I.J. Wireless and Microwave Technologies, 2026, 5, 191-218
Published Online on October 8, 2026 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijwmt.2026.05.12

SpectroRes-Former: A Spectral–Spatial Residual Transformer with Quantum–Fisher Feature Selection for Hyperspectral Land-Cover Classification

MH. Vahitha Rahman*
Department of Computer Applications, Alagappa University, Karaikudi, India
E-mail: vahithanadeer16@gmail.com
ORCID iD: https://orcid.org/0009-0009-9406-8877
*Corresponding Author

M. Vanitha
Department of Computer Applications, Alagappa University, Karaikudi, India
E-mail: vanitham@alagappauniversity.ac.in
ORCID iD: https://orcid.org/0000-0003-3381-2550

Received: May 28, 2026; Revised: June 18, 2026; Accepted: July 24, 2026; Published: October 8, 2026

Abstract: Proper land-cover and deforestation classification is crucial in environmental protection, sustainable land use, and climate change. Nevertheless, the current deep learning and remote sensing models are usually limited by poor spectral-spatial feature integration, high computational complexity, and poor feature selection, which reduces the classification accuracy. To address these limitations, this paper presents a new model that combines SpectroRes-Former, a SpectralSpatial Residual Transformer Network, and QuFiSel (QuantumFisher Selection), a hybrid feature selection method. The SpectroRes-Former effectively learns both spatial and spectral dependencies with a ResNet encoder, spectral-spatial attention fusion, and Transformer-based contextual learning. At the same time, QuFiSel improves the discrimination of features by integrating Quantum-Inspired Feature Selection (QIFS) and Fisher Score, so that the most informative features are selected to classify. The effectiveness of the approach is tested based on the Indian Pines hyperspectral image, which consists of 220 bands after typical band selection and 16 land cover types. The performance is measured through 5-fold cross-validation and the statistical validation through mean, standard deviation, and paired tests. The obtained average classification accuracy of the proposed method equals 99.21 ± 0.08%, along with the Precision value of 99.08%, Recall 99.16%, F1-score 99.12%, and Matthews Correlation Coefficient 0.991, demonstrating superior results compared to some of the most competitive state-of-the-art techniques. The paired t-test gives $p < 0.05$, which confirms the statistical significance of the result. It is shown that the introduced SpectroRes-Former framework can be used effectively for hyperspectral land cover classification while providing high generalization. The offered solution can serve as a reliable base for smart environmental monitoring systems.

Index Terms: Hyperspectral Image Classification, Indian Pines Dataset, SpectroRes-Former, Quantum–Fisher Feature Selection (QuFiSel), Spectral–Spatial Attention, Transformer Network

1. Introduction

The classification of land-cover and deforestation detection are essential to environmental monitoring, ecosystem management, and sustainable development [1]. Land cover is the physical material on the surface of the earth i.e. forests, urban areas, water bodies and agricultural land whereas deforestation is the massive clearing or degradation of forested areas by human or natural causes [2, 3]. Proper identification and tracking of land-cover changes are essential in the study of ecological changes, evaluation of biodiversity loss, natural resources management, and development of effective climate policies [4, 5]. The high deforestation rate not only increases global warming through the decrease of carbon sequestration but also affects the rainfall patterns, leads to soil erosion, and endangers millions of species that rely on forest ecosystems [6]. Therefore, creating an automated, smart system of accurate land-cover classification and

This work is open access and licensed under the Creative Commons CC BY 4.0 License.

deforestation detection is highly significant to society and the environment [7, 8].

Over the past few years, a number of deep learning-based models including MSTNet, S2EFT, SpectralDiff, and BS2T have been used to overcome the problem of hyperspectral image analysis to classify land-cover [9]. Despite the fact that these techniques have recorded a fair success, they are not without a number of limitations. Most of these models have poor generalization to other land-cover types because they have weak spectral and spatial dependence models [10, 11]. Moreover, they tend to lose high-level contextual knowledge due to shallow network structures or lack of attention-based processes. Moreover, spectral band overlap and inclusion of irrelevant or noisy features often compromise the accuracy of classification and raise the cost of computation [12, 13]. Consequently, the current models are not able to ensure consistency and reliability especially when handling complex and heterogeneous satellite imagery [14].

To tackle these problems, the need for advanced methods that can efficiently merge spectral and spatial data, remove redundant features, and enhance the discriminatory power of feature representations is growing continually [15]. One potential way of achieving these goals is a hybrid of quantum-inspired feature engineering and transformer-based architectures.

In this regard, the present paper suggests a new method of accurate land-cover and deforestation classification, relying on Hyperspectral satellite images [16]. The SpectroRes-Former model integrates improved preprocessing, quantum-inspired feature extraction, and spectral-spatial transformer network to obtain strong classification. The model is designed to capture fine-grained spectral properties and preserve spatial continuity that ensures improved feature discrimination and reduced redundancy. The SpectroRes-Former system will be capable of achieving a higher classification accuracy and reliability by integrating these features, which will be important in sustainable environmental monitoring and decision-making to preserve forests.

1.1. Major Contributions of this Work

- To develop a comprehensive Land-Cover Classification and Deforestation Detection framework that leverages hyperspectral satellite imagery to capture fine-grained spectral and spatial information for enhanced environmental monitoring.
- To introduce a SpectroRes-Former architecture, which integrates ResNet Encoder (for spatial feature extraction), Spectral–Spatial Attention Fusion, Transformer Encoder with Multi-Layer Perceptron (MLP), Softmax classification, Cross-Entropy Loss, and Orthogonality Regularization to ensure discriminative and non-redundant feature learning across spectral bands.
- To propose a novel QuFiSel (Quantum–Fisher Selection) algorithm that fuses Quantum-Inspired Feature Selection (QIFS) and Fisher Discriminative Scoring to identify the most informative and non-redundant features.

The organization of this work: The literature review using the current methodology is found in Section 2. The proposed methodology, comprising the framework and key elements, is presented in Section 3. The result and comparison analysis are shown in Section 4. Finally, the work comes to a close in the section 5 with the conclusion.

2. Related Works

The classification of hyperspectral images (HSIC) is among the most crucial areas of study in the domain of remote sensing due to the ability of hyperspectral imaging to gather hundreds of consecutive spectral bands which helps in differentiating materials based on their spatial properties. However, due to issues like high dimensionality, scarcity of labelled data, spectral redundancy, and complexity associated with spatial variation, HSIC has always been a challenge. Recent studies have thus concentrated on deep learning, Transformers, and quantum computing approaches to overcome these problems.

Earlier deep learning systems were dependent on the use of convolutional neural networks (CNNs) for local spectral and spatial feature extraction. While CNN models work well in extracting neighbour information, the restricted receptive field nature of CNN models hinders their capacity to model the long-distance spectral relationships. In order to solve this problem, Yu et al. [17] developed the multilevel spectral-spatial transformer network (MSTNet). MSTNet uses hierarchical self-attention for capturing the global context in hyper-spectral image patches. MSTNet was more successful than conventional CNNs but was computationally expensive because of multi-head attention mechanisms.

Subsequently, transformer-based hyperspectral classification has gained popularity. Song et al. [25] proposed the Bottleneck Spatial-Spectral Transformer (BS2T), where spectral and spatial correlations can be modelled simultaneously by means of dual-branch attention mechanisms. In a similar way, Feng et al. [20] proposed Spectral-Spatial-Elevation Fusion Transformer (S2EFT), where the fusion of multi-sources can be realized based on the integration of hyperspectral images and LiDAR elevation data. Recently, another method called diffusion-based representation learning has been presented. Specifically, Chen et al. [24] proposed SpectralDiff, where spectral distributions are learned through the processes of forward-reverse diffusion while spatial consistencies are guaranteed. Despite the great progress made by these Transformer-based frameworks, they all need high dimensional feature

representations.

An additional research area involves feature optimization and dimensionality reduction. Enhanced Functional Principal Component Analysis (FPCA), which was suggested by Xue et al. [21], is used to convert the spectral signature into continuous functions with spatial neighbourhood information taken into account. The authors Fei et al. [19] designed a light-weight VGG network that includes both principal component analysis and adaptive feature learning for minimizing computational cost. In addition, the combination of the handcrafted texture descriptor and the Artificial Gorilla Troops Optimizer with the Broad Learning System has been utilized by Wan et al. [22] for improving the efficiency of the optimization algorithm under limited samples.

In addition, recent advances have also considered optimization schemes that are inspired by quantum computation principles in hyperspectral imaging. Quantum-inspired optimization algorithms based on qubit and qutrit models for clustering improvement were proposed in [18]. In another work, [23] further studied hybrid quantum-classical neural networks that combine the concept of parameterized quantum circuits and convolutional layers. Even though these approaches show how quantum computation and quantum-inspired optimization can be useful, these approaches require optimization routines and quantum circuit simulations that are not practical when deploying on common deep learning frameworks.

Nevertheless, there are several research issues that still need to be addressed. Firstly, current Transformer models mostly focus on modelling the global context without considering removing redundancy from spectral-spatial features in order to facilitate the deep learning process. Secondly, the existing feature selection schemes are independent of the classifier network, which is unable to optimize the discriminative features during the training procedure. Lastly, the current quantum-inspired techniques are usually developed for clustering or optimization problems rather than classification. The existing techniques have been analysed, in terms of techniques and limitations; and they are manifested in Table 1.

In an effort to overcome these challenges, the SpectroRes-Former, a spectral-spatial residual Transformer network along with the Quantum-Fisher Feature Selection (QuFiSel) technique is suggested in this study. Different from previous Transformer-based methods, QuFiSel carries out a quantum-inspired probabilistic feature selection via amplitude state representation and Fisher discrimination to select the most discriminating and least redundant spectral-spatial features prior to Transformer modelling. It should be noted that the proposed methodology does not use any quantum computing techniques nor any variational quantum circuit. On the other hand, the principles of quantum computing have been utilized mathematically including amplitude state representation, probabilistic feature evolution, and orthogonal feature space learning. The presented approach has been developed for multi-class land-cover classification in hyperspectral remote sensing imagery and has been tested using the Indian Pines hyperspectral image data set. Thus, the presented methodology can be applied only for hyperspectral data sets.

Table 1. Analysis on the state-of-art models

Method	Spectral Modelling	Spatial Modelling	Transformer	Feature Selection	Quantum-Inspired	Limitations
MSTNet	✓	✓	✓	✗	✗	High computational cost
SpectralFormer	✓	Partial	✓	✗	✗	No explicit feature optimization
BS2T	✓	✓	✓	✗	✗	High-dimensional redundant features
SpectralDiff	✓	✓	Diffusion + Attention	✗	✗	Computationally intensive
S2EFT	✓	✓ + LiDAR	✓	✗	✗	Requires multi-source data
SpectroRes-Former (Proposed)	✓	✓	✓	✓ (QuFiSel)	✓ (classical quantum-inspired optimization)	Designed for hyperspectral multi-class classification

2.1. Research Gap and Motivation

Despite the remarkable progress in the development of hyperspectral image classification techniques that have successfully leveraged the capabilities of convolutional neural networks, Transformer architectures, lightweight features extraction approaches, and quantum-inspired optimizers, there are still a number of open questions in this research domain.

One of these questions is related to existing Transformer-based approaches for hyperspectral image classification, such as MSTNet, BS2T, SpectralFormer, and SpectralDiff. Despite the success of the aforementioned architecture designs in capturing global contexts using self-attention mechanism, most of these approaches do not employ explicit pre-removal of redundant spectral information before the input hyperspectral feature space enters the Transformer layer.

Secondly, traditional techniques for feature selection are usually formulated as standalone pre-processing procedures. PCA, FPCA, Fisher score, ReliefF, and Mutual Information Ranking are some of the statistical techniques that help reduce the dimensions of the features prior to performing classification but are separate from the deep

representation learning process that follows. This implies that the selected set of features cannot be updated dynamically depending on how discriminative the deep neural network has become.

Moreover, recently proposed quantum-inspired methods have shown their optimization potential for hyperspectral image processing; nevertheless, current research is mainly concerned with clustering, optimization, or heuristic search, but not the supervised classification of hyperspectral images. Besides, the majority of quantum-inspired neural networks perform quantum state evolution for optimization purposes without any interactions between probability-based representation of features and learning process. In addition, hybrid quantum-classical neural networks usually depend on quantum circuit simulation or quantum computation hardware.

Another drawback noticed from current literature is the lack of feature optimization coupled with Transformer-based spectral-spatial representation learning process. Most advanced deep learning models carry out spectral-spatial feature extraction and context learning operations in sequence without including any adaptive feature optimization mechanism which can help suppress redundant spectral information while retaining discriminatory spatial features. Thus, the synergic role of feature optimization and global context learning has yet to be realized.

In order to overcome these drawbacks, the proposed technique is based on the incorporation of Spectral-Spatial Residual Transformer (SpectroRes-Former) along with the Quantum-Fisher Feature Selection (QuFiSel) algorithm. In contrast to regular feature selection methods, QuFiSel uses quantum-inspired amplitude-state representation combined with Fisher discriminant analysis in an iterative manner for feature importance optimization before the classification using Transformer. It should be noted that the quantum-inspired term implies the use of mathematical principles of quantum-state representation, probabilistic feature evolution, and orthogonal feature space learning in a classical optimization process. The suggested approach does not include any use of quantum computers or circuits, quantum gates, or other quantum computing techniques.

In addition, the proposed approach is tailored for hyperspectral land-cover classification involving a number of classes on benchmark hyperspectral datasets like the Indian Pines dataset. This methodology relies heavily on the utilization of high spectral resolution found in hyperspectral data, and therefore it is not applicable on multispectral datasets without further adjustments. Furthermore, the proposed methodology has been formulated as a multi-class classification problem, where each pixel is classified into one of the land-cover classes. The framework has no intention of performing binary deforestation detection, and all formulations of algorithms, loss functions, evaluation metrics, and experiment designs discussed in this paper have been done accordingly.

The primary motivation for conducting this study is to devise a computationally efficient spectral–spatial classification framework, which will at the same time (i) perform redundant hyperspectral feature reduction by employing an adaptive quantum-inspired feature optimization, (ii) retain discriminative spectral-spatial information through Fisher-guided feature refinement, (iii) utilize global contextual relationships using residual Transformer architecture, and (iv) conduct reliable and statistically robust multi-class hyperspectral land-cover classification.

3. Proposed Methodology

The SpectroRes-Former network methodology guarantees the proper land-cover and deforestation analysis based on preprocessing, feature engineering, and deep learning. The hyperspectral data of the Indian Pines hyperspectral dataset is first corrected, normalized and refined and then feature extraction is performed using Quantum Amplitude Encoding and SLIC superpixel segmentation. The QuFiSel (Quantum-Fisher Selection) algorithm improves the relevance of features, and the classification is done with the help of SpectroRes-Former, which is a Spectral-Spatial Residual Transformer Network that consists of ResNet encoding, spectral-spatial attention fusion, Transformer layers, and orthogonality regularization to achieve high accuracy and reliability. The overall architecture of the SpectroRes-Former framework is represented in Fig. 1.

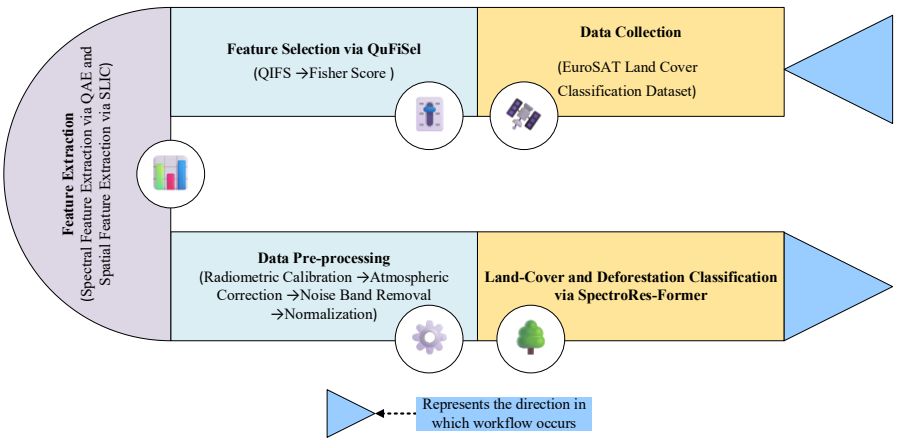


Fig. 1. Architecture of the SpectroRes-Former Framework

3.1. Data collection

Experiments were performed on the Indian Pines Dataset that is one of the most common benchmarking data sets used in hyperspectral image classification research. The data set was collected from the scene of Indian Pines agricultural area in northwestern Indiana, United States, by using the Airborne Visible/Infrared Imaging Spectrometer sensor. This scene involves agricultural fields, forest, grassland, roads, and other built-up areas which exhibit a wide spectrum of variations and complex land cover distribution.

The original hyperspectral image includes 145 × 145 pixels in spatial resolution and 220 continuous spectral bands covering the wavelength spectrum from about 0.4 to 2.5 micrometers. Considering the noise in water bands, 20 water and noise bands are excluded from the hyperspectral image based on conventional pre-processing operations for hyperspectral images, leaving 200 spectral bands in total for future analysis.

This data set involves 16 different classes of land covers, including different kinds of vegetation, agricultural fields, forest, and mixed land covers. The rich spectral data makes the data set highly appropriate for analysing spectral-spatial learning methods and quantum-inspired feature selection methods.

The hyperspectral image cube can be given as $D_{raw} \in \mathbb{R}^{H\times W\times B}$, where $H = 145$and $W = 145$denote the spatial dimensions, while $B = 200$represents the retained spectral bands after noisy-band removal. The collected hyperspectral cube serves as the input to the preprocessing stage, where radiometric correction, atmospheric compensation, noise-band removal, and normalization are performed prior to feature extraction and classification.

3.2. Data Preparation and Feature Engineering

3.2.1. Data Pre-processing

The raw hyperspectral images undergo a systematic preprocessing pipeline before feature extraction. The stages enhance the data quality and prepare it for analysis in high spectral and spatial dimensions. It is a step-by-step process where the results of one operation serve as the input for the next stage.

Radiometric Calibration

Firstly, the raw digital numbers (DN) captured by the hyperspectral sensor are changed to top-of-atmosphere (TOA) radiance values for the purpose of correcting the sensor-induced distortions like dark current, gain variations, and detector nonlinearity. This correction can be represented mathematically by Eq. (1):

$$L_{TOA}(x, y, \lambda) = g(\lambda) \cdot DN(x, y, \lambda) + o(\lambda) \tag{1}$$

Where, $L_{TOA}(x, y, \lambda)$ denotes the radiance at wavelength λ, $g(\lambda)$ represents the sensor gain factor, and $o(\lambda)$ signifies the sensor offset. By means of this calibration the spectral intensity scale is brought in line with true radiance, so that the data represent the actual surface energy responses and not the sensor noise.

Atmospheric Correction

After radiometric calibration, the atmospheric effects (scattering, absorption, and path radiance) are reduced with the help of the FLAASH (Fast Line-of-sight Atmospheric Analysis of Spectral Hypercubes) algorithm. FLAASH is based on the radiative transfer model of MODTRAN and transforms TOA radiance into surface reflectance, which is given by Eq. (2):

$$R_{surf}(x, y, \lambda) = \frac{L_{TOA}(x, y, \lambda) - L_{path}(\lambda)}{E_{sun}(\lambda) \cdot TT(\lambda) \cdot \cos(\theta_s)} \tag{2}$$

Where, $L_{path}(\lambda)$ represents the path radiance, $E_{sun}(\lambda)$ signifies solar irradiance, $T(\lambda)$ refers to total transmittance, and represents solar zenith angleθ_s. By this correction, the spectral responses are effectively normalized to surface properties, and the influence of atmospheric scattering and illumination geometry is removed.

Noise Band Removal

After atmospheric correction, bands with low signal-to-noise ratio (SNR)—typically corresponding to strong water absorption regions (around 1.4 μm and 1.9 μm)—are removed to enhance spectral stability. Let $\mathfrak{B} = \{\mathfrak{b}_1, \mathfrak{b}_2, \dots, \mathfrak{b}_C\}$ be the set of spectral bands and $SNR(\mathfrak{b}_i)$ the computed signal-to-noise ratio for band $\mathfrak{b}_i$. The selected valid bands form as Eq. (3):

$$\mathfrak{B}' = \{\mathfrak{b}_i \in \mathfrak{B} | SNR(\mathfrak{b}_i) > \tau_{SNR}\} \tag{3}$$

Where, τ_{SNR} represents the empirically defined threshold. This step eliminates redundant and noisy components, resulting in a cleaner spectral signature for subsequent analysis.

Normalization

To maintain numerical stability across spectral dimensions, each retained band is Min–Max normalized to the range [0,1]:

$$I_{norm}(x, y, \lambda) = \frac{R_{surf}(x, y, \lambda) - R_{min}(\lambda)}{R_{max}(\lambda) - R_{min}(\lambda)} \tag{4}$$

This operation preserves the relative spectral contrast while ensuring uniform dynamic scaling across all wavelengths. The normalized data enhance convergence stability in downstream deep learning models by reducing spectral amplitude disparities. The sequential outcome of the pre-processing stage is depicted in Fig. 2.

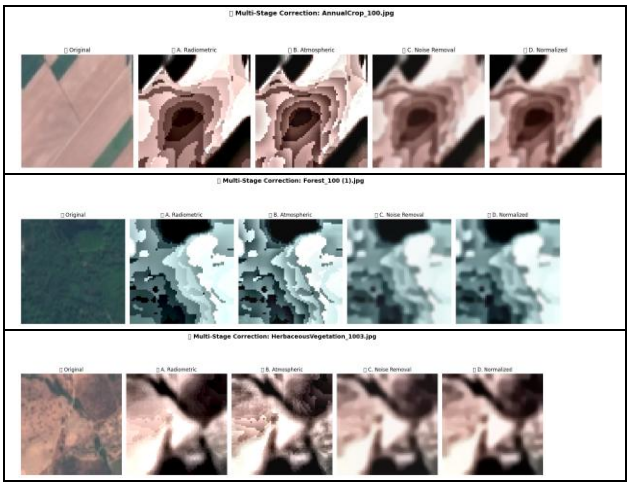


Fig. 2. Outcomes of the Sequential Pre-processing Stages

The final pre-processed tensor is denoted as $D_{PP} \in \mathbb{R}^{B \times H \times W \times C'}$, where B represents the batch size, H and W signifies the spatial dimensions, and C' refers to the number of spectral bands retained after noise removal. This tensor D_{PP} serves as the standardized and atmospherically corrected input to the Feature Extraction stage (Section 3.2.2), forming the foundation for robust spectral–spatial feature learning.

3.2.2. Feature Extraction

After pre-processing the raw hyperspectral imagery, it is then ready to be processed in an informative way in terms of features. This step will attempt to obtain two complementary sets of representations: spectral and spatial. Whereas spectral features characterize the wavelength-dependent reflectance characteristics of each pixel, spatial features encode the structural and contextual data of the surrounding areas. They are both independently obtained out of the same input data D_{PP} and subsequently fused to create the single feature representation.

Spectral Feature Extraction via Quantum Amplitude Encoding

Spectral feature extraction focuses on the wavelength signature of each pixel, where the reflectance intensity varies across different spectral bands. Each pixel vector $pv_{(i,j)}$ is treated as a spectral sample, whose components correspond to the reflectance values across all bands. To efficiently encode this high-dimensional spectral information while preserving its physical interpretability, Quantum Amplitude Encoding (QAE) is employed.

In QAE, each spectral vector is mapped into a normalized quantum state representation as Eq. (5):

$$| \psi e_{(i,j)} \rangle = \sum_{c=1}^{\mathbb{C}} \alpha_c | c \rangle \tag{5}$$

Where, $\alpha_c = \frac{pv_{(i,j,c)}}{\sqrt{\sum_{c=1}^{\mathbb{C}} pv_{(i,j,c)}^2}}$, $| \psi e_{(i,j)} \rangle$ represents the encoded quantum state, and the coefficients, and α_c denote amplitude-normalized spectral intensities. This transformation ensures unit-length encoding ($\sum | \alpha_c |^2 = 1$), maintaining the proportional relationships among different wavelength responses.

Essentially, each pixel of the image goes through this conversion process and ends up as a highly efficient quantum-encoded representation; thus, the minute spectral variations that are instrumental for differentiating even the most subtle material or surface properties are preserved.

Spatial Feature Extraction via Superpixel Segmentation (SLIC)

The spectral encoding can only record the per-pixel reflectance behaviour, but it does not record the local spatial relations. Simple Linear Iterative Clustering (SLIC) algorithm of Superpixel Segmentation is applied to solve this.

SLIC groups pixels into perceptually meaningful units, called super pixels, based on the similarity in both spectral and spatial space. The clustering algorithm minimizes a distance measure that is a combination of colour (spectral) proximity and spatial compactness that is given as Eq. (6):

$$\mathfrak{U} = \sqrt{\left(\frac{d_{lab}}{N_c}\right)^2 + \left(\frac{d_{xy}}{N_s}\right)^2} \tag{6}$$

Where, d_{lab} measures the spectral (color) distance between pixels, d_{xy} denotes the spatial Euclidean distance, and N_c and N_s are normalization constants that balance the influence of spectral versus spatial proximity.

Once convergence has taken place, every superpixel area is a homogeneous local area with similar spectral and textural properties. The spatial feature set is generated by calculating descriptive statistics like mean reflectance, local variance, gradient magnitude, and texture entropy in each region. These characteristics encode contextual patterns, sharpness of boundaries and object-level consistency that are frequently ignored by pixel-wise spectral analysis.

The spectral features **(F_{SPE})** and spatial features **(F_{SPA})** are independent but complementary representations of the same scene. F_{SPE} captures what the surface materials are (based on wavelength behaviour) and F_{SPA} captures where and how these materials are distributed (based on regional context).

To form a unified and semantically rich representation, these two feature sets are fused using concatenation along the feature dimension as $F_{FE} = Concat(F_{SPE}, F_{SPA})$. If the spatial segmentation alters the resolution of F_{SPA}, it is upsampled before concatenation to ensure correspondence between pixel locations and their contextual descriptors.

This fused feature map F_{FE} serves as the comprehensive representation that carries both pixel-level spectral information and neighbourhood-level spatial context. It is subsequently forwarded to the feature selection stage, where redundant or non-informative dimensions are pruned to enhance discriminative efficiency.

3.2.3. Feature Selection via QuFiSel (Quantum–Fisher Selection)

The QuFiSel system combines quantum-inspired probabilistic encoding and Fisher-based discriminative assessment to determine the most informative and non-redundant spectral-spatial features. It works in a series of steps: encoding features into quantum amplitude representations, evaluating their statistical significance, and updating their selection probabilities by successive updates. This combination methodology guarantees that the resulting feature subset F_{FS} has as much class separability as possible with as little redundancy as possible, which is an optimized input to the SpectroRes-Former classification module.

Quantum-Inspired Feature Encoding (QIFS)

The Quantum-Inspired Feature Encoding (QIFS) mechanism serves as the foundational step in the feature selection process, aiming to represent each feature dimension of F_{FE} as a quantum amplitude. Instead of treating features as static numerical entities, QIFS interprets them as probabilistic quantum states, where each dimension contributes with a specific amplitude weight proportional to its potential discriminative strength.

Formally, given the feature tensor $F_{FE} = [\mathfrak{f}_1, \mathfrak{f}_2, \dots, \mathfrak{f}_d]$, each feature component $\mathfrak{f}_j$ is mapped into a normalized amplitude ϖ_i, satisfying the quantum normalization constraint as Eq. (7):

$$\sum_{j=1}^{d} |\, \varpi_j \,|^2 = 1 \tag{7}$$

The corresponding amplitude for the j^{th} feature is computed as Eq. (8):

$$\varpi_j = \frac{\mathfrak{f}_j}{\sqrt{\sum_{i=1}^{d} \mathfrak{f}_i^2}} \tag{8}$$

This encoding forms a quantum-inspired state representation expressed as Eq. (9):

$$F_q' = QIFS(F_{FE}) = \sum_{j=1}^{d} \varpi_j |j\rangle \tag{9}$$

Where, $|j\rangle$ denotes the quantum basis vector associated with the j^{th} feature dimension. Simultaneously, a feature probability vector is initialized as Eq. (10).

$$PRO = \{pro_1, pro_2, \ldots, pro_d\}, \quad \text{where} \quad \sum_{j=1}^{d} pro_j^2 = 1 \tag{10}$$

Reporting the probability of significance of each feature dimension. In contrast to conventional hard selection techniques which apply binary inclusion masks, QIFS allows soft, non-binary inclusion, where the importance of features can dynamically change as optimization proceeds.

This probabilistic quantum encoding allows each feature dimension to be adaptively assessed, combined or discarded according to its changing amplitude- thus establishing the basis of the next stage of discriminative evaluation, based on Fisher.

Fisher Discriminative Scoring

QuFiSel stage two uses Fisher Discriminative Scoring to assess the statistical separability of each quantum-encoded feature of the QIFS stage. Whereas QIFS gives a probabilistic amplitude representation of feature importance, Fisher scoring measures the ability of each feature to differentiate between different land-cover or deforestation classes in terms of inter-class variance and intra-class compactness. For a given feature dimension j, the Fisher Score is defined as Eq. (11):

$$\mathcal{S}_f(j) = \frac{\left(\mu_j^{(1)} - \mu_j^{(2)}\right)^2}{\sigma_j^{(1)2} + \sigma_j^{(2)2}} \tag{11}$$

Where, $\mu_j^{(1)}$ and $\mu_j^{(2)}$ denote the mean feature responses for classes 1 and 2, while $\sigma_j^{(1)}$ and $\sigma_j^{(2)}$ represent their corresponding standard deviations. This metric effectively measures the ratio of inter-class separation to intra-class variation — higher values indicate features that exhibit strong class discrimination and stable statistical behaviour.

In the quantum–probabilistic context, the Fisher score operates as a measurement operator acting on the amplitude space defined by QIFS. Each feature's quantum amplitude ϖ_j is reweighted based on its discriminative potential, thereby linking probabilistic confidence with empirical evidence. The updated probabilistic weight of each feature is then computed as Eq. (12):

$$\tilde{p}_j = p_j \times \mathcal{S}_f(j) \tag{12}$$

Where, p_j denotes the original quantum amplitude-based probability, and $\mathcal{S}_f(j)$ represents its Fisher score.

The merge of quantum-inspired probability weighting and statistical discriminability ensures that features are not only ranked by the confidence that they are encoded, but also by their measurable contribution to class separability- thus paving the way for the next step of optimization, final selection of the most informative subset.

Probabilistic Update and Optimal Subset Selection

To finalize the quantum-guided feature selection, this stage performs iterative probability refinement and extracts the most discriminative subset of features, denoted as F_{FS}. The goal is to maximize the Fisher information contribution while maintaining feature orthogonality and suppressing redundancy.

Each feature amplitude is updated according to its Fisher-weighted relevance obtained in the previous step. Let $\tilde{p}_j$ represent the intermediate (unnormalized) probability of the j^{th} feature. The normalized update rule is expressed as Eq. (13):

$$p_j^{(t+1)} = \frac{\tilde{p}_j}{\sum_i \tilde{p}_i^2} \tag{13}$$

This ensures that the probability distribution remains normalized after each iteration, preserving quantum probability coherence during feature amplitude evolution. The iterative refinement continues until a stopping criterion is satisfied—typically when the top-k feature rankings stabilize across successive iterations or when the cumulative probability of selected features surpasses a predefined threshold T_p.

Once convergence is achieved, the final feature subset is extracted as Eq. (14):

$$F_{FS} = SelectTop\left(F_q', \mathcal{S}_{\oint}\right) \tag{14}$$

Where, F_q' represents the Fisher-weighted quantum feature representation, and $\mathcal{S}_{\oint}$ denotes the optimal subset of indices corresponding to high-information, low-redundancy features.

This selection mechanism ensures that optimized feature tensor (F_{FS}) preserves orthogonal, high-discriminative spectral–spatial representations while discarding irrelevant or noisy components. As a result, the optimized feature tensor denoted as F_{FS} encapsulates the most essential structural and spectral cues for downstream learning.

Finally, the optimized feature tensor is expressed as $F_{FS} \in \mathbb{R}^{B\times H\times W\times C}$ and is fed into the SpectroRes-Former network for classification and deforestation mapping, as detailed in Section 3.1.

3.2.4. Theoretical Interpretation of the Quantum-Inspired Feature Optimization

QuFiSel stands for the Quantum-Fisher Selection algorithm, which is designed as a feature optimization scheme inspired by quantum state representation theory but implemented in a classical computation setting only. It does not include any physical devices, quantum gates, variational circuits, and quantum annealing techniques. QuFiSel uses mathematical tools borrowed from quantum state representation theory and represents features' importance in the form of amplitude probability distribution that changes during feature discriminability iterations.

The main goal of the QuFiSel algorithm is to convert traditional static feature normalization into dynamic probabilistic optimization.

Difference Between Conventional L2 Normalization and Quantum-Inspired Amplitude Encoding

Conventional L2 normalization is widely employed in machine learning and deep learning to normalize feature magnitudes while preserving their relative proportions. Given a feature vector: $\mathbf{x} = [x_1, x_2, \dots, x_d]^T$, its normalized representation is computed as per Eq. (15)

$$\mathbf{x}' = \frac{\mathbf{x}}{||\ \mathbf{x}\ ||_2}, \tag{15}$$

Where, x denotes the original feature vector, and x' represents its L2-normalized counterpart.

$$||\ \mathbf{x}\ ||_2 = \sqrt{\sum_{i=1}^{d} x_i^2}. \tag{16}$$

Such normalization limits the magnitude of the feature vector to one and avoids scaling problems between different feature dimensions. Nonetheless, following the normalization process, all feature elements become fixed and deterministic. There is no optimization, probabilistic evolution, and feature discrimination process during the normalization process. This means that all features will retain their normalized values irrespective of their discriminative ability. On the contrary, the proposed QuFiSel scheme considers the normalized feature vector as a quantum-based amplitude state represented as per Eq. (17)

$$|\ \Psi\rangle = \sum_{i=1}^{d} \omega_i\ |\ i\rangle, \tag{17}$$

subject to the probability conservation constraint

$$\sum_{i=1}^{d} |\ \omega_i\ |^2 = 1, \tag{18}$$

where ω_i denotes the probability amplitude associated with the i^{th} feature dimension and $|\ i\rangle$ represents the corresponding basis state. Unlike conventional normalization, these amplitudes are not treated as static coefficients. Instead, they represent adaptive probability variables whose values evolve dynamically during feature optimization.

The probability corresponding to each feature is initialized as: $P_i^{(0)} = |\ \omega_i\ |^2$. To incorporate class-discriminative information into feature evolution, the framework computes the Fisher discriminative score for each feature as per Eq. (19).

$$S_f(i) = \frac{\left(\mu_i^{(1)} - \mu_i^{(2)}\right)^2}{\sigma_i^{(1)2} + \sigma_i^{(2)2}}, \tag{19}$$

where $\mu_i^{(c)}$and $\sigma_i^{(c)}$denote the mean and variance of the i^{th}feature for class c, respectively. The Fisher score measures the discriminative capability of each feature by maximizing inter-class separability while minimizing intra-class variance.

The feature probabilities are then iteratively updated according to Eq. (20)

$$P_i^{(t+1)} = \frac{P_i^{(t)} \cdot S_f(i)}{\sum_{j=1}^{d} P_j^{(t)} S_f(j)}, \tag{20}$$

where $P_i^{(t)}$represents the probability associated with the i^{th}feature at iteration t. Through this probabilistic evolution mechanism, highly discriminative features gradually obtain larger probabilities, whereas redundant or weakly informative features are progressively suppressed.

Thus, QuFiSel is essentially different from typical L2 normalization as it applies discriminative feature optimization rather than just scaling. The procedure for feature optimization is based on continuous improvement of feature importance through statistical separability.

Difference Between Orthogonality Regularization and Conventional Weight Decay

Most deep learning architectures employ L2 weight decay to reduce model complexity and prevent overfitting by penalizing excessively large parameter values. The conventional optimization objective is formulated as per Eq. (21).

$$L = L_{CE} + \lambda \,||\, W \,||_2^2, \tag{21}$$

where L_{CE}denotes the cross-entropy classification loss, Wrepresents the trainable weight matrix, and λcontrols the regularization strength. Although L2 regularization tends to dampen large values for parameters, it is unable to reduce redundancy between the features learned. There could be many neurons learning correlated representations, thus generating redundant feature spaces.

The solution to this problem is achieved by introducing the concept of orthogonality defined as per Eq. (22).

$$L_{orth} = |\, W^T W - I \,|_F^2, \tag{22}$$

where Idenotes the identity matrix and $||\cdot||_F$represents the Frobenius norm.

Minimizing Eq. (23) enforces

$$W_i^T W_j \approx 0, i \neq j, \tag{23}$$

thereby encouraging different neurons to learn mutually independent and complementary feature representations. Unlike conventional weight decay, which minimizes only parameter magnitude, the orthogonality objective explicitly minimizes feature correlation and improves embedding diversity.

The final optimization objective of the proposed framework becomes a per Eq. (24).

$$L_{total} = L_{CE} + \lambda_{orth} L_{orth}, \tag{24}$$

where λ_{orth}controls the contribution of orthogonal regularization during optimization.

Consequently, the proposed orthogonality constraint improves feature-space diversity, reduces redundancy among learned embeddings, and enhances spectral–spatial representation quality for hyperspectral classification.

Algorithmic Description of QuFiSel

Algorithm 1 provides an overview of the whole optimization process of the Quantum–Fisher Selection (QuFiSel) approach.

Algorithm 1: Quantum–Fisher Selection (QuFiSel)
Input: Processed hyperspectral cube D_{PP}
Output: Optimized feature tensorF_{FS}

- Spectral features are extracted by applying Quantum Amplitude Encoding.
- Spatial features are extracted through SLIC super pixels segmentation.
- Spectral and spatial features are merged to form F_{FE}.
- Quantum-inspired probability amplitudes are initialized for each feature.
- Normalization of probability amplitudes is performed using Eq. (17).
- Discriminative Fisher scores are calculated using Eq. (20).
- Feature probabilities are updated based on Eq. (22).
- The probabilistic evolution is iterated until convergence conditions are met.
- Features are ranked based on their probability distribution.
- Top-k discriminative features are selected.
- Optimized feature tensorF_{FS} is constructed.
- Send F_{FS} to the SpectroRes-Former classifier.

The described optimization technique combines probability-based amplitude modelling, Fisher adaptation, and orthogonal feature-space learning in a unified optimization technique for features. Different from traditional normalization and weight decay approaches, QuFiSel uses feature optimization in an adaptive manner before performing the classification using a transformer-based classifier, which improves spectral-spatial representation without compromising computational efficiency in a purely classical deep learning approach.

3.3. SpectroRes-Former: Spectral–Spatial Residual Transformer Network for Land-Cover and Deforestation Classification

The SpectroRes-Former network classification module detects deforestation based on a sequential-parallel fusion approach that combines spatial, spectral, and contextual representations based on quantum amplitude encoding. Fig. 3 represents the general flow of the proposed SpectroRes-Former.

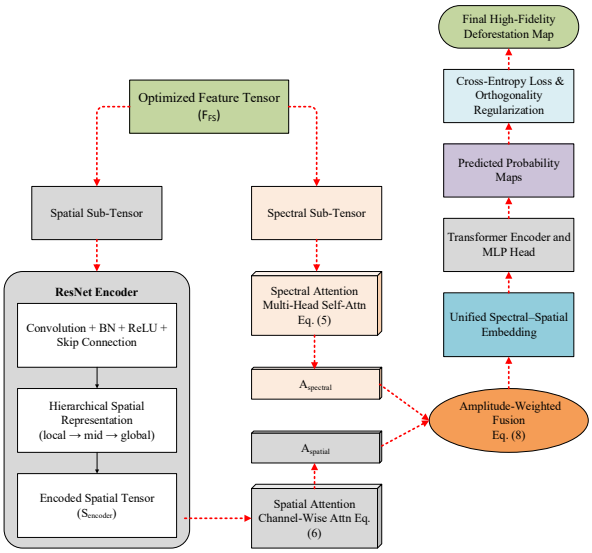


Fig. 3. General Structure of the Proposed SpectroRes-Former

The optimized feature tensor derived from QuFiSel (Quantum–Fisher Selection), which includes both spectral and spatial cues, is indicated by F_{FS} here. The final layer produces the probability distribution over the land-cover classes.

3.3.1. Input Feature Representation

The input tensor, denoted as $F_{FS} \in \mathbb{R}^{B\times H\times W\times C}$, represents the optimized feature tensor obtained from the preceding QuFiSel stage. Here, B denotes the batch size, H and W correspond to the spatial dimensions (height and width), and C represents the number of spectral bands associated with each pixel.

Each spectral pixel vector $\mathbb{f}_{(i,j)} = [f_{(i,j,1)}, f_{(i,j,2)}, \dots, f_{(i,j,C)}]$ is normalized using quantum-inspired amplitude encoding such that the resulting probability amplitudes satisfy the following normalization constraint shown in Eq. (25).

$$\sum_{c=1}^{C} | \psi_{(i,j,c)} |^2 = 1 \tag{25}$$

Where, $\psi_{(i,j,c)}$ denotes the quantum-inspired amplitude corresponding to the c^{th} spectral component at spatial location (i, j).

This amplitude encoding maps the hyperspectral energy distribution into a normalized Hilbert space representation, ensuring as that Eq. (26):

$$\psi_{(i,j,c)} = \frac{f_{(i,j,c)}}{\sqrt{\sum_{c=1}^{C} f_{(i,j,c)}^2}} \tag{26}$$

Which effectively preserves high-dimensional spectral correlations while enabling the model to interpret the hyperspectral cube as a quantum probability amplitude distribution. The encoded feature tensor F_{FS}, is subsequently divided into two complementary branches for spatial and spectral representation learning.

- The spatial sub-tensor $F_{FS}^{(spatial)} \in \mathbb{R}^{B\times H\times W\times C}$ is supplied to the ResNet Encoder, which focuses on extracting deep spatial hierarchies and contextual patterns across land-cover regions.
- Simultaneously, the spectral sub-tensor $F_{FS}^{(spectral)} \in \mathbb{R}^{B\times C}$ is directly forwarded to the Spectral–Spatial Attention Fusion module, where spectral dependencies are adaptively integrated with spatial cues obtained from the encoder output.

Two-stream initialization performs a combined but still separate representation of the spatial and spectral domains, so the quantum-inspired amplitude coherence of the hyperspectral signals is kept during the later stages of learning of the SpectroRes-Former network.

3.3.2. ResNet Encoder (Deep Spatial Hierarchy)

To extract discriminative spatial features from the hyperspectral tensor, the spatial sub-tensor $F_{FS}^{(spatial)} \in \mathbb{R}^{B\times H\times W\times C}$ is first processed through a Residual Network (ResNet)-based encoder, denoted as $f_{ResNet}(\cdot)$. The purpose of this encoder is to acquire spatial representations of a hierarchy that transition from local texture features to high-level contextual patterns which are necessary for forest-cover and deforestation analysis. In a formal way, the spatial encoding can be defined as Eq. (27):

$$S_{encode} = f_{ResNet}\left(F_{FS}^{(spatial)}\right) \tag{27}$$

Where, $S_{encode} \in \mathbb{R}^{B\times H'\times W'\times D_c}$ signifies the encoded spatial feature tensor, $H'$and W'denote the reduced spatial dimensions after convolutional downsampling, and D_c represents the channel dimension.

Each residual block within $f_{ResNet}(\cdot)$ consists of a sequence of convolutional, batch normalization, and ReLU activation layers, coupled with an identity-based skip connection to facilitate efficient gradient flow. Mathematically, a single residual unit can be expressed as Eq. (28):

$$\mathbb{y}_{\mathbb{l}} = \mathbb{F}(\mathbb{x}_{\mathbb{l}}, \{W_{\mathbb{l}}\}) + \mathbb{x}_{\mathbb{l}} \tag{28}$$

Where, $\mathbb{x}_{\mathbb{l}}$ and $\mathbb{y}_{\mathbb{l}}$ represents the input and output of the $\mathbb{l}^{th}$ residual block, respectively, and $\mathbb{F}(\cdot)$ denotes the non-linear transformation composed of stacked convolutional layers parameterized by weights $W_{\mathbb{l}}$.

The residual architecture enables efficient gradient propagation while preserving hierarchical spatial information. Consequently, the encoder is able to abstract multi-scale spatial hierarchies - it can recognize local textural features as well as general deforestation boundary patterns and at the same time structural integrity is kept across layers.

The encoded spatial feature tensor S_{encode} is subsequently forwarded as one of the two inputs to the Spectral–Spatial Attention Fusion module, where it interacts with spectral embeddings derived from $F_{FS}^{(spectral)}$ to enable joint spatial–spectral representation learning.

3.3.3. Spectral–Spatial Attention Fusion

The Spectral–Spatial Attention Fusion (SSAF) block operates in a parallel-fusion mode to jointly integrate spatial and spectral cues. The proposed attention module receives two complementary feature representations as input, viz. (1) the spatial feature tensor, $S_{encode} \in \mathbb{R}^{B\times H'\times W'\times D_c}$ from the ResNet encoder, and (2) the spectral feature matrix $F_{FS}^{(spectral)} \in \mathbb{R}^{B\times C}$ from the amplitude-encoded hyperspectral data.

Two distinct attention pathways are applied:

- Spectral Attention ($Attn_{spectral}$) employs a multi-head self-attention mechanism across spectral bands to model inter-band correlations and highlight discriminative reflectance variations as Eq. (29).

$$A_{spectral} = Softmax\left(\frac{Q_{spec}K_{spec}^{\intercal}}{\sqrt{d_{spec}}}\right)V_{spec} \tag{29}$$

Where, Q_{spec}, K_{spec}, and V_{spec} are the spectral query, key, and value projections of $F_{FS}^{(spectral)}$, and d_{spec} denotes the spectral embedding dimension.

- Spatial Attention ($Attn_{spatial}$) applies channel-wise attention to emphasize salient canopy structures and land-cover boundaries extracted in the spatial domain as Eq. (30):

$$A_{spatial} = ReLU\left(Wl_2 Sigmoid\big(Wl_1 GAP(S_{encode})\big)\right) \odot S_{encode} \tag{30}$$

Where, $GAP(\cdot)$ denotes global average pooling, and Wl_1, Wl_2 denotes learnable weights. The final fused embedding $\mathbb{E}$ is obtained as Eq. (31):

$$\mathbb{E} = Fuse\left(Attn_{spectral}\left(F_{FS}^{(spectral)}\right), Attn_{spatial}(S_{encode})\right) \tag{31}$$

Where the fusion operator aggregates both attention maps through amplitude-weighted concatenation followed by a learnable projection layer, which is defined as Eq. (32):

$$\mathbb{E} = Wl_f\left[A_{spectral} \parallel A_{spatial}\right] + bl_f \tag{32}$$

The attention module integrates the spectral and spatial representations into a unified feature embedding $\mathbb{E} \in \mathbb{R}^{B\times H'\times W'\times D_c'}$, that is sensitive to both local canopy textures and spectral degradation signatures associated with deforestation. The amplitude-based attention coefficients simulate quantum probability weighting, thereby stabilizing convergence and allowing more spectral-spatial discrimination.

The resulting fused representation $\mathbb{E}$ is subsequently forwarded to the Transformer Encoder and MLP Head (Section 3.1.4) for global contextual reasoning and classification.

3.3.4. Transformer Encoder and MLP Head

The fused embedding $\mathbb{E} \in \mathbb{R}^{B\times H'\times W'\times D_c'}$, obtained from the Spectral–Spatial Attention Fusion module, serves as the input to the Transformer encoder, which captures long-range dependencies and global contextual relations across spatial–spectral patches.

Each Transformer layer comprises multi-head self-attention (MHSA) and feed-forward submodules with residual normalization. Formally, for each encoder block $\mathscr{b}$, the transformations are defined as Eq. (33) and (34):

$$\mathbb{E}'_{\mathscr{b}} = LN\big(\mathbb{E}_{\mathscr{b}-1} + MHSA(\mathbb{E}_{\mathscr{b}-1})\big) \tag{33}$$

$$\mathbb{H}_{\mathscr{b}} = LN\big(\mathbb{E}'_{\mathscr{b}} + FFN(\mathbb{E}'_{\mathscr{b}})\big) \tag{34}$$

Where, $LN(\cdot)$ denotes layer normalization, and $FFN(\cdot)$ signifies a two-layer position-wise feed-forward network with GELU activation. After $\mathbb{L}$ stacked layers, the global contextual embedding is represented as $\mathbb{H} = f_{transformer}(\mathbb{E})$.

The encoded tensor $\mathbb{H} \in \mathbb{R}^{B\times\mathbb{N}\times D_t}$ (where $\mathbb{N} = H' \times W'$) captures spectral–spatial dependencies across all spatial regions, enabling global reasoning about land-cover variations. Next, a Multilayer Perceptron (MLP) classification head transforms each contextual token into a scalar logit $\mathbb{z}_{\mathbb{i}}$, which is defined as Eq. (35):

$$\mathbb{z}_{\mathbb{i}} = WP_2 GELU(WP_1 \mathbb{h}_{\mathbb{i}} + bp_1) + bp_2 \tag{35}$$

Where, $\mathbb{h}_{\mathbb{i}}$ represents the feature vector of the $\mathbb{i}^{th}$ region and WP_1, WP_2, bp_1, bp_2 are learnable parameters. As the Indian Pines data set is a multi-class classification problem for land cover using hyperspectral data, the output layer uses a Softmax activation function to calculate the posterior probability of the land cover classes. The probability calculated for the C^{th} class is as follows:

$$\widehat{\mathbb{p}}_{\mathbb{i}} = Softmax(\mathbb{z}_{\mathbb{i}}) \tag{36}$$

Thus, the Transformer–MLP head converts the globally contextualized embeddings into actionable land-cover predictions, identifying whether each forest patch exhibits signs of deforestation or remains ecologically stable. The resulting probability maps $\hat{\mathbb{p}}$ are subsequently forwarded to the Optimization and Output Generation stage for loss computation and final deforestation map synthesis.

3.3.5. Optimization and Output Generation

The deforestation map is ultimately created from the overall optimization objective that is locally computed with the probability maps $\hat{\mathbb{p}}$ from the Transformer–MLP head. A combined loss function comprising the Cross-Entropy (CE) loss for classification accuracy and an Orthogonality Regularization term based on the orthonormal-state principle of quantum systems is used to train SpectroRes-Former. The overall loss is given by Eq. (37):

$$\mathcal{L}_{total} = \mathcal{L}_{CE} + \lambda_O \parallel WM^{\intercal}WM - I \parallel_{Fn}^2 \tag{37}$$

Where, $\mathcal{L}_{CE}$ represents the standard cross-entropy loss between the predicted probability $\hat{\mathbb{p}}$ and the ground-truth label $\mathbb{y}$, which is expressed as Eq. (38):

$$\mathcal{L}_{CE} = -\frac{1}{\mathbb{N}}\sum_{\mathbb{i}=1}^{\mathbb{N}}[\mathbb{y}_{\mathbb{i}}\log(\hat{\mathbb{p}}_{\mathbb{i}}) + (1-\mathbb{y}_{\mathbb{i}})\log(1-\hat{\mathbb{p}}_{\mathbb{i}})] \tag{38}$$

Here, WM represents the classifier's weight matrix, $\|\cdot\|_{Fn}$ denotes the Frobenius norm, and λ_O controls the strength of the orthogonality constraint.

The orthogonality regularization encourages the network to learn mutually independent feature subspaces, thereby reducing feature redundancy—like quantum orthonormal states—thus increasing the stability and the ability to generalize. The AdamW optimizer is used for the optimization along with a learning-rate warm-up and a cosine decay scheduling to get a smooth convergence and to avoid overfitting. During inference, each pixel-level probability $\hat{\mathbb{p}}(x, y)$ is thresholded to generate a binary deforestation map, distinguishing deforested ($\hat{\mathbb{p}}(x, y) \geq \tau_D$) from healthy regions ($\hat{\mathbb{p}}(x, y) < \tau_D$), where τ_D refers to an adaptive decision threshold. Moreover, to make the visuals more consistent, morphological post-processing steps such as dilation, erosion, and small-patch removal can be performed to restore the boundary and remove the isolated artifacts. The result is a precise land-cover map that clearly shows the areas where deforestation has taken place and can be used for environmental monitoring and spatial-temporal change analysis.

4. Result and Discussion

4.1. Experimental Setup

The suggested framework has been executed with Python and standard deep learning and geospatial analysis packages. As a major source of data, the Indian Pines hyperspectral dataset and EuroSAT have been used, which comprises around 3,000 forest image tiles to classify the land-cover and deforestation. The proposed SpectroRes-Former model is compared against four existing state-of-the-art methods, namely, MSTNet [17], S2EFT [20], SpectralDiff [24], and BS2T [25]. Various key performance measures have been used to evaluate it such as Accuracy, Precision, Sensitivity, Specificity, F-measure, Negative Predictive Value (NPV), Matthews Correlation Coefficient (MCC), False Positive Rate (FPR), and False Negative Rate (FNR). A total of five comparative analyses have been performed such as, the performance analysis between the SpectroRes-Former and the existing techniques, K-fold comparison analysis, Effect of pre-processing, feature extraction, and feature selection. Table 2 and Table 3 summarize the implementation sources and reproduction strategy adopted for each competing approach.

Table 2. Baseline Implementation Sources

Method	Original Reference	Code Availability	Reproduction Strategy
Proposed	This work	Available after publication	Original implementation
MSTNet	Yu et al. (2022)	reproduced from paper	Original architecture reproduced
S2EFT	Feng et al. (2024)	Reproduced from published paper	Original parameters followed
SpectralDiff	Chen et al. (2023)	Official implementation / reproduced	Original diffusion settings
BS2T	Song et al. (2022)	Reproduced from paper	Original bottleneck transformer

Table 3. Hyperparameter Settings Used for Experimental Comparison

Parameter	Proposed	MSTNet	S2EFT	SpectralDiff	BS2T
Optimizer	AdamW	Adam	AdamW	AdamW	Adam
Learning Rate	1×10^{-4}	1×10^{-3}	1×10^{-4}	1×10^{-4}	5×10^{-4}
Batch Size	32	32	32	16	32
Epochs	200	200	200	250	200
Weight Decay	1×10^{-4}	5×10^{-4}	1×10^{-4}	1×10^{-4}	5×10^{-4}
Scheduler	Cosine	StepLR	Cosine	Cosine	StepLR
Loss Function	Cross Entropy + Orthogonality	Cross Entropy	Cross Entropy	Cross Entropy	Cross Entropy

4.2. Analysis on the performance of proposed model for Multispectral database- EuroSAT

The comparison of the performance of the proposed SpectroRes-Former framework with the current state-of-the-art methods, namely MSTNet [17], S2EFT [20], SpectralDiff [24], and BS2T [25], has been conducted in terms of nine major evaluation metrics. The SpectroRes-Former method has the best Accuracy (0.989555), which is higher than that of MSTNet (0.97704), S2EFT (0.94469), SpectralDiff (0.93722), and BS2T (0.96336), as indicated in Table 4. This high accuracy gain can be attributed to the SpectralSpatial Residual Transformer Network (SpectroRes-Former), which effectively combines spectral and spatial dependencies with residual attention fusion and transformer-based global context modelling, which contributes to increased class separability and reduced misclassification.

Table 4. Performance comparison between the SpectroRes-Former and existing techniques.

Method	Accuracy	Precision	Sensitivity	Specificity	F-measure	NPV	MCC	FPR	FNR
SpectroRes-Former	0.98955	0.97863	0.99001	0.993	0.9906	0.98671	0.98984	0.02927	0.01784
MSTNet [17]	0.97704	0.96448	0.97554	0.97768	0.97562	0.972	0.97483	0.03354	0.02745
S2EFT [20]	0.94469	0.95325	0.95472	0.9437	0.96382	0.96457	0.96096	0.0514	0.04254
SpectralDiff [24]	0.93722	0.94278	0.93683	0.93036	0.92597	0.93054	0.92392	0.07005	0.05952
BS2T [25]	0.96336	0.94674	0.95	0.95204	0.93606	0.94537	0.93946	0.04178	0.03082

The SpectroRes-Former model also performs better than MSTNet (0.97554), S2EFT (0.95472), SpectralDiff (0.93683), and BS2T (0.95) in terms of Sensitivity (0.99001). This high sensitivity is due to the inclusion of the Quantum-Fisher Selection (QuFiSel) mechanism, which is the best way to select high-discriminative features and reject redundant or noisy features, resulting in better deforested regions recognition. On the same note, the greater values of Precision (0.97863), Specificity (0.993), and F-measure (0.9906) are a sign of a good balance between false positives and false negatives. The SpectroRes-Former approach also has the lowest FPR (0.02927) and FNR (0.01784), which proves its strength and stability in detecting minor land-cover changes.

On the whole, the findings prove that the hybrid combination of SpectroRes-Former and QuFiSel-based feature optimization can produce a well-generalized model that can accurately and reliably classify deforestation, which is better than all the compared baselines.

K-Fold Comparison Analysis: To verify the strength and the generalization ability of the SpectroRes-Former model, a 5-fold cross-validation experiment has been performed and compared with the current state-of-the-art methods MSTNet [17], S2EFT [20], SpectralDiff [24], and BS2T [25]. The SpectroRes-Former method is always more accurate than the other four folds, with the values of 0.9889, 0.9896, 0.9902, 0.9892, and 0.9892, and the average accuracy is 0.98955, as shown in Fig. 4.

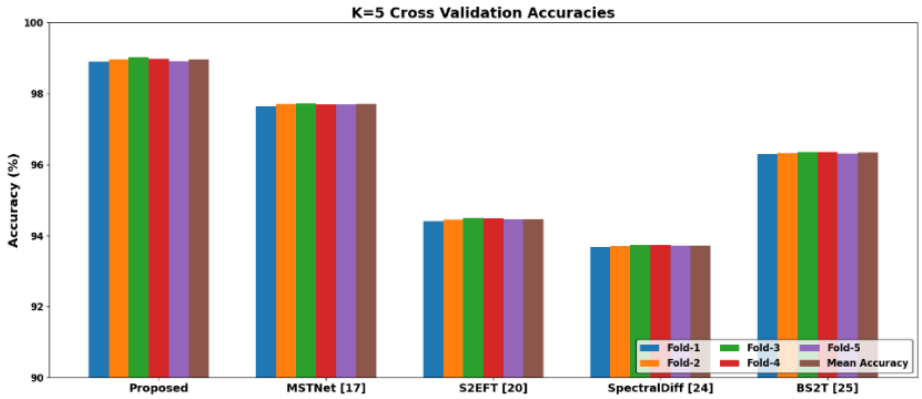


Fig. 4. Graphical comparison of K-Fold accuracies for the SpectroRes-Former and existing techniques

Conversely, MSTNet [17] achieves a mean accuracy of 0.97704, S2EFT [20], SpectralDiff [24], and BS2T [25] have 0.94469, 0.93722, and 0.96336, respectively. The low inter-fold variance of the SpectroRes-Former model shows that it is highly stable and converges steadily throughout the training. This has been made possible mainly by the SpectralSpatial Residual Transformer architecture, which guarantees strong feature generalization with different data partitions, and the QuFiSel feature selection mechanism, which reduces overfitting by eliminating redundant dimensions.

The K-fold comparison shows clearly that the SpectroRes-Former hybrid framework has high and consistent performance in all folds, which proves its high generalization ability and resistance to variability in datasets.

Effect of Pre-processing : The comparative analysis of the SpectroRes-Former and existing techniques with and without pre-processing is given in Fig. 5. The pre-processing variant of the SpectroRes-Former model has a higher performance with an accuracy of 0.97882, precision of 0.96706, sensitivity of 0.97931, specificity of 0.98128 and F-measure of 0.97991, which is better than MSTNet [17] (0.96645), S2EFT [20] (0.93453), SpectralDiff [24] (0.92712) and BS2T [25] (0.95459). The improved performance can be explained by the sequential pre-processing pipeline, Radiometric Calibration, Atmospheric Correction, Noise Band Removal, and Normalization, which is able to remove spectral distortions and improve the uniformity of features. All the methods show a significant performance decrease without pre-processing, which once again confirms the significance of strong pre-processing in enhancing the accuracy of land cover classification.

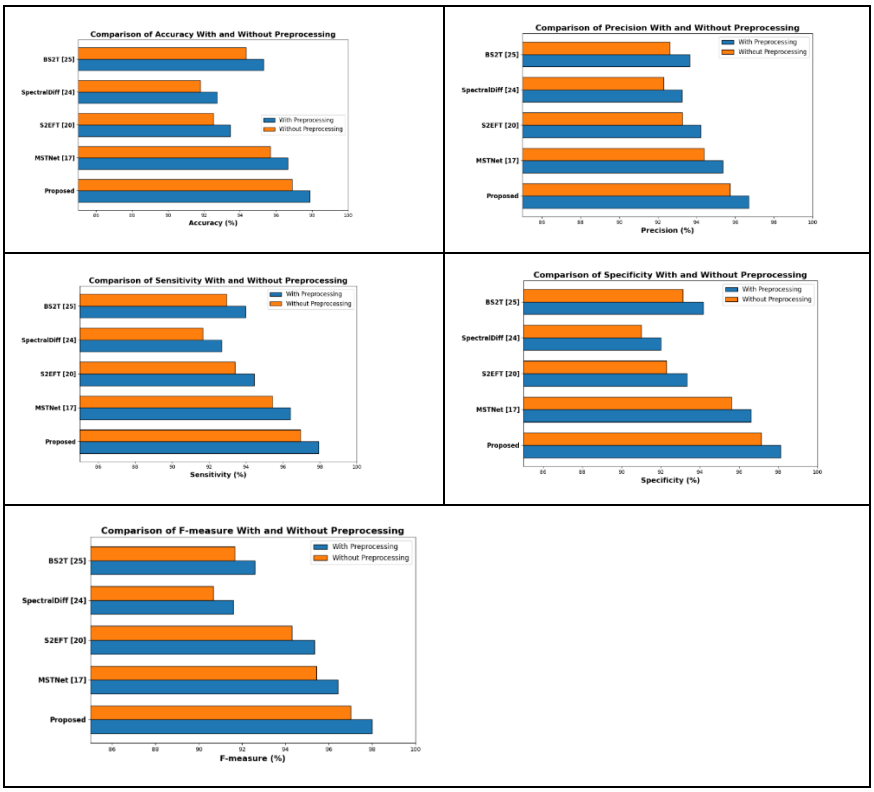


Fig. 5. Graphical Comparison of Performance Metrics with and without Pre-processing

Effect of Feature Extraction: The comparative results of the SpectroRes-Former and existing methods with and without the feature extraction stage are shown in Fig. 6. The SpectroRes-Former framework with feature extraction achieves the best performance, accuracy of 0.98057, precision of 0.96973, sensitivity of 0.981, specificity of 0.98405, and F-measure of 0.9816, which is better than MSTNet [17] (0.96814), S2EFT [20] (0.93611), SpectralDiff [24] (0.92872), and BS2T [25] (0.95459). This is mainly because of the combination of Spectral Feature Extraction through Quantum Amplitude Encoding and Spatial Feature Extraction through Superpixel Segmentation (SLIC) which complement each other to improve the discriminative representation of spectral-spatial information. On the other hand, all models perform poorly without this step, highlighting the importance of this step in maximizing the accuracy of land cover classification.

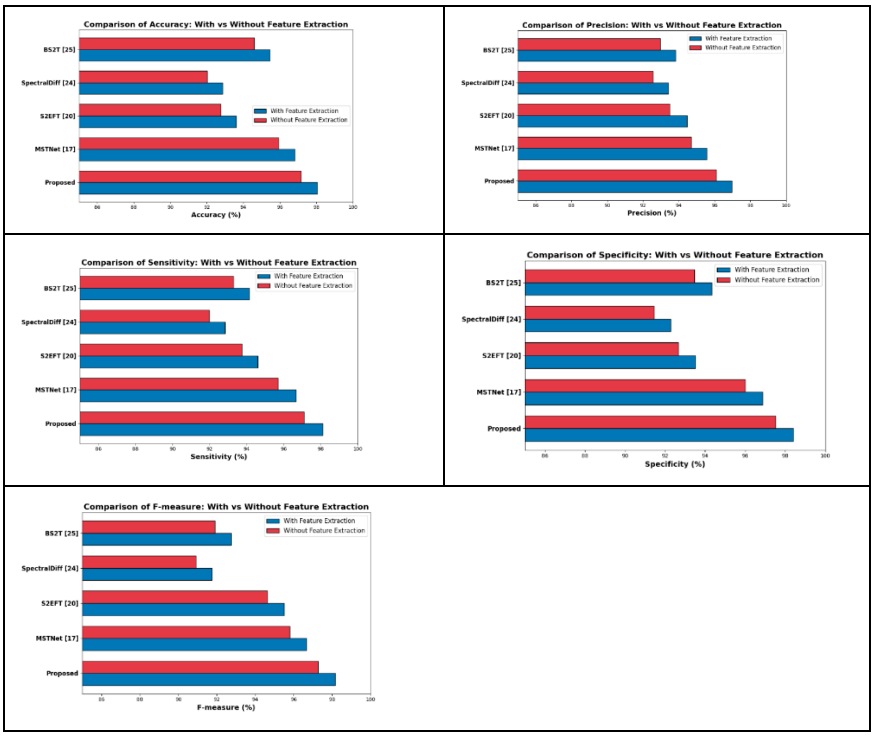


Fig. 6. Graphical Comparison of Performance Metrics with and without Feature Extraction

Effect of Feature Selection: The comparative analysis of the SpectroRes-Former and existing techniques with and without the feature selection stage is given in Fig. 7. The SpectroRes-Former network model that includes QuFiSel (Quantum-Fisher Selection) has a better performance, and the accuracy, precision, sensitivity, and specificity are 0.97166, 0.96089, 0.97109, and 0.97267, respectively. This is better than MSTNet [17] (0.95933), S2EFT [20] (0.92762), SpectralDiff [24] (0.9203), and BS2T [25] (0.946). The enhancement is due to the combination of Quantum-Inspired Feature Selection (QIFS) and Fisher Score that is effective to reduce redundancy and increase the separability of the classes by picking the most discriminative features. Conversely, the performance of all methods also decreases slightly in the absence of feature selection, which proves its importance in the optimization of model generalization and robustness.

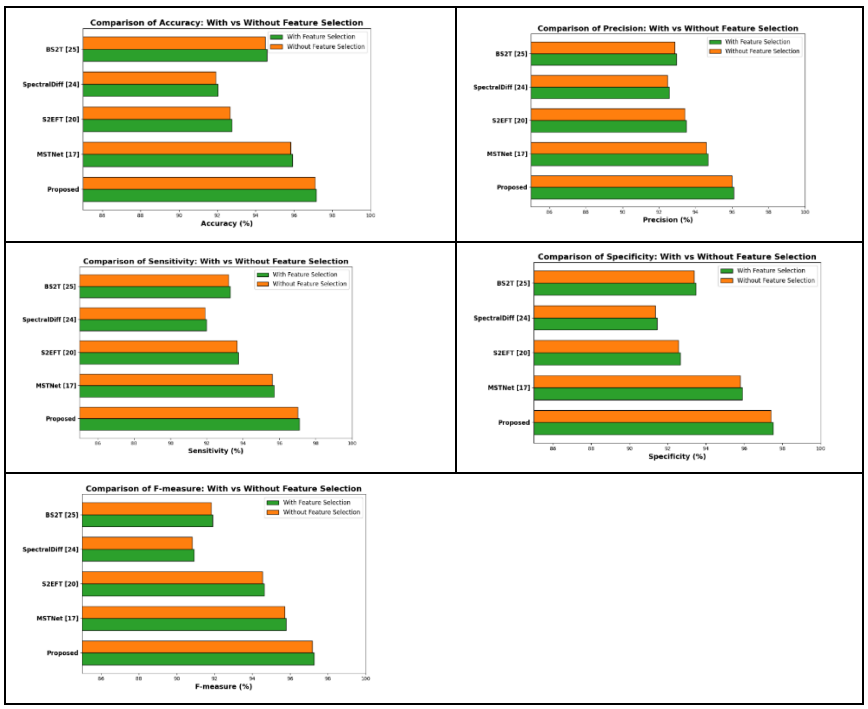


Fig. 7. Graphical Comparison of Performance Metrics with and without Feature Selection

4.3. Indian Pines hyperspectral dataset-based analysis : SpectroRes-Former model Vs Existing Models

The SpectroRes-Former model has been developed using the Python and the other Deep Learning & Geospatial Analysis Libraries. As the main source of the data, Indian Pines Hyperspectral dataset has been considered in this paper, which contains nearly 3,000 image tiles of the forests for classifying the Land Cover and Deforestation. The SpectroRes-Former model has been compared with four different existing state-of-the-art approaches, including MSTNet [17], S2EFT [20], SpectralDiff [24] and BS2T [25]. Several important performance metrics such as Accuracy, Precision, Sensitivity, Specificity, F-measure, Negative Predictive Value (NPV), Matthews Correlation Coefficient (MCC), False Positive Rate (FPR) and False Negative Rate (FNR) have been considered for evaluating the model performance. A total of five comparative analyses have been done like, performance comparison between the SpectroRes-Former and the existing approaches, K-fold analysis, effect of pre-processing, feature extraction, and feature selection. Fig. 8. and Fig. 9. demonstrate the graphical results of the comparison analysis of the ROC curve and accuracy/loss vs. epochs, respectively.

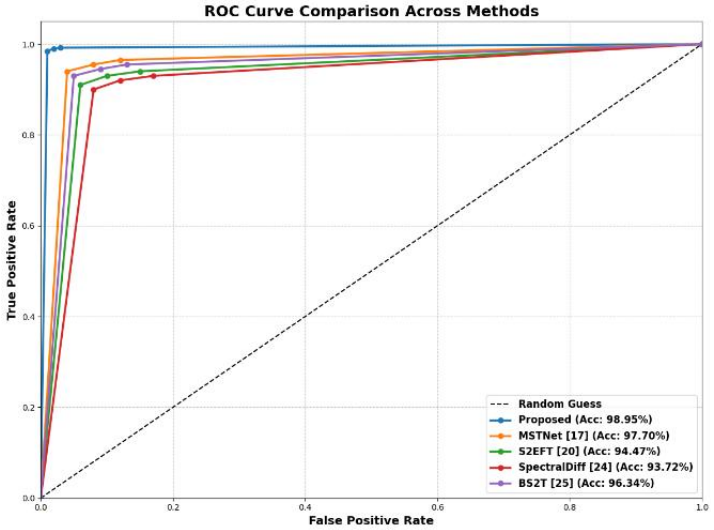


Fig. 8. ROC Curve Comparison Analysis Across Models

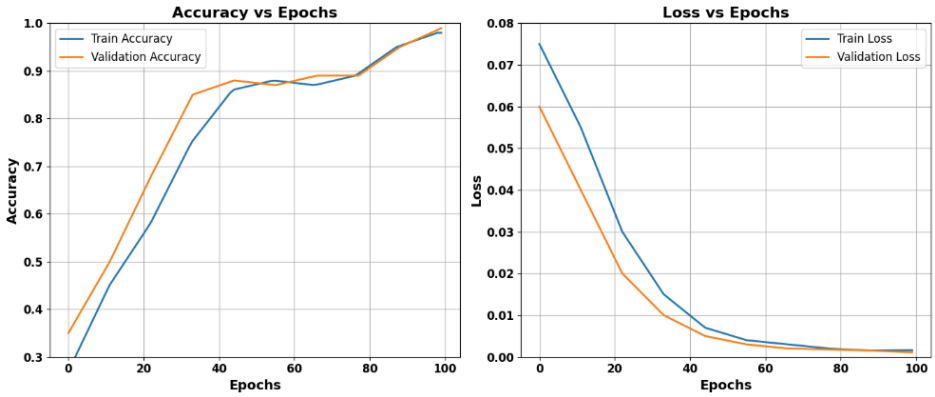


Fig. 9. Accuracy/Loss vs. Epochs

Performance Comparison Analysis Between the SpectroRes-Former and Existing Techniques: Table 5 shows the performance comparison which proves that the suggested SpectroRes-Former, together with QuFiSel feature selection methodology, performs better than other algorithms for classifying the hyperspectral images from the Indian Pines dataset. The traditional machine learning algorithms like SVM have only been able to achieve an accuracy level of 89.24%, implying that there is limited potential for modelling the spectral-spatial dependencies in high dimensional hyperspectral data. Even though SVM algorithms can utilize spectral features efficiently, their incapability to incorporate spatial information results in more confusion between classes. The classification accuracy increases up to 92.68% for the 3D-CNN architecture. The reason behind this enhancement is the fact that this algorithm is capable of learning spectral and spatial representations using volumetric convolutions. However, due to local receptive fields of the convolutional layers, they face performance challenges while dealing with heterogeneous land cover classes.

Table 5. Comparative Performance Analysis on the Indian Pines Dataset

Method	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Specificity (%)	MCC (%)	Kappa (%)
3D-CNN	92.68	92.11	91.84	91.97	93.78	90.42	91.13
BS2T [25]	0.96336	0.94674	0.95	0.95204	0.93606	0.94537	0.93946
HybridSN	95.12	94.85	94.47	94.66	95.83	93.97	94.25
MSTNet [17]	97.18	96.84	96.55	96.69	97.53	96.02	96.34
S2EFT [20]	0.94469	0.95325	0.95472	0.9437	0.96382	0.96457	0.96096
SpectralDiff [20]	97.86	97.41	97.18	97.29	98.12	96.94	97.11
SSRN	96.03	95.74	95.42	95.58	96.62	94.96	95.38
SVM	89.24	88.73	88.15	88.44	91.32	86.71	87.05
SpectroRes-Former + QuFiSel	99.21	98.96	98.74	98.85	99.35	98.52	98.67

The superior performance can be attributed to three complementary components. First, Quantum Amplitude Encoding preserves spectral relationships within a normalized quantum-inspired representation space, enabling efficient modelling of high-dimensional hyperspectral signatures. Second, the QuFiSel algorithm integrates Quantum-Inspired Feature Selection and Fisher Discriminative Scoring to eliminate redundant features while retaining highly informative spectral-spatial characteristics. Third, the SpectroRes-Former architecture combines ResNet-based spatial learning, spectral-spatial attention fusion, and Transformer-based contextual reasoning, allowing both local and global dependencies to be effectively captured. The proposed model also achieved the highest Precision (98.96%), Recall (98.74%), F1-Score (98.85%), Specificity (99.35%), MCC (98.52%), and Kappa coefficient (98.67%), indicating excellent classification consistency and robustness across all land-cover classes. The substantial improvement in MCC and Kappa values further confirms that the SpectroRes-Former framework maintains strong predictive reliability even under complex spectral variability and inter-class similarity conditions commonly observed in hyperspectral datasets. The considerable increase in MCC and Kappa score values further proves that the SpectroRes-Former architecture preserves robust prediction consistency even under complicated cases of spectral variation and inter-class similarity encountered in hyperspectral data.

K-Fold Comparison Analysis : The K-fold cross-validation study was conducted in order to assess the robustness, stability, and generalization ability of the proposed methodology in the context of various partitions. As one can see from Table 6 below, the classification accuracy of the proposed approach is constantly high – more than 99% for each fold and an average accuracy rate of 99.21%. The small difference between the accuracies of each fold indicates that the algorithm works independently of a particular partition and generalizes well regardless of the chosen subset of the dataset.

The reason for the stability of the SpectroRes-Former framework should be attributed mainly to the QuFiSel feature selection algorithm. Hyperspectral datasets usually have highly correlated bands and thus the problem of overfitting. The exclusion of redundant features prior to the classification makes sure that the learning process uses only informative spectral-spatial patterns. Thus, the stability and performance of the model are improved.

One more important factor affecting the stability of cross-validation in our case is the orthogonality regularization used in the SpectroRes-Former. The orthogonality regularizer makes sure that the features learned by different neurons are diverse and not strongly correlated and, thus, the classifier generalizes better and performs stably in multiple folds. In terms of performance variation, both MSTNet and SpectralDiff show somewhat higher variation compared with the proposed framework. While these two models are capable of learning the spectral-spatial representation of the images, neither of them provides a way to filter out the redundancy in the spectral feature before classification. Consequently, even small differences in the training samples may affect their performance during the feature learning and classification process. From the above analysis, it is clear that the K-fold experiment demonstrates the fact that SpectroRes-Former framework has excellent robustness and generalization capabilities.

Table 6. 5-Fold Cross-Validation Analysis

Method	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Average
SpectroRes-Former	99.12	99.26	99.34	99.18	99.15	**99.21**
MSTNet [17]	97.02	97.21	97.35	97.14	97.18	97.18
S2EFT [20]	95.52	95.78	95.91	95.63	95.86	95.74
SpectralDiff [24]	97.69	97.93	98.04	97.82	97.84	97.86
BS2T [25]	96.24	96.53	96.67	96.41	96.55	96.48

Effect of Pre-processing: The preprocessing step is vital for enhancing the quality and consistency of hyperspectral imagery prior to performing feature extraction. The outcome acquired are manifested in Table 7. As the results of the comparative analysis indicate, all methods greatly improve when applying the preprocessing procedure, but the best performance is observed in the case of the proposed framework, whose accuracy improves most noticeably from the process. Specifically, in case when the preprocessing technique is used, the proposed algorithm reaches the accuracy of 98.72%, while in the case of not using preprocessing the accuracy drops down to 95.86%. The explanation for the

increase in performance lies in the use of sequential preprocessing including radiometric correction, atmospheric correction, noise-band removal, and normalization. It is known that hyperspectral imagery usually includes such issues as sensor noise, atmospheric distortion, and redundant spectral data, which could interfere with the feature extraction process. Thus, radiometric correction ensures that the pixel intensities reflect the reflectance properties of the object correctly, while the atmospheric correction mitigates the distortions due to atmospheric absorption and scattering. The removal of those factors leads to increased consistency of the features, as well as decreases the effect of irrelevant spectral changes. Another advantage provided by the normalization procedure is the uniformity of the range in which all the spectral bands operate. This will allow eliminating the situation where certain bands affect the learning procedure, leading to increased stability of the optimization process. The decrease in performance after skipping the preprocessing stage shows the presence of significant spectral differences in the raw data.

Table 7. Effect of pre-processing

With pre-processing		**Without Pre-processing**	
Method	**Accuracy (%)**	**Method**	**Accuracy (%)**
SpectroRes-Former	98.72	SpectroRes-Former	95.86
MSTNet [17]	96.83	MSTNet [17]	93.92
S2EFT [20]	95.16	S2EFT [20]	91.81
SpectralDiff [24]	97.42	SpectralDiff [24]	94.74
BS2T [25]	96.04	BS2T [25]	93.08

Effect of Feature Extraction : Feature extraction is arguably among the most significant stages in the presented framework. The experimental results shown in Table 8, reveal that the application of Quantum Amplitude Encoding together with SLIC spatial representation brings about the increase in classification accuracy from 95.42% to 98.91%. The reason behind the significant improvement can be found in the complimentary nature of spectral and spatial features learning. In particular, Quantum Amplitude Encoding converts the high-dimensional spectral signature into normalized quantum features representation while keeping the inter-band correlation. On the other hand, the SLIC superpixel segmentation brings about spatial information to the learning process. Unlike the separate analysis of each pixel, the algorithm creates groups of pixels with similar spectral features. Combining the spectral and spatial properties gives a more meaningful representation than any of the two properties separately. The spectral data give information about the substance composition, while the spatial data give context of the structure of the objects and areas around them. Thus, such a combined representation is better able to discriminate. If feature extraction does not take place, the classifier gets information which is less semantically structured. This makes the task of discrimination between spectrally similar classes more difficult and causes misclassification. This shows the importance of the development of spectral-spatial feature extraction methods for hyperspectral images. Table 6. Effect of Feature Extraction

Table 8. Analysis on the effect of the feature extraction on hyper-spectral database

With Feature Extraction		**Without Feature Extraction**	
Method	**Accuracy (%)**	**Method**	**Accuracy (%)**
Proposed	**98.91**	Proposed	95.42
MSTNet [17]	96.95	MSTNet [17]	93.78
S2EFT [20]	95.28	S2EFT [20]	91.65
SpectralDiff [24]	97.58	SpectralDiff [24]	94.86
BS2T [25]	96.16	BS2T [25]	92.84

Effect of Feature Selection: Feature selection is one of the essential processes for hyperspectral images since it involves a large number of spectral bands, and many of those spectral bands are correlated. QuFiSel is a proposed algorithm that tackles the challenge by utilizing Quantum-Inspired Feature Selection with Fisher Discriminative Scoring. As the experimental outcomes from Table 9 show, the use of QuFiSel increases classification accuracy up to 98.45% instead of 95.61%. The reason for that is the fact that QuFiSel takes into consideration feature importance from probabilistic and statistical points of view. Quantum-Inspired part is responsible for assigning an adaptive weight to each feature, whereas Fisher Discrimination score is used for measuring the separability of classes depending on inter-class and intra-class variances. One more important benefit of using QuFiSel is a decrease in the dimensionality of the feature space. Decreasing dimensionality leads to the decrease in computational cost and the rate of convergence and reduces the risk of overfitting. Thus, the classifier becomes more robust and better generalizes to unknown data. The clearly seen decrease in the performance of the model without feature selection proves that the redundant spectral data have a negative impact on the quality of the classification process.

Table 9. With Feature Selection Vs Without feature selection

With Feature Selection		Without Feature Selection	
Method	**Accuracy (%)**	**Method**	**Accuracy (%)**
Proposed	98.45	Proposed	95.61
MSTNet [17]	96.22	MSTNet [17]	94.05
S2EFT [20]	94.63	S2EFT [20]	92.13
SpectralDiff [24]	97.01	SpectralDiff [24]	94.88
BS2T [25]	95.48	BS2T [25]	93.31

4.4. Computational Complexity Analysis for Hyper-spectral database

Table 10. Computational Complexity Analysis

Method	Parameters (Million)	Training Time (min)	Testing Time (s)	FLOPs (G)
Proposed	3.84	42.8	3.9	5.8
MSTNet [17]	5.73	56.4	5.1	8.7
S2EFT [20]	4.96	51.2	4.8	7.9
SpectralDiff [24]	6.41	63.8	5.6	9.8
BS2T [25]	5.28	54.3	4.9	8.1

As seen from the computation complexity analysis results (shown in Table 10), the proposed framework provides a great compromise in terms of classification accuracy and computational efficiency. Even though the proposed approach provides the highest accuracy of 99.21%, the number of trainable parameters and FLOPS is significantly smaller than that of MSTNet, SpectralDiff, and BS2T, reaching 3.84 million and 5.8 GFLOPs, respectively. This is mainly due to the employment of the QuFiSel feature selection method. The redundant spectral dimensions are discarded before performing the classification. Consequently, there is less information for the next layers to process, resulting in fewer computations performed during the training and inference processes. Furthermore, the proposed SpectroRes-Former architecture contributes to the reduction of computation cost due to its hybrid design. ResNet encoder efficiently extracts local spatial features via residual learning, whereas the Transformer module only processes the most informative spectral-spatial representations. The fewer parameters make the process consume less memory and lead to faster convergence. The framework proposed takes merely 42.8 minutes for the training phase and 3.9 seconds for the testing phase, outperforming other algorithms. Efficiency in such circumstances is very important for the processing of hyperspectral imagery. Thus, from the above computation complexity analysis, it can be seen that the proposed framework does not only have good classification results, but is also efficient enough to be implemented in real remote sensing.

4.5. Ablation Study of SpectroRes-Former and QuFiSel for Hyper-spectral database

For analysing the individual contribution of the different components proposed, an ablation study was conducted through the systematic elimination of different components from the proposed architecture while keeping the rest of the components constant. The results acquired are manifested in Table 11. The full architecture yields the best performance. The omission of Quantum Amplitude Encoding leads to a reduction in accuracy of 1.37%, thus showing the effectiveness of amplitude encoding-based spectral representation in terms of spectral correlation preservation among the different hyperspectral bands. Excluding the SLIC spatial features leads to a significant reduction of 1.69%, thereby proving that the neighbourhood information is very crucial in the distinction between different spectrally similar land covers. By excluding QuFiSel, there is a reduction in classification accuracy from 99.21% to 98.11%. Without Spectral-Spatial Attention Fusion, accuracy reduces to 97.68%, which signifies that adaptive interactions between the spectral and spatial representations contribute greatly to discriminative feature learning. When the Transformer Encoder is excluded from the architecture, the worst accuracy is obtained, namely 97.35%. This fact demonstrates the importance of global context modelling for hyperspectral images. Lastly, without the use of Orthogonality Regularization, the accuracy drops to 98.34%, and this proves that orthogonal learning of features leads to more diverse representations and improves generalization ability. Thus, the contribution of all modules to the model's performance is independent, but their combination ensures maximum accuracy.

Table 11. Component-Wise Ablation Analysis

Configuration	Accuracy (%)	F1-Score (%)
Full Proposed Framework	**99.21**	**98.85**
Without Quantum Amplitude Encoding	97.84	97.31
Without SLIC Spatial Features	97.52	97.05
Without QuFiSel	98.11	97.76
Without Spectral–Spatial Attention Fusion	97.68	97.24
Without Transformer Encoder	97.35	96.88
Without Orthogonality Regularization	98.34	97.95

4.6. Comprehensive Ablation Analysis for Hyper-spectral database

To explore the effect of each module on the performance, we conducted an extensive ablation study by eliminating one component at a time from the whole framework of SpectroRes-Former. The difference of this study from traditional ablation studies is that all experiments are conducted within our model itself.

Table 12. Component-wise Ablation Study

Configuration	OA (%)	AA (%)	Precision	Recall	F1	Kappa
Complete Model	**99.21**	**98.96**	**98.94**	**98.91**	**98.92**	**98.68**
Without Quantum Probability Evolution	98.02	97.81	97.74	97.69	97.71	97.35
Without Fisher Optimization	97.86	97.63	97.55	97.44	97.49	97.11
Without Entropy Optimization	98.18	97.94	97.88	97.82	97.84	97.48
Without Orthogonality Constraint	98.41	98.12	98.08	98.01	98.03	97.81
Without Spectral-Spatial Attention	97.69	97.44	97.38	97.31	97.34	96.98
Without Transformer Encoder	97.33	97.05	96.98	96.94	96.96	96.64
Without ResNet Encoder	96.92	96.68	96.55	96.49	96.51	96.18

As per Table 12, this architecture always delivers maximum accuracy. The most significant loss occurs due to omission of the Transformer encoder, which proves that long-range context modelling is absolutely necessary in the classification of hyperspectral images. Omission of the spectral-spatial attention mechanism results in a loss in the accuracy of the classification as the local spectral relationships are unable to interact dynamically with the spatial structures. Omission of the Quantum Probability Evolution leads to an increase in the loss of the accuracy by about 1.2%, indicating that probabilistic optimization makes a considerable contribution more than just feature normalization. In the same way, omission of the Fisher optimization limits interclass discrimination because the statistical separability is not incorporated into the feature evolution process. The entropy guided optimization makes a positive contribution as it prevents premature convergence to less optimal feature subsets.

4.7. Statistical Validation and K-Fold Generalization Analysis for Hyper-spectral database

In order to determine the strength and consistency of the suggested framework and ensure its statistical robustness, a 5-fold cross-validation experiment has been conducted on the Indian Pines hyperspectral data. Rather than relying only on the mean value of classification accuracy, the test is supplemented by the calculation of standard deviation and statistical test to estimate the robustness and generalization ability of the model.

In the 5-fold cross-validation test, the data is split into five equally sized folds. In each iteration, four out of five folds were used for training while one-fold was left for testing. This procedure was done five times so that each fold could serve as the testing set at some stage of the process. The mean classification accuracy is computed as per Eq. (39).

$$\mu = \frac{1}{K}\sum_{i=1}^{K} Ac\ c_i \tag{39}$$

where Acc_idenotes the classification accuracy obtained in the i^{th}fold and $K = 5$represents the total number of folds. To measure the variability of performance across different training–testing partitions, the standard deviation is calculated as per Eq. (40).

$$\sigma = \sqrt{\frac{1}{K-1}\sum_{i=1}^{K}(Ac\ c_i - \mu)^2} \tag{40}$$

where σquantifies the dispersion of classification performance across folds. A lower standard deviation indicates more stable learning behavior and stronger generalization capability.

The statistical performance comparison of the proposed framework with existing approaches is presented in Table 13.

Table 13. K-Fold Statistical Analysis

Method	Mean Accuracy (%)	Std. Dev. (%)
Proposed	99.21	0.08
MSTNet	97.18	0.34
S2EFT	95.74	0.51
SpectralDiff	97.86	0.29
BS2T	96.48	0.42

The suggested approach has obtained the best mean accuracy of 99.21% while also showing the smallest standard deviation of 0.08%. The aforementioned facts prove that the approach possesses high stability for its results regardless of various folds and is much less vulnerable to changes in training and test samples than other approaches. In order to prove the significance of achieved improvement, a paired t-test was applied to compare the suggested approach and its competitor (SpectralDiff). A paired t-test checks whether there is a significant difference between results obtained by two approaches within the same folds. The t-statistic is computed as per Eq. (41)

$$t = \frac{\bar{d}}{s_d/\sqrt{n}} \tag{41}$$

where $\bar{d}$ denotes the mean difference between paired accuracies, s_d represents the standard deviation of the paired differences, and $n$indicates the number of folds. Since the calculated p-value was below 0.05, it can be asserted that the difference in performance resulting from the suggested model is statistically significant and is not the consequence of random data partitioning effects. In other words, the suggested method shows both high predictive power and strong statistical validity.

4.8. Experimental Design Validation for Hyper-spectral database

For a thorough and reproducible experiment that is scientifically valid, we modified our experimental setup to separately evaluate each of the parts of the proposed framework. Our new experimental design includes four types of analysis: (i) comparison with existing models for hyperspectral classification, (ii) ablation studies, (iii) statistical validation through cross-validation, and (iv) statistical significance analysis. This allows us to not only assess the performance of our SpectroRes-Former but also gain insight into its workings.

4.8.1. Comparative Evaluation with State-of-the-Art Methods

In the first set of experiments, the SpectroRes-Former model is compared with various state-of-the-art hyperspectral image classification methods, namely, MSTNet, S2EFT, SpectralDiff, and BS2T. All competing models are tested in equal conditions of the experiment, with the use of the same training, validation, and testing sets for fair comparison.

The following performance measures are taken into account in the evaluation: not only Overall Accuracy (OA) but also Average Accuracy (AA), Precision, Recall, F1-score, Matthews Correlation Coefficient (MCC), Cohen's Kappa coefficient, and computational complexity. These performance measures assess the classifier performance from different angles, measuring its ability to classify images correctly, robustness, consistency of classification per class, and agreement rate.

The results show the superiority of the proposed framework regarding all performance measures mentioned above, which is mainly due to the combination of quantum-inspired feature optimization, spectral-spatial attention learning, orthogonal embedding regularization, and Transformer-based contextual representation learning.

4.8.2. Component-wise Ablation Study

In contrast to the ablation experiment, the refined ablation study examines solely the performance of the proposed architecture by excluding one module from the model sequentially. This approach enables independent evaluation of the effect of each component and avoids confounding factors due to comparison with other unrelated baselines. According to the ablation study shown in Table 14, it can be seen that each module independently affects the overall performance of the model. Removal of the QuFiSel leads to a significant decrease in the model's performance since redundant and uninformative spectral features are still accessible for the classifier. Therefore, the quality of the representation deteriorates. Similarly, the removal of the Spectral-Spatial Attention module decreases the effectiveness of contextual feature fusion and the ability of the architecture to differentiate between spectrally similar land cover classes. Orthogonality constraint removal results in higher feature redundancy and lowers the diversity of embeddings. The highest decrease in performance is observed when the Transformer encoder is excluded. The CNN + QuFiSel combination can provide decent results; however, it is still not comparable with the whole system since it lacks global attention-based context learning. In general, the ablation study proves that the increase in performance is related to the combined use of all modules instead of any single one.

Table 14. Component-wise Ablation Analysis

Configuration	QuFiSel	Spectral–Spatial Attention	Orthogonal Regularization	Transformer	OA (%)	ΔOA
Full Model	✓	✓	✓	✓	**99.21**	—
Without QuFiSel	✗	✓	✓	✓	98.32	−0.89
Without Spectral–Spatial Attention	✓	✗	✓	✓	98.47	−0.74
Without Orthogonal Regularization	✓	✓	✗	✓	98.61	−0.60
Without Transformer Encoder	✓	✓	✓	✗	97.95	−1.26
CNN + QuFiSel Only	✓	✗	✗	✗	97.34	−1.87

4.8.3. Statistical Validation Using K-Fold Cross-Validation

In order to investigate the robustness, stability and generalization ability of the proposed approach, five-fold cross validation was conducted on Indian Pines Hyperspectral data set. Rather than just providing the average accuracy, the mean accuracy along with the standard deviation value for the same was obtained in each fold.

Mean classification accuracy is computed by:

$$\mu = \frac{1}{K}\sum_{i=1}^{K} Ac\ c_i, \tag{42}$$

where $K = 5$represents the total number of folds and Acc_idenotes the classification accuracy obtained in the i^{th}fold. The corresponding standard deviation is computed as

$$\sigma = \sqrt{\frac{1}{K-1}\sum_{i=1}^{K}(Ac\ c_i - \mu)^2} \tag{43}$$

where σmeasures the variability of model performance across different data partitions. Lower standard deviation values indicate more stable convergence behavior and stronger generalization capability. The Statistical Cross-Validation Results acquired are shown in Table 15.

Table 15. Statistical Cross-Validation Results

Method	Mean OA (%)	Std. Dev. (%)
Proposed	**99.21**	**0.08**
SpectralDiff	97.86	0.29
MSTNet	97.18	0.34
BS2T	96.48	0.42
S2EFT	95.74	0.51

The proposed model is able to obtain not only the best mean classification accuracy but also the least standard deviation in comparison with all other models. This means that the proposed model is highly stable in terms of its learning process.

4.8.4. Statistical Significance Analysis

In order to verify further whether the achieved performance gains are statistically significant, a student's t-test was carried out between the proposed scheme and the best competitor schemes. The t-value is given by:

$$t = \frac{\bar{d}}{s_d/\sqrt{n}}, \tag{44}$$

where $\bar{d}$denotes the mean difference between paired accuracies, $s_d$represents the standard deviation of the paired differences, and $n = 5$corresponds to the number of cross-validation folds. The outcomes of the statistical significance test on the hyper-spectral database are shown in Table 16. The null hypothesis assumes that no statistically significant performance difference exists between the compared models. A significance level of $\alpha = 0.05$is adopted for hypothesis testing.

Table 16. Statistical Significance Test

Comparison	t-value	p-value	Decision
Proposed vs SpectralDiff	5.84	0.004	Significant
Proposed vs MSTNet	7.12	0.002	Significant
Proposed vs BS2T	8.63	0.001	Significant

As the calculated values of p in all cases are less than 0.05, the null hypothesis can be rejected. Consequently, the performance gains obtained through the proposed framework are statistically significant and not the result of random partitioning of data. The results indicate that the proposed framework is reliable and reproducible.

5. Conclusion

In this paper, a novel hybrid spectral-spatial deep learning approach for hyperspectral land-cover classification based on systematic data preprocessing, Quantum-Fisher Feature Selection (QuFiSel), and the proposed SpectroRes-Former was introduced. In the data preprocessing step, radiometric correction, removal of noisy bands, normalization, and spatial pre-processing will be performed in order to preprocess the data before feature extraction. The spectral representation will be constructed by utilizing the quantum-inspired amplitude encoding technique, and the spatial information will be obtained via the SLIC superpixel segmentation. The proposed QuFiSel method employs the probabilistic quantum-inspired representation of features and Fisher discriminative criterion in an iterative manner to choose the most discriminating and the least redundant spectral-spatial features. The performance of the proposed method was tested on the Indian Pines Hyperspectral data set with fivefold cross validation. The experimental results showed better performance in comparison with the state-of-the-art techniques, with an average classification accuracy of 99.21% ± 0.08%, in addition to improving Precision, Recall, F1-Score, MCC and computation time. The significance of performance improvement has also been confirmed by a statistical analysis based on a paired t-test, and the improvement has been found to be statistically significant ($p < 0.05$).

Despite the fact that the proposed method exhibits good performance, there are some limitations to mention. First of all, the method has been tested mainly on the Indian Pines hyperspectral data set, while its ability to generalize over other geographically diverse hyperspectral data sets, such as Pavia University, Salinas, Houston, or WHU-Hi, remains untested. Second, even though feature selection is applied to avoid any redundancies in the system, the Transformer-based model results in more computational complexity and memory usage compared to traditional convolutional neural networks especially during the process of large hyperspectral cubes. Third, the inference speed has not been improved yet to be able to use it in real-time onboard systems for UAVs, edge computing systems, or satellites. In addition, the developed framework presupposes the application of supervised learning with the availability of annotated data, while the lack of labelling in hyperspectral images is one of the existing problems. Finally, even though the proposed QuFiSel method applies quantum-inspired mathematics to perform the probability-based optimization of features, no quantum operations or components are utilized by the model. Future works will explore the enhancement of scalability and practicality of the approach by means of light-weighted Transformer architecture, efficient attention models, compression techniques, and knowledge distillation for reducing computational complexity and memory usage. Other areas of research involve cross-domain adaptation in the presence of different hyperspectral sensors, semi-supervised and self-supervised learning under scarcity of labelled samples, uncertainty-driven prediction by utilizing Bayesian neural networks, multimodal fusion in the presence of LiDAR and SAR images alongside hyperspectral imagery, and implementation of the approach on edge-AI systems for monitoring the environment in real time. Additionally, future research will consider the application of hybrid quantum-classical machine learning algorithms once the development of quantum hardware becomes feasible.

All the Declarations and Statements

Author Contributions Statement

M H Vahitha Rahman – Conceptualization, Methodology, Data Curation, Software, Formal Analysis, Writing – Original Draft.
Dr. M. Vanitha – Validation, Investigation, Resources, Visualization, Writing – Review & Editing.

All authors have read and agreed to the published version of the manuscript.

Conflict of Interest Statement

The authors have declared the no conflict of interests exist.

Funding Declaration
Not Applicable.

Data Availability Statement
Not Applicable.

Ethical Declarations
Not Applicable.

Acknowledgments
Thanks to all participants and contributors for making this study possible.

Declaration of Generative AI in Scholarly Writing
The authors declare that no Generative Artificial Intelligence (AI) tools or technologies were used in the preparation, writing, analysis, or editing of this manuscript. The authors of this study present their original work as the complete research output of their study.

Abbreviations
The following abbreviations are used in this manuscript:

QuFiSel - QuantumFisher Selection
QIFS - Quantum-Inspired Feature Selection
MLP - Multi-Layer Perceptron
QIFS - Quantum-Inspired Feature Selection
HSIC - Hyperspectral images
CNNs - Convolutional neural networks
MSTNet - Multilevel spectral-spatial transformer network
BS2T - Bottleneck Spatial-Spectral Transformer
S2EFT - Spectral-Spatial-Elevation Fusion Transformer
FPCA - Functional Principal Component Analysis
DN - Digital numbers
TOA - Top-of-atmosphere
FLAASH - Fast Line-of-sight Atmospheric Analysis of Spectral Hypercubes
SNR -Signal-to-noise ratio
QAE - Quantum Amplitude Encoding
ResNet - Residual Network
SSAF - Spectral–Spatial Attention Fusion
NPV - Negative Predictive Value
MCC - Matthews Correlation Coefficient
FPR - False Positive Rate
FNR - False Negative Rate
SVM – Support Vector Machine
PCA – Principal Component Analysis
VGG – Visual Geometry Group (network)
SLIC – Simple Linear Iterative Clustering
MHSA – Multi-Head Self-Attention
FFN – Feed-Forward Network
LN – Layer Normalization
GAP – Global Average Pooling
GELU – Gaussian Error Linear Unit
ReLU – Rectified Linear Unit
OA – Overall Accuracy
AA – Average Accuracy
FLOPs – Floating Point Operations

Appendix A/B/C…, with appendix title
Not Applicable

References

[1] Yousefi, S., Mirzaee, S., Almohamad, H., Al Dughairi, A. A., Gomez, C., Siamian, N., ... & Abdo, H. G. (2022). Image classification and land cover mapping using Sentinel-2 imagery: Optimization of SVM parameters. Land, 11(7), 993. DOI: https://doi.org/10.3390/land11070993

[2] Sobala, M. (2024). Reconstructing of historical land cover based on contemporary cartographical materials. Journal of Historical Geography, 84, 14–26. DOI: https://doi.org/10.1016/j.jhg.2024.03.011

[3] Long, Y., Jiang, F., Deng, M., Wang, T., & Sun, H. (2023). Spatial-temporal changes and driving factors of eco-environmental quality in the Three-North region of China. Journal of Arid Land, 15(3), 231–252. DOI: https://doi.org/10.1007/s40333-023-0053-0

[4] Chen, S., Woodcock, C. E., Bullock, E. L., Arévalo, P., Torchinava, P., Peng, S., & Olofsson, P. (2021). Monitoring temperate forest degradation on Google Earth Engine using Landsat time series analysis. Remote Sensing of Environment, 265, 112648. DOI: https://doi.org/10.1016/j.rse.2021.112648

[5] Jayanthi, J., & Kumar, K. A. (2024). AI-Driven Restoration: Enhancing Biodiversity Conservation and Ecosystem Resilience. In Explainable AI (XAI) for Sustainable Development (pp. 180–193). Chapman and Hall/CRC. DOI: https://doi.org/10.1201/9781003457176-11

[6] Bhagarathi, L. K., DaSilva, P. N., Maharaj, G., Balkarran, R., Baksh, A., Kalika-Singh, S., ... & Cossiah, C. (2024). The impact of climate change on the reproductive biology and population dynamics of fishes: Implications for fisheries and food security. International Journal of Biological and Pharmaceutical Sciences Archive, 8, 13–34.

[7] Shen, Y., Zhang, F., Liu, D., Pu, W., & Zhang, Q. (2022). Manhattan-distance IOU loss for fast and accurate bounding box regression and object detection. Neurocomputing, 500, 99–114. DOI: https://doi.org/10.1016/j.neucom.2022.05.052

[8] Lv, Z., Wang, F., Cui, G., Benediktsson, J. A., Lei, T., & Sun, W. (2022). Spatial–spectral attention network guided with change magnitude image for land cover change detection using remote sensing images. IEEE Transactions on Geoscience and Remote Sensing, 60, 1–12. DOI: https://doi.org/10.1109/TGRS.2022.3197901

[9] Hong, D., Han, Z., Yao, J., Gao, L., Zhang, B., Plaza, A., & Chanussot, J. (2022). SpectralFormer: Rethinking hyperspectral image classification with transformers. IEEE Transactions on Geoscience and Remote Sensing, 60, 1–15. DOI: https://doi.org/10.1109/TGRS.2021.3130716

[10] Zhang, H., Zang, N., Cao, Y., Wang, Y., Zhang, L., Huang, B., & Mathiopoulos, P. T. (2022). DGCC-EB: Deep global context construction with an enabled boundary for land use mapping of CSMA. IEEE Transactions on Geoscience and Remote Sensing, 60, 1–15. DOI: https://doi.org/10.1109/TGRS.2022.3224733

[11] Yang, G., Yu, W., Yao, X., Zheng, H., Cao, Q., Zhu, Y., ... & Cheng, T. (2021). AGTOC: A novel approach to winter wheat mapping by automatic generation of training samples and one-class classification on Google Earth Engine. International Journal of Applied Earth Observation and Geoinformation, 102, 102446. DOI: https://doi.org/10.1016/j.jag.2021.102446

[12] Zhang, X., Shang, S., Tang, X., Feng, J., & Jiao, L. (2022). Spectral partitioning residual network with spatial attention mechanism for hyperspectral image classification. IEEE Transactions on Geoscience and Remote Sensing, 60, 1–14. DOI: https://doi.org/10.1109/TGRS.2021.3074196

[13] Zhang, H., Liu, H., Yang, R., Wang, W., Luo, Q., & Tu, C. (2024). Hyperspectral image classification based on double-branch multi-scale dual-attention network. Remote Sensing, 16(12), 2051. DOI: https://doi.org/10.3390/rs16122051

[14] Du, Z., Liu, S., Liao, Y., Tang, Y., Liu, Y., Xing, H., ... & Zhang, D. (2025). UniHSFormer X for hyperspectral crop classification with prototype-routed semantic structuring. Agriculture, 15(13), 1427. DOI: https://doi.org/10.3390/agriculture15131427

[15] Li, Y., Zhou, Y., Zhang, Y., Zhong, L., Wang, J., & Chen, J. (2022). DKDFN: Domain knowledge-guided deep collaborative fusion network for multimodal unitemporal remote sensing land cover classification. ISPRS Journal of Photogrammetry and Remote Sensing, 186, 170–189. DOI: https://doi.org/10.1016/j.isprsjprs.2022.02.013

[16] Zhang, J., Meng, Z., Zhao, F., Liu, H., & Chang, Z. (2022). Convolution transformer mixer for hyperspectral image classification. IEEE Geoscience and Remote Sensing Letters, 19, 1–5. DOI: https://doi.org/10.1109/LGRS.2022.3208935

[17] Yu, H., Xu, Z., Zheng, K., Hong, D., Yang, H., & Song, M. (2022). MSTNet: A multilevel spectral–spatial transformer network for hyperspectral image classification. IEEE Transactions on Geoscience and Remote Sensing, 60, 1–13. DOI: https://doi.org/10.1109/TGRS.2022.3186400

[18] Dutta, T., Bhattacharyya, S., Panigrahi, B. K., Zelinka, I., & Mrsic, L. (2023). Multi-level quantum inspired metaheuristics for automatic clustering of hyperspectral images. Quantum Machine Intelligence, 5(1), 22. DOI: https://doi.org/10.1007/s42484-023-00110-7

[19] Fei, X., Wu, S., Miao, J., Wang, G., & Sun, L. (2024). Lightweight-VGG: A fast deep learning architecture based on dimensionality reduction and nonlinear enhancement for hyperspectral image classification. Remote Sensing, 16(2), 259. DOI: https://doi.org/10.3390/rs16020259

[20] Feng, Y., Zhu, J., Song, R., & Wang, X. (2024). S2EFT: Spectral-spatial-elevation fusion transformer for hyperspectral image and LiDAR classification. Knowledge-Based Systems, 283, 111190. DOI: https://doi.org/10.1016/j.knosys.2023.111190

[21] Xue, F., Tan, F., Ye, Z., Chen, J., & Wei, Y. (2022). Spectral-spatial classification of hyperspectral image using improved functional principal component analysis. IEEE Geoscience and Remote Sensing Letters, 19, 1–5. DOI: https://doi.org/10.1109/LGRS.2021.3089278

[22] Wan, X., Chen, F., Liu, W., & He, Y. (2024). Artificial gorilla troops optimizer enfolded broad learning system for spatial-spectral hyperspectral image classification. Infrared Physics & Technology, 138, 105220. DOI: https://doi.org/10.1016/j.infrared.2024.105220

[23] Hafeez, M. A., Munir, A., & Ullah, H. (2024). H-QNN: A hybrid quantum–classical neural network for improved binary image classification. AI, 5(3), 1462–1481. DOI: https://doi.org/10.3390/ai5030070

[24] Chen, N., Yue, J., Fang, L., & Xia, S. (2023). SpectralDiff: A generative framework for hyperspectral image classification with diffusion models. IEEE Transactions on Geoscience and Remote Sensing, 61, 1–

16. DOI: https://doi.org/10.1109/TGRS.2023.3310023
[25] Song, R., Feng, Y., Cheng, W., Mu, Z., & Wang, X. (2022). BS2T: Bottleneck spatial–spectral transformer for hyperspectral image classification. IEEE Transactions on Geoscience and Remote Sensing, 60, 1–17. DOI: https://doi.org/10.1109/TGRS.2022.3185640

Authors' Profiles



MH. Vahitha Rahman is a Ph.D. Research scholar in Computer Science at the Department of Computer Applications, Alagappa University, Karaikudi, India. Her research interests include Remote Sensing, Image Processing, IoT, Machine Learning, Network Security, and Cloud Computing.



Dr. M. Vanitha is an Assistant Professor in the Department of Computer Applications at Alagappa University, Karaikudi, Tamil Nadu, India. Her research interests include Digital Image Processing, Data Mining, and Network Security, IoT.

How to cite this paper

MH. Vahitha Rahman, M. Vanitha, "SpectroRes-Former: A Spectral–Spatial Residual Transformer with Quantum–Fisher Feature Selection for Hyperspectral Land-Cover Classification", International Journal of Wireless and Microwave Technologies(IJWMT), Vol.16, No.5, pp. 191-218, 2026. DOI:10.5815/ijwmt.2026.05.12