

<https://doi.org/10.1038/s40494-025-02102-1>

Mold spot detection for paper artifacts based on multimodal feature fusion



Xuexu Deng¹, Ya Zhao^{2,3} ✉, Dan Qin⁴ ✉, Zheng Ma⁴, Zhongyu Xiao¹, Xiling Luo¹, mingfu Zhao¹, Tao Song¹, Jianxu Wang¹, Bin Tang¹ & Huan Tang⁴

Chinese paper-based cultural artifacts are threatened by mold deterioration, necessitating precise mold identification for effective conservation. This study introduces a hyperspectral imaging-based triplePath multimodal feature fusion network (TPMFN) to improve the sensitivity and timeliness of traditional methods. It combines spatial and spectral features using an advanced 2-D CNN with spatial attention for RGB details, a SpectralFormer for analyzing 400–1000 nm hyperspectral data, and a hybridCNN for extracting spectral-spatial features. Various cross-modal fusions (concatenation, addition, multiplication) facilitate efficient integration and classification of multi-source data. Testing six mold types on paper artifacts, TPMFN significantly outperforms SVM, 1-D CNN, 2-D CNN, SSFTT, HybridSN, and SpectralFormer on OA (98.78%), AA (98.52%), and Kappa (98.21%), with a 3.83% improvement in detail recognition. Ablation studies confirmed a 7.5% feature discriminability enhancement from the dual attention mechanism. This research provides a high-precision, practical solution for paper artifact mildew detection, crucial for establishing preventative conservation.

As vital testaments to the brilliance of Chinese civilization, paper-based cultural heritage artifacts have significantly influenced intercultural dialogue and fostered the development of global civilizations. Ranging from classical paintings and canonical texts to personal correspondence and archival materials, these artifacts provide indispensable research resources owing to both their rich content and sophisticated papermaking craftsmanship^{1–3}. However, the inherent hygroscopic nature of paper makes it vulnerable to biodegradation and environmental degradation during storage and exhibition, thus increasing the challenges in preventive conservation. Owing to its high cellulose, hemicellulose, and nutrient content, paper is particularly susceptible to mold growth. Molds secrete cellulases to break down and absorb these constituents. This process progressively degrades the paper's structure, weakening the inter-fiber bonds and consequently diminishing its mechanical strength^{4,5}. This biodegradation process can also involve a series of intricate chemical reactions, which further contribute to the aging and embrittlement of the paper, ultimately leading to significant damage to the historical and artistic value of the cultural artifacts^{6–8}. Existing mold remediation strategies for cultural heritage are predominantly classified into three categories: mechanical, physical, and biochemical methods⁹. Mechanical interventions utilize instruments such as soft brushes, conservation scalpels, and HEPA-filtered vacuum systems to dislodge microbial colonization; however, complete microbial eradication is generally not achievable through

these methods alone¹⁰. Physical methods, such as ultraviolet (UV) irradiation and laser ablation, have shown efficacy in microbial decontamination; however, their effectiveness is limited against many bacterial species and dematiaceous fungi¹¹. Although biochemical methods demonstrate greater potential, the taxonomic diversity of molds necessitates species-level identification to develop targeted conservation strategies that mitigate secondary damage, as singular therapeutic interventions often fail to achieve complete eradication¹². Consequently, the accurate identification of fungal species is not only a fundamental prerequisite for effective mold remediation but, more importantly, a critical step in protecting paper-based cultural heritage from persistent fungal colonization. Existing conventional detection methodologies for paper-based artifacts include morphological characterization, molecular biology techniques¹³, and biochemical assays¹⁴. However, these methodologies rely exclusively on individual technological modalities, such as image-based or spectroscopic techniques, which hinders comprehensive mycological identification and consequently reveals inherent limitations. Thus, addressing the limitations of current individual technologies for the comprehensive identification of mold on paper-based cultural heritage to achieve more efficient and accurate detection remains a critical research challenge.

Hyperspectral imaging (HSI) technology, an advanced detection method that integrates both spectral and spatial information, has

¹Chongqing University of Technology, Chongqing Key Laboratory of Fiber Optic Sensing and Photoelectric Detection, Chongqing, China. ²School of Economics and Management, Chongqing, China. ³College of Tourism and Service Management, Chongqing Second Normal University, Chongqing, China. ⁴Research on Harmful Biological Control of Cultural Relics in Collections Key Scientific Research Base of the State Administration of Cultural Heritage (Chongqing China Three Gorges Museum), Chongqing, China. ✉e-mail: zhaoya0122@outlook.com; qindan_cctgm@163.com

demonstrated significant potential for mold detection due to its high sensitivity, capacity for multi-dimensional data acquisition, and non-destructive nature. Currently, this technology is predominantly employed in areas such as food quality assessment^{15–17} and crop disease detection^{18–20}, with initial advancements being made in its application for the identification and assessment of mold contamination within cultural heritage conservation. For instance, Lu et al.²¹ proposed an automated labeling method based on hyperspectral imagery for detecting mold damage in murals. This approach effectively addressed the limitations of traditional manual labeling methods, such as time consumption, subjectivity, and inconsistency. By integrating the spatial and spectral data from hyperspectral images, accurate identification and labeling of mold on mural surfaces were achieved, demonstrating significant practical value and research importance. Williams et al.²² conducted a study on the application of machine learning in conjunction with hyperspectral imaging for the non-invasive detection of aflatoxin contamination in pistachios. The results demonstrated that a residual network model attained high detection accuracy, suggesting the promising potential of this technology for aflatoxin detection. Ou et al.²³ employed hyperspectral imaging in conjunction with spectral fingerprints, vegetation indices, and various multi-dimensional features, along with machine learning techniques, to facilitate the early detection of gray mold on strawberry leaves. The results demonstrated that a convolutional neural network (CNN) model, which utilized fused features, exhibited superior performance, attaining a classification accuracy of 96.6%. This highlights the efficacy of fusion-based models in diminishing the dimensionality of classification data while simultaneously enhancing the predictive accuracy and precision of classification algorithms. Dai et al.²⁴ systematically investigated the spectral characteristics of simulated foxing on paper artifacts using hyperspectral imaging technology. By employing band arithmetic and the minimum noise fraction (MNF) method, the capacity to extract and differentiate features of the affected areas was effectively enhanced. Furthermore, a discriminative model, based on the K-nearest neighbor algorithm and backpropagation (BP) neural network, was developed, achieving a notable accuracy of over 79% in foxing detection. The aforementioned research offers significant technical support for the precise detection and scientific management of mold contamination, and further broadens the application potential of hyperspectral imaging technology within cultural heritage conservation. Although the aforementioned studies provide significant technical support for the accurate detection and scientific management of mold contamination, and highlight the potential application of hyperspectral imaging in cultural heritage conservation, the use of this technology for mold detection on paper-based artifacts remain in its early stages, with a particular deficiency in targeted and efficient methodologies.

Addressing a significant research gap in targeted and efficient methodologies for detecting mold on paper-based cultural heritage, this study introduces an innovative approach by integrating digital and hyperspectral image information within a multimodal feature fusion framework underpinned by machine learning principles. Furthermore, it strategically employs distinct feature extraction techniques specifically selected to address the varying morphological characteristics of the mold. Through a rigorous evaluation and validation of the influence of diverse feature extraction methods on mold identification, coupled with a comparative analysis of their accuracy and robustness, this research aims to determine the optimal detection model. This will address existing technological limitations in achieving efficient detection and further expand the application of multimodal feature fusion technology in cultural heritage conservation.

The subsequent structure of this paper is as follows: The “Methods” section provides a detailed account of the proposed mold detection methodology, sample preparation, and experimental setup. The “Results” section demonstrates the effectiveness of the proposed method through its application to simulated mold infestation samples and offers a comprehensive analysis of the experimental findings. Finally, the “Discussion” section concludes by summarizing the key contributions of this work and outlining potential directions for future research.

Methods

In this section, the paper first delineates the proposed TPMFN model and provides a detailed analysis of the feature extraction design logic for its three distinct pathways. Subsequently, the preparation process, acquisition system, and experimental setup of the simulated mold samples are described. Following data block division, the input hyperspectral image features are represented as $X \in \mathbb{R}^{B \times H \times W \times C}$, where B denotes the batch size, H and W denote the spatial dimension, and C denotes the number of channels, respectively.

TPMFN

The proposed model aims to enhance the accuracy and robustness of mold stain detection through the effective fusion of multimodal features. The core design principle of the TPMFN model is to fully leverage the distinctive characteristics of three different modalities for feature extraction: the spectral dimension of hyperspectral data, the joint spatial-spectral dimension, and the spatial dimension of RGB images. By employing a deep fusion mechanism, the model achieves a comprehensive representation of global features. The network architecture is illustrated in Fig. 1.

In the TPMFN model, the input hyperspectral image undergoes an initial processing stage to extract three RGB bands, thereby generating an RGB digital image. Subsequently, distinct feature extraction techniques are applied to the different modalities. For the RGB digital modality, a two-dimensional convolutional neural network (2-D CNN) is employed to capture spatial characteristics such as edge morphology and color variations in the generated pseudo-color image. Following this, a spatial attention module is incorporated to enhance the representation of spatial textures within the mold stain region while mitigating interference from non-mold stain areas. For the hyperspectral data modality, feature extraction proceeds through two distinct pathways: In the spectral feature extraction pathway, a SpectralFormer is initially employed to capture subtle spectral variations indicative of mold stains across the hyperspectral bands. Subsequently, Spectral Attention is applied to emphasize salient spectral bands, thereby generating spectral features. For spatial-spectral feature extraction, a Hybrid Convolutional Network is utilized to jointly analyze the spatial distribution and spectral characteristics of mold stains within the hyperspectral data.

Spatial features

The 2-D CNN is a deep learning model specifically designed for processing two-dimensional data. Its core mechanism employs convolutional operations for local feature extraction, where parameter-shared kernels capture low-level features (e.g., edges, textures, structures) within localized spatial regions, while progressively learning higher-level semantic patterns through hierarchical layers. Initially, the three RGB bands are extracted and flattened from X. Subsequently, a three-layer two-dimensional convolutional network hierarchically extracts spatial features from the hyperspectral data. A spatial attention mechanism is then applied to the features extracted by each convolutional layer to enhance the representation of key spatial locations. Finally, the resulting output is flattened, and its dimensionality is progressively reduced using a three-layer fully connected network, ultimately yielding the extracted spatial features. The spatial attention mechanism enhances salient regions by computing spatial weight distributions, with the attention energy defined in Eq. (1):

$$E = Q \cdot K \tag{1}$$

In Eq. (1), Q is the query matrix, and K is the key matrix.

Subsequently, a softmax operation is applied to each row of E to derive the attention weight A. The formula for calculating the attention weight is provided in Eq. (2):

$$A_{ij} = \frac{\exp(E_{ij})}{\sum_{k=1}^{H \cdot W} \exp(E_{ik})}, \forall i, j \in [1, H \cdot W] \tag{2}$$

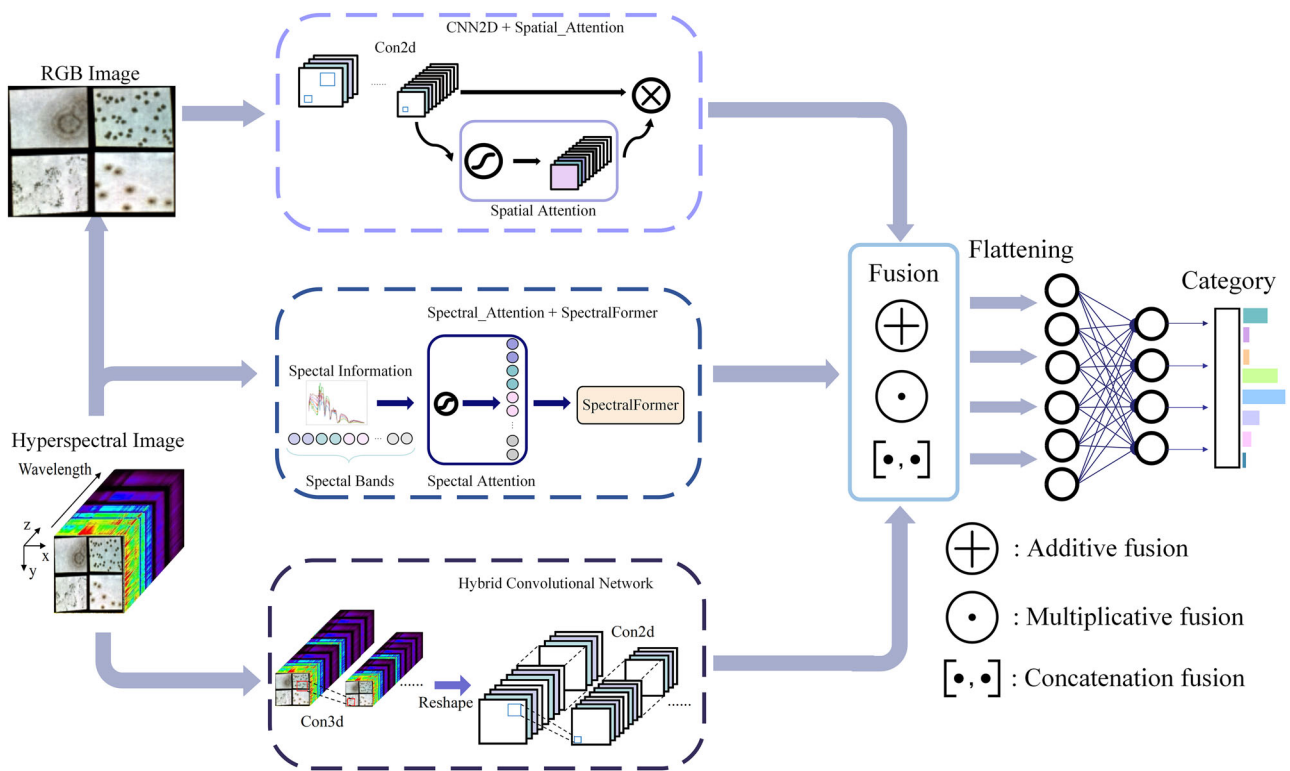


Fig. 1 | TPMFN network architecture, including three data fusion methods.

In Eq. (2), A_{ij} represents the weight distribution of the pixel spatial position at the i -th row and j -th column.

The specific network architecture of the spatial feature extraction algorithm is illustrated in Fig. 2.

Spectral features

SpectralFormer is an enhanced Transformer-based model designed to enable local spectral representations from multiple adjacent bands at each encoding position. This is achieved through the Group Smart Spectral Embedding (GSE) mechanism, which enhances the capture of subtle spectral variations, and the Cross-Layer Adaptive Fusion (CAF) mechanism, which improves the transfer of information between layers. Furthermore, SpectralFormer incorporates a cross-layer skip connection that adaptively learns to fuse the residuals, gradually propagating memory-like components from shallow to deep layers. While SpectralFormer excels at capturing global sequential information, it exhibits limitations in effectively modeling the local contextual information inherent in hyperspectral data. To address this specific limitation, this paper introduces an enhanced spectral feature extraction model based on SpectralFormer.

Its architecture is illustrated in Fig. 3. Initially, the spectral attention mechanism is incorporated, where global average pooling is performed to obtain a global representation of the spectral channels. Subsequently, it assigns weights to the spectral bands to emphasize the significant ones, followed by the application of linear projection for spectral embedding. Following this, classification tokens and positional embeddings are incorporated, followed by sequence modeling utilizing the Transformer module, and finally, the classification results are generated. The input to this process is the extracted spectral information $X \in \mathbb{R}^{B \times C \times N}$, where N represents the number of spectral features. The spectral attention process is formulated as shown in Eq. (3):

$$X' = X \cdot \text{softmax}(W_2 \cdot \sigma(W_1 \cdot z^T + b_1) + b_2) \tag{3}$$

In Eq. (3), d represents the hidden dimension, $W_1 \in \mathbb{R}^{d \times C}$, $W_2 \in \mathbb{R}^{C \times d}$, σ denotes the sigmoid activation function, b_1 and b_2 are bias parameters.

Spatial-spectral features

The hybrid convolutional network (hybridCNN) employs multi-dimensional feature extraction by applying 3D convolution to concurrently extract spectral and spatial features from hyperspectral data, while utilizing 2D convolution to further refine spatial feature extraction. The network then progressively fuses features across its layers, integrating high-dimensional spectral features into lower-dimensional representations and employing 2D convolution operations to aggregate and optimize spatial feature representations.

This architecture exhibits notable advantages in effectively integrating multi-dimensional features from hyperspectral data, thereby enhancing both the accuracy and robustness of classification tasks. By implementing a hierarchical feature extraction strategy, it optimizes computational resource allocation to address the inherent challenges of complexity and redundancy in hyperspectral data processing. The detailed architecture of the hybrid convolutional network model utilized in this study is illustrated in Fig. 4.

Simulated mold infestation samples

To ensure the diversity and accuracy of the experimental data, this study selected raw cotton-linen Xuan paper, cut into 5×5 cm specimens, as the substrate for simulating mold contamination. Based on a review of literature from the past five years concerning common mold species found in the preservation environments of paper-based cultural artifacts^{12,25}, six representative mold species from different genera were selected as experimental subjects, namely *Aspergillus niger*, *Penicillium citrinum*, *Trichoderma longibrachiatum*, *Alternaria alternata*, *Paecilomyces lilacinus*, and *Cladosporium cladosporioides*. These six mold species, representing distinct genera, underscore both the biodiversity and ecological adaptability characteristic of molds. Widely distributed in natural environments, they possess the

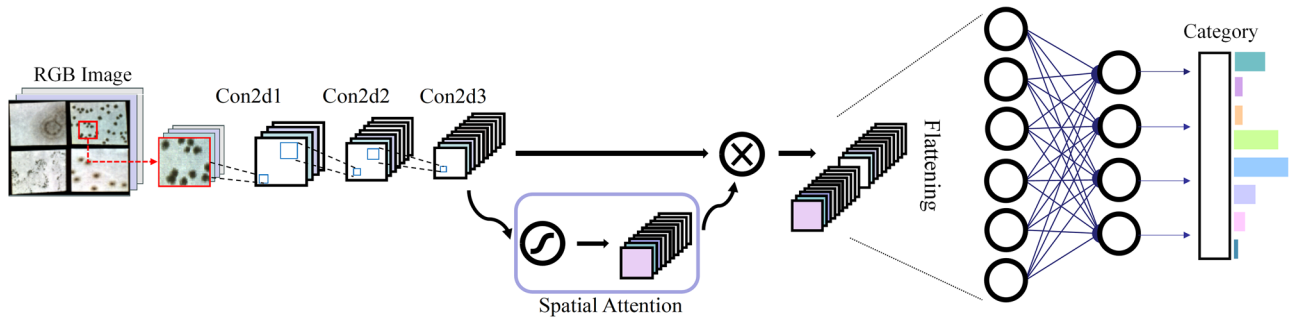


Fig. 2 | The network structure diagram of the spatial feature extraction model.

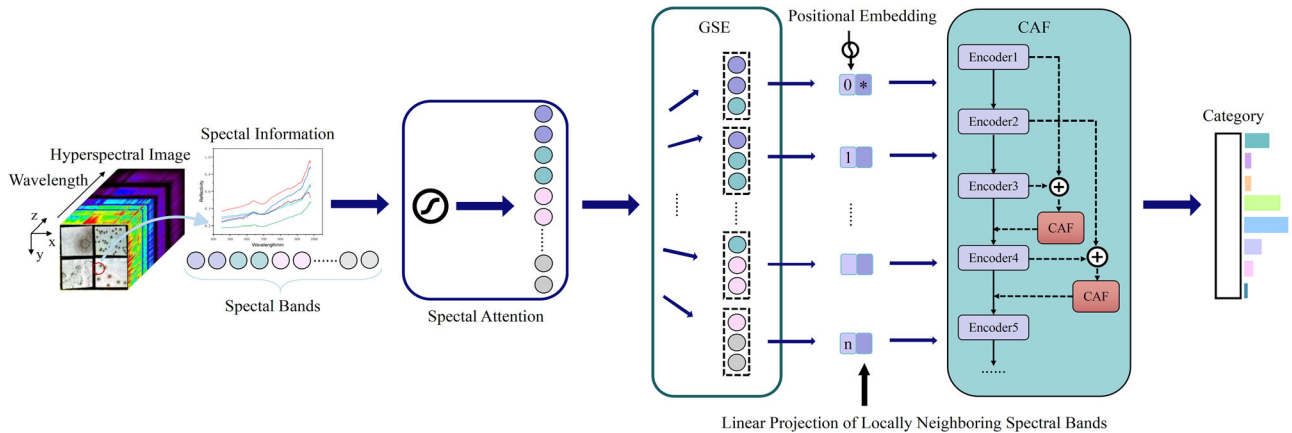


Fig. 3 | Network structure diagram of the spectral feature extraction model.

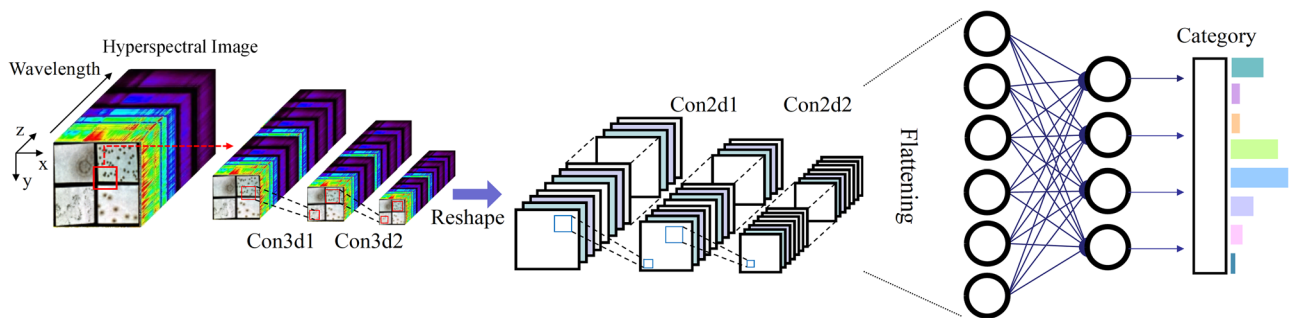


Fig. 4 | The network structure of the spatial-spectral feature extraction model.

potential to induce deterioration in cultural artifacts. Consequently, prioritizing the monitoring of these genera is essential during mold detection and conservation processes.

In the mold inoculation experiment, raw Xuan paper was initially cut to the desired size and sterilized. Subsequently, the sterilized paper sheets were placed in sterile Petri dishes and inoculated by spraying with a spore suspension of a specific concentration and a nutrient solution. To promote mold growth, the inoculated paper artifact specimens were cultured in an artificial climate chamber maintained at a temperature of 28 °C, a relative humidity of 80% RH, and under dark conditions for 7 days. During this period, the specimens were observed daily at regular intervals, and sterile water was replenished as needed to maintain appropriate paper moisture. The artifact paper and mold strains required for this experiment were both provided by the State Administration of Cultural Heritage Key Scientific Research Base for Research on the

Control of Harmful Organisms in Collection Artifacts. A photograph illustrating the simulated mold-infested specimens is depicted in Fig. 5. In this context, the ground truth images, considered the authentic and accurate reference images for machine learning purposes, were annotated using ENVI 5.6 software and subjected to manual correction to ensure data accuracy.

Mold stain acquisition system for paper-based cultural artifacts

In this study, a hyperspectral image acquisition system specifically designed for analyzing mold stains on paper-based cultural artifacts was constructed, as depicted in Fig. 6. The system comprises an iSpecHyper-VS1000 portable hyperspectral imager from Lyson Optical, two halogen light sources with adjustable brightness and color temperature, and a Canon RF 24 mm focal length lens. The detailed specifications of the hyperspectral imager are presented in Table 1. Following multiple imaging tests, the optimal

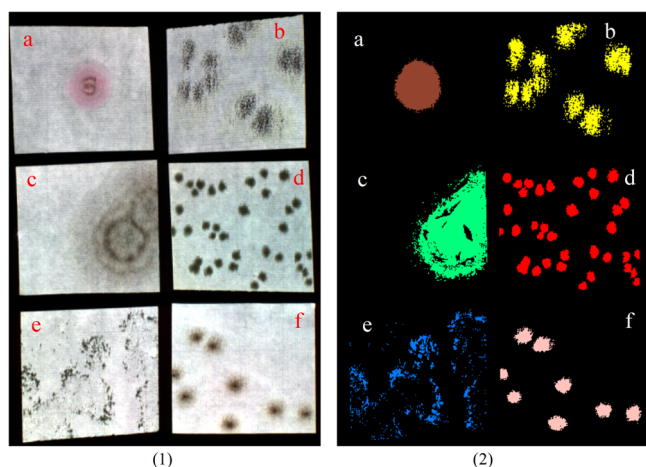


Fig. 5 | Images of simulated mold spots. (1) False-color images of the following species: a *Paecilomyces lilacinus*; b *Aspergillus niger*; c *Alternaria alternata*; d *Penicillium citrinum*; e *Trichoderma longibrachiatum*; f *Cladosporium cladosporioides*. (2) Corresponding ground truth images for each species.

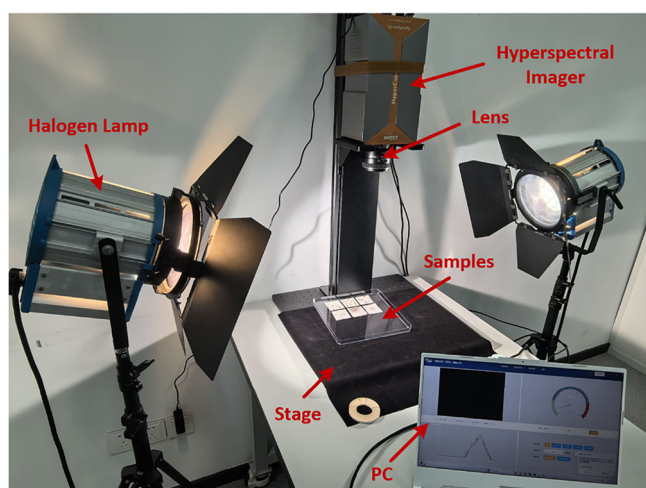


Fig. 6 | Mold stain image acquisition system.

Table 1 | iSpecHyper-VS1000 hyperspectral imager parameters

Main Technical Specifications	Specification Parameters
Spectral Range	400 nm–1000 nm
Spectral Resolution	<3nm
F-number	F/2.6
Imaging method	Pushbroom Imaging
Number of spectral channels	300
Detector	COMS/InGaAs(TE Cooled)
AD Dynamic Range	12bits
Frame Rate of Spectral Camera	50fps
Spatial Resolution of Image (Pixels)	1920 × 1200
Detector Pixel Size	5.8'6 μm × 5.86μm

parameters were determined as follows: a focal length of 4615 mm, a frame rate of 16 Hz s⁻¹ and an integration time of 56338μs. Images captured with these parameters exhibited excellent quality, with sharp edges and minimal distortion.

Table 2 | Class-specific mold coverage characteristics and training-test sample allocation

No.	Class	Training	Test
1	<i>Paecilomyces lilacinus</i>	250	2251
2	<i>Aspergillus niger</i>	358	3225
3	<i>Alternaria alternata</i>	652	5867
4	<i>Penicillium citrinum</i>	365	3289
5	<i>Trichoderma longiformis</i>	184	1650
6	<i>Cladosporium cladosporioides</i>	186	1674
Total		1995	17956

Experimental setup

Evaluation metrics. To evaluate the efficacy of the models, this paper quantitatively analyzes the classification results of each model using three commonly employed classification performance assessment metrics: Overall Accuracy (OA), Average Accuracy (AA), and Kappa Coefficient (Kappa). Among them, OA reflects the overall classification accuracy of the model across all categories, AA indicates the mean classification accuracy for each category, and the Kappa coefficient measures the consistency between the classification results and random classification results, thus providing a more comprehensive evaluation of the model’s performance. Meanwhile, to further visualize and compare the classification performance of each model, this paper also conducts qualitative analysis by visualizing classification maps to observe the spatial distribution characteristics and the accuracy of the classification boundaries, thereby comprehensively evaluating the classification performance of the models from both quantitative and qualitative perspectives.

Comparative analysis of models. In this paper, several state-of-the-art models are selected for comparative analysis with the proposed model, including: Support Vector Machine (SVM), 1-D-CNN²⁶, 2-D-CNN²⁷, SSFTT²⁸, SpectralFormer²⁹ and HybridSN³⁰. To ensure the uniformity and validity of the experimental data, a unified preprocessing approach is applied, where the Savitzky-Golay (SG) smoothing filter is used to remove noise, and principal component analysis (PCA) is employed to reduce the number of spectral bands to 30 dimensions. This helps eliminate redundant information and highly similar features, thereby improving recognition accuracy.

Experimental setup. In order to validate the multimodal feature fusion-based classification algorithm for hyperspectral mold detection proposed in this paper, all experiments were conducted using Python 3.12 and implemented in the PyCharm 2024.3.1 integrated development environment. The experiments were performed on a high-performance personal computer equipped with an Intel(R) Core i5-14600KF processor, an NVIDIA GeForce RTX 4070 GPU, and 32 GB of RAM.

During the experiments, all deep learning models employed Cross-Entropy Loss as the optimization objective, in conjunction with the Adam optimizer for parameter updates. Stochastic Gradient Descent (SGD) was used for model training optimization. The learning rate was uniformly set to 0.001, and the number of training epochs was fixed at 50 to ensure model stability and convergence, thereby enabling a reliable evaluation of each model’s performance.

Results

In this section, we first detail the hyperspectral image dataset utilized in our experiments. Subsequently, ablation experiments are conducted to systematically analyze the impacts of three distinct spectral-spatial feature extraction modules on model performance. Finally, the proposed model is benchmarked against established approaches, with both quantitative metrics and visual comparisons employed to evaluate the fungal lesion classification capabilities across different models.

Table 3 | Presents the analysis of the proposed model conducted on this dataset (Suboptimal results)

No.	Component				Indicators		
	2D-CNN Spatial_Att	SpectralFormer	Spectral_Att	HybridCNN	OA(%)	AA(%)	$\kappa \times 100$
1	√	×	×	√	90.31	85.25	87.68
2	√	×	√	√	93.58	91.04	92.83
3	√	√	×	√	92.91	90.19	92.27
4	×	√	√	√	91.01	88.81	89.58
5	√	√	√	×	93.45	90.39	91.46
6	√	√	√	√	97.12	93.24	95.16

Table 4 | Performance analysis of the proposed model on the benchmark dataset (Bold entries denote optimal results)

No.	Fusion Methods	OA(%)	AA(%)	$\kappa \times 100$
1	Additive Fusion	94.69	91.65	93.26
2	Multiplicative Fusion	98.78	98.52	98.21
3	Concatenation Fusion	97.46	96.67	96.81

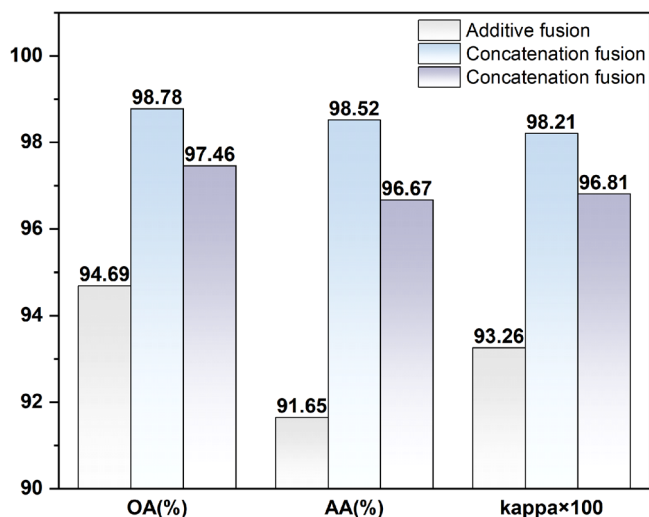


Fig. 7 | Comparative analysis of feature fusion methods.

Data-set description

The hyperspectral image dataset used in this study contains mold spots induced by the six aforementioned fungal infections, acquired at a sampling height of 40 cm. Each image has dimensions of 521 × 364 pixels, containing 19,951 valid pixels per spectral band across 300 spectral channels. The dataset was partitioned into training (10%) and test (90%) subsets. Table 2 details the class nomenclature and corresponding sample distribution between the training and test sets for the classification task.

Ablation study

To thoroughly validate the effectiveness of the proposed method, ablation experiments with various component combinations were conducted on the dataset described in this paper. Six configurations were considered, and the impact of each component on the overall model accuracy was analyzed by evaluating classification performance. All experimental results are presented in Table 3. Specifically, the model was divided into four modules: 2-D-CNN + Spectral_Att, SpectralFormer, Spectral_Att, and HybridCNN. The model output in Case 1 (excluding both SpectralFormer and Spectral_Att) achieved the lowest classification accuracy of 90.31%. In Case 4 (without the 2-D-CNN+Spectral_Att layer), the model’s classification accuracy

performed slightly better than the previous case, with an accuracy of 91.01%. In Case 5 (without the HybridCNN layer), relying solely on the separate processing of spatial and spectral features, the model’s classification accuracy was 93.45%. By comparing the performance of the two models (corresponding to Case 3 and Case 6) with and without the introduction of Spectral_Att to SpectralFormer, a significant increase in accuracy to 97.12% can be observed. This result indicates that Spectral_Att played a positive role in spectral feature processing, thereby effectively enhancing the classification accuracy of the model.

To further corroborate the effectiveness of the proposed algorithm, this study also conducted experiments on the same dataset utilizing different feature fusion methods. A total of three distinct fusion approaches were employed: additive fusion, multiplicative fusion, and concatenation fusion. The comprehensive experimental results are presented in Table 4. As illustrated in Fig. 7, the model employing multiplicative fusion consistently outperformed those utilizing additive fusion and concatenation fusion, achieving a superior accuracy of 98.78%. This indicates that multiplicative fusion possesses enhanced capabilities in feature interaction and information coupling, enabling a more thorough exploitation of the complementary relationships between multimodal features, thereby enhancing classification performance.

Quantitative analysis

Table 5 presents the OA, AA, Kappa coefficient, and per-class classification accuracies obtained by all methods detailed in Section 2. The optimal results are highlighted in bold. The evaluation data clearly demonstrate that the proposed TPMFN method achieves the best performance, yielding the highest OA, AA, and Kappa coefficient values, along with superior classification accuracies for specific categories. For instance, in the case of *Trichoderma longibrachiatum* (category 5), models such as SVM, 1D-CNN, 2D-CNN, SpectralFormer, and SSFTT exhibit limited effectiveness, potentially due to the small sample size and the spatially non-concentrated distribution of this class, which impedes robust feature learning. Furthermore, the percentage-based random sampling strategy may exacerbate existing class imbalance issues. In contrast, TPMFN demonstrates consistently high classification accuracies (above 96%) across all classes, indicating its strong capability in handling class imbalance for hyperspectral image classification tasks. Regardless of whether dealing with dominant or minority classes, the model maintains robust performance. This consistency likely arises from its synergistic multimodal feature fusion mechanism and effective hierarchical feature extraction.

Table 5 | Classification accuracy of different classification methods on the dataset (Bold data indicates the best results in each category)

No.	SVM	1-D-CNN	2-D-CNN	SSFTT	Spectralformer	HybirdSN	TPMFN
1	52.18	56.38	75.66	86.31	99.38	94.65	99.46
2	64.65	54.50	77.13	93.74	98.75	98.51	99.86
3	75.18	83.37	76.54	98.35	98.63	97.98	99.89
4	44.96	32.39	93.16	98.76	95.32	95.95	97.45
5	17.79	24.01	51.40	73.52	58.64	85.33	96.54
6	15.17	52.43	58.85	99.25	98.10	97.28	99.46
OA(%)	45.99	50.85	72.46	91.99	92.80	94.95	98.78
AA(%)	47.85	53.73	70.45	90.85	92.01	93.65	98.52
$\kappa \times 100$	43.92	48.35	71.94	90.26	89.90	93.11	98.21

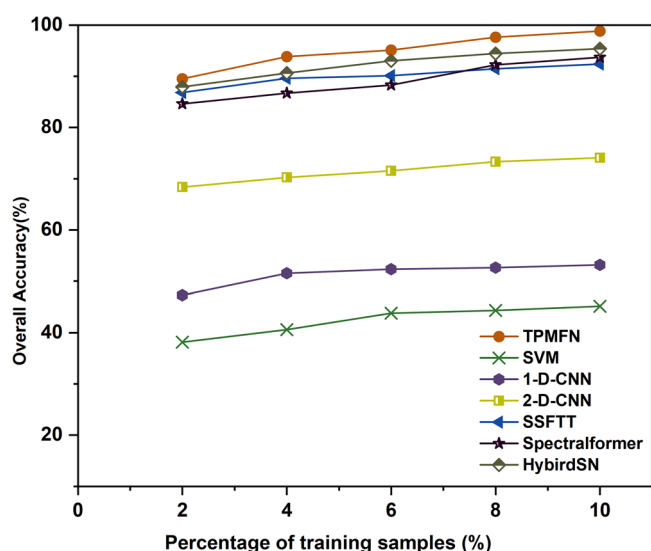


Fig. 8 | Comparison chart of overall accuracy with different proportions of training samples.

However, one exception exists: in the case of *Penicillium citrinum* (category 4), the classification accuracy of SpectralFormer was notably superior to that of TPMFN. The primary reason for this discrepancy is likely attributed to the sample distribution of this category, which is highly concentrated and aggregated across multiple locations, exhibiting a distinctly circular and compact characteristic, whereas the samples in other categories (e.g., category 5) are more dispersed. Consequently, our proposed method did not demonstrate a significant advantage in classifying this specific category. Nevertheless, TPMFN exhibits a notable advantage in datasets characterized by discrete and localized data points, as it can more effectively capture fine-grained local information. We also trained different models using varying proportions of the training datasets, as depicted in Fig. 8. TPMFN maintained robust performance even with a limited number of training samples. As the number of samples increased, the performance of SSFTT and HybridSN was only marginally lower than our method.

Visual evaluation

The classification result maps of the aforementioned comparison methods on this dataset are illustrated in Fig. 9. Visual inspection reveals that the classification map generated by TPMFN is the most refined and exhibits the closest resemblance to the ground truth map. In contrast, methods such as SVM, 1-D CNN, and 2-D CNN appear to

capture only limited spectral or superficial features, resulting in classification maps with substantial noise and significant misclassifications. This indirectly suggests that these models are unable to accurately identify object categories and exhibit suboptimal performance. While SpectralFormer, SSFTT, and HybridSN achieved classification accuracies exceeding 90% for the majority of the data, they still exhibit some misclassifications in complex scenarios, particularly within the regions of *Penicillium citrinum* and *Trichoderma longibrachiatum*. Notably, the method proposed in this paper largely and correctly identified these two mold regions, demonstrating the superior performance of TPMFN in terms of spatial classification.

Discussion

Addressing the challenges of accuracy and efficiency in mold detection on paper-based cultural artifacts, this paper proposes a multimodal feature fusion method based on hyperspectral imaging technology: TPMFN. This method comprehensively leverages spectral features, spatial features, and joint spectral-spatial features, employing a multi-pathway architecture for deep feature extraction and fusion. Specifically, TPMFN analyzes subtle spectral variations through a Spectral Transformer, combines a 2-D CNN with a spatial attention mechanism to capture mold spatial patterns, and simultaneously utilizes a hybrid convolutional network to integrate multi-dimensional features, thereby optimizing feature representation capabilities and enhancing computational efficiency. To further exploit this characteristic, this paper investigates several distinct fusion modules: additive, multiplicative, and concatenation. The experimental results demonstrate that the method employing multiplicative fusion significantly outperforms traditional classification methods and single-pathway networks in terms of precise capture of mold local feature variations, enhanced expression of key information regions, and overall classification performance. TPMFN’s innovative design provides new insights for hyperspectral image processing tasks and offers an efficient and intelligent solution for mold detection in the field of paper-based cultural heritage.

Despite the promising application potential of the TPMFN model, its performance may be constrained by its reliance on hyperspectral imaging technology, which necessitates sophisticated instrumentation with professional-grade parameter configurations and meticulous calibration. Furthermore, while the model aims to enhance computational efficiency, the inherent complexity of integrating spectral, spatial, and joint features within a multi-pathway architecture is anticipated to still result in a significant computational burden. Additionally, the robustness of the model to intrinsic variability in real-world environmental conditions, such as the effects of non-uniform illumination and diverse substrate properties, warrants further in-depth investigation and thorough evaluation.

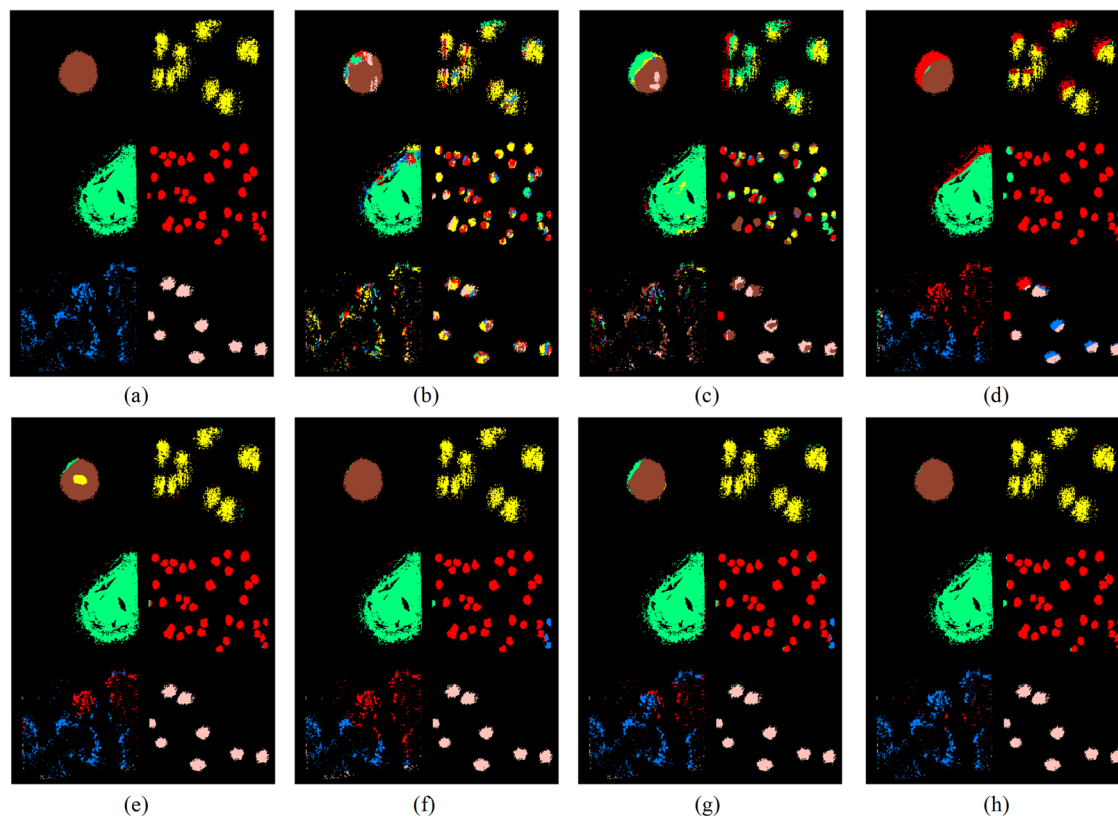


Fig. 9 | Comparison of classification results. **a** Ground truth map. **b** Result from SVM(OA = 45.99%). **c** Result from 1-D-CNN(OA = 50.85%). **d** Result from 2-D-CNN(OA = 72.46%). **e** Result from Spectralformer (OA = 91.99%). **f** Result from

SSFTT(OA = 92.80%). **g** Result from HybridSN (OA = 94.95%). **h** Result from TPMFN (OA = 98.78%).

Data availability

The data in this study is a unique hyperspectral image dataset. Currently, it is not suitable to fully disclose this data. If necessary, the corresponding author can be contacted to discuss possible forms of data sharing.

Received: 24 March 2025; Accepted: 5 October 2025;

Published online: 25 October 2025

References

- Zhang, X., Yan, Y., Yao, J., Jin, S. & Tang, Y. Chemistry directs the conservation of paper cultural relics. *Polym. Degrad. Stab.* **207**, 110256 (2023).
- Wang, S. & Jin, S. Recent research and prospect of deacidifying materials for paper and paper-based cultural relics. *Acta Chim. Sin.* **81**, 309 (2023).
- Yan, Y., Tang, Y. & Yang, Y. Chemical conservation of paper-based cultural heritage. *Molecules* **30**, 122 (2024).
- Carter, H. A. The chemistry of paper preservation: part 2. The yellowing of paper and conservation bleaching. *J. Chem. Educ.* **73**, 1068 (1996).
- Małachowska, E., Dubowik, M., Boruszewski, P. & Przybysz, P. Accelerated ageing of paper: effect of lignin content and humidity on tensile properties. *Herit. Sci.* **9**, 132 (2021).
- Pinheiro, A. C., Sequeira, S. O. & Macedo, M. F. Fungi in archives, libraries, and museums: a review on paper conservation and human health. *Crit. Rev. Microbiol.* **45**, 686–700 (2019).
- Yang, L. Causes of damage to paper artifacts and measures for their protection and restoration *Collection*, 110–112, (2023).
- Sahu, B. B. et al. Preserving cultural heritage: analyzing the antifungal potential of ionic liquids tested in paper restoration. *Plos ONE* **14**, e0219650 (2019).
- Wu, F., Jie, L., Li, R. & He, D. Research progress on the application of fungicides and antibacterial nanomaterials in cultural heritage conservation. *Sci. Conserv. Archaeol.* **35**, 115–127 (2023).
- Rivas, T., Pozo-Antonio, J. S., López de Silanes, M. E., Ramil, A. & López, A. J. Laser versus scalpel cleaning of crustose lichens on granite. *Appl. Surf. Sci.* **440**, 467–476 (2018).
- Pfendler, S. et al. UV-C as an efficient means to combat biofilm formation in show caves: evidence from the La Glacière Cave (France) and laboratory experiments. *Environ. Sci. Pollut. Res.* **24**, 24611–24623 (2017).
- Meng, Q., Li, X., Geng, J., Liu, C. & Ben, S. A biological cleaning agent for removing mold stains from paper artifacts. *Herit. Sci.* **11**, 243 (2023).
- Yan, L., Hong, W. & Lv, X. Isolation and molecular identification of molds on archived paintings and silk artifacts. *Cult. Relics Identif. Apprec.* **10**, 60–65 (2021).
- Wilson, R. A. et al. Identification of toxic mold species through Raman spectroscopy of fungal conidia. *Plos ONE* **15**, e0242361 (2020).
- Vashpanov, Y., Heo, G., Kim, Y., Venkel, T. & Son, J.-Y. Detecting green mold pathogens on lemons using hyperspectral images. *Appl. Sci.* **10**, 1209 (2020).
- Xia, Y. et al. Physicochemical properties and gel quality monitoring of surimi during thermal processing using hyperspectral imaging combined with deep learning. *Food Control* **175**, 111258 (2025).
- Chun, S.-W. et al. Deep learning algorithm development for early detection of *Botrytis cinerea* infected strawberry fruit using hyperspectral fluorescence imaging. *Postharvest Biol. Technol.* **214**, 112918 (2024).
- Zou, Z. et al. Research on nondestructive detection of sweet-waxy corn seed varieties and mildew based on stacked ensemble learning and hyperspectral feature fusion technology. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **322**, 124816 (2024).

19. Zhang, X. et al. A deep learning-based approach for automated yellow rust disease detection from high-resolution hyperspectral UAV Images. *Remote Sens.* **11**, 1554 (2019).
20. Siripatrawan, U. & Makino, Y. Assessment of food safety risk using machine learning-assisted hyperspectral imaging: classification of fungal contamination levels in rice grain. *Microb. Risk Anal.* **27**, 100295 (2024).
21. Lu, M. et al. Research on labeling method for fungal diseases in murals based on hyperspectral images. In *2024 3rd International Conference on Image Processing and Media Computing (ICIPMC)* 93–100 (IEEE, 2024).
22. Williams, L., Shukla, P., Sheikh-Akbari, A., Mahroughi, S. & Mporas, I. Measuring the level of aflatoxin infection in pistachio nuts by applying machine learning techniques to hyperspectral images. *Sensors* **25**, 1548 (2025).
23. Ou, Y., Yan, J., Liang, Z. & Zhang, B. Hyperspectral imaging combined with deep learning for the early detection of strawberry leaf gray mold disease. *Agronomy* **14**, 2694 (2024).
24. Dai, R., Tang, B., Zhao, M., Tang, H. & Liu, H. Study on detection method of foxing on paper artifacts based on hyperspectral imaging technology. *J. Phys. Conf. Ser.* **2010**, 012177 (2021).
25. Li, J. & Li, M. Mold prevention and treatment of textiles and paper cultural relics. *J. Suihua Univ.* **44**, 141–144 (2024).
26. Hu, W., Huang, Y., Wei, L., Zhang, F. & Li, H. Deep convolutional neural networks for hyperspectral image classification. *J. Sens.* **2015**, 1–12 (2015).
27. Zhao, W. & Du, S. Spectral–spatial feature extraction for hyperspectral image classification: a dimension reduction and deep learning approach. *IEEE Trans. Geosci. Remote Sens.* **54**, 4544–4554 (2016).
28. Sun, L., Zhao, G., Zheng, Y. & Wu, Z. Spectral–spatial feature tokenization transformer for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **60**, 1–14 (2022).
29. Hong, D. F. et al. SpectralFormer: rethinking hyperspectral image classification with transformers. *IEEE Trans. Geosci. Remote Sens.* **60**, 1–15 (2021).
30. 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 Geosci. Remote Sens. Lett.* **17**, 277–281 (2020).

Acknowledgements

This research was supported by grants from the Chongqing Talents Program (Grants No. cstc2021ycjh-bgzxm0287), the Self-Supporting Project of Chongqing China Three Gorges Museum (Grants No. 3GM2022-KTZ07), the Science and Technology Research Project of Chongqing Municipal Education Commission (Grants No. KJQN202201110), the

Research and Innovation Team Project of Chongqing University of Technology (Grants No. 2023TDZ014), and the Chongqing Postgraduate Research and Innovation Project (Grants No. CYS25748).

Author contributions

The study's conception and design were collaboratively developed by Xuexu Deng, Ya Zhao, and Dan Qin contributed to the study conception and design. The preparation of materials, as well as data collection and analysis, were preform conducted by Zheng Ma, Zhongyu Xiao, Xiling Luo, Tao Song, Bin Tang, and Jianxu Wang. Mingfu Zhao and Huan Tang are responsible for oversaw the supervision of the research activities. The first draft of the manuscript was written by Xuexu Deng. All authors reviewed and approved the final version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Ya Zhao or Dan Qin.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025