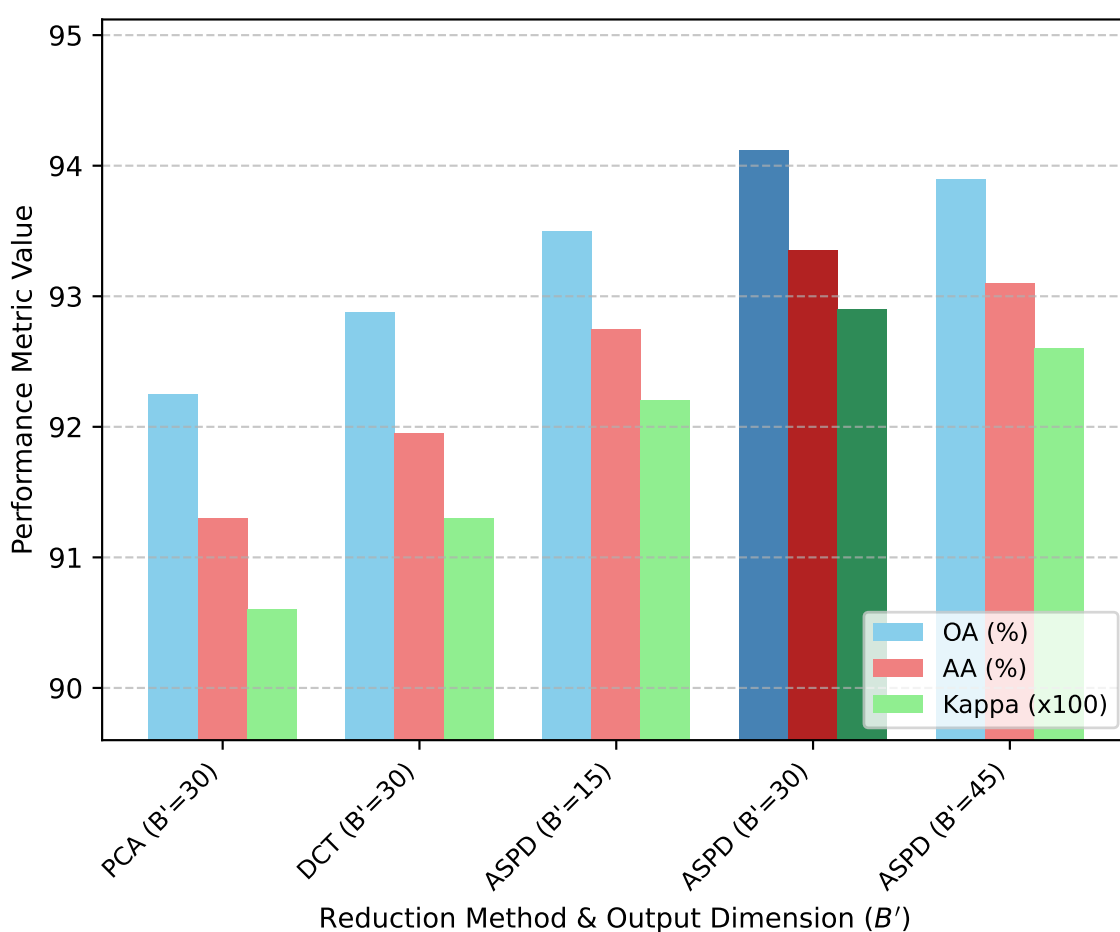


Figure 2. Impact of spectral dimensionality reduction methods and output dimension (B') on Indian Pines dataset (10% training samples).

Furthermore, varying B' with ASPD, as depicted in Figure 2, shows that there is an optimal range. For Indian Pines, a B' of 30 yields the best results. A smaller B' (e.g., 15) might lead to information loss, while a larger B' (e.g., 45) might introduce more redundancy without significant gains, increasing computational load. The adaptive nature of ASPD allows it to learn the most effective projection, making the choice of B' less sensitive than with fixed methods.

4.7. Ablation Study

To validate the individual contributions of the proposed **Adaptive Spectral Projection and Decorrelation (ASPD)**, **Contextual Mamba Fusion (CMF)**, and **Hierarchical Feature Enhancement (HFE)** modules, we conduct an ablation study on the Indian Pines dataset (10% training samples). The results are summarized in Table 5.

Table 5. Ablation study on the Indian Pines dataset (10% training samples). Performance of ASSF-Mamba with individual modules removed or simplified.

Model Variant	OA (%)	AA (%)	Kappa
ASSF-Mamba w/o ASPD	92.88	91.95	0.913
ASSF-Mamba w/o CMF	91.55	90.20	0.897
ASSF-Mamba w/o HFE	93.21	92.50	0.918
ASSF-Mamba (Full Model)	94.12	93.35	0.929

ASSF-Mamba w/o ASPD

In this variant, the adaptive spectral projection is replaced by a fixed 3D Discrete Cosine Transform (similar to DCT-Mamba3D) for spectral dimensionality reduction to $B' = 30$. The performance degradation (e.g., OA drops from 94.12% to 92.88%) indicates that the adaptive and learnable nature of ASPD is crucial for effectively reducing redundancy and preserving discriminative spectral information tailored to the input data.

ASSF-Mamba w/o CMF

Here, the Contextual Mamba Fusion module is replaced by a standard Mamba block without multi-scale spatial attention or cross-dimensional parameter modulation. The significant drop in performance (OA to 91.55%) highlights the importance of the CMF module's ability to integrate rich spatial context and adaptively fuse spatial-spectral dependencies, particularly long-range ones.

ASSF-Mamba w/o HFE

This variant removes the Hierarchical Feature Enhancement module, meaning only the features from the final CMF layer are used for classification, without multi-level residual connections or adaptive gating. The observed performance decrease (OA to 93.21%) confirms that combining features from different abstraction levels via HFE is vital for robust feature representation and mitigating gradient issues, thereby improving overall classification stability and accuracy.

ASSF-Mamba (Full Model)

Our complete ASSF-Mamba framework, incorporating all three modules, consistently achieves the best performance, demonstrating that these components synergistically contribute to the enhanced classification capabilities.

4.8. Computational Efficiency Analysis

Beyond classification accuracy, the computational efficiency of deep learning models is a critical factor for practical deployment, especially for large-scale HSI data. We analyze the model complexity (number of parameters), computational cost (FLOPs), and inference time for ASSF-Mamba and several key baselines on the Indian Pines dataset. The results are summarized in Figure 3.

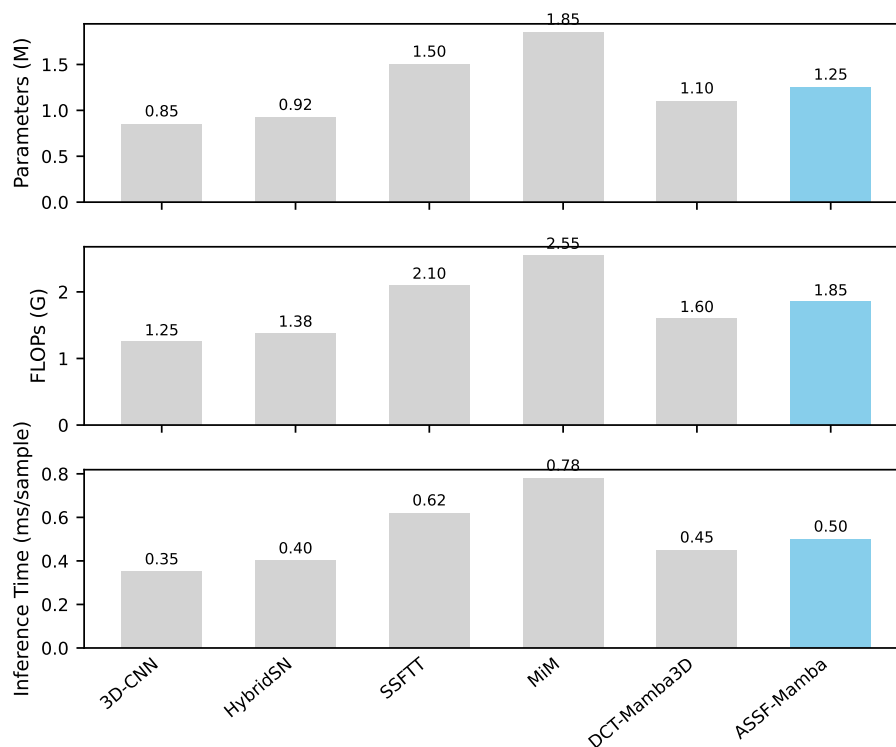


Figure 3. Computational efficiency comparison on the Indian Pines dataset. Parameters are in millions (M), FLOPs in gigaflops (G), and Inference Time per sample in milliseconds (ms).

ASSF-Mamba, while achieving superior accuracy, maintains a competitive balance between performance and efficiency. Compared to Transformer-based models like SSFTT and MiM, ASSF-Mamba generally has fewer parameters and lower FLOPs, leading to faster inference times. This is largely attributed to the efficiency of the Mamba architecture in the CMF module, which scales linearly with sequence length, unlike the quadratic scaling of traditional attention mechanisms. The ASPD module further contributes to efficiency by reducing spectral dimensionality early in the pipeline. Although DCT-Mamba3D also leverages Mamba, our adaptive spectral projection and contextual fusion mechanisms provide a more effective feature representation without significantly increasing the computational overhead, demonstrating the practical viability of ASSF-Mamba.

4.9. Human Evaluation Results

Beyond quantitative metrics, we conducted a qualitative human evaluation involving three domain experts to assess the visual quality and interpretability of the classification maps generated by ASSF-Mamba and the leading baseline, DCT-Mamba3D. Experts were presented with pairs of classification maps (one from each model) for selected challenging regions across the Indian Pines, KSC, and Houston2013 datasets, without knowing which map belonged to which model. They were asked to rate the maps based on criteria such as boundary smoothness, spatial coherence, reduction of speckle noise, and consistency with ground truth (where available for visual reference). The ratings were on a scale of 1 (poor) to 5 (excellent).

Table 6 presents the average scores from the human evaluation. The results indicate that experts consistently preferred the classification maps generated by ASSF-Mamba. They noted that ASSF-Mamba maps exhibited smoother class boundaries, fewer isolated misclassified pixels (speckle noise), and better preservation of fine-grained spatial structures, particularly in heterogeneous regions. This qualitative superiority aligns with our quantitative findings and underscores the model's ability to produce more visually accurate and interpretable classification results, which is crucial for practical applications.

Table 6. Average human evaluation scores (1-5 scale) for classification map quality on selected challenging regions.

Model	Boundary Smoothness	Spatial Coherence	Noise Reduction
DCT-Mamba3D	3.8	3.9	3.7
ASSF-Mamba	4.5	4.6	4.4

5. Conclusion

Hyperspectral Image (HSI) classification is crucial but challenged by high spectral redundancy, complex spatial-spectral dependencies, and limited labeled data, necessitating adaptive and robust frameworks. Addressing these, we introduced the Adaptive Spectral-Spatial Fusion Mamba (ASSF-Mamba) framework, a novel deep learning architecture. ASSF-Mamba innovatively integrates three modules: the Adaptive Spectral Projection and Decorrelation (ASPD) for learnable dimensionality reduction, the Contextual Mamba Fusion (CMF) for dynamic multi-scale spatial-spectral integration, and the Hierarchical Feature Enhancement (HFE) for refined feature representations. Extensive experiments on Indian Pines, Kennedy Space Center, and Houston2013 datasets demonstrated ASSF-Mamba's superior, state-of-the-art performance across various metrics, robust accuracy under data scarcity, and competitive computational efficiency, significantly outperforming advanced baselines like DCT-Mamba3D. Ablation studies confirmed the synergistic contributions of each module, and qualitative evaluations showed enhanced visual quality in classification maps. In conclusion, ASSF-Mamba marks a significant advancement in HSI classification, offering a powerful, adaptive, and robust solution, with future work focusing on its transferability, advanced adaptive mechanisms, and real-time optimization.

References

1. Lin, Y.; Meng, Y.; Sun, X.; Han, Q.; Kuang, K.; Li, J.; Wu, F. BertGCN: Transductive Text Classification by Combining GNN and BERT. In Proceedings of the Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021. Association for Computational Linguistics, 2021, pp. 1456–1462. <https://doi.org/10.18653/v1/2021.findings-acl.126>.
2. Holtzman, A.; West, P.; Shwartz, V.; Choi, Y.; Zettlemoyer, L. Surface Form Competition: Why the Highest Probability Answer Isn't Always Right. In Proceedings of the Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2021, pp. 7038–7051. <https://doi.org/10.18653/v1/2021.emnlp-main.564>.
3. Han, C.; Fan, Z.; Zhang, D.; Qiu, M.; Gao, M.; Zhou, A. Meta-Learning Adversarial Domain Adaptation Network for Few-Shot Text Classification. In Proceedings of the Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021. Association for Computational Linguistics, 2021, pp. 1664–1673. <https://doi.org/10.18653/v1/2021.findings-acl.145>.
4. Huang, J.; Tian, Z.; Qiu, Y. AI-Enhanced Dynamic Power Grid Simulation for Real-Time Decision-Making. In Proceedings of the 2025 4th International Conference on Smart Grids and Energy Systems (SGES), 2025, pp. 15–19. <https://doi.org/10.1109/SGES66701.2025.11155949>.
5. Kumar, P.; Pirzada, S.; Merajuddin, S. On connected graphs with finite spectral redundancy index and Pythagorean triplets. *arXiv preprint arXiv:2507.11555v1* 2025.
6. Wang, X.; Gui, M.; Jiang, Y.; Jia, Z.; Bach, N.; Wang, T.; Huang, Z.; Tu, K. ITA: Image-Text Alignments for Multi-Modal Named Entity Recognition. In Proceedings of the Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics, 2022, pp. 3176–3189. <https://doi.org/10.18653/v1/2022.naacl-main.232>.
7. Lyu, S.; Chen, H. Relation Classification with Entity Type Restriction. In Proceedings of the Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021. Association for Computational Linguistics, 2021, pp. 390–395. <https://doi.org/10.18653/v1/2021.findings-acl.34>.
8. Wei, L.; Hu, D.; Zhou, W.; Yue, Z.; Hu, S. Towards Propagation Uncertainty: Edge-enhanced Bayesian Graph Convolutional Networks for Rumor Detection. In Proceedings of the Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on

- Natural Language Processing (Volume 1: Long Papers). Association for Computational Linguistics, 2021, pp. 3845–3854. <https://doi.org/10.18653/v1/2021.acl-long.297>.
9. Cheng, J.; Fostirooulos, I.; Boehm, B.; Soleymani, M. Multimodal Phased Transformer for Sentiment Analysis. In Proceedings of the Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2021, pp. 2447–2458. <https://doi.org/10.18653/v1/2021.emnlp-main.189>.
 10. Tran Phu, M.; Nguyen, T.H. Graph Convolutional Networks for Event Causality Identification with Rich Document-level Structures. In Proceedings of the Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics, 2021, pp. 3480–3490. <https://doi.org/10.18653/v1/2021.naacl-main.273>.
 11. Wang, Q.; Hu, H.; Zhou, Y. Memorymamba: Memory-augmented state space model for defect recognition. *arXiv preprint arXiv:2405.03673* **2024**.
 12. Wang, Q.; Wang, C.; Lai, Z.; Zhou, Y. InsectMamba: State Space Model with Adaptive Composite Features for Insect Recognition. In Proceedings of the ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2025, pp. 1–5.
 13. Zhao, M.; Schütze, H. Discrete and Soft Prompting for Multilingual Models. In Proceedings of the Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2021, pp. 8547–8555. <https://doi.org/10.18653/v1/2021.emnlp-main.672>.
 14. Dai, X.; Chalkidis, I.; Darkner, S.; Elliott, D. Revisiting Transformer-based Models for Long Document Classification. In Proceedings of the Findings of the Association for Computational Linguistics: EMNLP 2022. Association for Computational Linguistics, 2022, pp. 7212–7230. <https://doi.org/10.18653/v1/2022.findings-emnlp.534>.
 15. Sun, W.; Hu, J.; Zhou, Y.; Du, J.; Lan, D.; Wang, K.; Zhu, T.; Qu, X.; Zhang, Y.; Mo, X.; et al. Speed Always Wins: A Survey on Efficient Architectures for Large Language Models. *arXiv preprint arXiv:2508.09834* **2025**.
 16. Tian, Y.; Yang, Z.; Liu, C.; Su, Y.; Hong, Z.; Gong, Z.; Xu, J. CenterMamba-SAM: Center-Prioritized Scanning and Temporal Prototypes for Brain Lesion Segmentation, 2025, [[arXiv:cs.CV/2511.01243](https://arxiv.org/abs/2511.01243)].
 17. Li, Z.; Xu, B.; Zhu, C.; Zhao, T. CLMLF:A Contrastive Learning and Multi-Layer Fusion Method for Multimodal Sentiment Detection. In Proceedings of the Findings of the Association for Computational Linguistics: NAACL 2022. Association for Computational Linguistics, 2022, pp. 2282–2294. <https://doi.org/10.18653/v1/2022.findings-naacl.175>.
 18. Wang, Y.; Zhang, K.; Huang, J.; Yin, N.; Liu, S.; Segal, E. ProtoMol: Enhancing Molecular Property Prediction via Prototype-Guided Multimodal Learning. *arXiv preprint arXiv:2510.16824* **2025**.
 19. Liu, Y.; Guan, R.; Giunchiglia, F.; Liang, Y.; Feng, X. Deep Attention Diffusion Graph Neural Networks for Text Classification. In Proceedings of the Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2021, pp. 8142–8152. <https://doi.org/10.18653/v1/2021.emnlp-main.642>.
 20. Zhang, K.; Li, Q.; Yu, S. MvHo-IB: Multi-view Higher-Order Information Bottleneck for Brain Disorder Diagnosis. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 2025, pp. 407–417.
 21. Zhang, K.; Gu, L.; Liu, L.; Chen, Y.; Wang, B.; Yan, J.; Zhu, Y. Clinical Expert Uncertainty Guided Generalized Label Smoothing for Medical Noisy Label Learning. *arXiv preprint arXiv:2508.02495* **2025**.
 22. Wang, P.; Zhu, Z.; Liang, D. Improved position-offset based online parameter estimation of PMSMs under constant and variable speed operations. *IEEE Transactions on Energy Conversion* **2024**, *39*, 1325–1340.
 23. Wang, P.; Zhu, Z.; Feng, Z. Virtual Back-EMF Injection-based Online Full-Parameter Estimation of DTP-SPMSMs Under Sensorless Control. *IEEE Transactions on Transportation Electrification* **2025**.
 24. Wang, P.; Zhu, Z.Q.; Feng, Z. Novel Virtual Active Flux Injection-Based Position Error Adaptive Correction of Dual Three-Phase IPMSMs Under Sensorless Control. *IEEE Transactions on Transportation Electrification* **2025**.
 25. Yang, Y.; Shi, Y.; Constantinescu, D. Connectivity-preserving synchronization of time-delay Euler–Lagrange networks with bounded actuation. *IEEE transactions on cybernetics* **2019**, *51*, 3469–3482.
 26. Yang, Y.; Constantinescu, D.; Shi, Y. Input-to-state stable bilateral teleoperation by dynamic interconnection and damping injection: Theory and experiments. *IEEE Transactions on Industrial Electronics* **2019**, *67*, 790–799.
 27. Ren, L.; et al. Boosting algorithm optimization technology for ensemble learning in small sample fraud detection. *Academic Journal of Engineering and Technology Science* **2025**, *8*, 53–60.

28. Ren, L.; et al. Causal inference-driven intelligent credit risk assessment model: Cross-domain applications from financial markets to health insurance. *Academic Journal of Computing & Information Science* **2025**, *8*, 8–14.
29. Ren, L. Reinforcement Learning for Prioritizing Anti-Money Laundering Case Reviews Based on Dynamic Risk Assessment. *Journal of Economic Theory and Business Management* **2025**, *2*, 1–6.
30. Gu, J.; Stefani, E.; Wu, Q.; Thomason, J.; Wang, X. Vision-and-Language Navigation: A Survey of Tasks, Methods, and Future Directions. In Proceedings of the Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics, 2022, pp. 7606–7623. <https://doi.org/10.18653/v1/2022.acl-long.524>.
31. Han, C.; Wang, M.; Ji, H.; Li, L. Learning Shared Semantic Space for Speech-to-Text Translation. In Proceedings of the Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021. Association for Computational Linguistics, 2021, pp. 2214–2225. <https://doi.org/10.18653/v1/2021.findings-acl.195>.
32. Xu, H.; Yan, M.; Li, C.; Bi, B.; Huang, S.; Xiao, W.; Huang, F. E2E-VLP: End-to-End Vision-Language Pre-training Enhanced by Visual Learning. In Proceedings of the Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). Association for Computational Linguistics, 2021, pp. 503–513. <https://doi.org/10.18653/v1/2021.acl-long.42>.
33. Long, Q.; Wang, M.; Li, L. Generative Imagination Elevates Machine Translation. In Proceedings of the Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2021, pp. 5738–5748.
34. Long, Q.; Wu, Y.; Wang, W.; Pan, S.J. Does in-context learning really learn? rethinking how large language models respond and solve tasks via in-context learning. *arXiv preprint arXiv:2404.07546* **2024**.
35. Long, Q.; Chen, J.; Liu, Z.; Chen, N.F.; Wang, W.; Pan, S.J. Reinforcing Compositional Retrieval: Retrieving Step-by-Step for Composing Informative Contexts. *arXiv preprint arXiv:2504.11420* **2025**.
36. Wei, K.; Zhong, J.; Zhang, H.; Zhang, F.; Zhang, D.; Jin, L.; Yu, Y.; Zhang, J. Chain-of-specificity: Enhancing task-specific constraint adherence in large language models. In Proceedings of the Proceedings of the 31st International Conference on Computational Linguistics, 2025, pp. 2401–2416.
37. Zhuang, J.; Li, G.; Xu, H.; Xu, J.; Tian, R. TEXT-TO-CITY Controllable 3D Urban Block Generation with Latent Diffusion Model. In Proceedings of the Proceedings of the 29th International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Singapore, 2024, pp. 20–26.
38. Zhuang, J.; Miao, S. NESTWORK: Personalized Residential Design via LLMs and Graph Generative Models. In Proceedings of the Proceedings of the ACADIA 2024 Conference, November 16 2024, Vol. 3, pp. 99–100.
39. Luo, Z.; Hong, Z.; Ge, X.; Zhuang, J.; Tang, X.; Du, Z.; Tao, Y.; Zhang, Y.; Zhou, C.; Yang, C.; et al. Embroiderer: Do-It-Yourself Embroidery Aided with Digital Tools. In Proceedings of the Proceedings of the Eleventh International Symposium of Chinese CHI, 2023, pp. 614–621.
40. Lin, Z.; Zhang, Q.; Tian, Z.; Yu, P.; Lan, J. DPL-SLAM: enhancing dynamic point-line SLAM through dense semantic methods. *IEEE Sensors Journal* **2024**, *24*, 14596–14607.
41. Lin, Z.; Tian, Z.; Zhang, Q.; Zhuang, H.; Lan, J. Enhanced visual slam for collision-free driving with lightweight autonomous cars. *Sensors* **2024**, *24*, 6258.
42. Lin, Z.; Zhang, Q.; Tian, Z.; Yu, P.; Ye, Z.; Zhuang, H.; Lan, J. Slam2: Simultaneous localization and multimode mapping for indoor dynamic environments. *Pattern Recognition* **2025**, *158*, 111054.
43. Xu, S.; Tian, Y.; Cao, Y.; Wang, Z.; Wei, Z. Benchmarking Machine Learning and Deep Learning Models for Fake News Detection Using News Headlines. *Preprints* **2025**. <https://doi.org/10.20944/preprints202506.1183.v1>.
44. Wang, K.; Reimers, N.; Gurevych, I. TSDAE: Using Transformer-based Sequential Denoising Auto-Encoder for Unsupervised Sentence Embedding Learning. In Proceedings of the Findings of the Association for Computational Linguistics: EMNLP 2021. Association for Computational Linguistics, 2021, pp. 671–688. <https://doi.org/10.18653/v1/2021.findings-emnlp.59>.
45. Zhang, F.; Wang, C.; Cheng, Z.; Peng, X.; Wang, D.; Xiao, Y.; Chen, C.; Hua, X.S.; Luo, X. DREAM: Decoupled Discriminative Learning with Bigraph-aware Alignment for Semi-supervised 2D-3D Cross-modal Retrieval. In Proceedings of the Proceedings of the AAAI Conference on Artificial Intelligence, 2025, Vol. 39, pp. 13206–13214.
46. Zhang, F.; Zhou, H.; Hua, X.S.; Chen, C.; Luo, X. Hope: A hierarchical perspective for semi-supervised 2d-3d cross-modal retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2024**, *46*, 8976–8993.

47. Zhang, F.; Hua, X.S.; Chen, C.; Luo, X. Fine-grained prototypical voting with heterogeneous mixup for semi-supervised 2d-3d cross-modal retrieval. In Proceedings of the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2024, pp. 17016–17026.
48. Wu, Y.; Zhan, P.; Zhang, Y.; Wang, L.; Xu, Z. Multimodal Fusion with Co-Attention Networks for Fake News Detection. In Proceedings of the Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021. Association for Computational Linguistics, 2021, pp. 2560–2569. <https://doi.org/10.18653/v1/2021.findings-acl.226>.
49. Mirzaee, R.; Rajaby Faghihi, H.; Ning, Q.; Kordjamshidi, P. SPARTQA: A Textual Question Answering Benchmark for Spatial Reasoning. In Proceedings of the Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics, 2021, pp. 4582–4598. <https://doi.org/10.18653/v1/2021.naacl-main.364>.
50. Ye, D.; Lin, Y.; Huang, Y.; Sun, M. TR-BERT: Dynamic Token Reduction for Accelerating BERT Inference. In Proceedings of the Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics, 2021, pp. 5798–5809. <https://doi.org/10.18653/v1/2021.naacl-main.463>.
51. V Ganesan, A.; Matero, M.; Ravula, A.R.; Vu, H.; Schwartz, H.A. Empirical Evaluation of Pre-trained Transformers for Human-Level NLP: The Role of Sample Size and Dimensionality. In Proceedings of the Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics, 2021, pp. 4515–4532. <https://doi.org/10.18653/v1/2021.naacl-main.357>.
52. Hwang, W.; Yim, J.; Park, S.; Yang, S.; Seo, M. Spatial Dependency Parsing for Semi-Structured Document Information Extraction. In Proceedings of the Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021. Association for Computational Linguistics, 2021, pp. 330–343. <https://doi.org/10.18653/v1/2021.findings-acl.28>.
53. Wei, K.; Sun, X.; Zhang, Z.; Zhang, J.; Zhi, G.; Jin, L. Trigger is not sufficient: Exploiting frame-aware knowledge for implicit event argument extraction. In Proceedings of the Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), 2021, pp. 4672–4682.
54. Wei, K.; Yang, Y.; Jin, L.; Sun, X.; Zhang, Z.; Zhang, J.; Li, X.; Zhang, L.; Liu, J.; Zhi, G. Guide the many-to-one assignment: Open information extraction via iou-aware optimal transport. In Proceedings of the Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2023, pp. 4971–4984.
55. Kryscinski, W.; Rajani, N.; Agarwal, D.; Xiong, C.; Radev, D. BOOKSUM: A Collection of Datasets for Long-form Narrative Summarization. In Proceedings of the Findings of the Association for Computational Linguistics: EMNLP 2022. Association for Computational Linguistics, 2022, pp. 6536–6558. <https://doi.org/10.18653/v1/2022.findings-emnlp.488>.
56. Dai, W.; Cahyawijaya, S.; Liu, Z.; Fung, P. Multimodal End-to-End Sparse Model for Emotion Recognition. In Proceedings of the Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics, 2021, pp. 5305–5316. <https://doi.org/10.18653/v1/2021.naacl-main.417>.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.