

## Article

# RS-Net: Hyperspectral Image Land Cover Classification Based on Spectral Imager Combined with Random Forest Algorithm

Xuyang Li <sup>1</sup> , Xiangsuo Fan <sup>1,2,\*</sup> , Qi Li <sup>1</sup> and Xueqiang Zhao <sup>3,4</sup><sup>1</sup> School of Automation, Guangxi University of Science and Technology, Liuzhou 545006, China; 221077062@stdmail.gxust.edu.cn (X.L.); liqi@gxust.edu.cn (Q.L.)<sup>2</sup> Guangxi Collaborative Innovation Centre for Earthmoving Machinery, Guangxi University of Science and Technology, Liuzhou 545006, China<sup>3</sup> School of Geography and Planning, Sun Yat-Sen University, Guangzhou 510275, China; zhaoxq28@mail2.sysu.edu.cn<sup>4</sup> China Water Resources Pearl River Planning Surveying and Designing Co., Ltd., Guangzhou 510610, China

\* Correspondence: 100002085@gxust.edu.cn

**Abstract:** Recursive neural networks and transformers have recently become dominant in hyperspectral (HS) image classification due to their ability to capture long-range dependencies in spectral sequences. Despite the success of these sequential architectures, mainstream deep learning methods primarily handle two-dimensional structured data. However, challenges such as the curse of dimensionality, spectral variability, and confounding factors in hyperspectral remote sensing images limit their effectiveness, especially in remote sensing applications. To address this issue, this paper proposes a novel land cover classification algorithm that integrates random forests with a spectral transformer network structure (RS-Net). Firstly, this paper presents a combination of the Gramian Angular Field (GASF) and Gramian Angular Difference Field (GADF) algorithms, which effectively maps the multidimensional time series constructed for each pixel onto two-dimensional image features, enabling precise extraction and recognition in the backend network algorithms and improving the classification accuracy of land cover types. Secondly, to capture the relationships between features at different scales, this paper proposes a SpectralFormer network architecture using the Context and Structure Encoding (CASE) module to effectively learn dependencies between channels. This architecture enhances important features and suppresses unimportant ones, thereby addressing the semantic gap and improving the recognition capability of land cover features. Finally, the final prediction results are determined by a voting mechanism from the Random Forest algorithm, which synthesizes predictions from multiple decision trees to enhance classification stability and accuracy. To better compare the performance of RS-Net, this paper conducted extensive experiments on three benchmark HS datasets obtained from satellite and airborne imagers, comparing various classic neural network models. Surprisingly, the RS-Net algorithm achieves high performance and efficiency, offering a new and effective tool for land cover classification.



**Citation:** Li, X.; Fan, X.; Li, Q.; Zhao, X. RS-Net: Hyperspectral Image Land Cover Classification Based on Spectral Imager Combined with Random Forest Algorithm. *Electronics* **2024**, *13*, 4046. <https://doi.org/10.3390/electronics13204046>

Academic Editor: Yosef Pinhasi

Received: 31 August 2024

Revised: 4 October 2024

Accepted: 9 October 2024

Published: 14 October 2024

**Keywords:** hyperspectral image classification; recursive neural networks; transformers; spectral transformer network structure (RS-Net); land cover classification accuracy



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Remote sensing technology plays a crucial role in land cover and land use analysis. It allows for detailed mapping of ecosystems, urban development, agriculture, and natural resources [1–4]. Hyperspectral remote sensing technology [5], as an advanced means of Earth observation, provides rich data support for surface cover classification, environmental monitoring, resource management [6–8], and other fields by recording reflectance, transmittance, and radiance information of landforms over hundreds of continuous spectral bands [9–12]. Hyperspectral remote sensing technology has been widely used in various fields. In agriculture, hyperspectral remote sensing monitors crop growth, assesses soil

nutrients, and detects pests. In forestry, it enables tree species identification, forest resource assessment, and fire monitoring [13–15]. In the field of forestry, hyperspectral remote sensing can be used for forest resource investigation, tree species identification, and forest fire monitoring [16–18]. In the field of environmental monitoring, hyperspectral remote sensing can be used for water quality assessment, air quality monitoring, and urban heat island effect research [19–21]. In the field of geological exploration, hyperspectral remote sensing can be used for mineral resources exploration, geological structure analysis, and lithology identification [22–24].

Hyperspectral images are distinguished by their high dimensionality, superior resolution, and robust signal-to-noise ratio, which empowers us to detect and track nuanced alterations on the earth's surface with greater precision. Nevertheless, the handling and interpretation of hyperspectral imagery are fraught with challenges, including the management of vast data volumes, the intricacy of processing, and the complexities associated with feature extraction [25–27]. The data footprint of hyperspectral images is considerable, as each pixel encapsulates data from numerous spectral channels, ranging into the hundreds. This magnitude of data imposes significant burdens on storage, transmission, and computational processing.

The basic principle of hyperspectral remote sensing technology is to utilize the electromagnetic wave radiation characteristics of sensors in different bands to obtain spectral information of features [28]. Hyperspectral images have high dimensionality, meaning they contain a large number of consecutive spectral bands. This richness of information enables precise identification and classification of features. High resolution, on the other hand, refers to hyperspectral images with small pixel sizes that can clearly depict the details and structures of the features. A high signal-to-noise ratio ensures the quality and reliability of hyperspectral images, enabling more accurate analysis and interpretation of image data.

Hyperspectral remote sensing [29–32] technology employs specialized sensors to capture geophysical information, with these sensors capable of recording extensive narrowband data across a spectrum ranging from the visible to the near-infrared wavelengths. Specifically, hyperspectral imagery typically encompasses dozens to hundreds of contiguous spectral bands, each with a width of a few to several tens of nanometers, endowing the data with an exceptionally high spectral resolution. For instance, certain hyperspectral imaging systems are capable of producing images with a wavelength range of 400 to 2500 nanometers, with band intervals of 10 nanometers or less. This fine spectral resolution imbues hyperspectral imagery with a “high-dimensional” characteristic, as each pixel corresponds to a spectral curve comprising hundreds of spectral measurements. In terms of spatial resolution, hyperspectral remote sensing imagery can offer ground resolutions ranging from several meters to several tens of meters, which is contingent upon the specific design of the sensor and the altitude of the flight platform.

Hyperspectral remote sensing image classification enhances target differentiation by leveraging the unique characteristics of hyperspectral data, extending beyond conventional multispectral methods. Traditional algorithms, however, are limited in their effectiveness for hyperspectral images [33]. Convolutional Neural Networks (CNNs) have emerged as a promising deep learning technique for this purpose, capable of autonomous feature learning and classification through multilayer networks [34]. Despite their potential, CNNs face challenges such as high computational requirements, overfitting due to the large number of spectral bands in hyperspectral data, and the need for complex preprocessing, which can affect model performance.

Researchers have turned to recurrent neural networks (RNNs) for classifying high-resolution and complex hyperspectral remote sensing images, especially for time-series data like multitemporal images [35]. RNNs excel in capturing temporal features, crucial for analyzing dynamic changes in ground cover types across time points, supporting applications in natural resource management and environmental monitoring. Despite their strength in handling complex data, RNNs' computational efficiency can be low, leading to investigations into optimizing their architecture and incorporating attention mechanisms

to enhance long-distance dependency capture [36]. U-Net, an advanced Convolutional Neural Network architecture [37], has become integral in hyperspectral remote sensing due to its strong feature extraction and adaptability to complex scenes [38]. Its unique structure, with contraction and expansion paths, allows for effective feature representation learning and precise image segmentation. While U-Net excels in recognizing and classifying surface cover types, challenges like processing efficiency and overfitting during training persist. Researchers are addressing these by enhancing the U-Net architecture and integrating advanced learning techniques.

ViT is a model for image classification that utilizes the Transformer architecture [39]. It divides an image into fixed-size blocks (tokens) and processes these blocks as input to the Transformer. There are architectures in the ViT model, such as the two proposed in the literature [40,41], that can integrate ViT with two large models and share the weights. This enables the same processing to be applied to different image locations, thereby utilizing the data more efficiently. The advantage of this processing is that it enables the Transformer to utilize the data more fully, which is valuable to learn. SpectralFormer is a framework based on the Transformer [42], which integrates the Vision Transformer (ViT) and Transformer in hyperspectral remote sensing technology. In hyperspectral remote sensing technology, SpectralFormer can process the continuous spectral information of hyperspectral images. It captures the dependencies between different spectral bands by using the self-attention mechanism to enhance the recognition and classification of surface coverage types. The application of SpectralFormer in hyperspectral remote sensing technology offers several advantages: SpectralFormer can automatically learn comprehensive feature representations from hyperspectral images, thereby enhancing the recognition and classification of surface coverage types. Recognition and classification capabilities. Adaptability to complex image scenes and efficient computational performance. However, SpectralFormer faces challenges in hyperspectral remote sensing, such as efficiently processing large-scale datasets and dealing with the overfitting problem during model training.

Overall, the utilization of SpectralFormer in hyperspectral remote sensing technology holds great significance. It offers new ideas and methods for remote sensing image classification and target detection [43]. Therefore, we should not abandon the SpectralFormer network in favor of developing a new framework to capitalize on its advantages. Therefore, there is an urgent need for a more refined method to overcome these limitations. This will help solve the problem by efficiently maintaining the richness of information during feature extraction and ensuring the consistency of spatial and semantic information during feature fusion. In this paper, we propose a new RS image semantic segmentation network framework called RS-Net, applied to hyperspectral remote sensing data for land cover classification. Firstly, we suggest that the combination of GASF and GADF efficiently maps the target information embedded in the one-dimensional signals processed in the Transformer to the features on the 2D image. This approach solves the problem of accurately extracting and recognizing information in the back-end network algorithm. Secondly, we introduce the SpectralFormer network architecture using the CASE module to effectively learn the dependencies between channels, enhancing important features and suppressing unimportant ones. The semantic gap refers to the discrepancy between the low-level spectral and spatial features directly extracted from the raw hyperspectral data and the intuitive understanding and interpretation of scenes by humans. This gap exists in the mapping process from raw pixel values to category labels with clear semantic meanings. The method is capable of capturing the relationships between features at different scales, thereby addressing the semantic gap issue. Ultimately, the prediction outcomes are refined through a voting process incorporating outputs from numerous decision trees [44], thereby enhancing the robustness and precision of the classification.

The main contributions of this paper are summarized as follows:

- (1) Proposing a fusion of Random Forest and SpectralFormer network architecture (RS-Net) algorithm for land cover classification to address the challenges of dimensional

catastrophe, spectral variability, and confusion arising from hyperspectral remote sensing images.

- (2) Enhancing characterization by combining GASF and GADF for the Grammy corner field.
- (3) Introducing more powerful contextual modeling and feature interaction with the CASE module.
- (4) Employing a dual-network architecture enhances robustness in handling noisy data and outliers, effectively capturing diverse texture features, preventing information loss and excessive feature fusion, and thereby improving the differentiation in land cover classification for remote sensing images.

## 2. Methodology and Data

### 2.1. Dataset

In order to determine if the model proposed in this paper achieves high accuracy and superior qualitative results for land cover classification across various satellite remote sensing image products that are universal and generalizable, we selected three benchmark hyperspectral (HS) datasets for experiments. We utilized an airborne acquisition platform to provide a comprehensive and realistic validation of our proposed algorithm. In the experiments of this paper, we selected three hyperspectral known public datasets. The details are summarized in Table 1. The Houston data, Indian Pines data, and Pavia University data are presented below. In this paper, the Indian Pines dataset, the Pavia University dataset, and the Houston dataset are divided such that the training set comprises 50% of all samples, the validation set comprises 25%, and the test set also comprises 25%.

**Table 1.** Summary of the three investigated HS datasets, including data collection information labels, feature classification information labels, and corresponding sample sizes.

Dataset	Houston		Indian Pines		Pavia University	
Sensor platform	ITERS CASI-1500		AVIRIS		ROSIS	
Loc. and Time	Aircraft-borne America, 2012		Aircraft-borne Northwest Indiana, 1992		Aircraft-borne Italy, 2003	
GSD	2.5 m		20 m		1.3 m	
Wavelength	364–1046 nm		400–2500 nm		430–860 nm	
Data size	349 × 1905 × 144		145 × 145 × 220		610 × 340 × 103	
#	Class	Samples	Class	Samples	Class	Samples
1	Unclassified		Alfalfa	46	Asphalt	6631
2	Healthy grass	1251	Corn-notill	1428	Meadows	18,649
3	Stressed grass	1254	Corn-mintill	830	Gravel	2099
4	Synthetic grass	697	Corn	237	Trees	3064
5	Trees	1244	Grass-pasture	483	Painted metal sheets	1345
6	Soil	1242	Grass-trees	730	Bare Soil	5029
7	Water	325	Grass-pasture-mowed	28	Bitumen	1330
8	Residential	1268	Hay-windrowed	478	Self-Blocking Bricks	3682
9	Commercial	1244	Oats	20	Shadows	947
10	Road	1252	Soybean-notill	972		
11	Highway	1227	Soybean-mintill	2455		
12	Railway	1235	Soybean-clean	593		
13	Parking Lot 1	1233	Wheat	205		
14	Parking Lot 2	469	Woods	1265		
15	Tennis Court	428	Buildings-Grass-Trees-Drives	386		
16	Running Track	660	Stone-Steel-Towers	93		

Table 1 summarizes the three investigated datasets on high school (HS), including information on data collection labels, feature classification labels, and corresponding sample sizes.

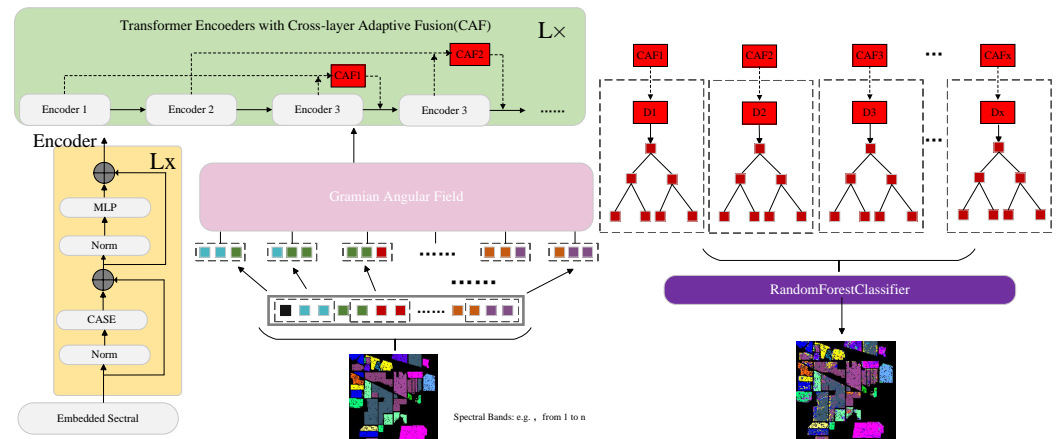
- (1) **Houston Data:** The first hyperspectral dataset was collected in June 2012 by the ITRES CASI-1500 sensor. It includes hyperspectral imagery of the University of Houston and its surrounding area in Texas, USA. The dataset was provided by the NSF-funded Center for Airborne Laser Mapping (NCALM) at the University of Houston. The images are  $349 \times 1905$  pixels in size and contain 144 bands distributed over a spectral range from 364 nm to 1046 nm, with the lidar modality containing one channel. Fifteen distinguishable land cover classes were studied in the scene, and the spatial resolution of the image reached 2.5 m. Additionally, we acquired specific details about this dataset. For example, the color quality of the images is high and contains a wealth of detailed information. In addition, the accuracy of these images has been verified, demonstrating their ability to faithfully represent the actual properties of the target objects. In addition, this dataset provides a wealth of geographic information, which is highly valuable for various research areas.
- (2) **Indian Pines Data:** The second hyperspectral dataset was collected in 1992 using an Airborne Visible Infrared Imaging Spectrometer (AVIRIS) to conduct comprehensive measurements of the Indian Pines region in northwestern Indiana, USA. The resulting image data totaled  $145 \times 145$  pixels in size, had a spatial resolution of 20 m, covered a broad spectral range of 400–2500 nm, and encompassed 220 bands. It is worth noting that among these bands, bands 104 to 108, 150 to 163, and 220 were identified as noisy bands. These bands were purposely excluded from the subsequent analysis process. The remaining 200 bands were finally selected for in-depth study.
- (3) **Pavia University Data:** The last dataset was collected in 2003 using the German ROSIS-03 Airborne Reflectance Optical Spectral Imager in the city of Pavia, Italy. The size of the images is  $610 \times 340$  pixels with a spatial resolution of 1.3 m. The dataset covers the wavelength range of 430–860 nm and contains 115 spectral channels. For the classification study, we specifically selected 103 bands and excluded the 12 bands interfered with by noise to ensure the accuracy of the results. This carefully selected subset helps improve data processing efficiency and reduce sensitivity to noise.

## 2.2. RS-Net Model

In response to the limitations of current deep learning methods, which predominantly handle two-dimensional data and struggle with issues such as dimensionality curse, spectral variability, and classification confusion in hyperspectral remote sensing images, this study introduces a novel algorithmic framework named RS-Net. This framework integrates Random Forest and the SpectralFormer network architecture for land cover classification, as depicted in Figure 1. The primary innovation of this paper lies in the effective translation of one-dimensional signal data into two-dimensional image features. To achieve this, we propose the use of the Gramian Angular Field (GASF) and Gramian Angular Difference Field (GADF) algorithms. These algorithms adeptly map the target information embedded within the one-dimensional signals, which are processed by the Transformer, onto a two-dimensional feature space. This approach significantly enhances the accuracy of feature extraction and recognition within the backend network algorithms. To address the challenge of capturing feature relationships across different scales, we introduce the SpectralFormer network equipped with the CASE module. This module is designed to effectively learn the inter-channel dependencies, thereby enhancing the prominence of important features while suppressing less relevant ones. This targeted feature refinement addresses the semantic gap issue commonly encountered in hyperspectral image analysis. The final component of the RS-Net algorithm is the Random Forest classifier, which is employed to aggregate the prediction results from the SpectralFormer network. By consolidating these results through a voting mechanism, the algorithm achieves pixel-level multispectral image classification. This strategy not only realizes a more precise classification but also effectively mitigates the semantic gap and dimensionality curse problems inherent in hyperspectral remote sensing imagery. In summary, this paper systematically develops and integrates a series of innovative techniques to overcome the challenges asso-



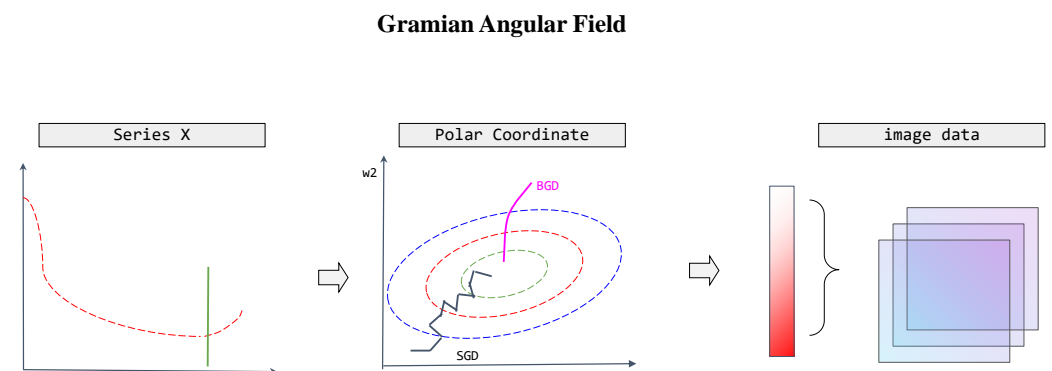
ciated with hyperspectral image classification. The proposed RS-Net algorithm leverages the strengths of both SpectralFormer for feature representation and Random Forest for robust classification, resulting in a comprehensive and effective solution for land cover classification in hyperspectral remote sensing images.



**Figure 1.** Schematic representation of the RS-Net architecture for hyperspectral image classification tasks.

### 2.2.1. Gramian Corner (GAFs)

The 1D sequence data are scaled and converted from the Cartesian coordinate system to the polar coordinate system. The temporal correlation of different points in time is then identified by analyzing the angle sum/difference between these points, as illustrated in Figure 2.



**Figure 2.** Schematic diagram of the structure of the Gramercy corner.

The green line on the far left represents 1. Given a time series  $X$ , it is scaled to fit between the green lines. The proposed mapping then generates a result in the polar coordinate system with a unique inverse function. The green, red, and blue circles in the middle diagram are provided to visually distinguish the polar coordinate angles. As time increases, the corresponding values twist between different angular points on the circle, akin to ripples in water.

Depending on whether it is an angular sum or an angular difference, there are two implementations: GASF (which corresponds to an angular sum) and GADF (which corresponds to an angular difference). Firstly, the data are scaled to the range of  $[-1, 1]$  or  $[0, 1]$ . The formula for scaling is as follows:

$$x_0^i = \frac{x_i - \min(x)}{\max(x) - \min(x)} \quad (1)$$

Swish  $x$  is an input. The scaled sequence data are converted to a polar coordinate system, i.e., the values are considered as the angle cosines and the timestamps as the radii, with the following formula:

$$\alpha = \arccos(x_i), -1 < (x_i) < 1 \quad (2)$$

Correspondence of angles and

$$GASF = x' \cdot x - \sqrt{I - x^2} \cdot \sqrt{I - x'^2} \quad (3)$$

Corresponding to the angle difference:

$$GASF = \sqrt{I - x'^2} \cdot x - x' \cdot \sqrt{I - x^2} \quad (4)$$

Then, we combine two methods GASF and GADF to generate images to achieve feature extraction through the network.

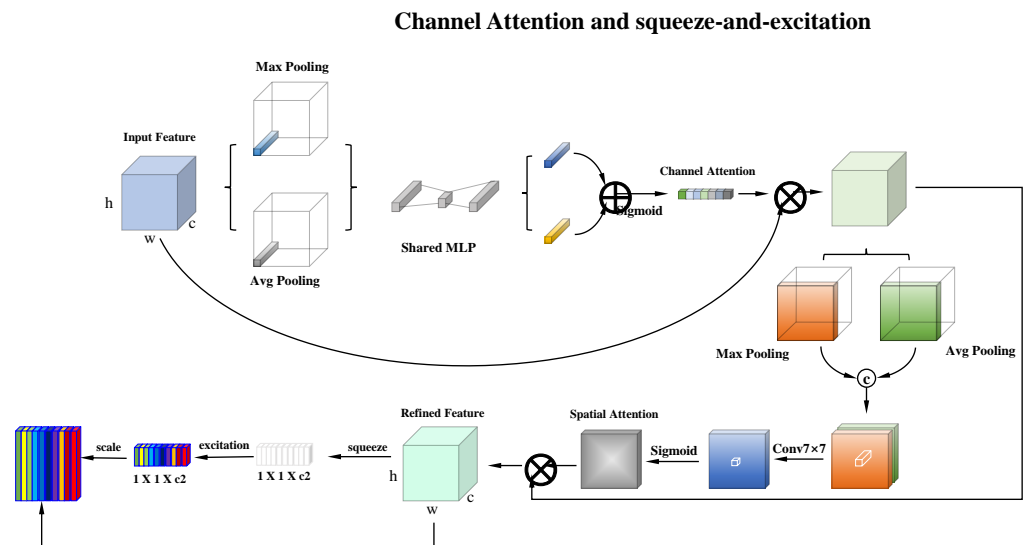
In the processing of time series data, the combination of the Gram Angle Sum Field (GASF) and Gram Angle Difference Field (GADF) enables a multidimensional characterization of the data, thus enhancing the learning and recognition capabilities of the model. The GASF focuses on the overall morphology and structure of the time series, which is crucial for understanding the fundamental static patterns and intrinsic characteristics of the dataset. On the other hand, the GADF concentrates on the dynamic changes and trends within the time series, which is vital for analyses concerned with the evolution and transformation of data over time. It can provide insights into the temporal variations in the data in classification recognition. By combining these two representations, the model can effectively capture diverse information about the time series data, thereby offering a more comprehensive feature set for future learning tasks.

This integrated representation enhances the model's capacity to identify crucial patterns in time series data and contributes to enhancing the model's performance, particularly when handling intricate time series data. In addition, this diversity of features helps reduce the risk of overfitting the model and enhances the model's capacity to generalize over unseen data, enabling it to adjust to various learning task demands. Pictorial representations of time series data can be combined with other types of data, such as text or images, to enhance the processing capabilities of the model. This approach makes multimodal learning more straightforward and effective, enhancing the overall performance of the model. In summary, the combination of GASF and GADF provides an effective feature extraction method that can offer robust support for processing and analyzing time series data. With this approach, the model can better capture the key information in time series data, thus improving its performance in various application scenarios.

### 2.2.2. Contextual and Structural Encoding (CASE)

In the field of deep learning, especially in computer vision tasks, feature extraction and channel recalibration are key steps to improve model performance. This study proposes to apply the Squeeze-and-Excitation (SE) module before and after integrating it with the Transformer architecture to incorporate channel attention, creating the CASE module, as illustrated in Figure 3. The CASE module introduces an extra processing step to improve channel selection and tuning capabilities during feature extraction, allowing for more precise control over the feature representation. In the Transformer architecture that incorporates channel attention, the SE module can be inserted after the self-attention layer to enhance the model's focus on crucial feature channels, conveying more valuable information in subsequent layers. The SE module can be applied after the multihead attention computation to further adjust the importance of the channels based on the attention-weighted features, enabling the model to prioritize features that are more relevant to the current task. We also propose the idea that the SE module can be applied after each head of

multihead attention separately, adjusting the importance of the channels according to the different subspace information captured by each head.



**Figure 3.** CASE module.

The width, height, and channel numbers are represented by the symbol  $\times$ , which denotes multiplication, while the sigmoid symbol represents the fusion of two features.

Specifically, given an input  $x$  with the number of feature channels  $c_1$ , a feature with the number of feature channels  $c_2$  is obtained after a series of general transformations, such as convolution. Unlike traditional CNNs, we next recalibrate the previously obtained features through three operations.

The first operation is the Squeeze operation, which converts each two-dimensional feature channel into a real value by compressing the spatial dimension of the feature map. This operation simulates the global receptive field to some extent, enabling the model to capture the global response distribution across the feature channels. The feature dimensions after the Squeeze operation align with the number of channels in the original input feature map, enabling the model to incorporate extra global contextual information while retaining the original data.

Next is the Excitation operation, which generates a weight for each feature channel via the parameter. These weights are learned to explicitly model the correlation between feature channels, similar to a gating mechanism in a recurrent neural network. The Excitation operation identifies and enhances the feature channels that are most important for the task at hand, while suppressing the less important ones, thus improving the feature representation of the model.

Finally, there is the Reweight operation, which applies the weights generated by the Excitation operation to each channel of the original feature map. This operation realizes the recalibration of the feature map in the channel dimension through a channel-by-channel weighting operation. This operation not only adjusts the importance of feature channels but also dynamically adapts the representation of features based on the task requirements.

The Squeeze-and-Excitation (SE) module enables fine-grained control and adjustment of feature channels through three operations: Squeeze, Excitation, and Reweight. This enhances the feature representation capability and performance of the model. In this way, Computer-Aided Software Engineering (CASE) can further enhance the performance of the model when processing data such as images and text. This fusion strategy enables the model to effectively capture and utilize the relationships between feature channels and subspaces, leading to improved performance across a variety of tasks.



### 2.2.3. Random Forest

Random Forest (RF), as an ensemble learning method, comprises multiple decision trees, each independently trained on different subsets of data. In the field of hyperspectral agricultural remote sensing image segmentation, this method demonstrates unique advantages, particularly when handling high-dimensional datasets. Hyperspectral remote sensing data typically contains hundreds or even thousands of bands, leading to extremely high data dimensionality. Random forests can effectively handle high-dimensional data, without the need for dimensionality reduction or feature selection, by directly utilizing information from all bands. This feature gives Random Forest a significant advantage in processing complex and dynamically changing agricultural remote sensing images.

In agricultural remote sensing image segmentation tasks, the number of samples from different categories may vary significantly, leading to an imbalanced dataset. Random Forest balances errors by adjusting the weights of each decision tree to enhance the model's performance when handling unbalanced datasets. This approach allows the model to maintain good performance under various data distributions. The process of constructing the Random Forest model not only helps generate predictions but also evaluates the importance of features, providing a basis for further feature selection and optimization. In addition, the parallel integration feature of the RF framework effectively controls the risk of overfitting while maintaining the engineering implementation simplicity of model construction and fast training speed.

For unbalanced datasets, the Random Forest model demonstrates good robustness in balancing errors and enhancing model performance across various data distributions. At the same time, its robustness to features ensures that high accuracy is maintained even in the presence of missing features or noise interference.

In practice, model fusion can be achieved by integrating multiple Transformer models with an RF framework. First, Transformer models are utilized to predict the input data. Subsequently, the prediction results are employed as inputs for the Random Forest (RF) model to achieve the final prediction. This approach enhances the overall accuracy and stability of the model, particularly when working with intricate and dynamically changing agricultural remote sensing images. This approach combines the sensitivity of deep learning to local features and the robustness of random forests to global features. It is expected to achieve a more efficient, accurate, and reliable solution for segmenting agricultural remote sensing images.

### 2.2.4. H-Swish Activation Function

The activation function determines the output characteristics of a neural network and influences its overall performance. In recent years, researchers have proposed a series of novel activation functions aiming to address the training challenges in neural networks, particularly the issue of gradient vanishing. Among them, ReLU (Rectified Linear Unit), as a classical activation function, has attracted a lot of attention due to its simple and efficient gradient computation. The original design of the Rectified Linear Unit (ReLU) activation function aims to address the issue of gradient vanishing during neural network training, which is particularly noticeable in recurrent neural network (RNN) models like Long Short-Term Memory networks (LSTM). The gradient calculation of the ReLU function is uncomplicated and direct, assigning a gradient of 0 to inputs less than 0 and 1 to inputs greater than 0. This design avoids the numerical decay caused by the successive multiplication of gradients and ensures that the gradient in the forward propagation can be calculated in the same way as the gradient in the backward propagation. Stable transfer of the gradient in the forward propagation. However, the sparsity of the ReLU function also brings certain limitations, especially in cases where the input is consistently negative. This situation may cause the gradient to remain at 0 during the backpropagation process, thereby hindering the adjustment of the weights and bias parameters. Consequently, this can trigger the phenomenon known as neuron death.

The Swish function not only demonstrates excellent performance but also effectively alleviates the issue of vanishing gradients, making it widely utilized in neural networks. The design of this function takes into account the nonlinear characteristics and the stability of the gradient multiplication, ensuring good performance in training neural networks. The mathematical formula for the Swish activation function is defined as follows:

$$\text{Swish}(x) = x\sigma(\beta x) \quad (5)$$

Swish  $x$  is a special case of the Sigmoid contraction function when positive values are considered.

In a 2017 study, it was found that the Swish function, as an activation function, significantly enhances the performance of neural network models when analyzing ImageNet datasets compared with the ReLU and sigmoid functions. The effectiveness of the Swish function in the backpropagation process is a key factor contributing to its performance improvement as it helps alleviate the gradient vanishing problem.

The Swish function mitigates the gradient vanishing problem. It is able to produce larger gradients during forward propagation, which helps mitigate the gradient vanishing problem and thus improves the training efficiency of the model. The Swish function is non-monotonic, enhancing the expressive power of the model in certain intervals, contributing to improved performance. When the input value is large, the Swish function behaves close to a linear function, allowing it to smoothly interpolate between linear and ReLU functions in neural networks, thereby enhancing the generalization ability of the model.

Hard Swish is a variant of Swish that was developed to simplify the computation of the formula. The original Swish formulation includes a Sigmoid function, which is computationally complex. Hard Swish replaces the Sigmoid function with a segmented linear function, making the computation much simpler. The mathematical formula is defined as follows:

$$H\text{-swish} = x \frac{\text{ReLU6}(x + 3)}{6} \quad (6)$$

where  $x$  is the input value of the activation function.

H-Swish (Hard Swish), as an enhanced activation function, offers significant advantages over the original Swish function. H-Swish simplifies the computation process and reduces computational complexity by utilizing segmented linear functions instead of Sigmoid functions. This optimization makes H-Swish more computationally efficient in its execution and is suitable for scenarios that require a large number of parallel computations. Although H-Swish simplifies the computational process, it still maintains similar performance advantages as the original Swish, such as mitigating the gradient vanishing problem and nonmonotonicity. These features enable H-Swish to maintain a performance comparable to the original Swish while ensuring efficient computation. In summary, H-Swish, with its enhanced computational efficiency, ease of integration, and consistent performance, is an optimized choice of activation function, particularly in scenarios that demand efficient computation and rapid model training.

#### 2.2.5. Evaluation Metrics

For the evaluation of hyperspectral pixel classification, in addition to the three traditional evaluation metrics—Overall Accuracy (OA), Average Accuracy (AA), and Kappa based on the confusion matrix—we introduce the F1 and Params metrics to assess the effectiveness of various methods. The Params metrics represent the parameters of each network. In our experiments, we set the batch size of all methods to 64 in order to accurately compute and compare their parameters. The smaller the value of Params, the fewer computational resources are used for the corresponding model.

#### 2.2.6. Comparative Methods

The primary goal of our experiments is to assess whether the proposed RS-Net can be deemed a state-of-the-art tool for hyperspectral image classification tasks. Therefore,

we have chosen four popular deep learning-based solutions for comparison: CNN, RNN, Transformer, and SpectralFormer's model. The details of our competitors are as follows. (1) A CNN consists of a convolutional layer, a corresponding 1-D or 2-D BN layer, ReLU composition, a Fully Connected (FC) layer, a Maximum Pooling layer, and an output layer. (2) In an RNN, two layers of loops and cascaded gated loop units are included. (3) The Transformer model is based on the Vision Transformer (ViT) architecture, consisting of 5 encoder blocks with a dimension of 64 for each grouped spectral embedding. Each encoder block comprises 4 self-attentive layers, 8 hidden layers of Multilayer Perceptrons (MLPs), and a dropout layer that deactivates 10% of the neurons. (4) SpectralFormer is based on the ViT model, enhanced with grouped spectral embeddings to improve local spectral details and additional FC layers to encode spatial information of flat image blocks. The dimension of each grouped spectral embedding is 64, and the grouped spectral nesting is set to 2.

### 2.2.7. Realization Details

All of our experiments were primarily based on Python 3.9 implemented in PyTorch 1.12.1 using a workstation with an Nvidia GeForce GTX 3060 Laptop GPU card. The number of epochs was set to 300. An Adam optimizer was utilized to reduce the learning rate by multiplying it with a decay factor of 0.9 every 30 epochs. The initial learning rate was  $5 \times 10^{-4}$ .

## 3. Experimental Results

### 3.1. Ablation Study

To evaluate the performance of the network structure proposed in this paper, ablation experiments were conducted on the Indian Pines dataset. The results are shown in Table 2. The original Overall Accuracy (OA) was 73.73%. After the application of the Grammy Corner Field, the OA increased to 74.91%, with a reduction of 114.18 parameters. This improvement demonstrates the effectiveness of the Grammy Corner Field scheme, which enhances the tight connections between channels and improves the network's feature learning capability. This improvement may be due to the scheme's ability to extract and transfer local feature information more effectively, thus providing a richer context for classification. The reduction in the number of parameters also indicates a decrease in model complexity, which helps to reduce overfitting and improve generalization. The introduction of the CASE attention mechanism further enhanced the classification accuracy. This is likely because it can weight the key features within the network, emphasizing spectral information that is more important for the classification task while suppressing irrelevant noise. The attention mechanism allows the network to focus more on the subtle differences between land cover types, thus improving the precision of classification. The integration of the Random Forest with the CASE module resulted in an OA of 77.31%, confirming its advantage for hyperspectral image pixel classification. This indicates the effectiveness of the multimodel fusion strategy. Random Forest, as a powerful ensemble learning method, can handle nonlinear relationships and complex feature interactions in hyperspectral data, while the CASE module enhances feature expression through attention mechanisms. The combination of both may have improved the model's ability to recognize the spectral and spatial heterogeneity of different land cover types. However, when all modules were tested together, the OA accuracy improved to 77.51%, an increase of 3.78%. The further improvement in overall accuracy suggests a complementarity between these modules. Each module enhances the model's performance in different aspects: the Grammy Corner Field scheme optimizes feature transfer, the CASE attention mechanism enhances feature significance, and the Random Forest provides robust classification decisions. This multifaceted enhancement works together to improve the model's ability to parse complex scenes. Clearly, all modules proposed in this paper are effective for agricultural land cover classification and address the corresponding issues, thereby optimizing the overall classification accuracy.

**Table 2.** Results of ablation experiments on Indian Pines dataset using RS-Net with different module combinations.

Module	Implementation					
GADS + GADF	×	✓	×	✓	×	✓
CASE	×	×	✓	✓	✓	✓
RF	×	×	×	×	✓	✓
OA (%) ↑	73.73	74.91 (+1.18)	75.66 (+1.93)	76.48 (+2.75)	77.31 (+3.58)	<b>77.51</b> (+3.78)
AA (%) ↑	81.57	83.12 (+1.55)	82.59 (+1.02)	84.98 (+3.41)	84.70 (+3.13)	<b>85.34</b> (+3.77)
Kappa ↑	0.7017	0.7137 (+0.0120)	0.7234 (+0.0217)	0.7334 (+0.0317)	0.7418 (+0.0401)	<b>0.7455</b> (+0.0438)
F1 ↑	70.03	72.44 (+2.41)	73.84 (+3.81)	76.66 (+6.63)	78.12 (+8.09)	<b>78.32</b> (+8.29)
params ↓	451.83	<b>337.65</b> (−114.18)	346.23 (−105.60)	371.25 (−80.58)	386.68 (−65.15)	444.36 (−7.47)

The check mark represents the module used, the wrong mark represents the module not used, the up arrow represents the larger the better, and the down arrow represents the smaller the better. Bold font means the best effect in the same module.

### 3.2. Multimethod Comparison

In order to determine if the model proposed in this paper demonstrates high accuracy and superior qualitative results for land cover classification of multiple satellite remote sensing image products with universality and generalizability, our classification method is applied to three hyperspectral datasets: the Houston dataset, Indian Pines dataset, and Pavia University dataset. We compare our proposed model with other advanced and representative models to produce qualitative results. The classification results of different models on the Houston dataset are presented in Table 3.

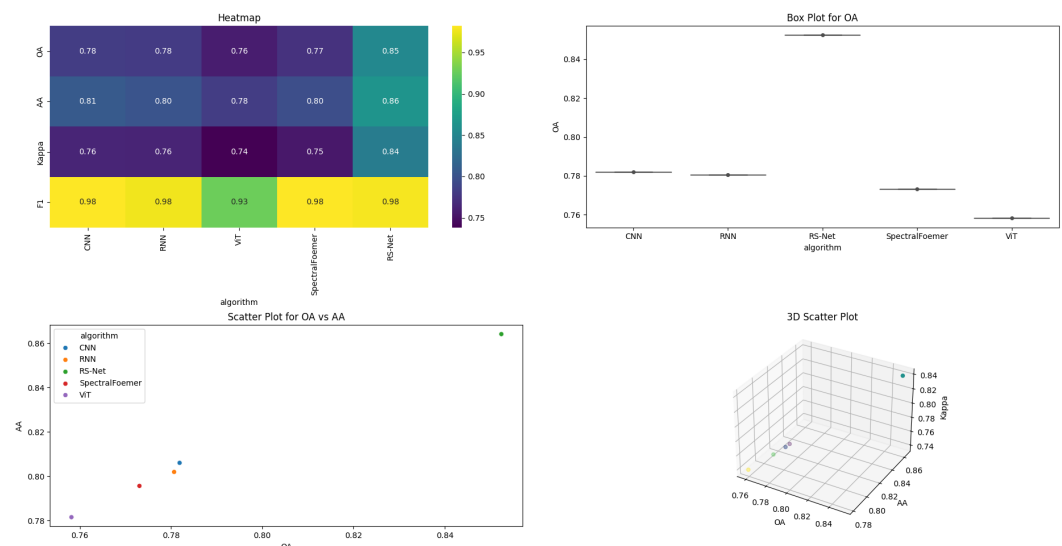
**Table 3.** Classification Accuracy of Different Classification Models in the Houston Dataset. The best results for each row are shown in bold.

#	Class	Method				
		CNN	RNN	ViT	SpectralFoemer	RS-Net
1	Healthy grass	<b>87.08</b>	83.29	84.14	83.38	85.75
2	Stressed grass	96.33	<b>98.12</b>	92.76	97.37	96.89
3	Synthetic grass	99.20	99.41	98.81	99.41	<b>99.80</b>
4	Trees	93.84	<b>98.11</b>	96.12	97.92	98.10
5	Soil	97.25	95.08	96.40	<b>97.25</b>	96.96
6	Water	97.90	97.02	94.41	<b>99.30</b>	98.60
7	Residential	81.43	76.49	76.77	75.40	<b>88.43</b>
8	Commercial	49.28	38.08	47.77	47.10	<b>74.73</b>
9	Road	62.32	71.67	72.99	68.84	<b>77.90</b>
10	Highway	71.33	66.41	47.68	52.32	<b>73.06</b>
11	Railway	69.25	75.33	80.46	80.55	<b>85.38</b>
12	Parking Lot 1	54.65	60.61	40.92	52.16	<b>63.68</b>
13	Parking Lot 2	57.54	51.58	44.91	46.32	<b>60.35</b>
14	Tennis Court	99.19	<b>100.00</b>	99.19	97.17	98.78
15	Running Track	92.38	91.54	<b>98.94</b>	98.52	97.67
	OA (%) ↑	78.19	78.07	75.82	77.31	<b>85.24</b>
	AA (%) ↑	80.60	80.19	78.15	79.56	<b>86.41</b>
	Kappa ↑	0.7637	0.7625	0.7383	0.7541	<b>0.8398</b>
	F1 ↑	98.12	97.65	92.76	<b>98.21</b>	97.84
	Params ↓	<b>128.22</b>	151.44	226.16	328.31	276.94

It can be concluded that, overall, VIT performs the worst. OA, AA, and Kappa are lower than in other models. The classical recurrent neural networks, RNN, CNN, and SpectralFormer, stand out more.

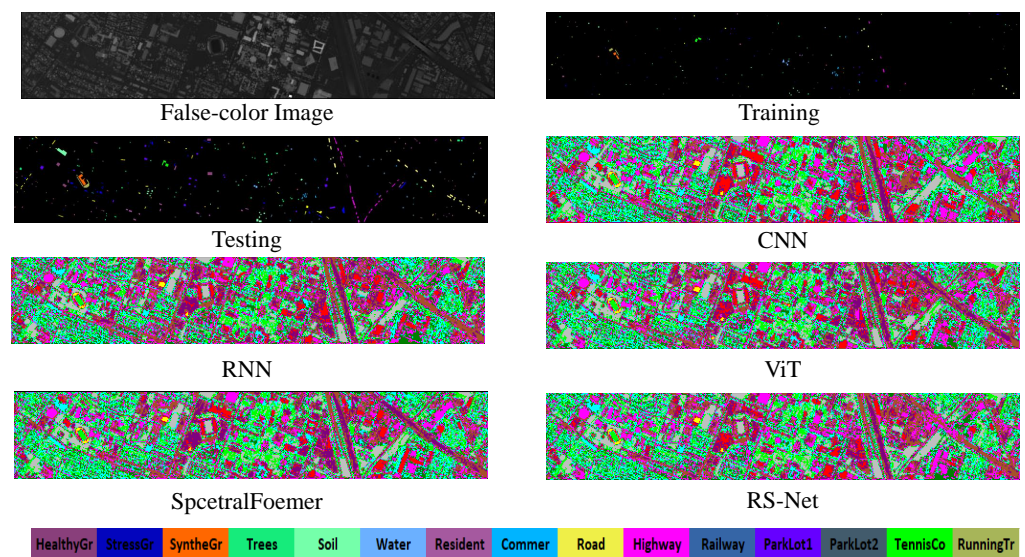
It can be concluded that, overall, the Vision Transformer (VIT) model performs the worst in hyperspectral image classification tasks, showing a significant disadvantage in accuracy metrics compared with other models. We speculate that this result may be due to the VIT model's insufficient adaptation to the small sample size characteristic of hyperspectral datasets, as VIT models are more suited for handling large-scale datasets. Consequently, in key evaluation metrics such as Overall Accuracy (OA), Average Accuracy (AA), and Kappa coefficient, the performance of VIT is inferior to that of other models. In contrast, classical models such as recurrent neural networks (RNN), Convolutional Neural Networks (CNN), and SpectralFormer have shown more impressive performance in hyperspectral image classification. These models are more effective at capturing the spatiotemporal features within hyperspectral data, resulting in higher accuracy rates in classification tasks. Particularly, SpectralFormer, with its unique spectral transformation network structure, is able to delve deeper into the hierarchical nature of spectral information, which is crucial for improving classification performance under small sample conditions. Additionally, RNN and CNN models, thanks to their inherent capabilities in sequence processing and spatial feature extraction, demonstrate strong robustness and generalization abilities in the classification of hyperspectral images. These models handle the dimensionality curse and spectral variability in hyperspectral data more effectively, thus maintaining classification accuracy while also improving model efficiency. Overall, the application potential of these traditional deep learning models on small hyperspectral datasets provides valuable references and insights for future research. The RS-Net algorithm proposed in our paper is able to better extract time-series information and spatial features from spectral data. For classifications with a limited number of training samples, its accuracy outperforms other methods. RS-Net exhibits superior performance capabilities.

A visualization example of the evaluation metrics data is presented, demonstrating how various chart types can be utilized to present and analyze the data. Each chart serves a specific purpose, such as heatmaps for displaying data matrices, box-and-whisker plots for illustrating distributions, scatter plots for showing relationships between variables, and 3D scatter plots for visualizing three-dimensional data, as depicted in Figure 4.



**Figure 4.** Evaluating indicator data visualization at Houston HS.

Figure 5 illustrates the color images, transformations, test labels, and classification maps obtained through the comparison method on the Houston HS dataset.



**Figure 5.** Illustration of color images, transformation and test labels, and classification maps obtained by comparative methods on the Houston HS dataset.

Table 4 presents the classification results of various models on the Indian Pine dataset.

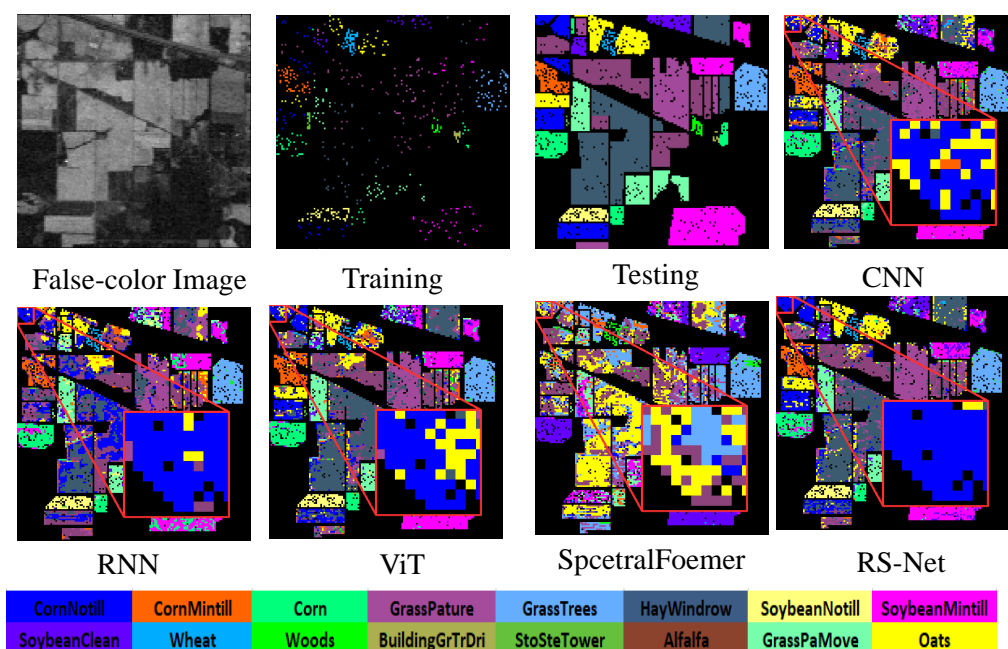
**Table 4.** Classification Accuracy of Different Classification Models in Indian Pine Dataset. The best results for each row are shown in bold.

#	Class	Method				
		CNN	RNN	ViT	Spcetralfomer	RS-Net
1	Brocoli_green_weeds_1	62.64	<b>73.63</b>	47.98	60.04	64.95
2	Brocoli_green_weeds_2	42.73	42.98	38.90	70.02	<b>78.31</b>
3	Fallow	89.67	27.72	71.74	89.13	<b>93.47</b>
4	Fallow_rough_plow	82.10	24.38	76.06	89.26	<b>91.72</b>
5	Fallow_smooth	82.64	79.91	72.45	<b>77.90</b>	80.63
6	Stubble	96.13	<b>97.27</b>	95.67	88.61	94.53
7	Celery	72.98	6.97	57.52	<b>81.48</b>	80.82
8	Grapes_untrained	65.22	42.18	30.07	<b>66.99</b>	67.82
9	Soil_vinyard_develop	65.07	18.44	25.18	57.80	<b>76.06</b>
10	Corn_senesced_green_weeds	96.91	91.98	95.68	<b>98.76</b>	<b>98.76</b>
11	Lettuce_romaine_4wk	91.88	<b>93.89</b>	69.21	90.75	90.91
12	Lettuce_romaine_5wk	62.42	26.06	18.48	53.63	<b>68.18</b>
13	Lettuce_romaine_6wk	<b>100.00</b>	97.78	95.56	<b>100.00</b>	<b>100.00</b>
14	Lettuce_romaine_7wk	74.36	28.21	17.95	89.74	<b>97.43</b>
15	Vinyard_untrained	63.64	18.18	45.45	<b>90.90</b>	81.81
16	Vinyard_vertical_trellis	<b>100.00</b>	80.00	40.00	<b>100.00</b>	<b>100.00</b>
	OA (%) ↑	71.74	53.27	50.64	73.73	<b>77.51</b>
	AA (%) ↑	78.03	53.10	56.12	81.57	<b>85.34</b>
	Kappa ↑	0.6787	0.4673	0.4486	0.7017	<b>0.7455</b>
	F1 ↑	50.89	36.99	38.90	70.03	78.32
	Params ↓	160.96	151.44	337.65	451.83	346.23

It can be concluded that, overall, the ViT still performs the worst, but there has been a decline across all algorithms this time. We suspect this is due to the complexity of the dataset's categories and the high degree of similarity among the types of ground objects. Even the classic RNN algorithm does not seem to be very efficient. We speculate that this may be because RNNs have limited capabilities in extracting complex spectral features and spatial information from hyperspectral images, which has impacted their classification



performance when faced with diversified and highly similar types of ground objects. CNNs and SpectralFormer, however, have still demonstrated high performance, which we guess is mainly due to the following reasons. In terms of feature extraction capability, CNNs can effectively extract local features from images through their convolutional layers, which are crucial for distinguishing similar types of ground objects. SpectralFormer may be able to better capture spectral information due to its unique spectral attention mechanism, thus maintaining high classification accuracy in complex datasets. In terms of spatial context understanding, both CNNs and SpectralFormer can better understand the spatial relationships between pixels, which is particularly important for hyperspectral image classification because the spatial distribution of ground objects often provides additional classification information. The model architectures of CNNs and SpectralFormer may be more suitable for handling high-dimensional data, such as hyperspectral images, and they can effectively reduce the impact of the curse of dimensionality while maintaining classification task performance. Therefore, they have maintained a good effect. As for the model we proposed, it combines their advantages without exception, so naturally, the effect is the best. Figure 6 illustrates color images, transformed and tested labels, and classification maps of various classification models obtained by comparing methods on the Indian Pines HS dataset.



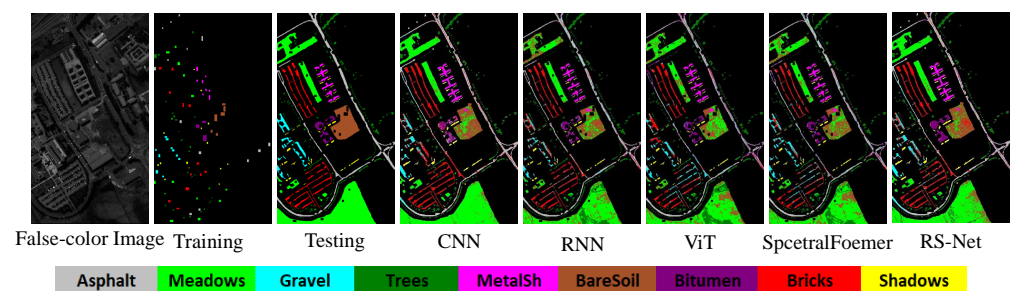
**Figure 6.** Illustration of color images, transformation and test labels, and classification maps obtained by comparative methods on the Indian Pines HS dataset.

Table 5 presents the classification results of various models on the Pavia University dataset.

The learning skill of deep learning is powerful, and the classical Convolutional Neural Network (CNN) has shown comparable classification accuracy to the RS-Net algorithm proposed in this paper on the Pavia University dataset. Figure 7 illustrates the color images, transformed and tested labels, and classification maps of various classification models obtained through the comparative method on the Pavia University HS dataset.

**Table 5.** Classification Accuracy of Different Classification Models in the Pavia University Dataset. The best results for each row are shown in bold.

#	Class	Method				
		CNN	RNN	ViT	SpectralFoemer	RS-Net
1	Asphalt	73.89	79.92	64.70	75.21	<b>86.80</b>
2	Meadows	<b>82.04</b>	72.21	65.74	69.43	76.52
3	Gravel	66.94	63.47	53.11	71.46	<b>83.41</b>
4	Trees	95.88	<b>98.11</b>	96.02	97.91	96.77
5	Painted metal sheets	99.37	98.47	99.28	99.01	<b>99.46</b>
6	Bare Soil	72.35	78.26	51.66	67.61	<b>87.31</b>
7	Bitumen	93.58	82.10	91.34	92.15	<b>94.08</b>
8	Self-Blocking Bricks	<b>91.97</b>	87.51	77.50	77.44	81.03
9	Shadows	<b>99.87</b>	96.48	<b>99.87</b>	99.50	87.79
OA (%) ↑		81.93	78.35	68.83	74.95	<b>82.84</b>
AA (%) ↑		86.21	84.05	77.69	83.30	<b>88.14</b>
Kappa ↑		0.7628	0.7223	0.6018	0.6797	<b>0.7787</b>
F1 ↑		82.92	71.56	64.60	68.82	<b>83.37</b>
Params ↓		<b>127.18</b>	151.44	167.77	261.68	201.19



**Figure 7.** Illustration of color images, transformation and test labels, and classification maps obtained by comparative methods on the Pavia University HS dataset.

### 3.3. Land Cover Classification Analysis

On the Houston dataset, our algorithm demonstrates superior classification performance for categories such as Road, Highway, and Parking Lot 1 and 2, as evidenced by the more accurate predictions in the prediction maps. In the Indian Pines dataset, our overall land cover classification outperforms other experimental algorithms. Similarly, on the Pavia University dataset, our algorithm achieves notable improvements in classifying Asphalt, Gravel, and other categories compared with other algorithms. This clearly indicates that our algorithm reduces misclassification errors in land cover classification tasks, providing a more precise classification, which is a critical capability for land classification analysis and detection tasks.

## 4. Discussion

The RS-Net algorithm proposed in this paper excels in hyperspectral image classification tasks, particularly in the context of land cover classification. The proposed RS-Net algorithm effectively mitigates challenges related to high-dimensional hyperspectral data, spectral variability, and classification errors. The results of ablation experiments demonstrate that the introduction of the combination of Gramian Angular Field (GASF and GADF) algorithms significantly improves classification accuracy while reducing model parameters, highlighting the advantages of GASF and GADF in feature extraction for land cover analysis.

Furthermore, the incorporation of the CASE module enhances the model's ability to identify crucial features and mitigates the semantic gap issue, which is particularly

beneficial for distinguishing between different land cover types. The RS-Net algorithm's ability to capture feature relationships at different scales is validated through its superior performance compared with other deep learning methods, such as CNNs, RNNs, and Transformer, on several publicly available datasets. Notably, RS-Net shows a higher classification accuracy and better generalization ability, especially for categories with small sample sizes, indicating its effectiveness in handling unbalanced datasets common in land cover classification tasks.

GASF and GADF effectively map the target information embedded in one-dimensional signals to features in two-dimensional images, which are accurately extracted and recognized by back-end network algorithms. The recognition accuracy of the fused GASF/GADF generated images is higher than that of direct recognition using only the original 1D sequence signals, proving the effectiveness of the GASF/GADF method in enhancing land cover classification.

However, the Gramian Angular Field method, while effective in characterizing one-dimensional sequences into two-dimensional images and aiding in the adaptation of current network algorithms for signal processing tasks, has some limitations. The construction of GASF/GADF involves specific computational time costs that must be considered in the system's overall processing, leading to additional delays in interactive output at the back-end. During the GASF/GADF conversion process, some signal detail information may be lost, which could impact the fine-grained classification of land cover types. Additionally, the data captured by many complex systems may not be simple one-dimensional sequences, limiting the applicability of GASF/GADF.

Therefore, exploring the application of Generalized Autoregressive Conditional Heteroskedasticity (GAF) in signal analysis across various scales to unveil more detailed relationships and patterns, and closely integrating it with current deep learning techniques is a significant future direction for signal processing researchers, particularly in the context of land cover and land use analysis. This approach could lead to more nuanced and accurate land cover classification, supporting sustainable land management and environmental conservation efforts. Additionally, the MSNAT extracts features at different scales to capture the multiscale information in hyperspectral images. This means it can simultaneously attend to the fine-grained details and broader contextual information within the images. This concept inspires us to conduct further research in this area, and we plan to carry out a study in this direction in the future [45].

## 5. Conclusions

After comprehensive research and experimental validation, we have successfully developed the RS-Net algorithm, which has demonstrated outstanding performance in hyperspectral image classification tasks. RS-Net effectively addresses the challenges of dimensionality curse, spectral variability, and classification ambiguity in hyperspectral image processing by integrating random forests with the SpectralFormer network architecture, along with the Gram Angular Field (GASF and GADF) algorithms and the CASE module. Overall, compared with other algorithms, the Vision Transformer (ViT) underperforms in hyperspectral image classification, particularly when dealing with complex and similar ground object types. The decline in performance across all algorithms indicates the challenges posed by the complexity of the dataset. While recurrent neural networks (RNNs) struggle with extracting complex spectral and spatial features, Convolutional Neural Networks (CNNs) and SpectralFormer show strong adaptability due to their effective feature extraction and spatial context understanding capabilities. CNNs excel at capturing local image features, while SpectralFormer improves classification accuracy with its spectral attention mechanism. Both architectures are well suited for handling high-dimensional data, mitigating the impact of the dimensionality curse. Our proposed model, which integrates the strengths of CNNs and SpectralFormer, achieves the best performance, demonstrating the superiority of a combined approach in hyperspectral image classification. RS-Net has outperformed existing models on multiple public datasets, providing an innovative

solution for the classification of hyperspectral remote sensing images with significant application value, especially in the field of land cover classification.

**Author Contributions:** Conceptualization, X.L., X.F., Q.L. and X.Z.; methodology, X.L., X.F., Q.L. and X.Z.; software, X.L., X.F., Q.L. and X.Z.; validation, X.L., X.F., Q.L. and X.Z.; formal analysis, X.L., X.F., Q.L. and X.Z.; investigation, X.L., X.F., Q.L. and X.Z.; resources, X.L., X.F., Q.L. and X.Z.; data curation, X.L., X.F., Q.L. and X.Z.; writing—original draft preparation, X.L., X.F., Q.L. and X.Z.; writing—review and editing, X.L., X.F., Q.L. and X.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work is the result of the research project funded by the Guangxi Key Research and Development Program: AB24010312.

**Data Availability Statement:** Readers in need can contact the corresponding author.

**Conflicts of Interest:** Author Xueqiang Zhao was employed by the company China Water Resources Pearl River Planning Surveying and Designing Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

HS	hyperspectral
RS-Net	Random Forest and SpectralFormer network
GASF	Gramian Angular Field
GADF	Gramian Angular Difference Field
CASE	Contextual and Structural Encoding
CNNs	Convolutional Neural Networks
RNN	Recurrent Neural Network
ViT	Vision Transformer
MSI	Multispectral Instrument
NCALM	NSF-funded Center for Airborne Laser Mapping
Aviris	Airborne Visual Infrared Imaging Spectrometer
OA	Overall Accuracy
AA	average accuracy

## References

1. Dhanaraj, K.; Angadi, D.P. Land use land cover mapping and monitoring urban growth using remote sensing and GIS techniques in Mangaluru, India. *GeoJournal* **2022**, *87*, 1133–1159. [\[CrossRef\]](#)
2. Campbell, A.D.; Fatoyinbo, T.; Charles, S.P.; Bourgeau-Chavez, L.L.; Goes, J.; Gomes, H.; Halabisky, M.; Holmquist, J.; Lohrenz, S.; Mitchell, C.; et al. A review of carbon monitoring in wet carbon systems using remote sensing. *Environ. Res. Lett.* **2022**, *17*, 025009. [\[CrossRef\]](#)
3. Sánchez-Azofeifa, G.A.; Castro-Esau, K.L.; Kurz, W.A.; Joyce, A. Monitoring carbon stocks in the tropics and the remote sensing operational limitations: From local to regional project. *Ecol. Appl.* **2009**, *19*, 480–494. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Asadzadeh, S.; de Oliveira, W.J.; de Souza Filho, C.R. UAV-based remote sensing for the petroleum industry and environmental monitoring: State-of-the-art and perspectives. *J. Pet. Sci. Eng.* **2022**, *208*, 109633. [\[CrossRef\]](#)
5. Hongjun, S. Dimensionality reduction for hyperspectral remote sensing: Advances, challenges, and prospects. *Natl. Remote Sens. Bull.* **2022**, *26*, 1504–1529.
6. Zhang, Y.; Migliavacca, M.; Penuelas, J.; Ju, W. Advances in hyperspectral remote sensing of vegetation traits and functions. *Remote Sens. Environ.* **2021**, *252*, 112121. [\[CrossRef\]](#)
7. Veraverbeke, S.; Dennison, P.; Gitas, I.; Hulley, G.; Kalashnikova, O.; Katagis, T.; Kuai, L.; Meng, R.; Roberts, D.; Stavros, N. Hyperspectral remote sensing of fire: State-of-the-art and future perspectives. *Remote. Sens. Environ.* **2018**, *216*, 105–121. [\[CrossRef\]](#)
8. Hieronymi, M.; Bi, S.; Müller, D.; Schütt, E.M.; Behr, D.; Brockmann, C.; Lebreton, C.; Steinmetz, F.; Stelzer, K.; Vanhellemont, Q. Ocean color atmospheric correction methods in view of usability for different optical water types. *Front. Media SA* **2023**, *10*, 1129876.
9. Huang, Y.; Peng, J.; Chen, N.; Sun, W.; Du, Q.; Ren, K.; Huang, K. Cross-scene wetland mapping on hyperspectral remote sensing images using adversarial domain adaptation network. *ISPRS J. Photogramm. Remote Sens.* **2023**, *203*, 37–54. [\[CrossRef\]](#)



10. Vali, A.; Comai, S.; Matteucci, M. Deep learning for land use and land cover classification based on hyperspectral and multispectral earth observation data: A review. *Remote Sens.* **2020**, *12*, 2495. [\[CrossRef\]](#)
11. Sun, W.; Liu, S.; Wang, M.; Zhang, X.; Shang, K.; Liu, Q. Soil copper concentration map in mining area generated from AHSI remote sensing imagery. *Sci. Total Environ.* **2023**, *860*, 160511. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Sharma, L.K.; Gupta, R.; Pandey, P.C. Future aspects and potential of the remote sensing technology to meet the natural resource needs. In *Advances in Remote Sensing for Natural Resource Monitoring*; Wiley-Blackwell: Hoboken, NJ, USA.
13. Tao, H.; Feng, H.; Xu, L.; Miao, M.; Long, H.; Yue, J.; Li, Z.; Yang, G.; Yang, X.; Fan, L. Estimation of crop growth parameters using UAV-based hyperspectral remote sensing data. *Sensors* **2020**, *20*, 1296. [\[CrossRef\]](#)
14. Yu, H.; Kong, B.; Wang, Q.; Liu, X.; Liu, X. Hyperspectral remote sensing applications in soil: A review. In *Hyperspectral Remote Sensing*; Elsevier: Amsterdam, The Netherlands, 2020; pp. 269–291.
15. Abd El-Ghany, N.M.; Abd El-Aziz, S.E.; Marei, S.S. A review: Application of remote sensing as a promising strategy for insect pests and diseases management. *Environ. Sci. Pollut. Res.* **2020**, *27*, 33503–33515. [\[CrossRef\]](#) [\[PubMed\]](#)
16. White, J.C.; Coops, N.C.; Wulder, M.A.; Vastaranta, M.; Hilker, T.; Tompalski, P. Remote sensing technologies for enhancing forest inventories: A review. *Can. J. Remote Sens.* **2016**, *42*, 619–641. [\[CrossRef\]](#)
17. Pang, Y.; Räsänen, A.; Wolff, F.; Tahvanainen, T.; Männikkö, M.; Aurela, M.; Korpelainen, P.; Kumpula, T.; Virtanen, T. Comparing multispectral and hyperspectral UAV data for detecting peatland vegetation patterns. *Int. J. Appl. Earth Obs. Geoinf.* **2024**, *132*, 104043. [\[CrossRef\]](#)
18. Zhang, H.; Song, H.j.; Yu, B.c. Application of hyper spectral remote sensing for urban forestry monitoring in natural disaster zones. In Proceedings of the 2011 International Conference on Computer and Management (CAMAN), Wuhan, China, 19–21 May 2011; IEEE: Piscataway, NJ, USA, 2011; pp. 1–4.
19. Adjovu, G.E.; Stephen, H.; James, D.; Ahmad, S. Overview of the application of remote sensing in effective monitoring of water quality parameters. *Remote Sens.* **2023**, *15*, 1938. [\[CrossRef\]](#)
20. Liu, C.; Xing, C.; Hu, Q.; Wang, S.; Zhao, S.; Gao, M. Stereoscopic hyperspectral remote sensing of the atmospheric environment: Innovation and prospects. *Earth-Sci. Rev.* **2022**, *226*, 103958. [\[CrossRef\]](#)
21. Liu, K.; Su, H.; Zhang, L.; Yang, H.; Zhang, R.; Li, X. Analysis of the urban heat island effect in Shijiazhuang, China using satellite and airborne data. *Remote Sens.* **2015**, *7*, 4804–4833. [\[CrossRef\]](#)
22. Pour, A.B.; Zoheir, B.; Pradhan, B.; Hashim, M. Editorial for the special issue: Multispectral and hyperspectral remote sensing data for mineral exploration and environmental monitoring of mined areas. *Remote Sens.* **2021**, *13*, 519. [\[CrossRef\]](#)
23. Carrino, T.A.; Crósta, A.P.; Toledo, C.L.B.; Silva, A.M. Hyperspectral remote sensing applied to mineral exploration in southern Peru: A multiple data integration approach in the Chapi Chiara gold prospect. *Int. J. Appl. Earth Obs. Geoinf.* **2018**, *64*, 287–300. [\[CrossRef\]](#)
24. Pal, M.; Rasmussen, T.; Porwal, A. Optimized lithological mapping from multispectral and hyperspectral remote sensing images using fused multi-classifiers. *Remote Sens.* **2020**, *12*, 177. [\[CrossRef\]](#)
25. Foerster, S.; Brosinsky, A.; Koch, K.; Eckardt, R. Hyperedu online learning program for hyperspectral remote sensing: Concept, implementation and lessons learned. *Int. J. Appl. Earth Obs. Geoinf.* **2024**, *131*, 103983. [\[CrossRef\]](#)
26. Upadhyay, V.; Kumar, A. Hyperspectral remote sensing of forests: Technological advancements, opportunities and challenges. *Earth Sci. Inform.* **2018**, *11*, 487–524. [\[CrossRef\]](#)
27. Jaiswal, G.; Rani, R.; Mangotra, H.; Sharma, A. Integration of hyperspectral imaging and autoencoders: Benefits, applications, hyperparameter tuning and challenges. *Comput. Sci. Rev.* **2023**, *50*, 100584. [\[CrossRef\]](#)
28. Chen, B.; Liu, L.; Zou, Z.; Shi, Z. Target detection in hyperspectral remote sensing image: Current status and challenges. *Remote Sens.* **2023**, *15*, 3223. [\[CrossRef\]](#)
29. Fan, J.; Masini, R.P.; Medeiros, M.C. Bridging factor and sparse models. *Ann. Stat.* **2023**, *51*, 1692–1717. [\[CrossRef\]](#)
30. Gunasekaran, H.; Azizi, L.; van Wassenhove, V.; Herbst, S.K. Characterizing endogenous delta oscillations in human MEG. *Sci. Rep.* **2023**, *13*, 11031. [\[CrossRef\]](#)
31. Sousa, D.; Small, C. Global cross-calibration of Landsat spectral mixture models. *Remote Sens. Environ.* **2017**, *192*, 139–149. [\[CrossRef\]](#)
32. Wang, Y.; Sun, J.; Wei, Z.; Plaza, J.; Plaza, A.; Wu, Z. Cloud-edge selective background energy constrained filter for real-time hyperspectral target detection. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5523215. [\[CrossRef\]](#)
33. Geiß, C.; Pelizari, P.A.; Tunçbilek, O.; Taubenböck, H. Semi-supervised learning with constrained virtual support vector machines for classification of remote sensing image data. *Int. J. Appl. Earth Obs. Geoinf.* **2023**, *125*, 103571. [\[CrossRef\]](#)
34. Kattenborn, T.; Leitloff, J.; Schiefer, F.; Hinz, S. Review on Convolutional Neural Networks (CNN) in vegetation remote sensing. *ISPRS J. Photogramm. Remote Sens.* **2021**, *173*, 24–49. [\[CrossRef\]](#)
35. Chadha, G.S.; Panambilly, A.; Schwung, A.; Ding, S.X. Bidirectional deep recurrent neural networks for process fault classification. *ISA Trans.* **2020**, *106*, 330–342. [\[CrossRef\]](#)
36. Jain, R.; Jain, A.; Mauro, E.; LeShane, K.; Densmore, D. ICOR: Improving codon optimization with recurrent neural networks. *BMC Bioinform.* **2023**, *24*, 132. [\[CrossRef\]](#)
37. Siddique, N.; Paheding, S.; Elkin, C.P.; Devabhaktuni, V. U-net and its variants for medical image segmentation: A review of theory and applications. *IEEE Access* **2021**, *9*, 82031–82057. [\[CrossRef\]](#)

38. Zunair, H.; Hamza, A.B. Sharp U-Net: Depthwise convolutional network for biomedical image segmentation. *Comput. Biol. Med.* **2021**, *136*, 104699. [\[CrossRef\]](#)
39. Liu, Y.; Zhang, Y.; Wang, Y.; Hou, F.; Yuan, J.; Tian, J.; Zhang, Y.; Shi, Z.; Fan, J.; He, Z. A survey of visual transformers. *IEEE Trans. Neural Netw. Learn. Syst.* **2023**, *35*, 7478–7498. [\[CrossRef\]](#)
40. Fan, X.; Li, X.; Yan, C.; Fan, J.; Yu, L.; Wang, N.; Chen, L. MARC-Net: Terrain Classification in Parallel Network Architectures Containing Multiple Attention Mechanisms and Multi-Scale Residual Cascades. *Forests* **2023**, *14*, 1060. [\[CrossRef\]](#)
41. Fan, X.; Li, X.; Yan, C.; Fan, J.; Chen, L.; Wang, N. Converging Channel Attention Mechanisms with Multilayer Perceptron Parallel Networks for Land Cover Classification. *Remote Sens.* **2023**, *15*, 3924. [\[CrossRef\]](#)
42. Hong, D.; Han, Z.; Yao, J.; Gao, L.; Zhang, B.; Plaza, A.; Chanussot, J. SpectralFormer: Rethinking hyperspectral image classification with transformers. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 5518615. [\[CrossRef\]](#)
43. Shirazi, A.; Hezarkhani, A.; Beiranvand Pour, A.; Shirazy, A.; Hashim, M. Neuro-Fuzzy-AHP (NFAHP) technique for copper exploration using Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) and geological datasets in the Sahlabad mining area, east Iran. *Remote Sens.* **2022**, *14*, 5562. [\[CrossRef\]](#)
44. Chen, Q.; Li, X.; Zhang, Z.; Zhou, C.; Guo, Z.; Liu, Z.; Zhang, H. Remote sensing of photovoltaic scenarios: Techniques, applications and future directions. *Appl. Energy* **2023**, *333*, 120579. [\[CrossRef\]](#)
45. Qiao, X.; Roy, S.K.; Huang, W. Multiscale Neighborhood Attention Transformer with Optimized Spatial Pattern for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5523815. [\[CrossRef\]](#)

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