EnergyMamba: Energy Attention Modulated Mamba for Hyperspectral Image Classification

Faiq Ahmad, Saad Sohail, Muhammad Usama, Usman Ghous
Department of Computer Science, National University of Computer and Emerging Sciences (NUCES), Pakistan
{faiqahmad430;saadsohail5104}@gmail.com; {m.usama;usman.ghous}@nu.edu.pk

Manuel Mazzara
Institute of Software Development and Engineering
Innopolis University
Innopolis, 420500, Russia
m.mazzara@innopolis.ru

Danish Shehzad
Computer Science Department
New Uzbekistan University
Tashkent, 100007 Uzbekistan
d.shehzad@newuu.uz

Muhammad Ahmad
SDAIA-KFUPM, Joint Research Center for Artificial Intelligence (JRCAI)
King Fahd University of Petroleum and Minerals
Dhahran, 31261, Saudi Arabia.
mahmad00@gmail.com

***Abstract*—Hyperspectral image classification is challenged by extremely high spectral dimensionality, severe inter-band redundancy, and the need to model long-range spectral–spatial dependencies under limited labeled data. Although Transformer-based models provide powerful global context modeling, their quadratic complexity restricts scalability, while recent Mamba-based state-space models, despite their linear complexity, lack explicit mechanisms for adaptive feature discrimination. This paper proposes EnergyMamba, a hybrid architecture that integrates selective state-space modeling with an energy-based attention mechanism to address these limitations in a principled manner. The Mamba branch efficiently captures long-range spectral–spatial dependencies with linear computational complexity, whereas the Energy Attention branch incorporates Hopfield-inspired local energy functions and multi-head energy attention to dynamically reweight features, suppress spectral redundancy, and enhance class separability. By coupling global state-space dynamics with energy-guided associative feature refinement, EnergyMamba provides a discriminative and stable inductive bias for hyperspectral representation learning. Extensive experiments on the Salinas, WHU-Hi-HanChuan, and OHID-1 datasets demonstrate that EnergyMamba achieves state-of-the-art performance, attaining overall accuracies of 99.84%, 99.36%, and 94.59%, respectively, using only 5% of labeled samples, with particularly strong gains on the challenging OHID-1 dataset.**

***Index Terms*—Hyperspectral Image Classification, Mamba, Energy Attention, Multi-Head Energy Attention, Hopfield Network, Spectral-Spatial Feature Extraction.**

I. INTRODUCTION

Hyperspectral imaging (HSI)[1] is an advanced remote sensing modality that acquires continuous spectral signatures across hundreds of narrow and contiguous bands spanning the visible to infrared spectrum, providing fine-grained spectral–spatial information for each pixel [1]. HSI classification aims to assign a semantic land-cover or material label to every pixel, serving as a fundamental task for applications such as environmental monitoring, urban planning, and precision agriculture [2], [3]. Despite its rich information content, effective HSI classification remains challenging due to the inherent high dimensionality of spectral data, strong inter-band redundancy, and pronounced spatial–spectral heterogeneity, which collectively complicate discriminative feature extraction and robust model learning [4], [5]. These challenges are further exacerbated in practical scenarios where annotated samples are scarce, making the design of efficient and adaptive modeling strategies critical for reliable HSI analysis.

Recent advances in deep learning have substantially improved HSI classification performance, with convolutional neural networks (CNNs) and vision Transformers (ViTs) emerging as dominant paradigms. CNN-based methods exploit localized receptive fields and hierarchical feature extraction to capture spatial–spectral patterns effectively, benefiting from translation invariance and parameter sharing [6], [7]. Representative architectures such as GITANet [6] enhance spectral discrimination through group interactive attention, while CNN-GCN fusion frameworks [7] incorporate graph-based reasoning to model spatial relationships. However, the fixed receptive fields of CNNs inherently limit their ability to capture long-range dependencies, which are essential for modeling global spectral correlations and extended spatial structures in hyperspectral scenes [8], [9]. ViTs address this limitation by employing self-attention mechanisms that explicitly model global spectral–spatial interactions, as exemplified by CAT [10] with its center-focused attention and SClusterFormer [11] using cluster-based attention with deformable convolutions. Despite their strong representational capacity, Transformer-based models incur quadratic computational and memory complexity with respect to sequence length [12], rendering them inefficient for high-resolution HSIs and vulnerable to overfitting and redundancy amplification under limited training data [13].

In recent years, Mamba architectures based on selective state space models have emerged as an efficient alternative for long-sequence modeling [14], offering linear computational complexity while retaining the ability to capture long-range

[1]HSI denotes both hyperspectral imaging and hyperspectral image, where appropriate.

dependencies [15]. In the hyperspectral domain, MambaHSI exploits separate spatial and spectral Mamba blocks to model pixel-level interactions and spectral group dependencies [12], while DBMGNet integrates bidirectional Mamba with graph convolutional networks to jointly encode sequential and non-Euclidean spatial structures [16]. Subsequent extensions incorporate additional inductive biases, including morphological priors in MorpMamba [17], frequency-aware hierarchical processing in FAHM [18], and wavelet-based multiscale feature fusion in WaveMamba [19]. These works collectively demonstrate the promise of Mamba for scalable hyperspectral modeling. Nevertheless, existing Mamba-based approaches often rely on fixed architectural designs and uniform spectral processing, which can limit adaptability across datasets with varying spectral complexity and spatial resolution. Moreover, the lack of explicit mechanisms for dynamically prioritizing semantically informative features can weaken discriminative capability, particularly in scenes with subtle inter-class differences. Many models also struggle to balance local detail preservation with global contextual modeling, leading to over-smoothed spectral transitions and blurred class boundaries [20].

To overcome these limitations, recent studies have explored hybrid strategies that combine Mamba with complementary architectures. For example, HSI-MFormer integrates Mamba with Transformer experts to leverage both linear-complexity sequence modeling and attention-based global reasoning [15], while DualMamba adopts a lightweight convolution–Mamba fusion to enhance spatial feature extraction [21]. However, such hybrid designs often increase architectural complexity and computational overhead, undermining Mamba's efficiency advantages [12], and may still fall short in capturing fine-grained spectral-channel interactions or mitigating high-dimensional feature redundancy [22]. Additionally, harmonizing state space dynamics with spatially constrained attention remains challenging, frequently introducing trade-offs between cross-spectral dependency modeling and spatial resolution preservation. These issues are further amplified under spectral variability caused by atmospheric and illumination changes [23], highlighting the need for a unified and adaptive framework that enhances discriminative power without sacrificing scalability.

To address these limitations, this work proposes EnergyMamba, a hybrid architecture that unifies selective state space modeling with energy-based associative feature learning for HSI classification. EnergyMamba is composed of two tightly coupled branches with complementary inductive biases. The Mamba branch models long-range spectral and spatial dependencies via input-modulated state transitions, enabling stable global context propagation with linear time and memory complexity, which is critical for high-dimensional hyperspectral sequences. In parallel, the Energy Attention branch formulates feature interaction as an energy minimization process inspired by modern Hopfield networks, where semantically consistent spectral bands and spatial tokens correspond to low-energy states. Through multi-head energy attention, this branch computes local energy landscapes that drive dynamic gating vectors, adaptively reweighting features according to their discriminative relevance. Unlike conventional attention mechanisms that rely solely on similarity scores, the proposed energy-guided gating explicitly suppresses redundant spectral responses and stabilizes feature selection under limited supervision. By integrating energy-based associative refinement with state space dynamics, EnergyMamba achieves a principled balance between global dependency modeling and local discriminative enhancement, resulting in improved class separability, sharper decision boundaries, and increased robustness to spectral variability.

II. Proposed Methodology

The input HSI is represented as a data cube $\mathbf{X} \in \mathbb{R}^{H\times W\times B}$, where H and W denote the spatial dimensions and B the number of spectral bands. To reduce spectral redundancy and computational burden while preserving the most informative spectral components, Principal Component Analysis (PCA) is applied along the spectral dimension, yielding a transformed cube $\mathbf{X} \in \mathbb{R}^{H\times W\times C}$ with $C \ll B$. From this reduced representation, spatial–spectral patches of size $P \times P \times C$ are extracted, each centered at spatial location (i, j). For each patch $\mathbf{P}_{i,j}$, the corresponding label $y_{i,j}$ is assigned according to the ground-truth class of its central pixel, ensuring spatial consistency in supervision.

Each patch $\mathbf{P}_{i,j} \in \mathbb{R}^{P\times P\times C}$ is reshaped into a sequence of hyperspectral tokens $\mathbf{X}_{\text{patch}} \in \mathbb{R}^{P^2\times C}$, where each token corresponds to a spatial pixel with an associated spectral feature vector. This sequential representation enables the joint modeling of spatial locality and spectral dependency, and serves as the common input to both the Mamba branch and the Energy branch for subsequent feature extraction.

A. Mamba Branch

The token sequence $\mathbf{X}_{\text{patch}}$ is first projected into an inner latent space via a linear transformation, producing $\mathbf{U} \in \mathbb{R}^{D_{\text{inner}}\times P^2}$ according to:

$$\mathbf{U} = \mathbf{W}_{\text{in}}\mathbf{X}_{\text{patch}}^{\top} + \mathbf{b}_{\text{in}} \tag{1}$$

where $\mathbf{W}_{\text{in}}$ and $\mathbf{b}_{\text{in}}$ are learnable parameters. The Mamba module then models the resulting sequence using selective state space dynamics, which enable efficient long-range dependency modeling with linear complexity. At each discrete time step t, corresponding to the t-th spatial token, the internal state evolution and output generation are governed by:

$$\mathbf{x}_t = \mathbf{A}\mathbf{h}_{t-1} + \mathbf{B}_t\mathbf{u}_t; \quad \mathbf{h}_t = \mathbf{x}_t \tag{2}$$

$$\mathbf{y}_t = \mathbf{C}_t\mathbf{h}_t + \mathbf{D}\mathbf{u}_t \tag{3}$$

where $\mathbf{h}_t$ denotes the latent state, $\mathbf{u}_t$ the input at step t, and $\mathbf{A}$, $\mathbf{B}_t$, $\mathbf{C}_t$, and $\mathbf{D}$ are learnable parameters representing the state transition, input projection, output projection, and skip connection, respectively. The input-dependent matrices

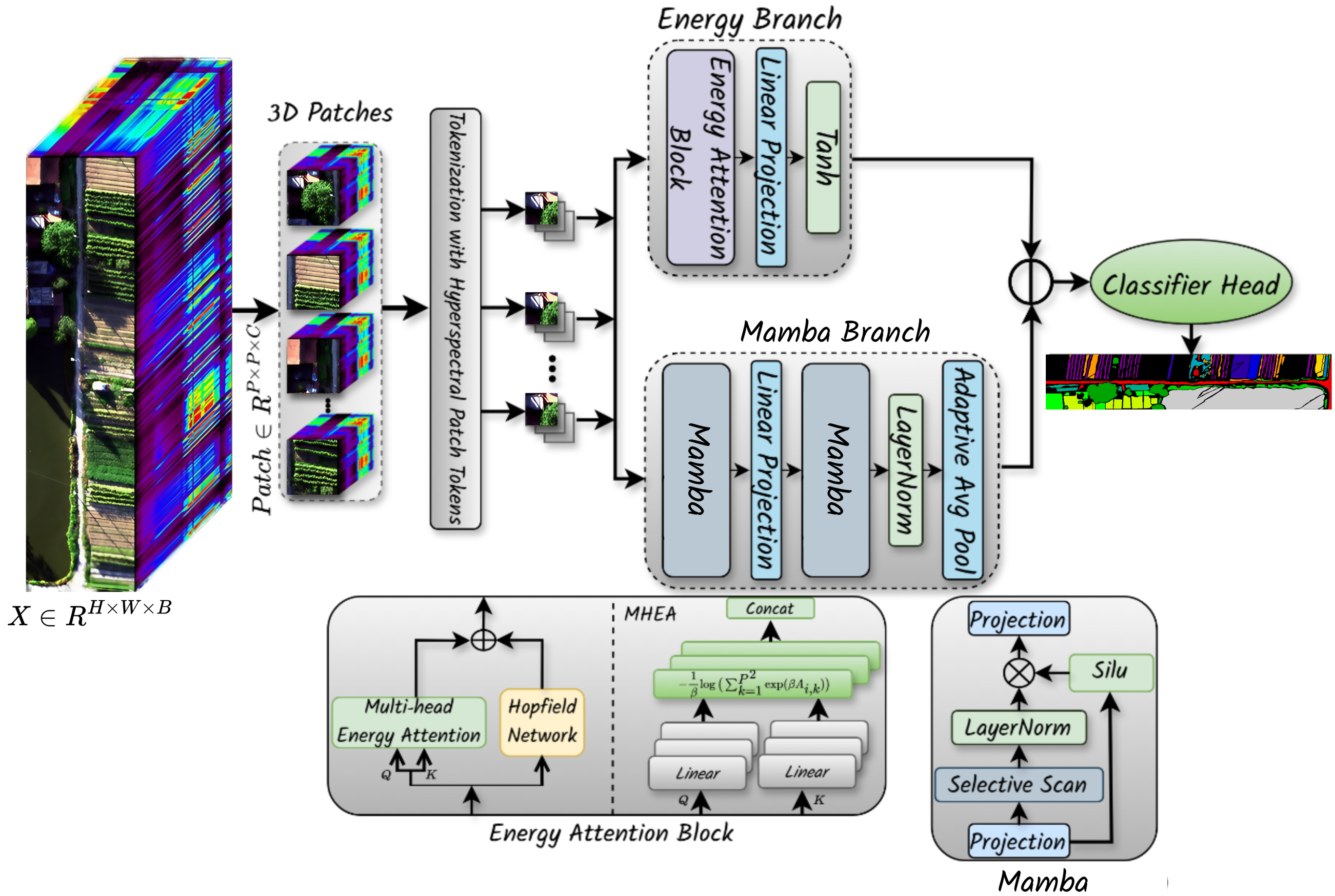

Fig. 1: Overview of the proposed EnergyMamba architecture for HSI classification. The input hyperspectral cube $\mathbf{X} \in \mathbb{R}^{H\times W\times B}$ is partitioned into 3D spectral-spatial patches and tokenized into hyperspectral patch tokens. These tokens are processed in parallel by two complementary branches. The Mamba branch employs selective state space modeling with input-dependent state transitions to capture long-range spectral and spatial dependencies with linear computational complexity. The Energy branch performs energy-based associative feature refinement using Hopfield-inspired local energy functions and multi-head energy attention to generate dynamic gating signals that emphasize discriminative spectral bands and spatial tokens. The outputs of both branches are fused to produce energy-guided global representations, which are subsequently fed into a classifier head for final pixel-wise classification.

$\mathbf{B}_t$ and $\mathbf{C}_t$ are modulated by a learnable time-step parameter Δ_t, enabling adaptive state transitions conditioned on the local spectral–spatial content. These dynamics are efficiently computed using the Selective Scan operation, allowing parallelized sequence processing without sacrificing temporal expressiveness.

The resulting output sequence $\mathbf{Y} \in \mathbb{R}^{D_{\text{inner}}\times P^2}$ is normalized via layer normalization (LN)) and gated through a Hadamard product ($\odot$) with a SiLU-activated projection of the input sequence,

$$\mathbf{Y}_{\text{out}} = \text{LN}(\mathbf{Y}) \odot \text{SiLU}(\mathbf{W}_{\text{gate}}\mathbf{U}) \tag{4}$$

which enhances informative responses while suppressing irrelevant activations. The gated features are further projected, normalized, and aggregated through adaptive average pooling to obtain a fixed-dimensional representation $\mathbf{f}_{\text{Mamba}} \in \mathbb{R}^{D}$, summarizing global spectral–spatial dependencies captured by the state space model.

B. Energy Branch

The Energy branch is designed to perform energy-based associative feature refinement, producing a gating vector that adaptively modulates the representations learned by the Mamba branch. It consists of a Hopfield module and a multi-head Energy Attention module, which jointly model local consistency and global interactions from an energy minimization perspective.

*a) **Hopfield Module**:* The input sequence $\mathbf{X}_{\text{patch}}$ is linearly projected to obtain:

$$\mathbf{G} = \mathbf{W}_{\text{hn}}\mathbf{X}_{\text{patch}}^{\top} + \mathbf{b}_{\text{hn}} \tag{5}$$

where $\mathbf{G} \in \mathbb{R}^{D_{\text{hn}} \times P^2}$. Inspired by modern Hopfield networks, a local energy value is computed for each spatial token as

$$e_{\text{hn},i} = -\frac{1}{2} \sum_{d=1}^{D_{\text{hn}}} \big(\text{ReLU}(G_{d,i})\big)^2, \quad i = 1, \ldots, P^2 \tag{6}$$

yielding an energy vector $\mathbf{e}_{\text{hn}} \in \mathbb{R}^{P^2}$. Lower energy values correspond to more stable and semantically consistent feature patterns, providing a measure of local discriminative confidence.

*b) **Energy Attention Module:*** To capture global interactions, a multi-head Energy Attention mechanism is applied. For each head, queries and keys are computed as:

$$\mathbf{Q} = \mathbf{W}_Q \mathbf{X}_{\text{patch}}^{\top}; \quad \mathbf{K} = \mathbf{W}_K \mathbf{X}_{\text{patch}}^{\top} \tag{7}$$

with $\mathbf{Q}, \mathbf{K} \in \mathbb{R}^{D_{\text{qk}} \times P^2}$. The affinity matrix is given by:

$$\mathbf{A} = \frac{\mathbf{Q}^{\top}\mathbf{K}}{\sqrt{D_{\text{qk}}}} \tag{8}$$

and an energy value for each token is derived via a Log-Sum-Exp formulation:

$$e_{\text{attn},i} = -\frac{1}{\beta} \log \left(\sum_{k=1}^{P^2} \exp(\beta A_{i,k}) \right) \tag{9}$$

where β is a learnable temperature parameter controlling energy sharpness. Energies from all heads are concatenated and combined with the Hopfield energies according to:

$$\mathbf{e} = \mathbf{e}_{\text{attn}} + \alpha \mathbf{e}_{\text{hn}} \tag{10}$$

where α is a learnable scalar. The resulting energy vector is projected and passed through a hyperbolic tangent activation to produce the final gating vector:

$$\mathbf{g}_{\text{EA}} = \tanh\left(\mathbf{W}_{\text{proj}}\mathbf{e} + \mathbf{b}_{\text{proj}}\right) \tag{11}$$

which adaptively encodes the relative importance of spectral–spatial features.

C. Feature Fusion and Classification

The global representation from the Mamba branch and the energy-guided gating vector from the Energy branch are fused via concatenation:

$$\mathbf{f}_{\text{fused}} = \text{concat}\left[\mathbf{f}_{\text{Mamba}}, \mathbf{g}_{\text{EA}}\right] \tag{12}$$

yielding a joint feature vector $\mathbf{f}_{\text{fused}} \in \mathbb{R}^{2D}$. This fused representation integrates long-range dependency modeling with adaptive energy-based feature selection and is mapped to class logits through a linear classifier:

$$\hat{\mathbf{y}} = \mathbf{W}_{\text{cls}}\mathbf{f}_{\text{fused}} + \mathbf{b}_{\text{cls}} \tag{13}$$

where $\hat{\mathbf{y}} \in \mathbb{R}^{K}$, $\mathbf{W}_{\text{cls}} \in \mathbb{R}^{K \times (2 \cdot D)}$, and $\mathbf{b}_{\text{cls}} \in \mathbb{R}^{K}$ are the classifier's weights and biases which enable end-to-end training for HSI classification.

III. Experimental Results and Discussion

This section presents a comprehensive evaluation of the proposed EnergyMamba framework for HSI classification. Its performance is systematically compared against representative state-of-the-art Transformer- and Mamba-based methods to assess both accuracy and robustness under limited supervision. Quantitative evaluation is conducted using overall accuracy (OA), average accuracy (AA), per-class accuracy, and Cohen's κ coefficient, providing complementary perspectives on classification performance, class balance, and statistical agreement. In addition, qualitative analyses are presented to visually examine classification maps and error distributions, offering further insight into the discriminative capability and spatial consistency of the proposed approach.

EnergyMamba was trained using the Adam optimizer with a batch size of 64 and a learning rate of 5×10^{-4}. Only 5% of the data was used for training and validation each, while the remaining 90% was reserved for testing to ensure fair comparison with state-of-the-art methods and to simulate data-scarce HSI scenarios. Spatial patches of size 12×12 were extracted around each labeled pixel, and PCA was applied to retain 15 principal components prior to training. Training was performed for up to 50 epochs using cross-entropy loss with early stopping (patience 10) and a ReduceLROnPlateau scheduler (factor 0.5, patience 5). The SSM branch employed a hidden dimension of 64, state dimension 16, and expansion ratio 2, while the Energy Attention branch used 8 heads with query-key dimension 64 and a Hopfield multiplier of 4.0.

A. Ablation Study

To isolate the contribution of the proposed Energy Attention mechanism, we conduct an ablation study by integrating different attention strategies into the same Mamba backbone. Table I reports the quantitative results on three benchmark datasets. The baseline Mamba model already achieves competitive performance on relatively homogeneous datasets such as Salinas, but its effectiveness degrades on more complex scenes, particularly OHID-1, where severe class imbalance and spectral similarity lead to a marked drop in average accuracy. Incorporating conventional attention mechanisms consistently improves performance, confirming that explicit feature interaction benefits state space modeling. Among these variants, multi-head attention achieves the strongest baseline improvement, indicating the importance of multi-view feature aggregation.

Despite these gains, all conventional attention mechanisms remain limited by similarity-based weighting, which inadequately suppresses redundant spectral responses in high-dimensional hyperspectral data. In contrast, the proposed Energy Attention consistently yields the highest performance across all datasets and metrics. On OHID-1, EnergyMamba improves average accuracy by 20.33% over the baseline Mamba and surpasses sparse attention by 1.99% in overall accuracy, highlighting its superior ability to handle complex class distributions. The substantial improvement in the κ coefficient, particularly on WHU-Hi-HanChuan (HC) and

TABLE I: Comparative analysis of EnergyMamba variants incorporating different attention mechanisms.

Model	WHU-Hi-HanChuan			Salinas			OHID-1		
	OA (%)	AA (%)	κ (%)	OA (%)	AA (%)	κ (%)	OA (%)	AA (%)	κ (%)
Mamba	97.49	94.48	97.06	99.59	99.69	99.54	91.43	67.47	80.06
Mamba + Simple Attention	98.12	96.51	97.80	99.46	99.67	99.40	92.61	80.15	82.73
Mamba + Multihead Attention	98.53	96.95	98.28	99.69	99.69	99.65	93.84	83.57	85.60
Mamba + Cross Attention	97.83	96.01	97.46	99.60	99.66	99.55	93.06	78.91	83.80
Mamba + Sparse Attention	97.96	96.10	97.61	99.58	99.64	99.53	93.08	83.22	84.05
Mamba + Energy Attention	**99.36**	**98.61**	**99.26**	**99.84**	**99.79**	**99.83**	**95.07**	**87.80**	**88.63**

OHID-1, further indicates that energy-guided gating enhances classification reliability and inter-class separability. These results demonstrate that modeling feature interactions through energy minimization provides a more effective inductive bias than similarity-driven attention when coupled with state space dynamics.

*a) **Effects of Patch Size and Training Samples**:* The influence of spatial–spectral patch size on EnergyMamba is evaluated in Table II. As the patch size increases, performance initially improves across all datasets, reflecting enhanced contextual awareness and more effective modeling of local spectral–spatial structures. Peak performance is achieved at intermediate patch sizes, most notably at 12×12 for OHID-1, beyond which accuracy gradually declines. This degradation can be attributed to the introduction of redundant or irrelevant spatial information, which dilutes discriminative features and increases spectral mixing, particularly in heterogeneous scenes. These observations indicate that EnergyMamba benefits from a balanced receptive field that preserves local detail while providing sufficient contextual support.

TABLE II: EnergyMamba accuracy across patch sizes, showing gains up to an optimal size before declining due to redundancy.

Patch	SA			HC			OHID-1		
	κ(%)	OA(%)	AA(%)	κ(%)	OA(%)	AA(%)	κ(%)	OA(%)	AA(%)
8×8	99.66	99.69	99.77	98.01	98.30	96.77	87.59	94.59	85.03
10×10	99.65	99.68	99.65	98.83	99.00	98.19	87.90	94.71	87.21
12×12	99.83	99.84	99.79	99.26	99.36	98.61	**88.63**	**95.07**	**87.80**
14×14	99.81	99.83	99.74	99.39	99.48	98.72	86.92	94.19	87.04
16×16	99.89	99.91	99.84	**99.42**	**99.50**	**98.76**	85.55	93.71	83.88
18×18	99.90	99.91	99.84	99.36	99.45	98.54	83.94	93.14	78.66
20×20	**99.91**	**99.92**	**99.88**	99.36	99.45	98.61	84.44	93.29	79.45

Table III examines the effect of varying the proportion of labeled training samples. EnergyMamba demonstrates strong robustness under extremely limited supervision, achieving high accuracy even with only 1% of labeled data. Performance improves steadily as the training ratio increases, with diminishing returns observed beyond 7%, suggesting early saturation of the model's representational capacity. The consistently strong performance in low-label regimes highlights the effectiveness of energy-guided feature selection in stabilizing learning and mitigating overfitting, making EnergyMamba particularly suitable for practical hyperspectral applications where annotated data are scarce.

B. Comparison with Transformer and Mamba Models

To benchmark the effectiveness of the proposed model, we compare EnergyMamba (EM) with representative Transformer- and Mamba-based methods, including Pyramid

TABLE III: EnergyMamba accuracy vs. training data %, with improvements plateauing due to saturation.

Train %	SA			HC			OHID-1		
	κ(%)	OA(%)	AA(%)	κ(%)	OA(%)	AA(%)	κ(%)	OA(%)	AA(%)
1%	97.68	97.92	98.05	95.69	96.32	92.82	80.05	91.27	70.56
3%	99.51	99.56	99.64	98.51	98.72	97.58	85.06	93.57	79.88
5%	99.68	99.71	99.77	99.26	99.37	98.78	87.65	94.69	84.86
7%	99.86	99.87	99.87	99.45	99.53	99.18	89.20	95.31	85.85
9%	99.89	99.90	99.88	99.60	99.65	99.34	90.96	96.07	89.23
10%	99.89	99.90	99.92	99.70	99.74	99.44	90.35	95.81	89.63

Spatial-Spectral Transformer (PF) [24], Differential Spatial-Spectral Transformer (DF) [25], Morphological Spatial-Spectral Transformer (MF) [26], Spatial-Spectral Transformer (SF) [27], Multihead Spatial-Spectral Mamba (MHM) [22], Morphological Spatial-Spectral Mamba (MM) [17], and Wavelet Spatial-Spectral Mamba (WM) [19].

Tables IV–VI and Figures 2–4 provide a comprehensive comparison between the proposed EnergyMamba and representative Transformer- and Mamba-based models across three benchmark hyperspectral datasets with varying degrees of spectral complexity and spatial heterogeneity. Overall, EnergyMamba consistently achieves superior performance in terms of overall accuracy (OA), average accuracy (AA), and κ coefficient, indicating improved class separability, robustness, and statistical reliability.

Across all benchmarks, EnergyMamba consistently achieves superior OA, AA, and κ coefficient, indicating improved class separability, robustness, and reliability. Compared with Transformer-based approaches such as PF, DF, MF, and SF, EnergyMamba demonstrates clear advantages in both quantitative metrics and spatial consistency. Although Transformers can model global dependencies via self-attention, their similarity-driven weighting can amplify spectral redundancy and noise in high-dimensional hyperspectral inputs, particularly under limited supervision, which commonly leads to overconfident confusion among spectrally similar classes and degraded performance on minority categories. This behavior is reflected in the lower AA and κ values of competing Transformer models on HC and OHID-1, where strong spectral overlap and intra-class variability make purely similarity-based attention less effective. In contrast, EnergyMamba performs energy-guided feature selection that explicitly suppresses redundant responses and promotes semantically stable representations, yielding higher per-class consistency and markedly improved κ scores.

When compared with recent Mamba-based methods such as MHM, MM, and WM, EnergyMamba further improves performance, particularly on challenging scenes. Existing Mamba

TABLE IV: Performance comparison on the **Salinas dataset** across class-wise accuracies and aggregate metrics.

Class	PF	DF	MF	SF	MHM	MM	WM	***EM***
Broccoli 1	100.00	100.00	98.51	100.00	99.00	99.66	99.12	100.00
Broccoli 2	99.02	91.05	100.00	100.00	99.79	99.97	99.52	100.00
Fallow	97.64	99.88	98.71	99.09	99.61	99.38	98.26	100.00
Fallow Rough	82.55	99.12	99.28	99.55	96.41	96.89	94.82	97.69
Fallow Smooth	100.00	98.83	95.85	99.41	97.14	96.97	98.38	99.71
Stubble	99.83	100.00	100.00	100.00	99.94	99.97	98.65	100.00
Celery	99.16	100.00	99.47	99.97	99.60	99.68	98.20	100.00
Grapes	95.69	94.51	91.49	94.11	94.10	98.43	97.36	99.79
Soil Vinyard	99.39	99.92	99.73	99.78	99.46	99.87	98.66	100.00
Corn Senesced	98.71	98.88	97.83	97.43	98.92	98.77	99.19	99.63
Lettuce 4wk	92.10	99.89	92.00	96.94	99.69	98.54	97.61	100.00
Lettuce 5wk	95.33	99.94	99.08	99.95	99.94	99.82	98.21	100.00
Lettuce 6wk	99.27	100.00	97.45	98.85	98.91	97.33	94.42	100.00
Lettuce 7wk	100.00	100.00	92.43	100.00	98.34	98.44	94.29	100.00
Vinyard Untrained	78.94	88.68	89.15	88.53	95.96	92.43	80.61	99.88
Vinyard Vertical	98.52	100.00	96.93	98.14	95.82	99.13	99.14	100.00
κ	94.50	96.16	95.28	96.47	97.30	97.92	95.24	**99.83**
OA	95.07	96.55	95.76	96.83	97.57	98.14	95.74	**99.84**
AA	96.01	98.17	97.37	98.23	98.29	98.45	96.65	**99.79**

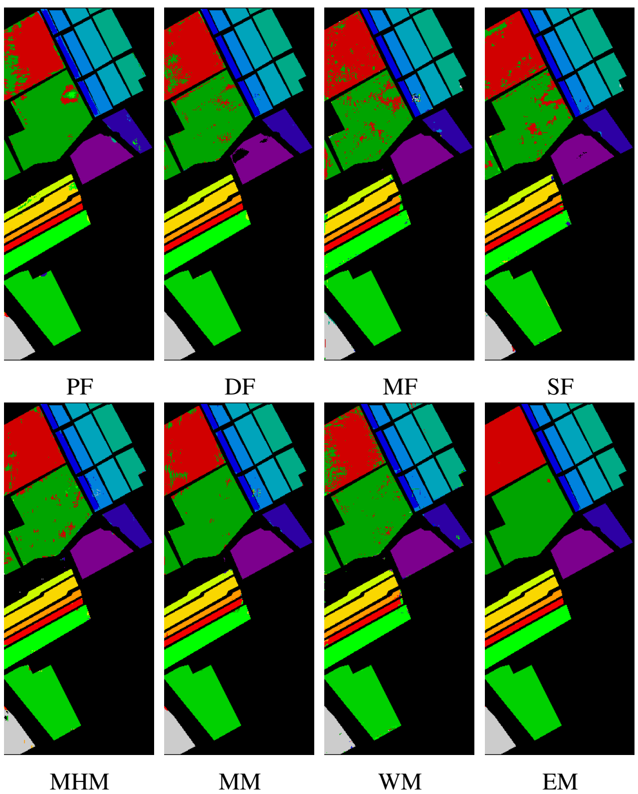

Fig. 2: Classification maps for the **Salinas dataset**, highlighting spatial variability and class-specific performance.

TABLE V: Performance comparison on the **HC dataset** across class-wise accuracies and aggregate metrics.

Class	PF	DF	MF	SF	MHM	MM	WM	***EM***
Strawberry	97.90	94.92	98.32	99.45	93.26	96.59	97.35	99.59
Cowpea	97.58	93.24	96.02	98.20	83.11	92.91	89.73	99.03
Soybean	81.83	85.53	97.47	98.64	27.21	90.35	90.63	99.51
Sorghum	98.15	98.90	97.51	98.72	87.48	97.51	95.86	99.21
Water spinach	86.20	71.76	92.97	93.51	19.35	86.57	87.03	99.91
Watermelon	32.97	59.29	78.90	85.51	5.93	52.84	70.39	98.38
Greens	94.00	80.78	96.42	95.59	11.56	92.45	92.37	99.57
Trees	84.75	85.33	94.15	95.42	62.42	89.31	86.00	99.30
Grass	90.45	89.13	95.19	96.50	9.36	88.71	83.32	98.28
Red roof	99.08	95.00	98.39	98.99	64.52	97.46	98.71	99.61
Gray roof	98.82	94.73	98.72	99.05	75.11	93.46	96.82	99.65
Plastic	93.75	80.91	94.93	97.42	2.23	83.93	75.17	99.91
Bare soil	69.03	68.83	90.26	89.24	64.31	79.51	77.00	96.66
Road	95.68	89.94	97.39	96.79	70.79	92.74	92.31	99.41
Bright object	87.49	74.36	85.63	83.13	47.16	80.04	65.85	89.93
Water	99.49	99.04	99.88	99.91	98.09	99.45	99.51	99.84
κ	93.03	90.83	96.82	97.63	71.14	92.81	92.45	**99.26**
OA	94.05	92.16	97.28	97.98	75.51	93.86	93.55	**99.36**
AA	87.94	85.11	94.51	95.38	51.37	88.37	87.38	**98.61**

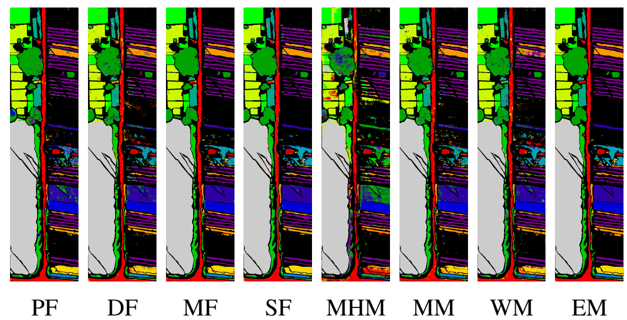

Fig. 3: Classification maps for the **HC dataset**, highlighting spatial variability and class-specific performance.

TABLE VI: Performance comparison on the **OHID-1 dataset** across class-wise accuracies and aggregate metrics.

Class	PF	DF	MF	SF	MHM	MM	WM	***EM***
Building	93.54	82.26	86.90	86.31	79.64	70.05	81.78	91.28
Farmland	75.04	78.27	83.01	87.89	64.90	55.85	77.56	85.94
Forest	92.37	77.06	91.37	90.68	75.35	72.76	85.18	83.38
Road	0.29	65.75	77.71	83.92	67.26	58.84	69.58	78.60
Water	97.26	96.70	98.03	97.05	96.87	94.29	97.81	97.87
Bareland	0.00	69.52	69.42	64.10	37.45	30.04	52.97	71.88
Fishpond	0.00	61.47	70.09	74.19	37.06	41.10	48.62	72.84
κ	75.14	80.31	87.24	87.00	75.14	68.30	82.73	**87.42**
OA	89.20	91.40	94.45	94.23	90.26	86.32	92.58	**94.59**
AA	51.21	75.86	82.36	83.45	65.51	60.42	73.36	**83.11**

variants primarily enhance representation learning by injecting additional priors such as multi-head feature aggregation, morphological operators, or wavelet-based multiscale decomposition, yet they typically lack an explicit, adaptive mechanism to prioritize discriminative spectral bands and spatial tokens conditioned on scene content. This limitation becomes critical in complex datasets such as OHID-1, which exhibits abrupt land-cover transitions, strong inter-class spectral similarity, and pronounced class imbalance. While the Mamba baselines tend to misclassify classes with subtle boundaries such as Road, Bareland, and Fishpond, EnergyMamba achieves higher class-wise accuracies and substantially higher AA, indicating improved robustness across both majority and minority classes.

The advantage of EnergyMamba is most pronounced on OHID-1, where it achieves the best OA and AA among all compared methods. The improvement in AA is particularly important because it reflects consistent per-class performance rather than dominance by majority classes, highlighting EnergyMamba's ability to mitigate class imbalance and enhance discrimination for underrepresented categories. These gains arise from the synergistic coupling of selective state space modeling and energy-based associative refinement, where the Mamba branch preserves global spectral-spatial context with linear complexity, and the Energy branch constructs an energy landscape that dynamically gates features to emphasize informative bands and tokens while suppressing noisy or redundant

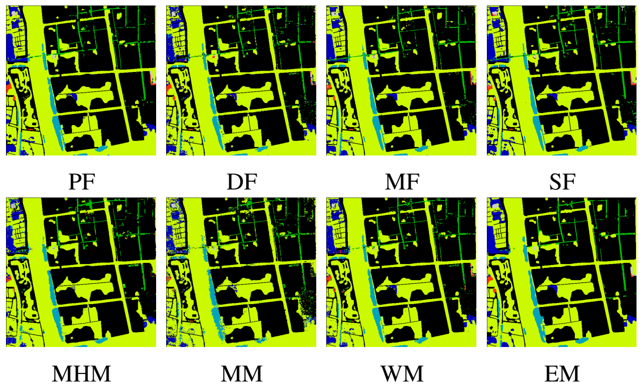

Fig. 4: Classification maps for the **OHID-1 dataset**, highlighting spatial variability and class-specific performance.

responses. Visual comparisons in Figure 4 corroborate this analysis, showing that EnergyMamba produces cleaner spatial delineation and fewer errors at class interfaces, especially in Farmland–Bareland and Road regions, where competing models exhibit evident spectral confusion.

On SA and HC, EnergyMamba maintains near-saturated accuracy while preserving fine spatial detail, demonstrating that its improvements are not limited to a single dataset regime. The consistently higher κ coefficients further indicate that the proposed energy-guided refinement increases agreement beyond chance and stabilizes predictions across heterogeneous regions. Overall, by integrating linear-complexity state space dynamics with energy-based feature selection, EnergyMamba provides a tradeoff between global dependency modeling and local discriminative enhancement, explaining its consistent superiority over both Transformer-based attention models and existing Mamba variants.

IV. Conclusion and Future Research Directions

This paper presented EnergyMamba, a hybrid architecture for hyperspectral image classification that combines selective state space modeling with energy-based associative feature refinement. By leveraging Mamba's linear-complexity state space dynamics, the proposed framework effectively captures long-range spectral–spatial dependencies, while the Energy Attention mechanism, integrating Hopfield-inspired energy modeling and multi-head energy attention, adaptively suppresses spectral redundancy and enhances discriminative feature representations. The hybrid model enables the proposed method to balance global contextual modeling with local discriminative enhancement, yielding stable and robust representations under limited supervision. Extensive experiments on the Salinas, WHU-Hi-HanChuan, and OHID-1 datasets demonstrate that EnergyMamba consistently outperforms state-of-the-art Transformer- and Mamba-based methods, achieving overall accuracies of 99.84%, 99.36%, and 94.59%, respectively, using only 5% labeled samples, along with significantly higher κ coefficients, particularly on the challenging OHID-1 dataset characterized by complex class distributions and spectral ambiguity. These results confirm the effectiveness and scalability of energy-guided state space modeling for hyperspectral analysis. Future research will focus on extending EnergyMamba through self-supervised and contrastive pre-training to further reduce annotation dependence, as well as domain adaptation and cross-scene generalization to enhance robustness under varying sensor, illumination, and atmospheric conditions.

References

[1] M. Ahmad, M. Mazzara, S. Distefano, A. Mehmood Khan, M. Hassaan Farooq Butt, and D. Hong, "Policymamba: Localized policy attention with state space model for land cover classification," IEEE Transactions on Neural Networks and Learning Systems, vol. 36, no. 10, pp. 17 814–17 825, 2025.

[2] F. Luo, Y. Liu, X. Gong, Z. Nan, and T. Guo, "Emvcc: Enhanced multi-view contrastive clustering for hyperspectral images," in Proceedings of the 32nd ACM International Conference on Multimedia, 2024, pp. 6288–6296.

[3] Z. Li, W. He, J. Li, F. Lu, and H. Zhang, "Learning without exact guidance: Updating large-scale high-resolution land cover maps from low-resolution historical labels," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024, pp. 27 717–27 727.

[4] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Spectral–spatial hyperspectral image segmentation using subspace multinomial logistic regression and markov random fields," IEEE Transactions on Geoscience and Remote Sensing, vol. 50, no. 3, pp. 809–823, 2011.

[5] L. He, J. Li, C. Liu, and S. Li, "Recent advances on spectral–spatial hyperspectral image classification: An overview and new guidelines," IEEE Transactions on Geoscience and Remote Sensing, vol. 56, no. 3, pp. 1579–1597, 2017.

[6] Y. Duan, C. Chen, M. Fu, X. Gong, Y. Niu, and F. Luo, "Gitanet: Group interactive threshold-based attention network for hyperspectral image classification," IEEE Transactions on Multimedia, pp. 1–14, 2025.

[7] C. Shi, D. Liao, and L. Wang, "Hybrid cnn-gcn network for hyperspectral image classification," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 18, pp. 10 530–10 546, 2025.

[8] M. Ahmad, M. Mazzara, S. Distefano, A. M. Khan, M. H. F. Butt, M. Usama, and D. Hong, "Graphmamba: Graph tokenization mamba for hyperspectral image classification," IEEE Transactions on Emerging Topics in Computing, vol. 13, no. 4, pp. 1510–1521, 2025.

[9] S. Mei, C. Song, M. Ma, and F. Xu, "Hyperspectral image classification using group-aware hierarchical transformer," IEEE Transactions on Geoscience and Remote Sensing, vol. 60, pp. 1–14, 2022.

[10] J. Feng, Q. Wang, G. Zhang, X. Jia, and J. Yin, "Cat: Center attention transformer with stratified spatial–spectral token for hyperspectral image classification," IEEE Transactions on Geoscience and Remote Sensing, vol. 62, pp. 1–15, 2024.

[11] Y. Fang, L. Sun, Y. Zheng, and Z. Wu, "Deformable convolution-enhanced hierarchical transformer with spectral-spatial cluster attention for hyperspectral image classification," IEEE Transactions on Image Processing, vol. 34, pp. 701–716, 2025.

[12] Y. Li, Y. Luo, L. Zhang, Z. Wang, and B. Du, "Mambahsi: Spatial–spectral mamba for hyperspectral image classification," IEEE Transactions on Geoscience and Remote Sensing, vol. 62, pp. 1–16, 2024.

[13] R. Xu, X.-M. Dong, W. Li, J. Peng, W. Sun, and Y. Xu, "Dbctnet: Double branch convolution-transformer network for hyperspectral image classification," IEEE Transactions on Geoscience and Remote Sensing, vol. 62, pp. 1–15, 2024.

[14] M. Ahmad, M. Mazzara, S. Distefano, and A. M. Khan, "Byte latent mamba with state space and knowledge distillation for hyperspectral image classification," IEEE Transactions on Geoscience and Remote Sensing, vol. 63, pp. 1–15, 2025.

[15] Y. He, B. Tu, B. Liu, J. Li, and A. Plaza, "Hsi-mformer: Integrating mamba and transformer experts for hyperspectral image classification," IEEE Transactions on Geoscience and Remote Sensing, pp. 1–1, 2025.

[16] H. Wang, P. Zhuang, X. Zhang, and J. Li, "Dbmgnet: A dual-branch mamba-gcn network for hyperspectral image classification," IEEE Transactions on Geoscience and Remote Sensing, pp. 1–1, 2025.

[17] M. Ahmad, M. H. F. Butt, A. M. Khan, M. Mazzara, S. Distefano, M. Usama, S. K. Roy, J. Chanussot, and D. Hong, "Spatial–spectral morphological mamba for hyperspectral image classification," Neurocomputing, vol. 636, p. 129995, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925231225006678

[18] P. Zhuang, X. Zhang, H. Wang, T. Zhang, L. Liu, and J. Li, "Fahm: Frequency-aware hierarchical mamba for hyperspectral image classification," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 18, pp. 6299–6313, 2025.

[19] M. Ahmad, M. Usama, M. Mazzara, and S. Distefano, "Wavemamba: Spatial-spectral wavelet mamba for hyperspectral image classification," IEEE Geoscience and Remote Sensing Letters, vol. 22, pp. 1–5, 2025.

[20] N. Li, Z. Wang, F. A. Cheikh, and M. Ullah, "S3am: A spectral-similarity-based spatial attention module for hyperspectral image classification," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 15, pp. 5984–5998, 2022.

[21] J. Sheng, J. Zhou, J. Wang, P. Ye, and J. Fan, "Dualmamba: A lightweight spectral-spatial mamba-convolution network for hyperspectral image classification," IEEE Transactions on Geoscience and Remote Sensing, 2024.

[22] M. Ahmad, M. H. F. Butt, M. Usama, H. A. Altuwaijri, M. Mazzara, S. Distefano, and A. M. K. and, "Multi-head spatial-spectral mamba for hyperspectral image classification," Remote Sensing Letters, vol. 16, no. 4, pp. 339–353, 2025. [Online]. Available: https://doi.org/10.1080/2150704X.2025.2461330

[23] J. Wu, J. Zhao, and H. Long, "Advanced hyperspectral image classification via spectral–spatial redundancy reduction and tokenlearner-enhanced transformer," IEEE Transactions on Geoscience and Remote Sensing, vol. 63, pp. 1–12, 2025.

[24] M. Ahmad, M. H. F. Butt, M. Mazzara, S. Distefano, A. M. Khan, and H. A. Altuwaijri, "Pyramid Hierarchical Spatial-Spectral Transformer for Hyperspectral Image Classification," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 17, pp. 17 681–17 689, 2024.

[25] M. Ahmad, M. Mazzara, S. Distefano, A. M. Khan, and S. L. Ullo, "Diffformer: a differential spatial-spectral transformer for hyperspectral image classification," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2025.

[26] S. K. Roy, A. Deria, C. Shah, J. M. Haut, Q. Du, and A. Plaza, "Spectral–spatial morphological attention transformer for hyperspectral image classification," IEEE Transactions on Geoscience and Remote Sensing, vol. 61, pp. 1–15, 2023.

[27] D. Hong, Z. Han, J. Yao, L. Gao, B. Zhang, A. Plaza, and J. Chanussot, "Spectralformer: Rethinking hyperspectral image classification with transformers," IEEE Transactions on Geoscience and Remote Sensing, vol. 60, pp. 1–15, 2021.