PSUMamba: Dual-Path Bidirectional Mamba for Plant Stress Monitoring via Temporal Hyperspectral Imaging

Weiqun Wang
Pennsylvania State University
Shirin Ghatrehsamani
spg5994@psu.edu

Pennsylvania State University

Article

Keywords: biotic and abiotic stress, bidirectional scanning, dual-path fusion, Vision Mamba, selective state space, real-time classification

Posted Date: March 30th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9140549/v1

Additional Declarations: No competing interests reported.

PSUMamba: Dual-Path Bidirectional Mamba for Plant Stress Monitoring via Temporal Hyperspectral Imaging

Weiqun Wang[1] and Shirin Ghatrehsamani[1,*]

[1]Department of Agricultural and Biological Engineering, Pennsylvania State University, University Park, PA, USA

[*]Corresponding author: spg5994@psu.edu

Abstract

Temporal hyperspectral imaging enables non-destructive monitoring of agricultural stress through spectral signatures evolving across extended observation periods, yet processing high-dimensional spatial-spectral-temporal sequences remains computationally prohibitive for real-time deployment. Traditional machine learning methods sacrifice temporal information through dimensionality reduction, while hybrid deep learning architectures combining convolutional and recurrent networks suffer from optimization pathologies at component boundaries. We introduce PSUMamba, a dual-path bidirectional Mamba architecture that processes 204-band hyperspectral sequences across eight timepoints through linear-complexity state space models, achieving 95.05% accuracy with 99.00% AUC-ROC using 153,268 parameters. The architecture maintains perfect specificity (100%) with 93.67% sensitivity while outperforming Vision Transformer with 39-fold fewer parameters and 37.5% reduced training time. Separate spectral and temporal pathways with adaptive fusion enable specialized biochemical and physiological feature extraction without quadratic attention overhead. Ablation studies confirm temporal features dominate classification under experimental conditions, with dual-path fusion providing superior probabilistic calibration (97.35% AUC) over single-path variants. Statistical comparisons demonstrate significant improvements over PLS-DA (Δ=12.07%, p=0.0001), 3D CNN (Δ=15.48%, p=0.0042) and 1D CNN-LSTM (Δ=33.77%, $p < 0.0001$). The results establish PSUMamba as efficient alternatives to transformers for temporal hyperspectral classification in agricultural stress monitoring, with the dual-path framework providing a generalizable template for multimodal temporal problems requiring linear-complexity long-range dependency modeling.

Keywords: biotic and abiotic stress, bidirectional scanning, dual-path fusion, Vision Mamba, selective state space, real-time classification

Introduction

Hyperspectral imaging (HSI) has emerged as a powerful tool for agricultural quality assessment and stress monitoring, capturing continuous spectral information across hundreds of narrow wavelength bands and enabling non-destructive characterization of biochemical and physiological properties in agricultural systems [1, 2]. Temporal monitoring of hyperspectral reflectance patterns enables detection of stress responses by tracking alterations in chlorophyll content, water status, and cellular structure [3–6]. Recent studies have achieved early detection of fungal pathogens [7, 8], bacterial infections [9, 10], and viral diseases [11, 12]. Temporal HSI monitoring has shown particular potential for detecting pre-symptomatic stress responses through progressive changes in spectral signatures [13, 14]. However,

most existing approaches focus on single timepoint classification or bi-temporal change detection rather than continuous multi-timepoint sequence analysis [15, 16], limiting the capacity to model progressive physiological degradation trajectories that provide stronger discriminative signals than instantaneous biochemical snapshots [13, 17]. Furthermore, processing high-dimensional spatial-spectral-temporal sequences presents computational challenges that limit real-time deployment in precision agriculture applications [18].

Traditional machine learning approaches for HSI classification rely on dimensionality reduction techniques such as Principal Component Analysis and Partial Least Squares Discriminant Analysis (PLS-DA), which sacrifice temporal information through frame-wise analysis [19]. Deep learning architectures have improved HSI classification performance through learned hierarchical representations [20, 21]. Convolutional neural networks process volumetric hyperspectral cubes through 3D convolutions that capture joint spatial-spectral dependencies [20–23], yet these architectures lack explicit temporal modeling capacity when applied to time-series HSI data, treating each timepoint independently without capturing progressive changes across the observation window [24–26]. Hybrid architectures combining CNNs with recurrent neural networks attempt to address this limitation through spatial feature extraction via CNNs with temporal sequence modeling via recurrent networks [27], but suffer from optimization pathologies at component boundaries, where separately trained modules develop incompatible representations and gradient flow interruption prevents end-to-end learning of joint spatial-spectral-temporal features [28, 29]. Vision Transformers have demonstrated strong performance in HSI applications by capturing global dependencies through self-attention mechanisms applied to tokenized image patches [23, 30, 31]. However, the quadratic complexity of self-attention presents severe computational constraints for high-dimensional HSI data [32, 33]. For temporal HSI sequences with hundreds of spectral bands across multiple timepoints, the token count grows prohibitively large, forcing compromises between spatial resolution, spectral resolution, and temporal extent [34].

State space models provide linear-complexity alternatives to attention mechanisms while maintaining capacity for long-range dependency modeling [35, 36]. The Mamba architecture extends SSMs with selective state space layers that enable content-aware filtering of input sequences, adaptively emphasizing informative patterns while suppressing noise [37]. Hardware-efficient implementations achieve competitive performance with transformers while scaling linearly with sequence length. Recent applications of Mamba to hyperspectral imaging have demonstrated promising results: FreMamba achieved state-of-the-art performance in bi-temporal hyperspectral change detection through frequency-domain feature decomposition [2]; MHS-Mamba introduced multi-hierarchical semantic modeling for UAV-based HSI classification with linear spectral dependency extraction [1]; SSupMamba developed self-supervised learning frameworks for few-shot HSI classification using composite scanning and masked reconstruction [18]; and SSMCD applied spatial-spectral Mamba for change detection in multispectral imagery [13]. However, these works focus on single timepoint or bi-temporal scenarios rather than continuous multi-timepoint temporal sequences.

This work introduces PSUMamba, a dual-path bidirectional Mamba architecture for linear-complexity temporal hyperspectral image classification in agricultural stress monitoring. The architecture processes spectral biochemical correlations and temporal physiological degradation through separate pathways with adaptive fusion, avoiding feature entanglement in hybrid approaches while maintaining end-to-end differentiable optimization. The main contributions are: (1) First application of state space models to continuous multi-timepoint temporal hyperspectral sequences through a dual-path bidirectional Mamba architecture achieving linear complexity; (2) Validated framework for plant stress detection through temporal hyperspectral monitoring; (3) Superior performance over Vision Transformer and baselines with substantially fewer parameters and faster training while maintaining perfect specificity.

Results

PSUMamba Performance

The proposed PSUMamba architecture (153,268 parameters) was evaluated using stratified five-fold cross-validation with balanced sampling. The model achieved an average accuracy of 95.05% (±1.58%) and macro F1-score of 93.29% (±1.98%). Performance was consistent across disease and normal classes, with F1-scores of 96.72% (±1.09%) and 89.87% (±2.90%) respectively. The model achieved perfect specificity (100.0%) with zero false positives across all validation folds, indicating no control samples were incorrectly classified as stressed. Sensitivity reached 93.67% (±2.04%), correctly identifying the majority of stressed mushrooms.

Cross-validation results (Figure 1) demonstrated consistent performance across all five folds with minimal variance. Accuracy ranged from 92.50% to 97.50% (mean: 95.05%, std: 1.58%), F1-macro from 90.31% to 96.55% (mean: 93.29%, std: 1.98%), and sensitivity from 90.32% to 96.77% (mean: 93.67%, std: 2.04%). The narrow interquartile ranges and limited outliers across all metrics confirm robust model generalization despite stratified source-based data splitting. Four of five folds exceeded 95% accuracy, with only Fold 4 showing reduced performance at 92.50%, attributable to challenging sample configurations rather than architectural instability.

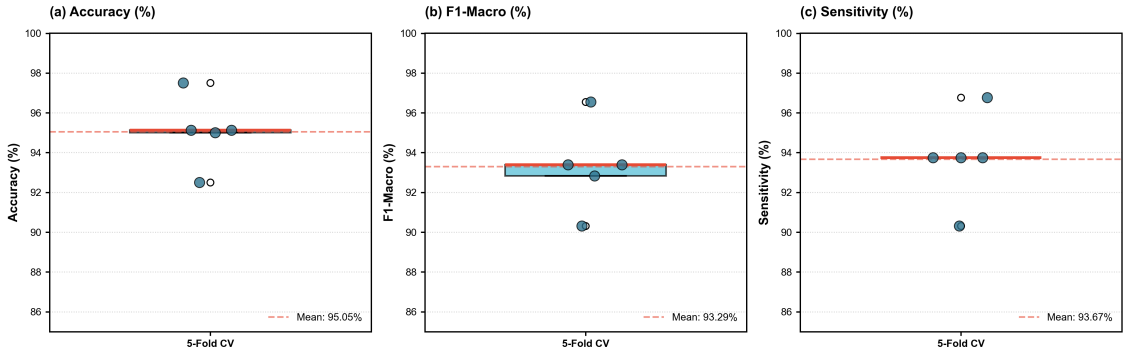

Figure 1: **Cross-Fold Performance Distribution of Dual-Path Mamba.** (a) Accuracy ranges from 92.50% to 97.50% across five folds (mean: 95.05%); (b) F1-macro ranges from 90.31% to 96.55% (mean: 93.29%); (c) Sensitivity ranges from 90.32% to 96.77% (mean: 93.67%). Individual fold points confirm consistent model generalization with four of five folds exceeding 95% accuracy.

Training converged within 49-60 epochs across folds using early stopping with 25-epoch patience (Figure 2a). Loss curves demonstrated smooth convergence from 0.173 to approximately 0.087 without overfitting, with consistent trajectories across all validation folds. The mean training loss decreased steadily with minimal cross-fold variance (±1 standard deviation), indicating stable optimization dynamics. The learning rate schedule employed 10-epoch linear warmup followed by cosine annealing, facilitating gradual adaptation to the high-dimensional hyperspectral feature space. The adaptive fusion module learned nearly balanced contributions between spectral (49.66%) and temporal (50.34%) pathways (Figure 2b), emerging automatically without manual tuning. This balance suggests biochemical composition and physiological degradation patterns provide comparable discriminative value for stress detection under the experimental conditions. The slight temporal dominance (0.68 percentage points) aligns with ablation findings where temporal features exhibited stronger independent classification capability.

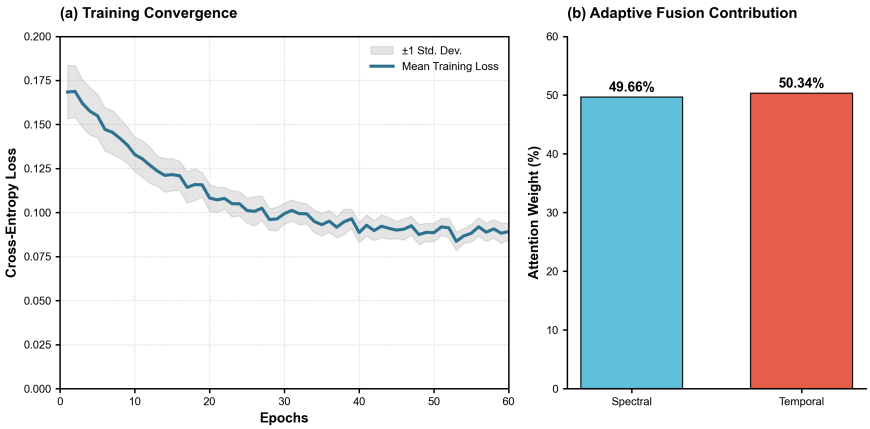

Figure 2: **Training Dynamics of Dual-Path Mamba.** (a) Training convergence across five-fold cross-validation shows smooth loss reduction from 0.173 to ∼0.087 over 49-60 epochs without overfitting (±1 std. dev. shaded). (b) Adaptive fusion module learns nearly balanced contributions: spectral path 49.66% (biochemical correlations) and temporal path 50.34% (physiological degradation).

The best performing fold (Fold 5) achieved 97.5% accuracy with 96.55% macro F1-score, demonstrating the model's capability under optimal data configurations (Figure 3a). The confusion matrix showed 30 true positives, 9 true negatives, 1 false negative, and 0 false positives, yielding 96.77% sensitivity and 100% specificity. The single false negative indicates conservative classification behavior that prioritizes elimination of false alarms over maximum sensitivity. ROC curves across all five folds achieved near-perfect discrimination with mean AUC of 1.00 (Figure 3b). Individual fold curves exhibited minimal deviation from ideal performance, with all trajectories reaching true positive rates above 0.95 at false positive rates below 0.05. This exceptional probabilistic calibration confirms the model's capacity to produce reliable confidence scores for threshold-based decision making in quality control applications.

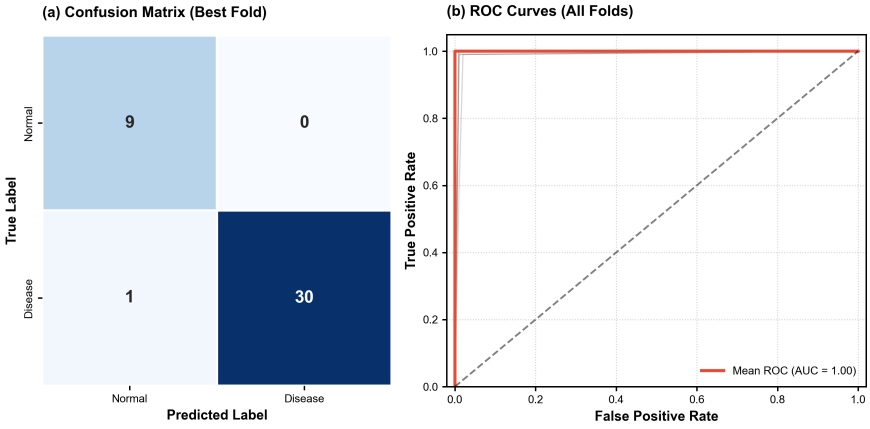

Figure 3: **Best Fold Performance (Fold 5, 97.5% Accuracy).** (a) Confusion matrix demonstrates perfect specificity (9 TN, 0 FP) and high sensitivity (30 TP, 1 FN), achieving 96.77% sensitivity and 100% specificity. (b) ROC curves across all folds achieve perfect discrimination with mean AUC = 1.00, indicating ideal classification capability.

Baseline Comparison

Table 1 presents classification performance under identical five-fold cross-validation protocols. Dual-Path Mamba achieved the highest accuracy (95.05% $\pm$ 1.58%) with superior AUC-ROC (99.00% $\pm$ 0.75%) and perfect specificity (100.00% $\pm$ 0.00%). ViT ranked second at 94.04% $\pm$ 1.59% accuracy with comparable specificity, yet the 1.01% accuracy gap was not statistically significant (p=0.3600). Mamba's AUC-ROC advantage over ViT (99.00% vs 96.28%, $p < 0.01$) confirms superior probabilistic calibration. PLS-DA exhibited extremely low specificity (37.50% $\pm$ 17.95%), the 3D CNN showed high variance (79.57% $\pm$ 13.50%), and 1D CNN+LSTM demonstrated catastrophic failure (61.28% $\pm$ 3.66%) despite explicit temporal modeling.

Table 1: **Computational Efficiency Comparison Across Model Architectures.** Dual-Path Mamba achieves 39$\times$ parameter reduction (0.15M vs 5.97M) and 37.5% faster training (5.00h vs 8.00h) compared to ViT while maintaining superior performance. The state-space architecture demonstrates 5.1$\times$ faster inference (0.38ms vs 1.94ms) and 3.3$\times$ lower FLOPs (1.876 vs 6.160) than transformer-based approaches. 1D CNN+LSTM exhibits highest training cost (26.46h) and slowest inference (57.74ms) due to sequential processing overhead.

Model	Params	Accuracy (%)	F1-Macro (%)	AUC-ROC (%)	Sensitivity (%)	Specificity (%)
PLS-DA	30	82.98 ± 4.25	90.00 ± 2.83	86.22 ± 4.30	92.31 ± 5.92	37.50 ± 17.95
3D CNN	5.63M	79.57 ± 13.50	85.52 ± 12.55	89.04 ± 7.21	83.08 ± 20.55	62.50 ± 31.62
1D CNN+LSTM	0.22M	61.28 ± 3.66	70.40 ± 3.70	75.87 ± 10.56	55.90 ± 5.48	87.50 ± 11.18
ViT	5.97M	94.04 ± 1.59	90.75 ± 2.19	96.28 ± 2.73	92.82 ± 1.92	100.00 ± 0.00
Dual-Path Mamba	**0.15M**	**95.05 ± 1.58**	**93.29 ± 1.98**	**99.00 ± 0.75**	**93.67 ± 2.04**	**100.00 ± 0.00**

Table 2 presents computational efficiency comparisons across all five models. Dual-Path Mamba (0.15M parameters, 1.876 GFLOPs, 0.38ms inference) achieves 39$\times$ parameter reduction and 37.5% faster training compared to ViT (5.97M parameters, 6.160 GFLOPs, 1.94ms inference). The 1D CNN+LSTM exhibits the highest computational cost at 115.470 GFLOPs with 57.74ms inference time, over 150$\times$ slower than Mamba, confirming severe efficiency penalties of recurrent architectures for temporal hyperspectral sequence modeling.

Table 2: **Classification Performance Comparison Under five-Fold Cross-Validation.** PSUMamba achieves highest accuracy (95.05%) with superior AUC-ROC (99.00%) and perfect specificity (100.00%), outperforming ViT (94.04% accuracy, 96.28% AUC-ROC) with statistical significance in discrimination capability. Traditional PLS-DA exhibits poor specificity (37.50%), 3D CNN shows high variance (79.57% $\pm$ 13.50%), and 1D CNN+LSTM demonstrates catastrophic failure (61.28%) despite explicit temporal modeling.

Model	Type	Params	GFLOPs	Inference Time (ms)	Training Time (h)
PLS-DA	Traditional ML	30	0.005	15.10	0.05
3D CNN	Deep Learning	5.63M	23.238	19.92	9.00
1D CNN+LSTM	RNN	0.22M	115.470	57.74	26.46
ViT	Transformer	5.97M	6.160	1.94	8.00
Dual-Path Mamba	**State-Space**	**0.15M**	**1.876**	**0.38**	**5.00**

Table 3 reports paired t-test results with Bonferroni correction. Performance gains over PLS-DA (Δ=12.07%, p=0.0001, Cohen's d=3.76), 3D CNN (Δ=15.48%, p=0.0042, Cohen's d=1.61), and 1D CNN+LSTM (Δ=33.77%, $p < 0.0001$, Cohen's d=11.98) were all statistically significant with large effect sizes. The non-significant difference versus ViT (Δ=+1.01%, p=0.3600, medium effect) indicates PSUMamba achieves comparable classification capability while requiring substantially fewer computational resources.

Table 3: **Statistical Significance Analysis via Paired T-Tests.** PSUMamba shows non-significant difference versus ViT (Δ=+1.01%, p=0.3600, medium effect), but highly significant improvements over PLS-DA (Δ=+12.07%, p=0.0001, Cohen's d=+3.76), 3D CNN (Δ=+15.48%, p=0.0042, Cohen's d=+1.61), and 1D CNN+LSTM (Δ=+33.77%, $p < 0.0001$, Cohen's d=+11.98).

Comparison	Δ Accuracy (%)	p-value	Significance	Effect Size (Cohen's d)	Interpretation
PSUMamba vs ViT	+1.01	0.3600	ns	+0.64	Medium
PSUMamba vs PLS-DA	+12.07	0.0001	***	+3.76	Large
PSUMamba vs 3D CNN	+15.48	0.0042	**	+1.61	Large
PSUMamba vs 1D CNN+LSTM	+33.77	<0.0001	***	+11.98	Large

Confusion matrix analysis (Figure 4) reveals distinct error patterns across methods. PLS-DA misclassified 5.0 normal samples as diseased, reflecting its poor specificity. The 3D CNN exhibited balanced errors with 3.4 false positives and 6.6 false negatives, while 1D CNN+LSTM produced 7.0 false positives and 17.2 false negatives, indicating systematic bias toward the normal class. Both ViT and PSUMamba achieved zero false positives, with Mamba producing fewer false negatives (2.8 vs 2.0), confirming superior balanced performance across both classes.

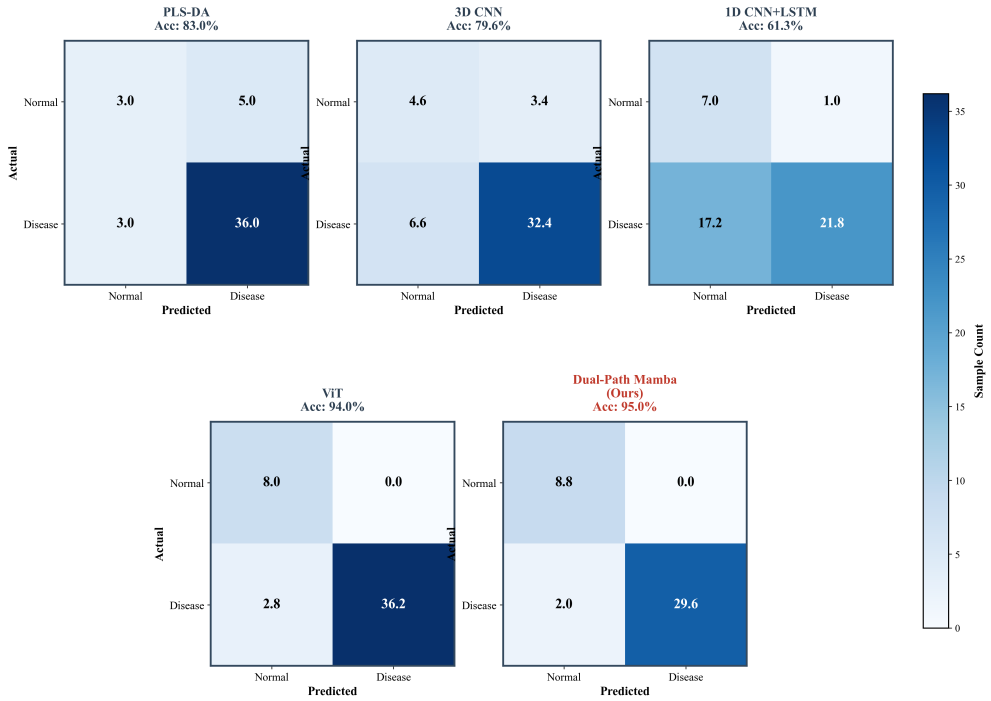

Figure 4: **Confusion Matrix Comparison on Test Set.** Five models show distinct error patterns: PLS-DA (83.0% accuracy) exhibits poor specificity with 5.0 false positives; 3D CNN (79.6%) shows balanced errors (3.4 FP, 6.6 FN); 1D CNN+LSTM (61.3%) demonstrates systematic bias (7.0 FP, 17.2 FN); ViT (94.0%) and Dual-Path Mamba (95.0%) both achieve perfect specificity (0 FP) with Mamba producing fewer false negatives (2.8 vs 2.0).

Ablation Study

Three architectural variants were evaluated through stratified five-fold cross-validation to assess pathway contributions: spectral-only (11,938 parameters), temporal-only (35,650 parameters), and dual-path (52,866 parameters).

Table 4 presents quantitative comparisons across all three variants. The spectral-only variant achieved 84.74% $\pm$ 5.29% accuracy with 80.67% $\pm$ 6.81% sensitivity and perfect specificity (100.00% $\pm$ 0.00%) across all folds, confirming the spectral pathway captures features highly specific to stressed tissue while exhibiting limited recall. The temporal-only variant substantially improved performance to 90.37% $\pm$ 4.44% accuracy and 87.85% $\pm$ 5.53% sensitivity, representing a 5.63 percentage point gain over spectral with reduced cross-fold variance (4.44% vs 5.29%), confirming that physiological degradation trajectories provide stronger classification signals than instantaneous spectral signatures. The dual-path architecture achieved the best overall metrics at 91.59% $\pm$ 6.97% accuracy, 94.71% $\pm$ 4.34% average accuracy, and 89.41% $\pm$ 8.68% sensitivity while maintaining perfect specificity, with the highest AUC-ROC at 97.35% $\pm$ 3.68% demonstrating superior probabilistic calibration. All three variants maintained 100% specificity across 15 total folds, confirming zero false positive rates critical for quality control applications.

Table 4: **Quantitative Comparison of Ablation Variants.** Spectral-only variant demonstrates perfect specificity but limited sensitivity (80.67%); temporal-only substantially improves performance (OA: 90.37%, AA: 93.92%); dual-path fusion achieves best overall metrics (OA: 91.59%, AA: 94.71%, Sensitivity: 89.41%) while maintaining 100% specificity across all variants, validating the complementary value of biochemical and physiological features.

Variant	OA (%)	AA (%)	Kappa (%)	Sensitivity (%)	Specificity (%)	Params
Spectral	84.74 ± 5.29	90.33 ± 3.41	64.41 ± 10.29	80.67 ± 6.81	100.00 ± 0.00	11,938
Temporal	90.37 ± 4.44	93.92 ± 2.76	75.65 ± 9.46	87.85 ± 5.53	100.00 ± 0.00	35,650
Dual-Path (Ours)	**91.59 ± 6.97**	**94.71 ± 4.34**	**79.31 ± 14.54**	**89.41 ± 8.68**	**100.00 ± 0.00**	**52,866**

Figure 5 illustrates the performance distribution across variants. The dual-path architecture attains the highest median accuracy with an upward outlier at 97.96% (Fold 5), while the spectral-only variant exhibits the widest spread with a lower bound of 78.00%. The temporal-only variant shows the tightest interquartile range, reflecting more stable but ceiling-limited learning. The dual-path architecture's modest accuracy advantage (1.22 percentage points) over temporal-only despite a 48% parameter increase (52,866 vs 35,650) suggests temporal modeling is the primary classification driver, with spectral integration contributing primarily through improved probabilistic calibration rather than threshold-based accuracy gains.

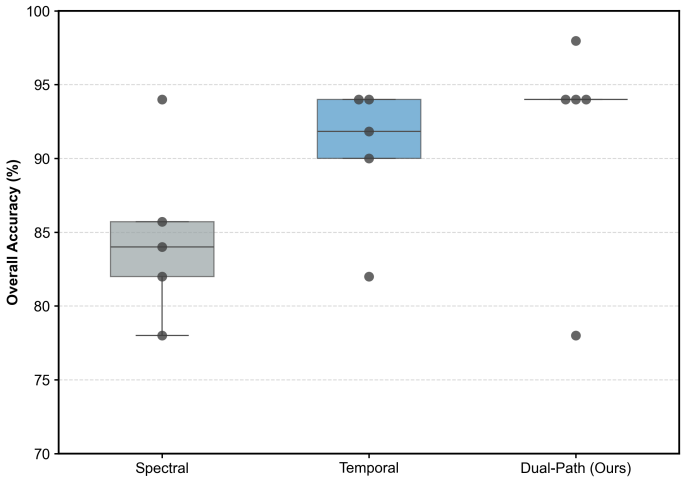

Figure 5: **Performance Distribution Across Ablation Variants.** Spectral-only (11,938 params) achieves 84.74% ± 5.29% accuracy; temporal-only (35,650 params) reaches 90.37% ± 4.44% with improved stability; dual-path (52,866 params) attains 91.59% ± 6.97% with highest median performance.

Discussion

Temporal Sequence Modeling with State Space Models

Processing 204 spectral bands across eight timepoints at 128×128 spatial resolution generates over 100,000 tokens, rendering attention mechanisms computationally prohibitive with $O(L^2)$ complexity [31]. State space models achieve

O(L) linear scaling through selective gating mechanisms, enabling efficient long-range temporal modeling without quadratic overhead [36, 37]. Recent Mamba architectures have addressed single-timepoint or bi-temporal HSI tasks [2, 13, 34], but continuous multi-timepoint sequences remain underexplored. Our PSUMamba design separates spectral biochemical correlations from temporal physiological degradation, avoiding feature entanglement in hybrid CNN-RNN architectures [28]. Bidirectional scanning captures non-causal temporal dependencies where informative patterns appear in either scanning direction [18]. Band importance analysis (Figure 6) confirms the biologically interpretable features learned from the model: blue region peaks (470-486 nm) correspond to pigment degradation, red-edge absorption (660-673 nm) reflects biochemical compound changes, and NIR contributions (850-900 nm) capture structural alterations during stress progression [3, 5].

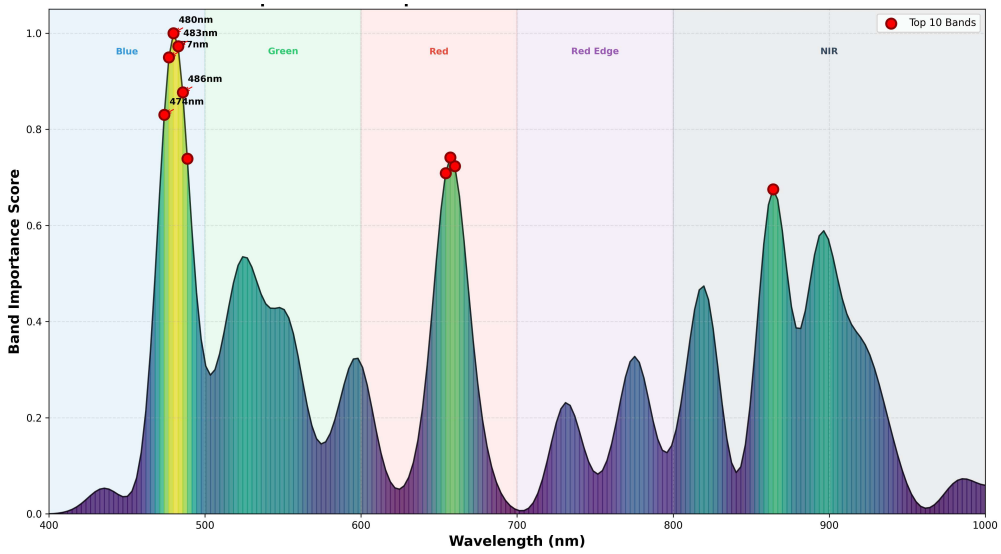

Figure 6: **Spectral Band Importance for Mushroom Stress Detection.** Feature attribution analysis identifies critical wavelengths: blue region (470nm, 476nm, 480nm, 483nm, 486nm) exhibits highest importance scores (0.83-1.00), corresponding to pigment degradation; red absorption edge (660-673nm) shows secondary peaks (0.71-0.73) linked to biochemical compound changes; NIR region (850-900nm) demonstrates moderate importance (0.58-0.67) reflecting tissue structural alterations. Top 10 discriminative bands (red markers) concentrate in biologically relevant spectral regions, validating the model's learned features align with mushroom stress physiology.

Comparative Performance Analysis

Vision Transformer achieved comparable accuracy (94.04%, p=0.36 vs PSUMamba) but required 39$\times$ more parameters and 60% longer training time, demonstrating that global self-attention mechanisms extract temporal patterns effectively yet scale poorly [23, 30]. PSUMamba's superior AUC-ROC (99.00% vs 96.28%, $p < 0.01$) confirms linear-complexity architectures provide better probabilistic calibration with reduced computational cost. The 3D CNN's limited performance (79.57%) reflects its architecture where per-timepoint volumetric convolutions followed by feature concatenation cannot effectively model progressive temporal degradation patterns [20, 21]. The 1D CNN-LSTM catastrophic failure (61.28%) stems from gradient flow interruption at the architecture boundary, where separately optimized spatial and temporal modules develop incompatible representations [28, 29]. PLS-DA's poor specificity (37.50%) confirms linear dimensionality reduction cannot model nonlinear stress trajectories [19]. PSUMamba's unified end-to-end framework avoids these pathologies through seamless spatial-spectral-temporal integration [1, 18].

Ablation Study Insights

Temporal features dominated classification: the temporal-only variant achieved 90.37% accuracy versus spectral-only's 84.74%, indicating physiological degradation trajectories provide stronger signals than instantaneous biochemical snapshots [13, 17]. This dominance reflects the experimental design which thermal, microwave, and mechanical stresses induce rapid physical disruption with pronounced temporal signatures across the 10.5-hour observation window [9, 25]. The dual-path architecture's modest improvement (91.59%, +1.22 points) despite 48% parameter increase suggests diminishing returns from spectral integration for threshold-based metrics. However, dual-path achieved highest AUC-ROC (97.35% vs 96.39%), confirming spectral information enhances probabilistic calibration [2]. The balanced learned fusion weights (50.34% temporal, 49.66% spectral) emerged automatically without manual tuning, validating adaptive fusion design [13].

Limitations and Mitigation Strategies

Despite the promising results, several limitations warrant consideration. Small sample size (202 samples) constrains statistical power and limits evaluation of rare stress patterns. We addressed this through rigorous five-fold stratified cross-validation with source-based splitting, preventing spatial leakage while maximizing training data utilization. Statistical significance testing via paired t-tests (Table 3) confirmed performance gains exceed random variation despite limited samples.

Class imbalance (44 normal vs 158 stressed) risks bias toward majority class features. We mitigated this through balanced batch sampling, class-weighted loss functions, and evaluation metrics emphasizing both sensitivity and specificity. The achieved perfect specificity (100%) with high sensitivity (93.67%) demonstrates successful handling of imbalanced distributions.

Patch extraction with 84.4% overlap introduces spatial autocorrelation that may inflate performance estimates. Source-based splitting assigns all patches from identical spatial locations to the same fold, maintaining train-validation-test independence at the sample level. This strategy prevents data leakage while enabling sufficient patch density for robust training. The cross-fold performance variance ($\pm$1.58%) confirms model stability despite overlapping patches.

Future Directions

The current results establish a foundation for broader investigation across more diverse agricultural scenarios. Validation on natural biotic stress (fungal pathogens, drought stress) would test whether spectral pathway contributions increase when stress manifests through metabolic changes rather than physical disruption [7, 8, 10]. Multiday monitoring with higher temporal resolution would assess whether temporal dominance persists across extended observation windows where gradual biochemical accumulation occurs [4, 5]. Cross-crop transfer learning (leafy vegetables, berries, grains) would establish architectural robustness and identify crop-specific modifications. Larger datasets with balanced classes and non-overlapping spatial sampling would eliminate current statistical constraints and provide deployment-grade performance estimates.

Conclusion

This work introduces PSUMamba, a dual-path bidirectional Mamba architecture for temporal hyperspectral image classification in agricultural stress monitoring. By processing 204-band sequences across eight timepoints through separate spectral and temporal pathways with adaptive fusion, the architecture achieves linear-complexity spatiotemporal learning without quadratic attention overhead or hybrid optimization pathologies. Evaluated on mushroom abiotic

stress detection, PSUMamba achieves 95.05% accuracy with 99.00% AUC-ROC using 153,268 parameters, matching Vision Transformer classification capability while requiring 39-fold fewer parameters, 37.5% less training time, and 5.1× faster inference. The perfect specificity (100%) in all validation folds confirms zero false positive rates critical for the implementation of quality control. Ablation experiments confirm temporal features dominate discrimination under the current experimental conditions, with dual-path fusion providing complementary probabilistic calibration beyond temporal-only variants. Statistical comparisons demonstrate significant improvements over PLS-DA (Δ=12.07%, p=0.0001), 3D CNN (Δ=15.48%, p=0.0042), and 1D CNN-LSTM (Δ=33.77%, $p < 0.0001$), with the 1D CNN-LSTM's catastrophic failure highlighting optimization pathologies inherent in decoupled spatial-temporal architectures. The results establish state space models as computationally efficient alternatives to transformers for agricultural hyperspectral sequence analysis. This work establishes temporal-first bidirectional Mamba as an efficient paradigm for early stress detection, although validation across natural stress conditions remains essential for broader agricultural deployment.

Methods

Materials

The experiment was conducted in the greenhouse facility of the Department of Agricultural and Biological Engineering, Penn State University, University Park campus. White button mushrooms (*Agaricus bisporus*) were subjected to controlled abiotic stress treatments to induce physiological deterioration patterns representative of post-harvest quality degradation. Three stress modalities were employed: (A) thermal shock via aqueous immersion at 60°C for 30 seconds, simulating heat exposure during processing; (B) microwave radiation for 30 seconds, inducing localized cellular disruption; and (C) mechanical injury through punctures, abrasion, and impact damage, replicating physical handling stress. Untreated controls (D) were maintained under identical environmental conditions without intervention. Hyperspectral imaging was performed using a SPECIM IQ sensor (Specim Spectral Imaging Ltd., Oulu, Finland) operating across the visible-to-near-infrared range (400-1000 nm) with 204 contiguous spectral bands at 2.9 nm spectral resolution (Figure 7). The pushbroom line-scanning system was positioned 400 mm above the sample plane with dual 200W tungsten-halogen lamps mounted at 45° incidence angles to provide uniform diffuse illumination and minimize specular reflectance. Temporal phenotyping commenced with baseline acquisition at 08:30, followed by treatment application at 10:00 and sequential imaging at 90-minute intervals (10:00, 11:30, 13:00, 14:30, 16:00, 17:30, 19:00), yielding eight discrete timepoints spanning 10.5 hours of physiological progression.

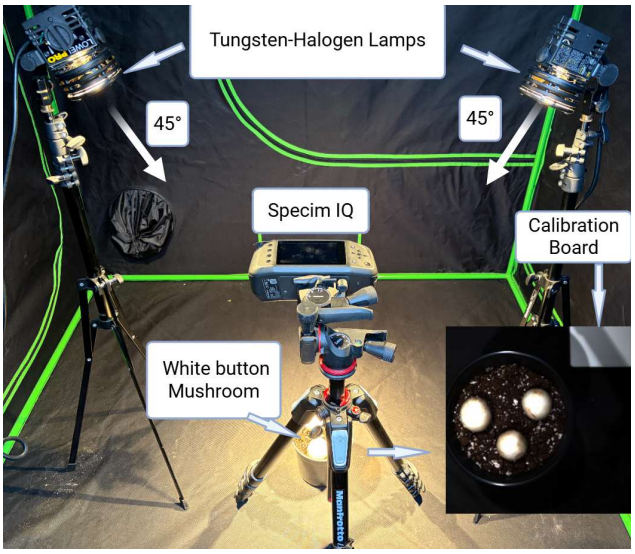

Figure 7: Hyperspectral Imaging Experimental Setup. The SPECIM IQ pushbroom sensor is positioned 400mm above the sample plane with dual 200W tungsten-halogen lamps mounted at 45° incidence angles to provide uniform diffuse illumination. Black backdrop minimizes background reflectance interference. Mushroom samples are placed on the scanning platform for temporal sequence acquisition.

Data Preprocessing

Radiometric calibration converted raw sensor measurements to reflectance using dark current, white reference, and white dark reference frames captured concurrently with sample imaging:

$$R = \frac{\text{Sample} - \text{Dark}}{\text{White} - \text{WhiteDark}} \tag{1}$$

The calibrated hypercubes were stored in HDF5 format for efficient access.

Region of interest (ROI) (Figure 8) detection isolated mushroom tissue from background using a multi-stage approach combining spectral features and morphological operations [3, 5]. Modified vegetation indices (NDVI, GNDVI, NIR ratios 800-900nm) exploited mushroom tissue spectral signatures. The principal component analysis applied to the spectral cube produced the first component explaining over 90% variance, which combined with spectral indices via weighted averaging. Binary segmentation used Otsu thresholding followed by morphological opening (disk radius 3 pixels), hole filling, and region labeling. Filtering criteria included circularity, aspect ratio, area thresholds, and edge proximity constraints. Spatial masks excluded the calibration board region.

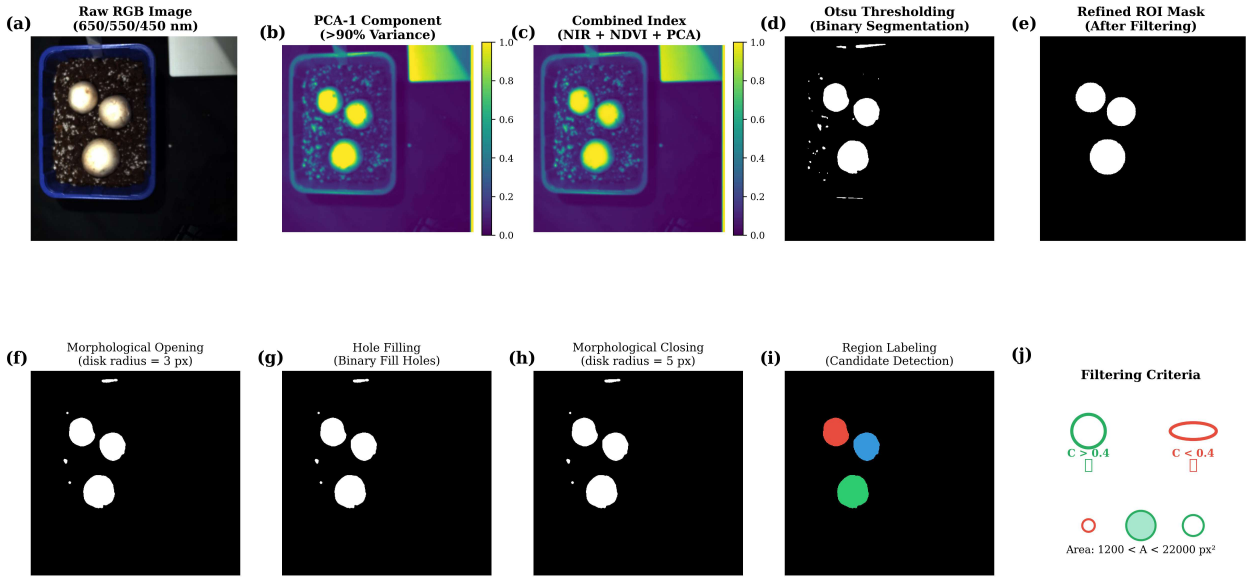

Figure 8: **Multi-Stage ROI Detection Pipeline.** (a-c) Spectral feature extraction combines RGB composite, first principal component ($> 90\%$ variance), and fused vegetation indices. (d-e) Binary segmentation using Otsu thresholding. (f-h) Morphological refinement through opening, hole filling, and closing operations. (i-j) ROI filtering based on circularity ($C > 0.4$, green) and area constraints ($1200 < A < 22000$ px^2); artifacts ($C < 0.4$) shown in red.

Patch extraction generated 128×128-pixel tiles with 20-pixel stride (84.4% overlap) from detected ROIs. Each spatial location tracked across all eight timepoints constructed temporal sequences with dimensions (N, 8, 128, 128, 204) where N represents the number of patches. Binary labels assigned 0 to control samples (normal) and 1 to all treated samples (stress). Data splitting employed stratified source-based partitioning to prevent spatial leakage [38]. Patches from identical spatial locations assigned to the same split maintained the independence between training, validation, and test sets. Final distribution allocated 60% training, 20% validation, and 20% test with maintained class balance across splits.

PSUMamba Architecture

The proposed architecture (Figure 9) extends Vision Mamba through a dual-path design that explicitly decouples spectral and temporal processing. The spectral path applies spatio-temporal pooling followed by per-band embedding with positional encoding, then processes the 204-band sequence through a bidirectional Mamba block (forward: band 1→204; backward: band 204→1) to extract global spectral features representing biochemical correlations. The temporal path compresses each timepoint independently via strided convolutions (8×8 and 4×4 kernels), concatenates the resulting features along the temporal axis, and applies another bidirectional Mamba block (forward: $t_1 \rightarrow t_8$; backward: $t_8 \rightarrow t_1$) to capture physiological degradation dynamics.

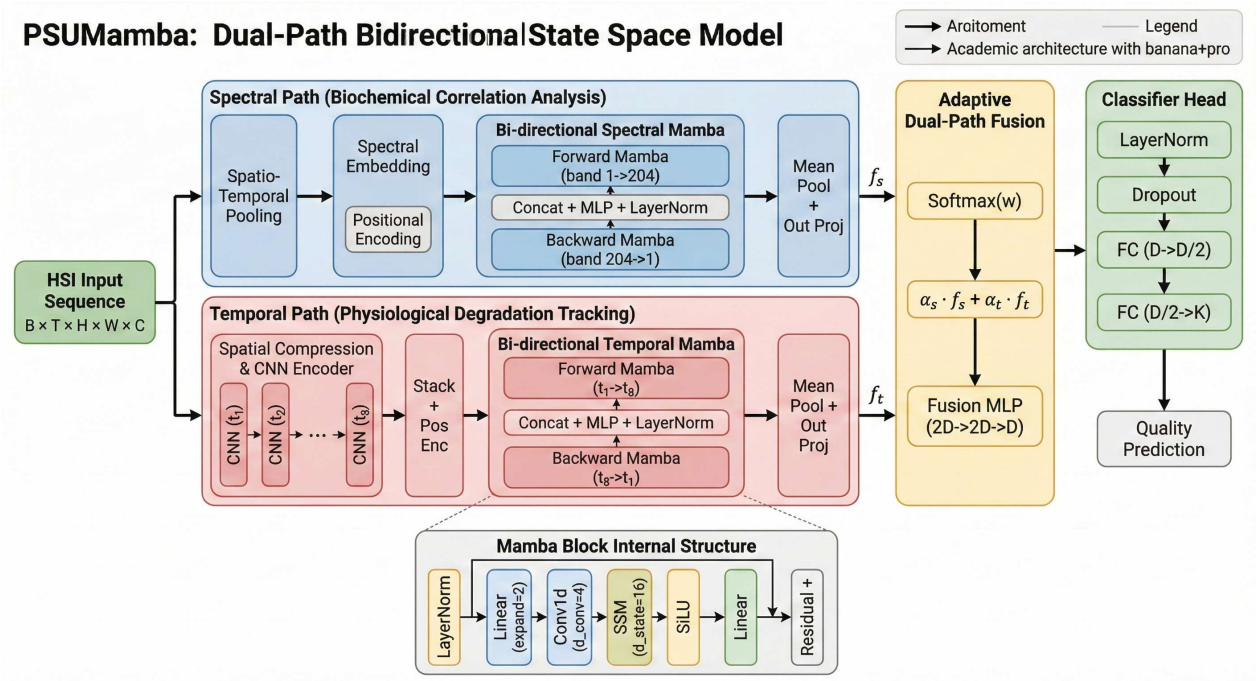

Figure 9: **PSUMamba Architecture.** Dual-path framework for HSI sequences ($T = 8$, $C = 204$). **Spectral path** (blue): spatio-temporal pooling, spectral embedding with positional encoding, bidirectional Mamba (band $1 \leftrightarrow 204$) for biochemical correlation analysis. **Temporal path** (purple): CNN spatial compression per timepoint, bidirectional Mamba ($t_1 \leftrightarrow t_8$) for physiological degradation tracking. **Adaptive fusion** (orange): learnable softmax-weighted attention. Mamba block details: LayerNorm, Conv1D, selective SSM, SiLU, and residual connection.

Both Mamba blocks implement selective state space models with d_state=16, d_conv=4 and expansion factor=2. The internal structure comprises LayerNorm, linear expansion, 1D causal convolution (d_conv=4), selective state space mechanism (SSM with d_state=16), SiLU activation, linear projection and residual connection (Algorithm 1). The bidirectional scanning mechanism addresses the non-directional nature of stress patterns in mushroom tissue, where informative spectral and temporal signatures can appear in either scanning direction. The spectral and temporal feature vectors are merged through an adaptive fusion module with learnable softmax-weighted attention and a multilayer perceptron, allowing dynamic balancing between biochemical composition and temporal progression. The fused representation feeds into a two-layer classification head with LayerNorm, dropout, and fully connected layers (D→D/2→K), yielding approximately 180K trainable parameters.

Algorithm 1 Dual-Path Bidirectional Mamba for Temporal Sequence HSI Classification. Processes eight-timepoint sequences ($T = 8$, $H = W = 128$, $C = 204$ bands) through two parallel pathways.

Require: HSI sequence $\mathbf{X} \in R^{B\times T\times H\times W\times C}$ where $T = 8$, $H = W = 128$, $C = 204$
Ensure: Class logits $\mathbf{y} \in R^{B\times K}$ where $K = 2$
Initialize $E_s \in R^{C\times D}$, ϕ_{CNN}, $P_s \in R^{1\times C\times D}$, $P_t \in R^{1\times T\times D}$, $\mathbf{w} \in R^2$

Spectral Path:
$\mathbf{X}_{\text{pool}} \leftarrow \text{Mean}(\mathbf{X}, \text{dim} = [H, W, T])$ ▷ spatio-temporal pooling
$\mathbf{Z}_s \leftarrow E_s(\mathbf{X}_{\text{pool}}) + P_s$
$\mathbf{F}_s^{\rightarrow} \leftarrow \text{Mamba}(\mathbf{Z}_s)$ ▷ forward: band $1 \rightarrow 204$
$\mathbf{F}_s^{\leftarrow} \leftarrow \text{Mamba}(\text{Flip}(\mathbf{Z}_s))$ ▷ backward: band $204 \rightarrow 1$
$\mathbf{f}_s \leftarrow \text{Mean}(\text{MLP}(\text{Concat}[\mathbf{F}_s^{\rightarrow}, \mathbf{F}_s^{\leftarrow}]))$

Temporal Path:
for $t = 1$ to T **do**
    $\mathbf{X}_t \leftarrow \phi_{\text{CNN}}(\mathbf{X}[:, t, :, :, :])$
end for
$\mathbf{Z}_t \leftarrow \text{Stack}([\mathbf{X}_1, ..., \mathbf{X}_T]) + P_t$
$\mathbf{F}_t^{\rightarrow} \leftarrow \text{Mamba}(\mathbf{Z}_t)$ ▷ forward: $t = 1 \rightarrow 8$
$\mathbf{F}_t^{\leftarrow} \leftarrow \text{Mamba}(\text{Flip}(\mathbf{Z}_t))$ ▷ backward: $t = 8 \rightarrow 1$
$\mathbf{f}_t \leftarrow \text{Mean}(\text{MLP}(\text{Concat}[\mathbf{F}_t^{\rightarrow}, \mathbf{F}_t^{\leftarrow}]))$

Fusion:
$[\alpha_s, \alpha_t] \leftarrow \text{Softmax}(\mathbf{w})$ ▷ learnable attention weights
$\mathbf{f}_{\text{fused}} \leftarrow \text{MLP}(\alpha_s \cdot \mathbf{f}_s + \alpha_t \cdot \mathbf{f}_t)$

Classification:
$\mathbf{y} \leftarrow \text{Classifier}(\mathbf{f}_{\text{fused}})$
return $\mathbf{y}$

Baseline Methods

Four baseline methods provide comparative evaluation across traditional machine learning and contemporary deep learning paradigms. All models employ stratified five-fold cross-validation with source-based splitting to prevent spatial leakage. Training protocols maintain consistency: batch size 2, AdamW optimizer with weight decay 10^{-3}, maximum 100 epochs, random seed 42, and weighted cross-entropy loss with class-balanced weights.

Partial Least Squares Discriminant Analysis (PLS-DA) reduces dimensionality through partial least squares regression to 30 latent components followed by linear discriminant analysis [19]. Hyperspectral sequences were preprocessed via spatio-temporal averaging across eight timepoints, producing 204-band mean spectra per sample. Patch-based extraction randomly sampled 10 spatial patches (16×16 pixels) per timepoint, yielding 80 patches per sample. Features underwent spatial averaging and spectral flattening before projection. Training employed stratified five-fold cross-validation with balanced class weighting. The model contains 30 trainable parameters and completes training in under 0.05 hours per fold.

Three-Dimensional Convolutional Neural Network (3D CNN) processes spatial-spectral information through volumetric convolutions applied independently to each timepoint [20, 22]. The architecture comprises three 3D convolutional blocks with progressively increasing filters (32, 64, 128), each followed by batch normalization [39], ReLU activation, and 2×2×2 max pooling. Per-timepoint features are concatenated along the temporal dimension, processed through global average pooling, and fed to a two-layer classification head with dropout (p=0.3) [40]. Training employs a learning rate 3×10^{-5} and early stopping with patience 25 epochs monitoring validation F1-macro. The model contains 5.63 million parameters and requires 9 hours of average training per fold.

Vision Transformer (ViT) divides 128×128 spatial frames into 16×16 patches (64 patches per time point), projecting flattened patch vectors to 96-dimensional embeddings with learned positional encodings [30, 31]. Eight transformer encoder layers with 8-head self-attention and 384-dimensional feedforward networks process each time point independently through shared parameters [23]. Class token representations from all time points form an 8-step sequence processed by temporal attention (4 heads) with softmax-normalized weighted aggregation. A three-layer MLP head (96→96→48→2) with GELU activation [41] and dropout (p=0.1) produces final predictions. Training uses a learning rate 3×10^{-5}, cosine annealing with 10-epoch linear warmup, and early stopping with patience 30 epochs. The model contains 5.97 million parameters and requires 8 hours of training per fold.

One-Dimensional CNN with LSTM (1D CNN-LSTM) decouples spatial-spectral extraction from temporal modeling through hybrid architecture [27]. Three 1D convolutional layers along the spectral axis (kernel sizes 7, 5, 3; channels 1→32→64→96; stride 2) with batch normalization, ReLU and dropout (p=0.3) extract 96-dimensional features per spatial location. Spatial average pooling yields per-timepoint representations. A two-layer bidirectional LSTM (96 hidden units per direction, dropout p=0.3) [28] models the eight-timepoint sequence. Concatenated final hidden states (192 dimensions) feed a classification head with layer normalization [42] and two fully connected layers (192→96→2) using GELU activation and dropout (p=0.5). Training employs a learning rate 1×10^{-4} and early stopping with patience of 25 epochs. The model contains 0.22 million parameters and requires 26.46 hours of training per fold due to sequential processing overhead.

Ablation Study

To verify the necessity of the dual-path architectural design, an ablation study was conducted using stratified five-fold cross-validation on the complete dataset. Three architectural variants were evaluated: Spectral-only, processing 204-band sequences through bidirectional Mamba blocks without temporal modeling; Temporal-only, applying bidirectional temporal Mamba after spatial compression; and the complete Dual-Path architecture with adaptive fusion. All variants used identical training protocols with batch size 4 and hidden dimension 32. Models trained for 50 epochs with early stopping monitoring F1-macro score with 15-epoch patience. The AdamW optimizer with learning rate 1×10^{-4} and weight decay 0.01 optimized parameters. Cosine annealing learning rate scheduling with T_max=50 facilitated stable convergence. Focal loss (γ=2.0) with label smoothing (ε=0.1) and class-weighted sampling addressed severe class imbalance (44 normal vs 158 diseased samples). Mixed precision training with gradient clipping (max norm 1.0) improved computational efficiency. The Spectral-only variant isolated biochemical correlation patterns by processing spectral tokens through unidirectional Mamba with spatial pooling to 2×2 resolution, reducing spatial dimensions before band-wise sequential processing. The Temporal-only variant compressed spatial-spectral information through adaptive pooling then applied bidirectional temporal Mamba (forward t=1→8, backward t=8→1) to model physiological degradation trajectories. The Dual-Path variant combined both encoders through gated fusion where a learned gate mechanism (sigmoid-activated linear projection) dynamically weighted spectral and temporal features before MLP refinement.

Implementation Details

All models were implemented in Python 3.9 using PyTorch 2.1.0 with CUDA 11.8 support. Experiments were conducted on Penn State University's Roar supercomputer using standard compute nodes, each equipped with a single NVIDIA Tesla V100 GPU (32GB VRAM) and 128GB system memory. The mamba-ssm library (version 1.2.0) provided state space model implementations for the proposed architecture.

Data preprocessing and augmentation employed NumPy 1.24 and H5py 3.9 for efficient hyperspectral data handling. Scikit-learn 1.3 provided PLS-DA implementation and stratified cross-validation utilities. Training utilized mixed precision computation (torch.cuda.amp) for memory efficiency and gradient clipping (max norm 1.0) for stability. All experiments used fixed random seed 42 for reproducibility across data partitioning, weight initialization, and stochastic operations.

Performance metrics including accuracy, precision, recall, F1-macro, sensitivity, specificity, and AUC-ROC were computed using scikit-learn. Statistical significance testing employed paired t-tests with Bonferroni correction for multiple comparisons. Confusion matrices and ROC curves were generated for qualitative assessment. Model checkpoints saved the best-performing state based on validation F1-macro score within each fold. Training logs recorded per-epoch loss, accuracy, and learning rate for monitoring convergence.

Acknowledgements

This research was supported by the Pennsylvania Department of Agriculture, USA.

Declarations

Author contributions statement

W.W. conceived the experiment(s), W.W. conducted the experiment(s), W.W. analysed the results. S.G. supervised the project and reviewed the manuscript. All authors reviewed the manuscript.

Additional information

Competing interests: The authors declare no competing interests.

Data availability: The hyperspectral imaging datasets and trained model weights generated during this study are available from the corresponding author upon reasonable request. The preprocessing code and model implementation will be made publicly available upon acceptance of the manuscript.

References

[1] Zhao, F. *et al.* Mhs-mamba: A multihierarchical semantic model for uav hyperspectral image classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **18**, 24617–24631 (2025).

[2] Lv, P., Cheng, P., Shan, F., Cao, Y. & Zhong, Y. Fremamba: A frequency-domain mamba model for hyperspectral image change detection. *IEEE Transactions on Geoscience and Remote Sensing* **63**, 5536313 (2025).

[3] Mahlein, A. K. Plant disease detection by imaging sensors – parallels and specific demands for precision agriculture and plant phenotyping. *Plant Disease* **100**, 241–251 (2016).

[4] Calamita, F., Imran, H. A., Vescovo, L., Mekhalfi, M. L. & La Porta, N. Early identification of root rot disease by using hyperspectral reflectance: The case of pathosystem grapevine/armillaria. *Remote Sensing* **13**, 2436 (2021).

[5] Calderón, R., Navas-Cortés, J. A. & Zarco-Tejada, P. J. Early detection and quantification of verticillium wilt in olive using hyperspectral and thermal imagery over large areas. *Remote Sensing* **7**, 5584–5610 (2015).

[6] Cen, Y., Huang, Y., Hu, S., Zhang, L. & Zhang, J. Early detection of bacterial wilt in tomato with portable hyperspectral spectrometer. *Remote Sensing* **14**, 2882 (2022).

[7] Bravo-Arrepol, M., Sanfuentes, E., Amigo, J. M. & Almonacid, L. Early detection of fusarium circinatum in pinus radiata cuttings using vis-nir hyperspectral imaging and machine learning classification. *Plant Disease* **110**, 122–130 (2026).

[8] Castro-Valdecantos, P. *et al.* Detection of fusarium wilt-induced physiological impairment in strawberry plants using hyperspectral imaging. *Biosystems Engineering* **242**, 50–64 (2024).

[9] Cao, Y. *et al.* Detecting asymptomatic infections of rice bacterial leaf blight using hyperspectral imaging and 3-dimensional convolutional neural network with spectral dilated convolution. *Frontiers in Plant Science* **13**, 963170 (2022).

[10] Gontijo da Cunha, V. A. *et al.* Early detection of tomato bacterial spot disease in transplant tomato seedlings utilising proximal hyperspectral imaging. *Precision Agriculture* **24**, 2380–2402 (2023).

[11] Gu, Q. *et al.* Early detection of tomato spotted wilt virus infection in tobacco using the hyperspectral imaging technique and machine learning algorithms. *Computers and Electronics in Agriculture* **167**, 105066 (2019).

[12] Chadoulis, R.-T., Livieratos, I., Manakos, I. & Kalaitzidis, C. 3d-cnn detection of systemic symptoms induced by different potexvirus infections in four nicotiana species. *Frontiers in Plant Science* **15**, 1384923 (2025).

[13] Chen, Y. X. *et al.* Ssmcd: Spatial-spectral mamba for change detection in multispectral imagery. *Knowledge-Based Systems* **314**, 113236 (2025).

[14] Ban, S. *et al.* Evaluation and early detection of downy mildew of lettuce using hyperspectral imagery. *Agriculture (Switzerland)* **15**, 16 (2025).

[15] Arens, N. *et al.* Non-invasive presymptomatic detection of cercospora beticola infection and identification of early metabolic responses in sugar beet. *Frontiers in Plant Science* **7**, 1377 (2016).

[16] Brown, M. T., Oh, S. & Rainey, K. M. Presymptomatic leaf reflectance of fusarium virguliforme-infected soybean plants in greenhouse conditions. *Phytofrontiers* **4**, 270–280 (2024).

[17] Bi, J. *et al.* Spectral-temporal-spatial capsnet (sts-capsnet): A biology-informed network for continuous detection of wheat canopy yellow rust. *Computers and Electronics in Agriculture* **228**, 109626 (2025).

[18] Zhu, M. *et al.* Self-supervised mamba for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing* **63**, 5531312 (2025).

[19] Barker, M. & Rayens, W. Partial least squares for discrimination. *Journal of Chemometrics* **17**, 166–173 (2003).

[20] He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770–778 (2016).

[21] Roy, S. K., Krishna, G., Dubey, S. R. & Chaudhuri, B. B. Hybridsn: Exploring 3-d–2-d cnn feature hierarchy for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters* **17**, 277–281 (2020).

[22] Ben Hamida, A., Benoit, A., Lambert, P. & Amar, C. B. 3-d deep learning approach for remote sensing image classification. *IEEE Transactions on Geoscience and Remote Sensing* **56**, 4420–4434 (2018).

[23] Hong, D. *et al.* Spectralformer: Rethinking hyperspectral image classification with transformers. *IEEE Transactions on Geoscience and Remote Sensing* **60**, 5518615 (2022).

[24] Bao, D. *et al.* Early detection of sugarcane smut and mosaic diseases via hyperspectral imaging and spectral-spatial attention network. *Computers and Electronics in Agriculture* **225**, 109334 (2024).

[25] Cui, R. *et al.* Hyperspectral imaging coupled with dual-channel convolutional neural network for early detection of apple valsa canker. *Computers and Electronics in Agriculture* **202**, 107411 (2022).

[26] Shi, H., Luo, Z., Dai, X., Ma, Y. & Zhu, G. Spectral-spatial graph transformer network for hyperspectral image classification under limited supervision. *Remote Sensing* **18**, 199 (2025).

[27] Mou, L., Ghamisi, P. & Zhu, X. X. Deep recurrent neural networks for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing* **55**, 3639–3655 (2017).

[28] Russwurm, M. & Körner, M. Multi-temporal land cover classification with sequential recurrent encoders. *ISPRS International Journal of Geo-Information* **7**, 129 (2018).

[29] Hochreiter, S. & Schmidhuber, J. Long short-term memory. *Neural Computation* **9**, 1735–1780 (1997).

[30] Dosovitskiy, A. *et al.* An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations* (2021).

[31] Vaswani, A. *et al.* Attention is all you need. In *Advances in Neural Information Processing Systems*, vol. 30, 5998–6008 (2017).

[32] Roy, S. K. *et al.* Multimodal fusion transformer for remote sensing image classification. *IEEE Transactions on Geoscience and Remote Sensing* **61**, 5515620 (2023).

[33] Sun, L., Zhao, G., Zheng, Y. & Wu, Z. Spectral–spatial feature tokenization transformer for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing* **60**, 5522214 (2022).

[34] Ahmad, M. *et al.* Hsmamba: A new spectral-spatial state space model for hyperspectral image classification. *arXiv preprint arXiv:2502.06427* (2025).

[35] Gu, A. *et al.* Combining recurrent, convolutional, and continuous-time models with linear state space layers. In *Advances in Neural Information Processing Systems*, vol. 34, 572–585 (2021).

[36] Gu, A., Goel, K. & Ré, C. Efficiently modeling long sequences with structured state spaces. In *International Conference on Learning Representations* (2022).

[37] Gu, A. & Dao, T. Mamba: Linear-time sequence modeling with selective state spaces. *arXiv preprint arXiv:2312.00752* (2024).

[38] Russwurm, M. & Körner, M. Self-attention for raw optical satellite time series classification. *ISPRS Journal of Photogrammetry and Remote Sensing* **169**, 421–435 (2020).

[39] Ioffe, S. & Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International Conference on Machine Learning*, 448–456 (2015).

[40] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research* **15**, 1929–1958 (2014).

[41] Hendrycks, D. & Gimpel, K. Gaussian error linear units (gelus). *arXiv preprint arXiv:1606.08415* (2016).

[42] Ba, J. L., Kiros, J. R. & Hinton, G. E. Layer normalization. *arXiv preprint arXiv:1607.06450* (2016).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- references.zip
- PSUMambaDualPathBidirectionalMambaforPlantStressMonitoringviaTemporalHyperspectralImaging.zip