


## Article

# Multi-Source Image Fusion Based Regional Classification Method for Apple Diseases and Pests

Hengzhao Li <sup>1,†</sup>, Bowen Tan <sup>1,†</sup>, Leiming Sun <sup>1</sup>, Hanye Liu <sup>2</sup>, Haixi Zhang <sup>1,\*</sup>  and Bin Liu <sup>1,3,4,5</sup>

<sup>1</sup> College of Information Engineering, Northwest A&F University, Xianyang 712100, China; liubin0929@nwfau.edu.cn (B.L.)

<sup>2</sup> School of Information Engineering, Yulin University, Yulin 719000, China

<sup>3</sup> Key Laboratory of Agricultural Internet of Things, Ministry of Agriculture and Rural Affairs, Xianyang 712100, China

<sup>4</sup> Shaanxi Engineering Research Center of Agricultural Information Intelligent Perception and Analysis, Xianyang 712100, China

<sup>5</sup> Shaanxi Key Laboratory of Agricultural Information Perception and Intelligent Service, Xianyang 712100, China

\* Correspondence: zh.haixi@nwfau.edu.cn

† These authors contributed equally to this work.

**Abstract:** Efficient diagnosis of apple diseases and pests is crucial to the healthy development of the apple industry. However, the existing single-source image-based classification methods have limitations due to the constraints of single-source input image information, resulting in low classification accuracy and poor stability. Therefore, a classification method for apple disease and pest areas based on multi-source image fusion is proposed in this paper. Firstly, RGB images and multispectral images are obtained using drones to construct an apple diseases and pests canopy multi-source image dataset. Secondly, a vegetation index selection method based on saliency attention is proposed, which uses a multi-label ReliefF feature selection algorithm to obtain the importance scores of vegetation indices, enabling the automatic selection of vegetation indices. Finally, an apple disease and pest area multi-label classification model named AMMFNet is constructed, which effectively combines the advantages of RGB and multispectral multi-source images, performs data-level fusion of multi-source image data, and combines channel attention mechanisms to exploit the complementary aspects between multi-source data. The experimental results demonstrated that the proposed AMMFNet achieves a significant subset accuracy of 92.92%, a sample accuracy of 85.43%, and an F1 value of 86.21% on the apple disease and pest multi-source image dataset, representing improvements of 8.93% and 10.9% compared to prediction methods using only RGB or multispectral images. The experimental results also proved that the proposed method can provide technical support for the coarse-grained positioning of diseases and pests in apple orchards and has good application potential in the apple planting industry.

check for  
updates

**Citation:** Li, H.; Tan, B.; Sun, L.; Liu, H.; Zhang, H.; Liu, B. Multi-Source Image Fusion Based Regional Classification Method for Apple Diseases and Pests. *Appl. Sci.* **2024**, *14*, 7695. <https://doi.org/10.3390/app14177695>

Academic Editor: Atsushi Mase

Received: 2 August 2024

Revised: 24 August 2024

Accepted: 25 August 2024

Published: 31 August 2024

**Keywords:** multi-source images; apple diseases and pests; vegetation indices; feature fusion; deep learning



**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

Apples are one of the most nutritious fruits, rich in vitamins, micronutrients, and fiber. China is the largest producer and exporter of apples [1], with apple production accounting for one-fourth of the total fruit production in the country and exports accounting for more than one-third of the total fruit exports [2]. Due to its advantageous geographical environment and unique natural climate conditions, Shaanxi Province has been recognized by the Food and Agriculture Organization of the United Nations as the best eugenic area for apples in the world [3]. By the end of 2022, the apple planting area in Shaanxi Province reached 6160 km<sup>2</sup>, accounting for 31.45% of the total apple planting area in the country. However, the occurrence of diseases and pests can lead to a decrease in apple yield and

quality, causing immeasurable losses [4]. The untimely and inaccurate discovery of diseases and insect pests will lead to the misuse and abuse of pesticides, causing pollution and seriously affecting the healthy and sustainable development of the apple industry in China.

When diseases and pests occur in apple orchards, the traditional solution is that fruit growers inspect the orchard on foot or by vehicle, visually judge the types and degree of infection of diseases and pests on fruit trees, and take management measures to curb the development of diseases and pests and reduce the losses caused by them [5]. However, the manual judgment method is time-consuming and laborious and often limited by the height and shade of apple trees, resulting in a lack of sight and omission of diseases and insect pests and a wrong judgment of the type and severity of apple diseases and insect pests. Therefore, realizing fast and efficient diagnosis of apple pests and diseases is crucial to the sustainable development of the apple industry.

Due to its excellent algorithmic performance, deep learning has been widely used in various fields, especially in computer vision tasks. For example, deep learning models can autonomously learn to understand image features at different levels, enabling the automated diagnosis of various crop diseases and pests [6–13]. At the same time, many theories and methods such as MLOps have been proposed to facilitate the development, deployment, and maintenance of deep learning models, providing assistance for the application of deep learning in agriculture [14]. The structure of convolutional neural networks (CNNs) is capable of adapting to the extraction of spatial features from images and exhibits outstanding performance in the fields of image processing and computer vision. Ahad et al. used a CNN to extend transfer learning and ensemble models to identify nine common rice diseases with 98 percent accuracy [7]. Liu et al. designed the Inception structure for feature extraction and used the dense connectivity strategy to realize feature propagation, which improved the accuracy of grape leaf disease recognition [8]. Thakur et al. combined the advantages of a VGG network and the Inception block to propose a lightweight CNN model, VGG-ICNN, which can significantly reduce the model parameters while achieving 99.16% classification accuracy on a variety of plant disease datasets. Singha Amit et al. proposed an end-to-end method based on convolutional neural networks (CNNs) and combined it with MLOps to improve the efficiency of development and deployment, realizing accurate classification of potato fusarium wilt and providing a solution for agricultural disease management [15]. Additionally, CNN networks can also be applied to object detection tasks, accurately locating the positions of disease occurrences while recognizing the types of diseases in images. Xie et al. introduced the Inception-v1 module, Inception-ResNet-v2 module, and Squeeze-and-Excitation module to propose a DR-IACNN for six faster common grape leaf disease detection models, achieving an average accuracy of 81.1% Mean (mAP) and a detection frame rate of 15.01 frames/second (FPS) [9]. Zeng et al. refactored the lead backbone of YOLOv5 using the neck module of MobileNetV3 and reduced the size of the model through pruning and quantification methods while achieving a detection efficiency of 26.5FPS and a 93% mAP, realizing the real-time accurate detection of tomato diseases [10]. Tang et al. combined the efficient channel attention (ECA) mechanism and the Transformer encoder to improve the YOLO model and proposed a real-time detection model, Pest-YOLO, which achieved 73.4% on multi-class pest detection tasks [11]. Liu et al. proposed an early apple leaf disease detection model RE-RCNN [13] by introducing a small spot feature enhancement branch and an improved SCMLoss to balance the differences between similar and large spots. However, in apple orchards, relying solely on leaf-scale disease and pest identification and detection methods is limited by perspective and field of view, making it difficult to quickly capture disease and pest information across the entire orchard. Thus, there is still a significant gap before practical application can occur in production. In comparison, the canopy-scale approach has significant advantages. Canopy-scale models can observe orchards from a broader perspective, capturing more spatial information, which is beneficial for comprehensively understanding the overall health of the orchard. Therefore, research on automatic classification methods for apple diseases and pest areas using canopy-scale spectral images has great significance.

As an important component of remote sensing tools, unmanned aerial vehicles (UAVs) have the advantages of low manufacturing cost, portability, and ease of operation, bringing new opportunities to the area of precision agriculture. They have a wide range of applications in crop growth monitoring [16–20], yield estimation [21–24], disease diagnosis [25–28], and pesticide spraying [29–31]. Due to their close-to-earth flight characteristics, UAVs can quickly and flexibly complete monitoring tasks and obtain data with ultra-high spatio-temporal resolution under suitable weather conditions [32]. In addition, the multispectral images collected by the multispectral cameras carried by UAVs can reflect specific pest information in each band and then can be targeted for the detection of the corresponding pest. For example, near-infrared spectroscopy images can detect corn pathogens [26] and peanut bacterial wilt [33], visible light (red, green, and blue) images can detect wheat leaf rust [34] and grape yellow flower disease [35], etc. These studies have confirmed the feasibility of using UAVs equipped with multispectral cameras to diagnose various crop diseases and pests. However, manually discovering and verifying the correlation between multispectral data and crop diseases and pests is inefficient, requires high professional expertise, and often lacks universality. Vegetation indices (VIs), due to their ability to significantly reflect crop conditions and their ease of calculation, have received widespread attention from researchers studying crop diseases and pests. Due to the differences in the composition of background substances such as plants and soil, as well as the changes in the material structure of plants in the process of crop erosion by diseases and insect pests, the differences and changes in spectral wavelengths at the corresponding wavelengths will occur. Therefore, the VIs can be used to distinguish between soil background, healthy plants, and infected plants. Mahmud et al. screened seven VIs related to apple fire blight through the clustering method and proposed an image segmentation method for diagnosing apple fire blight based on canopy multispectral images taken by drone [5]. Wu et al. proposed a recognition method for the gray mold of strawberry leaves by exploring the band and texture and calculated VIs of hyperspectral images of strawberry leaves, combining these various characteristics with existing machine learning models [36]. Most existing research establishes diagnostic methods based on the correlation between a single data source image and crop diseases and pests. However, it is difficult to comprehensively capture the growth status of vegetation solely through RGB images. At the same time, analyzing apple biomass solely through VIs may ignore the appearance and texture changes in the pathological process, resulting in the low accuracy of disease and insect pest diagnosis. Therefore, it is urgent to diagnose apple diseases and pests by combining multi-source image data.

This study reports work to solve the problems of low diagnostic accuracy and poor stability of the existing single-source image-based disease and pest regional classification methods, focusing on the following issues:

- Firstly, the visible light and multispectral image data of an apple orchard canopy are collected by UAV, and the multi-source multi-label apple canopy image dataset of apple diseases and pests is constructed, which made up for the shortage of multi-label classification training data in large-scale orchards.
- Secondly, a VI selection method based on the saliency attention module is proposed. Aiming at the habitat information of apple plants, a feature selection method is used to weight the selected 22 vegetation indices in order to improve the accuracy of regional multi-classification.
- Finally, a multi-label classification model called AMMFNet, based on the joint prediction of RGB and multispectral images, is constructed. It can effectively utilize the complementary features of RGB images and VIs so that the model can automatically pay attention to the features related to diseases and pests and reduce the impact of redundancy in multi-source data on prediction performance.

In summary, this paper aims to design a deep learning model based on multi-source image fusion by combining visible light and multispectral image data. This model aims to realize multi-label classification of apple disease and insect pest regional images. It further seeks to provide technical support for the fast and efficient coarse-grained positioning of apple

disease and insect pest canopy areas in order to provide new ideas and methods for the healthy development of the apple industry.

The structure of this paper is as follows: Section 2 introduces the detailed structure of AMMFNet. Section 3 presents the experiment results and corresponding analysis. Finally, Section 4 summarizes the research findings and outlines the corresponding research contributions.

## 2. Materials and Methods

### 2.1. Diseases and Pests Multi-Source Image Dataset of Apple Canopy

#### 2.1.1. Image Data Collection

The canopy multi-source image data of apple diseases and insect pests were collected at the Baishui Apple Test Demonstration Station of Northwest A&F University in Tongji Village, Du Kang Town, Baishui County, Weinan City, Shaanxi Province (Figure 1). Two data acquisitions were conducted on the orchard on 11 June and 31 August 2023. The data were captured at noon on sunny and windless days to minimize interference from factors such as light intensity and angle on the prediction results. The data acquisition equipment used was the DJI Phantom 4 Pro multispectral UAV. During the image capturing process, the aircraft was manually controlled to designate a cruise area and encompass the boundaries of the orchard to ensure data availability. The shooting angle of the camera was  $-90^\circ$  (directly downwards), the shooting height was 40FT, and the horizontal and lateral overlap rate was 75%. And the UAV kept hovering during shooting. In addition, at the same time as the UAV shooting, the areas where diseases and insect pests occur were manually recorded or marked. The visible light and multispectral image information obtained by UAV shooting is shown in Table 1.

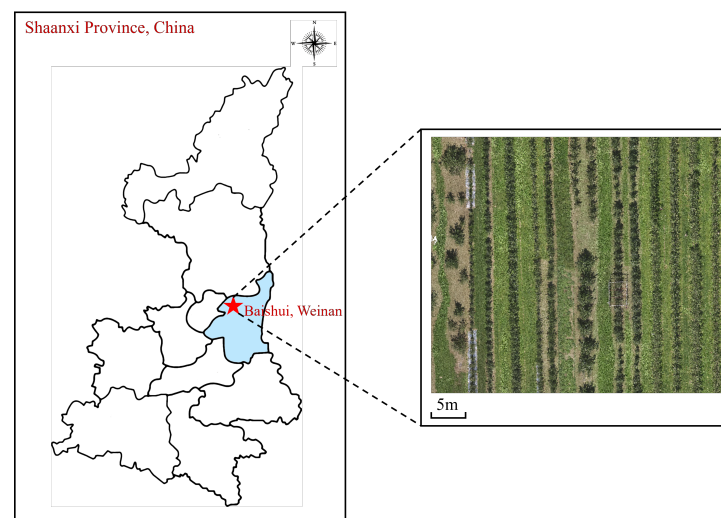


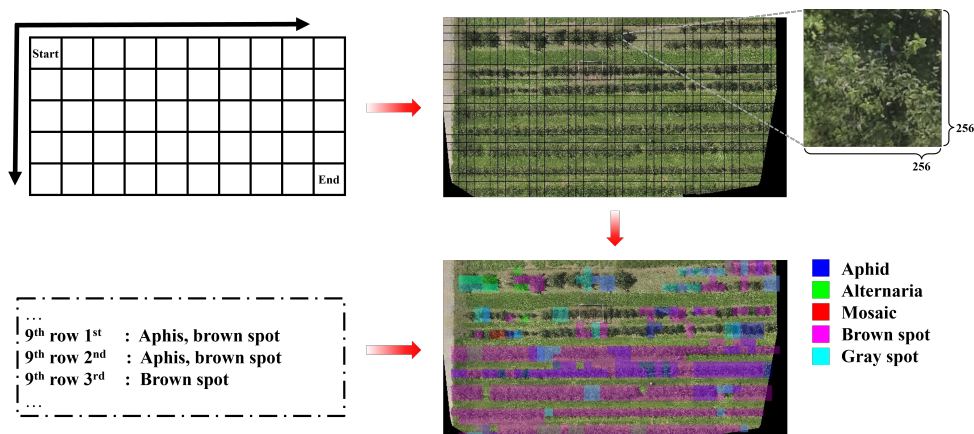
Figure 1. Data collection locations: Baishui, Weinan, Shaanxi, China.

Table 1. The parameters of the multispectral image.

Camera Category	Band Name	Wavelength/nm	Minimum	Mean	Maximum	Standard Deviation	Variance
RGB	Red	660	5.07	109.74	255.99	33.32	1110.50
	Green	550	14.84	115.87	254.94	27.86	776.29
	Blue	470	0.00	76.97	254.80	26.20	686.36
Multispectral	$R_{blue}$	450	0.20	2.42	87.85	1.07	1.15
	$R_{green}$	560	0.62	12.18	359.16	4.85	23.56
	$R_{red}$	650	0.21	11.01	457.65	7.49	56.06
	$R_{rededge}$	730	1.27	35.67	397.91	11.09	122.95
	$R_{nir}$	840	1.77	50.89	458.85	16.35	267.19

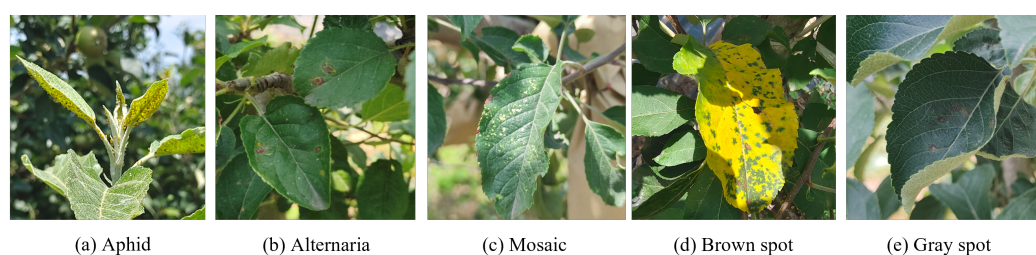
### 2.1.2. Construction of Multi-Source Image Dataset of Apple Diseases and Pests Canopy

To optimize the canopy-scale apple diseases and pests regional classification model, a self-built multi-source image pest dataset based on the RGB and multispectral canopy images taken by UAV was constructed. After that, the original image was divided into area images that had the same size, and each area image had multiple tags, which respectively indicated that the area was suffering from one or more types of apple diseases and pests, healthy, or had a soil background. The specific image preprocessing and labeling process is shown in Figure 2.



**Figure 2.** Image segmentation method and manual labeling process.

The original image needed radiation correction and geometric correction beforehand to eliminate the remote sensing imaging error caused by environmental factors and was then spliced to obtain a multispectral reflectance map. Angle correction and channel splicing were also performed on the reflectance map. Then, manual annotation was performed on each area image. Firstly, the entire orchard was divided into a grid-like coarse-grained area shown in Figure 2, and the image was cut into an image block with a size of  $256 \times 256$  pixels. Secondly, under the guidance of experts in the field of plant protection, each individual fruit tree was identified as the smallest unit to determine which type or types of diseases and pests it was suffering from, or whether it was healthy, and these were recorded. Based on this information, multi-label annotation was performed on the regional images. Finally, a total of 1901 area images were obtained, including 5 common apple diseases and pests. As shown in Figure 3, they are Aphids, Alternaria, mosaic, brown spot, and gray spot. Furthermore, it was necessary to clean the original dataset to eliminate the impact of outliers on the training effect. Traditional image data augmentation techniques such as flipping, rotation, sharpening, and whitening were adopted to simulate the possible shooting effects encountered in reality, enhance the robustness of the model, and prevent overfitting during the model training process. Finally, the dataset was divided into a training set and testing set in a ratio of 8:2, and the multi-source apple canopy image dataset of apple diseases and pests with multi-label information was constructed. The total number of RGB and multispectral images in the augmented data is 28,515, and the distribution information is shown in Table 2.



**Figure 3.** Five common apple diseases and pests images.

**Table 2.** Apple diseases and pests canopy multispectral image dataset.

Disease and Pest	Train Dataset	Val Dataset
Aphids	6012	1503
Alternaria	600	150
Mosaic	432	108
Brown spot	6156	1539
Gray spot	1116	279

Notes: This dataset is a multi-label dataset; thus, the number of images for each disease category may not be consistent with the total number of images. The distribution of RGB and multispectral images is consistent.

## 2.2. Vegetation Indices Calculation and Selection

### 2.2.1. Vegetation Indices of Apple Diseases and Pests

Vegetation indices, as indices for analyzing crop growth in agricultural remote sensing, can effectively analyze crop growth and provide accurate data support for agricultural production. The vegetation information on remote sensing images is mainly reflected by the spectral characteristics of and spectral differences in vegetation canopy leaves, and there is a certain correlation between the spectra of different bands and different elements or characteristic states of vegetation. Using multispectral remote sensing data, through linear or nonlinearity combination operations, some values that have certain indicative significance for vegetation growth and biomass are generated, which is the vegetation index. The visible light spectrum is controlled by the chlorophyll content of the leaves, the short-wave near-infrared spectrum is controlled by the intracellular structure of the leaves, and the medium-wave near-infrared spectrum is controlled by the water content in the leaves. Green light is effective for distinguishing plant categories, while red light is effective for assessing vegetation coverage and plant growth conditions. Healthy green vegetation has a large difference in reflectance between the near-infrared spectrum  $R_{nir}$  and the red spectrum  $R_{red}$ , because  $R_{red}$  is strongly absorbed by green plants, while  $R_{nir}$  exhibits high reflection and high transmission. When diseases and pests occur, bacteria and pests have varying degrees of influence on the changes in the chlorophyll, biomass, cellulose, and protein content in the invaded parts of the vegetation, which further affects the spectral information observable in the canopy of apple trees. Therefore, a total of 22 types of relevant vegetation indices, such as NDVI, GNDVI, RDVI, TNDVI, SR, and MTVI2, which are sensitive to vegetation and background differentiation, and MSAVI, OSAVI, and DVI, which are related to various biomass calculations, were selected and calculated. The calculation formulas for each vegetation index are shown in Table 3.

**Table 3.** Equations used for vegetation indices calculation.

Index Name	Equation	Introduction
NDVI [37]	$NDVI = \frac{R_{nir} - R_{red}}{R_{nir} + R_{red}}$	Normalized Differential Vegetation Index: a commonly used index to measure vegetation growth and coverage
TNDVI [38]	$TNDVI = \text{SQRT}\left(\frac{R_{nir} - R_{red}}{R_{nir} + R_{red}} + 0.5\right)$	Transformational Normalized Difference Vegetation Index: a modified NDVI more sensitive to soil brightness for arid and semi-arid regions.
GNDVI [39]	$GNDVI = \frac{R_{nir} - R_{green}}{R_{nir} + R_{green}}$	Green Normalized Difference Vegetation Index: an index emphasizing the green vegetation component, which is suitable for high-biomass areas
RENDVI [40]	$RENDVI = \frac{R_{rededge} - R_{red}}{R_{rededge} + R_{red}}$	Modified Red Edge Normalized Difference Vegetation Index: using red edge bands to be more sensitive to vegetation cover and chlorophyll content

Table 3. Cont.

Index Name	Equation	Introduction
MSAVI2 [41]	$MSAVI2 = \frac{1}{2}(2R_{nir} + 1 - \sqrt{((2R_{nir} + 1)^2 - 8(R_{nir} - R_{red})))}$	Modified Soil-Adjusted Vegetation Index 2: a vegetation index to reduce the influence of soil background, which is suitable for bare soil and vegetation mixed areas
RVI [42]	$RVI = \frac{R_{red}}{R_{blue}}$	Vegetation Index: the ratio of vegetation reflectance to soil reflectance, reflecting the degree of vegetation density
DVI [43]	$DVI = R_{nir} - R_{red}$	Differential Vegetation Index: simple vegetation and non-vegetation reflectance differences for rapid assessment of vegetation cover
GDVI [44]	$DVI = R_{nir} - R_{green}$	Global Differential Vegetation Index: considering differences in multiple bands for the assessment of complex vegetation environments
GRVI [44]	$GRVI = \frac{R_{nir}}{R_{green}}$	Green-Red Vegetation Index: using the ratio of green and red bands to reflect chlorophyll content
WDRVI [45]	$WDRVI = \frac{\alpha R_{nir} - R_{red}}{\alpha R_{nir} + R_{red}}$	Weighted Differential Vegetation Index: adjusting NDVI by weighting to be more sensitive to high vegetation cover areas
MSAVI [41]	$MSAVI = \frac{R_{nir} - R_{red}}{R_{nir} + R_{red} + L} (1 + L)$ $L = 1 - 2\alpha \times NDVI \times WDRVI$ $WDRVI = R_{nir} - \alpha R_{red}$	Modified Soil-Adjusted Vegetation Index: an improved version to reduce the influence of soil background on vegetation index
OSAVI [46]	$OSAVI = \frac{R_{nir} - R_{red}}{R_{nir} + R_{red} + 0.16}$	Optimizing Soil Adjustment Vegetation Index: further optimizing the correction of soil background by introducing soil factors
GOSAVI [47]	$GOSAVI = \frac{R_{nir} - R_{green}}{R_{nir} + R_{green} + 0.16}$	Global Optimizing Soil Adjustment Vegetation Index: an improved version of OSAVI considering the effects of multiple bands to improve accuracy
NDRE [48]	$NDRE = \frac{R_{nir} - R_{rededge}}{R_{nir} + R_{rededge}}$	Normalization Differential Red Edge Index: using red edge band, sensitive to vegetation structure and health status
SR [49]	$SR = \frac{R_{nir}}{R_{rededge}}$	Structure Ratio Index: using to reflect complex trade-offs in vegetation canopy structure
NLI [50]	$NLI = \frac{R_{nir}^2 - R_{red}}{R_{nir}^2 + R_{red}}$	Normalization of Leaf Canopy Index: using to assess vegetation canopy health and chlorophyll content
RDVI [51]	$RDVI = \sqrt{NDVI \times (R_{nir} - R_{red})} = \sqrt{\frac{(R_{nir} - R_{red})^2}{(R_{nir} + R_{red})}}$	Re-normalized Difference Vegetation Index: improving the accuracy of vegetation cover assessment by optimizing weights
MSR [52]	$MSR = \frac{\frac{R_{nir}}{R_{red}} - 1}{\sqrt{\frac{R_{nir}}{R_{red}} + 1}}$	Multispectral Ratios: comprehensive assessment of vegetation growth using ratios of multiple bands
NG [53]	$NG = \frac{R_{green}}{R_{nir} + R_{red} + R_{green}}$	Normalized Greenness: a vegetation index emphasizing green band reflectance for assessing chlorophyll content
NR [53]	$NR = \frac{R_{red}}{R_{nir} + R_{red} + R_{green}}$	Normalized Redness: a vegetation index based on red band reflectance that reflects the health status of vegetation
IPVI [54]	$IPVI = \frac{R_{nir}}{R_{R_{nir}} + R_{red}}$	Infrared Vegetation Index: using the infrared band to reflect the moisture content and health status of vegetation
MTVI2 [55]	$MTVI2 = \frac{1.5 \times (1.2(R_{nir} - R_{green}) - 2.5(R_{red} - R_{green}))}{\sqrt{(2R_{nir} + 1)^2 - (6R_{nir} - 5\sqrt{R_{red}}) - 0.5}}$	Modified Triangular Vegetation Index 2: vegetation index sensitive to chlorophyll content, suitable for high-biomass areas

### 2.2.2. Vegetation Indices Selection Algorithm

We applied salient attention to calculate the importance weight of the 22 calculated plant cover indices and significantly weighted the vegetation index calculated in the previous section, aiming to make the model pay more attention to the vegetation index that is beneficial to the classification of apple diseases and pests. The ReliefF algorithm for multi-label feature selection (RF-ML) [56] is used to calculate the importance of different

vegetation indices when calculating vegetation index weights. RF-ML is proposed for multi-label tasks based on the ReliefF and RReliefF [57] algorithms. ReliefF is one of the commonly used feature selection algorithms. By measuring the distance between adjacent samples and the difference between features within a certain range, each feature is assigned a specific weight. On this basis, RF-ML further considers the differences between multiple labels, so it is more suitable for the multi-label classification task. The calculation formula is shown in Equation (1).

$$W(X_j) = \frac{W_{dYX}(X_j)}{W_{dY}} - \frac{W_{dX}(X_j) - W_{dYX}(X_j)}{c - W_{dY}} \quad (1)$$

where  $c$  is the total number of iterations, i.e., the total number of randomly selected samples. For the  $j$ -th sample  $X_j$ , the weight  $W_{dX}$ , the tag weight  $W_{dY}$ , and the joint weight  $W_{dYX}$  calculation methods are shown in Equations (2)–(4).

$$W_{dX}(X_j) = \sum_{i=1}^c \sum_{z=1}^k \text{diff}(X_j, E_i, EK_z) \times d(E_i, EK_z) \quad (2)$$

$$W_{dY} = \sum_{i=1}^c \sum_{z=1}^k \text{mld}(E_i, EK_z) \times d(E_i, EK_z) \quad (3)$$

$$W_{dYX}(X_j) = \sum_{i=1}^c \sum_{z=1}^k \text{mld}(E_i, EK_z) \times \text{diff}(X_j, E_i, EK_z) \times d(E_i, EK_z) \quad (4)$$

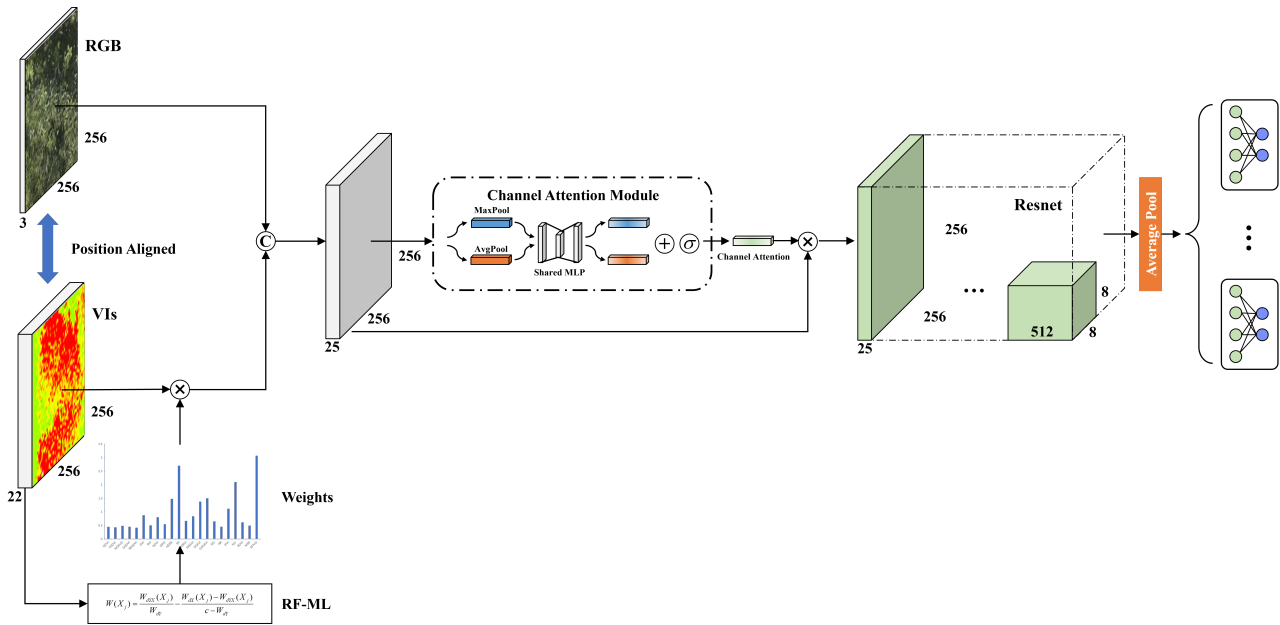
where  $k$  represents the number of neighbors, i.e., the total number of  $EK$  calculated for each iteration. The  $d$  is used to represent the distance between two samples. The Euclidean distance between all the samples is calculated and the first  $k$  samples are taken as the neighbor sample set  $EK$  of the current sample after sorting by distance. The  $\text{diff}(X_j, E_i, EK_z)$  means the difference in characteristic  $X_j$  between sample  $E_i$  and  $EK_z$ . The  $\text{mld}(E_i, EK_z)$  uses the Hamming Distance (HD) to measure the difference between sample  $q$  labeled  $Y_a$  and  $Y_b$ , and its calculation method is shown in Equation (5).

$$\text{mld}(E_i, EK_z) = \text{HD}(Y_a, Y_b) = \frac{|Y_a \cup Y_b| - |Y_a \cap Y_b|}{q} \quad (5)$$

The overall flow of the RF-ML algorithm is shown in Figure 4. First, the algorithm selects  $c$  samples by random sampling, where  $c$  is the total number of samples, and  $k$  neighbor weights are calculated for each sample. Second, the  $k$  nearest neighbor (KNN) algorithm is used to calculate the  $k$  nearest neighbors for each sample. KNN is an instance-based classification algorithm that determines sample categories by measuring the distance between samples. We used the KNN algorithm to calculate the Euclidean distance between samples and sort the first  $k$  nearest neighbor samples of the current sample. Based on the characteristics and labels of these samples, the importance weights of  $M$  vegetation indexes were calculated, respectively. Finally, the calculated weight was weighted to the corresponding vegetation index by point multiplication through the softmax function to realize the significant attention weighting.







**Figure 5.** The overall structure of AMMFNet.

### 2.3.1. Data Input Layer

To solve the problem of the low accuracy of single-source images using the diseases and pest classification algorithm, RGB and multispectral images were used as the data input of the model. This enhances the expression of the data information by using complementary information from multi-source images. Because multispectral images can reflect the physiological information of apples, to fully utilize the information in multispectral images, AMMFNet adopts the vegetation index selection algorithm introduced in Section 2.1 to process multispectral images into weighted vegetation indices. Furthermore, RGB images focus on intuitive effects and contain richer texture information. Therefore, AMMFNet uses RGB and multispectral images as multi-source inputs and transmits the processed RGB images and vegetation indices to the data fusion layer to obtain a more comprehensive representation of the characteristics of the apple orchard.

### 2.3.2. Data Fusion Layer

In the canopy images of apple diseases and pests, RGB images can reflect the texture information of vegetation, while vegetation indices reflect the physiological characteristics of vegetation. To make more comprehensive use of this information and fully excavate the common features beneficial to the prediction of apple diseases and pests, the above two kinds of images need to be fused.

In the multi-source image fusion task, according to different stages of fusion, the fusion method can be divided into data-level fusion, feature-level fusion, and decision-level fusion, which, respectively, correspond to the original data source stage, feature representation stage, and decision-making stage. The RGB and multispectral images of the canopy apple pests and diseases were spatially aligned due to simultaneous shooting at the same time and position, thus ensuring the spatial consistency between the RGB image and the calculated vegetation indices. Data-level fusion can ensure data integrity and help reduce the situations in which details are ignored due to the macro features obtained from feature extraction. This approach can mine the common features beneficial for disease and pest prediction from both types of data more effectively and can comprehensively describe the growth status and disease and pest conditions of apple trees. Therefore, the data-level fusion method was adopted to combine the two kinds of image data. The process can be expressed as Equation (6).

$$X = Cat(X_{RGB}, Dot(X_{VIs}, W_{RF-ML})) \tag{6}$$

wherein  $X$  is the fused data,  $X_{RGB}$  is the RGB images,  $X_{VIs}$  is the vegetation indices, and  $W_{RF-ML}$  is the importance weight calculated by RF-ML.  $Cat(\cdot)$  represents the data tensor stitching in the channel dimension, and  $Dot(\cdot)$  represents the data tensor and weight vector dot product in the channel dimension.

The fused data need to further integrate the texture information contained in the RGB image and the physiological information reflected by the vegetation index. Each channel of the fused data may carry features related to diseases and insect pests. For example, the red channel of the RGB image highlights the edge and veins of the leaves, and the green channel can better capture the changes in the surface texture and details of the leaves. The vegetation index can calculate the correlation with diseases and pests through the feature selection algorithm as shown in Figure 6. MTVI2, SR, and NLI can capture the changes in apple leaves caused by the invasion of diseases and insect pests such as brown spot or gray spot by monitoring physiological parameters such as vegetation growth status, chlorophyll content, and coverage, which has strong sensitivity to diseases and insect pests. Therefore, to reintegrate the pest and disease-sensitive channels in the data before inputting into the model, channel attention was used to ensure that the model automatically focuses on the feature channels sensitive to pests and diseases, fully explores the complementary information among multi-source features, and enhances the final feature representation. Channel attention [59] is a common attention mechanism in deep learning, aiming at improving the model attention to different channels of the input data. By learning the weights of each channel, the model can adaptively adjust the importance of the channels to capture the features of the input data more effectively. Specifically, the channel attention calculates the global average-pooling and global max-pooling of the input image, then passes through a shared multi-layer perceptron (MLP), and finally adds the two output results and calculates the channel attention score through the activate function. It then multiplies the attention score with the original image tensor point to enhance the results. The calculation is shown in Equation (7).

$$X' = Dot(\sigma(MLP(MaxPool(X)) + MLP(AvgPool(X))), X) \tag{7}$$

where  $X$  is the fused image after weighting, and  $\sigma(\cdot)$  is the activation function. The *Sigmoid* function is used as the activation function.

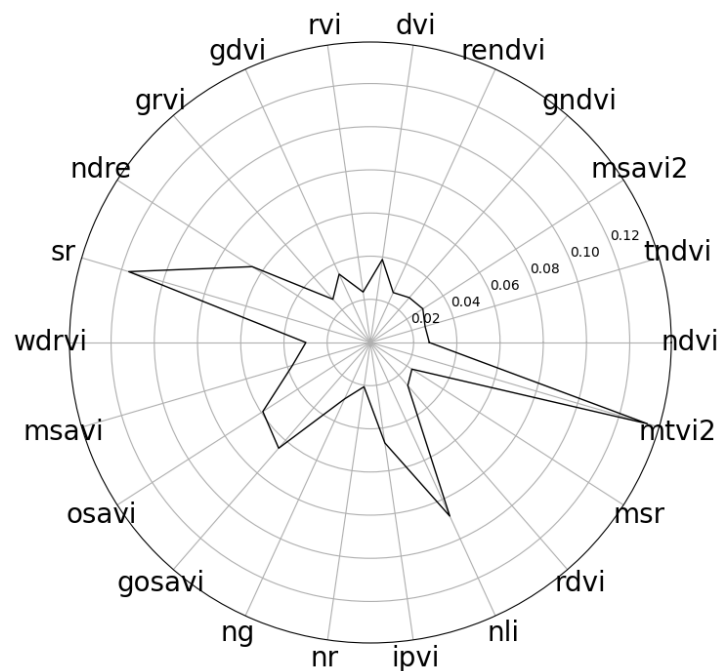


Figure 6. The weight of the vegetation indices calculated by RF-ML.

### 2.3.3. Feature Extraction Layer

Apple canopy fusion data contain rich disease and pest information; therefore, ResNet with its deep network structure advantages was adopted as the feature extraction network of the model in dealing with multi-label classification tasks of apple pests and diseases. In ResNet, the low-level convolution layer is mainly responsible for extracting the low-level features such as the texture, edges and colors of the disease and pest images. As the depth of the network increases, the residual blocks gradually extract more abstract mid-level features, such as texture variations, morphological features, and structural information from smaller regions, which helps the model understand more complex and similar types of diseases and pests. In the deep convolutional layers, the network focuses more on global context information, thereby extracting high-level semantic features, including information such as the overall shape and distribution patterns of pest and disease regions, which enables the model to better understand the image content. In general, the hierarchical feature extraction mechanism of ResNet provides a strong feature learning ability for diseases and pest classification tasks, which can effectively distinguish different types of diseases and pests and improve the performance and generalization ability of the model.

### 2.3.4. Classifier

Apple diseases and pests often exhibit different geometric shapes and visual characteristics, and there can be multiple types of pests and diseases present within a single region. Therefore, it is difficult for a single classifier to capture the complex characteristics of various types of diseases and pests at the same time, resulting in poor classification performance. For the features output by the feature extraction network, AMMFNet uses multiple binary classifiers to realize multi-label classification tasks. As shown in Figure 5, six binary classifiers (background and five types of diseases and insect pests) were adopted. Each binary classifier was specifically aimed at the classification of a type of disease or pests, thus realizing the multi-label classification task. This enables the model to carry out flexible learning according to the characteristics of different diseases and insect pests and finally improves the recognition accuracy and robustness of multi-label diseases and pests. Furthermore, the usage of multiple binary classifiers can also enhance the prediction interpretability of the model, as the output of each binary classifier can directly reflect the confidence level or probability of the model to the corresponding pest labels, which is convenient to understand and explain the prediction process of the model.

## 3. Results and Discussion

Extensive experiments and comparative analyses were carried out on the constructed multi-source apple canopy image dataset of apple diseases and pests to verify the effectiveness of the proposed method.

### 3.1. Experimental Environment Configuration

In the experiment, the training and evaluation of all the models were conducted on a setup comprising four NVIDIA Tesla T4 GPUs, with the server configuration parameters detailed in Table 4. The optimization for all the models was carried out using the Stochastic Gradient Descent (SGD) algorithm, with the momentum value set to 0.9. The batchsize was 128, and weight decay was implemented at a rate of  $5 \times 10^{-5}$ . The learning rate adjustment was accomplished through a cosine annealing schedule, initialized at  $1 \times 10^{-2}$  and culminating at a terminal rate of  $1 \times 10^{-4}$ . Pre-training for all the models was executed on the CIFAR100 dataset, followed by a fine-tuning process over 100 epochs on a multi-source image dataset depicting apple tree canopies afflicted with diseases and pests.

**Table 4.** Experimental setup.

Configuration Item	Value
CPU	Intel(R) Xeon(R) CPU E5-2650 v3 (Intel, Santa Clara, CA, USA)
GPU	NVIDIA Tesla T4 16 GB (NVIDIA, Santa Clara, CA, USA)
CUDA	11.3
Operating system	Ubuntu 18.04.2 LTS (64-bit)
Memory	128 GB
Hard drive	2 TB
Deep learning framework	PyTorch 1.11.0
Language	Python3.8

### 3.2. Evaluation Metrics

To verify the validity of the model, the following evaluation metrics were used to validate the model on the validation set. The evaluation metrics are divided into subset-based accuracy  $subsetacc$ , as shown in Equation (8), and sample-based evaluation metrics, accuracy, recall, precision, and F1 scores, calculated according to the Equations (9), (10), (11), and (12), respectively.

$$Subsetacc(h) = \frac{1}{p} \sum_{i=1}^p [[h(x_i) = Y_i]] \quad (8)$$

$$Accuracy(h) = \frac{1}{p} \sum_{i=1}^p \frac{|Y_i \cap h(x_i)|}{|Y_i \cup h(x_i)|} \quad (9)$$

$$Recall(h) = \frac{1}{p} \sum_{i=1}^p \frac{|Y_i \cap h(x_i)|}{|Y_i|} \quad (10)$$

$$Precision(h) = \frac{1}{p} \sum_{i=1}^p \frac{|Y_i \cap h(x_i)|}{|h(x_i)|} \quad (11)$$

$$F1(h) = \frac{2 \cdot precision_{exam}(h) \cdot Recall_{exam}(h)}{Precision_{exam}(h) + Recall_{exam}(h)} \quad (12)$$

where  $h(\cdot)$  presents the model function,  $p$  is the total number of samples,  $X_i$  is the  $i$ -th sample, and  $Y_i$  is the corresponding label for the sample. The calculation is correct only if the prediction result is exactly consistent with the sample multi-label.

#### 3.2.1. Comparison of Model Classification Accuracy

To verify the effectiveness of AMMFNet in apple disease and pest diagnosis, under the same experimental conditions, the constructed apple disease and insect pest multi-source canopy image dataset was used to compare the experimental performance of AMMFNet with other classification models. The experimental results show that the model accuracy is the best when the AMMFNet feature extraction network uses ResNet-18. Therefore, the feature extraction networks of AMMFNet in the comparative experiments all used ResNet-18. The comparison models used in the experiment included the multi-category classification model residual error attention network (Attention-92) [60] and NFNet [61] and the multi-label classification model Query2Label [62]. All the comparison models used a combination of single-source RGB images and RGB + VIs as inputs to verify the improved performance of AMMFNet relative to single-source input prediction and the effectiveness of multi-source feature fusion.

The experimental results are shown in Tables 5 and 6. The experimental results show that the accuracy of AMMFNet proposed in this paper is significantly higher than that of all the models using single-source input RGB images, and at the same time, it has lower

computational overhead. AMMFNet performs significantly better in accuracy than models that use only a single RGB input. When using RGB + VIs combined input, although it should theoretically provide more abundant information, not all models can effectively integrate these multi-source data. Among them, due to its excessive dependence on the attention mechanism, the Attention-92 model pays too much attention to redundancy data, which fail to effectively learn and integrate multi-source features, and its performance is even lower than that of the model using only RGB input. At the same time, each attention module of Attention-92 needs to perform calculations such as weight generation and feature recalibration, which takes up a large amount of computational resources and memory, resulting in a large forward pass size. The Query2Label model, due to the computational complexity of its self-attention mechanism, has a significantly higher forward pass size and parameter size than other models. In comparison, AMMFNet, with a smaller forward pass and parameter size, not only excels among all the models using combined inputs but also surpasses the high-performing NFNet model by 4.7% in subset accuracy. Although Weight Standardization is used in NFNet to replace the traditional BN layer, due to the large distribution differences between data channels, this standardization method has limited effect in processing multi-source data. However, before feature extraction, AMMFNet performs global feature selection on the input data, effectively enhancing the fusion of multi-source data and achieving higher model performance. In addition, AMMFNet is also more accurate than other models in predicting single apple diseases and insect pests. This shows that the proposed model can effectively fuse multi-source data and improve the accuracy of apple pest classification.

**Table 5.** Comparison of apple disease classification performance and computational complexity among different classification models.

Input Image	Model	Subsetacc/%	Accuracy/%	Recall/%	Precision/%	F1/%	Forward Pass Size (MB)	Params Size (MB)
RGB	Attention-92	39.19	61.29	89.71	61.30	72.81	723.53	192.37
	NFNet	80.04	75.61	76.62	77.11	76.92	199.65	<b>161.76</b>
	Query2Label	79.71	77.95	80.14	79.91	79.95	924.47	367.24
RGB + VIs	Attention-92	35.49	59.51	88.80	59.73	71.36	723.53	192.64
	NFNet	88.18	82.00	82.69	83.60	83.11	199.65	162.03
	Query2Label	86.16	81.49	82.89	83.22	83.10	924.47	367.51
	<b>AMMFNet</b>	<b>92.92</b>	<b>85.43</b>	<b>85.89</b>	<b>86.54</b>	<b>86.21</b>	<b>192.75</b>	164.63

The bold content is the column optimal value.

**Table 6.** Comparison of single-class disease and pest classification accuracy (%) among different classification models.

Input Image	Model	Aphid	Alternaria	Mosaic	Brown Spot	Gray Spot
RGB	Attention-92	68.51	87.31	99.60	72.02	86.33
	NFNet	94.91	97.92	100.0	89.60	96.82
	Query2Label	94.95	97.31	100.0	90.59	96.45
RGB + VIs	Attention-92	73.82	84.54	99.41	72.33	81.60
	NFNet	96.50	98.59	100.0	95.11	97.67
	Query2Label	95.81	98.33	99.90	93.86	97.17
	<b>AMMFNet</b>	<b>97.54</b>	<b>99.01</b>	<b>100.0</b>	<b>97.41</b>	<b>98.47</b>

The bold content is the column optimal value.

### 3.2.2. Effectiveness Analysis of Multi-Source Fusion Prediction Model

This section focuses on validating the effectiveness of AMMFNet by comparing its prediction accuracy with models trained using three different data sources: visible light images, multispectral images, and calculated vegetation indices. Additionally, this experiment also evaluates the classification accuracy of the AMMFNet model for individual categories to verify its effectiveness in identifying specific apple pest and disease categories.

As shown in Table 7, the RGB, MS, and VIs in the table represent the RGB image, the multispectral image, and the calculated vegetation indices, respectively, as the input model recognition result obtained. The experiments show that AMMFNet is superior to single-source input on ResNet18, ResNet34, and ResNet50, because RGB images and multispectral images carry different features, and these features are all conducive to pest classification. AMMFNet effectively uses the complementary information in the RGB and multispectral images taken by the UAV to form an effective feature representation. Among them, the recognition model using ResNet18 as the feature extraction network performed best, achieving a 92.92% subset accuracy, 85.43% sample accuracy, and 86.21% F1 score. This is because both the vegetation indices themselves and the multi-source data after feature selection and fusion already have more significant pest-sensitive characteristics. Compared with the deeper network structure, ResNet18 can converge faster and learn these significant data characteristics, but it will not be disturbed by noise in the high feature dimension. In addition, the subset accuracy of AMMFNet exceeds the subset accuracy of single-source RGB images, multispectral images, and vegetation indices.

The accuracy rate of AMMFNet for five types of apple diseases and insect pests is shown in Table 8. It can be found that the accuracy rate of AMMFNet for each type of disease and insect pest is higher than 97%. Among them, the recall rate of brown spot disease is the highest, reaching 95.6%. The recall rate of mosaic disease is the lowest. The above experimental results are related to the distribution of the dataset, and the number of positive samples is small. These experimental results prove that the proposed method can effectively combine the characteristics of RGB images and vegetation indices to learn the characteristics of various diseases and pests and accurately identify five common apple diseases and insect pests.

**Table 7.** Comparison of classification results between AMMFNet and single-source input model.

Model	Input	Subsetacc/%	Accuracy/%	Recall/%	Precision/%	F1/%
ResNet18	RGB	83.99	78.26	78.53	79.92	79.22
	MS	82.02	76.34	77.69	77.59	77.64
	VIs	87.58	80.71	80.46	81.77	81.11
	<b>AMMFNet</b>	<b>92.92</b>	<b>85.43</b>	<b>85.89</b>	<b>86.54</b>	<b>86.21</b>
ResNet34	RGB	82.38	77.64	78.06	79.83	78.93
	MS	83.10	77.37	78.04	79.38	78.70
	VIs	86.20	80.15	80.95	81.93	81.44
	<b>AMMFNet</b>	<b>87.02</b>	<b>81.88</b>	<b>82.71</b>	<b>83.99</b>	<b>83.34</b>
ResNet50	RGB	83.83	78.56	79.56	81.04	80.29
	MS	85.80	79.51	80.10	81.20	80.72
	VIs	86.40	80.28	80.97	82.09	81.53
	<b>AMMFNet</b>	<b>87.38</b>	<b>81.80</b>	<b>82.61</b>	<b>83.65</b>	<b>83.13</b>

The bold content is the column optimal value.

**Table 8.** Single-class apple disease and pest classification results of AMMFNet.

Evaluation Metrics	Aphid	Alternaria	Mosaic	Brown Spot	Gray Spot
Accuracy/%	97.54	99.01	<b>100.0</b>	97.41	98.47
Recall/%	79.63	78.75	66.77	<b>95.64</b>	86.09
Precision/%	94.02	95.91	<b>100.0</b>	97.49	85.64

The bold content is the column optimal value.

### 3.2.3. Ablation Experiment

To further verify the effectiveness of AMMFNet, the ablation experiments were designed to verify the effectiveness of each part of the model. The experimental results are shown in Table 9. When the RGB image is used as the input reference model, the vegetation indices can reflect more sufficient physiological information of apple plants than the

RGB image. Therefore, the subset accuracy rate predicted by the 22 vegetation indices is 3.59% higher than the RGB single-source image input result, and the improvement is the largest. And the use of the RF-ML algorithm can further improve the accuracy of model recognition. And RGB images can compensate for the local texture details that vegetation indices lack. Therefore, fusing these two types of images improves the subset accuracy by 6.21% compared to the baseline model. In addition, for two kinds of multi-source images, using data-level feature fusion can improve the quality of feature representation, while using feature-level fusion will cause the lack of detailed features of the diseases and insect pests, which will reduce the accuracy of the model, as shown in Table 10. After feature fusion, channel attention is used to re-weight the features, which further improves the classification accuracy by 2.72%. Ultimately, AMMFNet achieves a total improvement of 8.93% in subset accuracy compared to the baseline model.

**Table 9.** Ablation experiment.

Model	RGB	VI <sub>s</sub>	RF-ML	Channel Attention	Data-Level Fusion	Subsetacc/%
ResNet18	✓					83.99
		✓				87.58
	✓	✓				86.40
		✓	✓			88.14
	✓	✓	✓		✓	90.20
	✓	✓	✓	✓	✓	<b>92.92</b>

✓ represents the items retained in the ablation experiment.

**Table 10.** Classification results among RGB, multispectral (MS), and vegetation indices (VIs).

Model	Input Image	Subsetacc/%	Accuracy/%	Recall/%	Precision/%	F1/%
ResNet18	RGB	83.99	78.26	78.53	79.92	79.22
	MS	82.02	76.34	77.69	77.59	77.64
	VIs	<b>87.58</b>	<b>80.71</b>	<b>82.05</b>	<b>81.97</b>	<b>81.90</b>
ResNet34	RGB	82.38	77.64	78.06	79.83	78.93
	MS	83.10	77.37	78.04	79.38	78.70
	VIs	<b>86.20</b>	<b>80.15</b>	<b>80.95</b>	<b>81.93</b>	<b>81.44</b>
ResNet50	RGB	83.83	78.56	79.56	81.04	80.29
	MS	85.80	79.51	80.10	81.20	80.72
	VIs	<b>86.40</b>	<b>80.28</b>	<b>80.97</b>	<b>82.09</b>	<b>81.53</b>
ResNet101	RGB	86.61	81.14	81.85	83.01	82.50
	MS	87.32	80.67	81.43	82.05	81.74
	VIs	<b>89.26</b>	<b>82.29</b>	<b>82.83</b>	<b>83.61</b>	<b>83.22</b>

The bold content is the column optimal value.

### 3.2.4. Impact of Input Data on Model Performance

To verify the effectiveness of the calculated vegetation indices combination, this experiment trained ResNet series models on RGB, multispectral, and vegetation indices datasets, respectively. The universal effectiveness of the selected vegetation indices was validated through ResNet18, ResNet34, ResNet50, and ResNet101.

The experimental results are shown in Table 10. The results show that the accuracy of the proposed vegetation indices combination on the four models is higher than that of the RGB and multispectral images. Especially when the model structure is simpler, such as ResNet18, the vegetation indices combination has greater performance improvement. The reason is that a simpler model structure often implies fewer parameters and less computational load. Therefore, the model can quickly learn effective feature representations from the vegetation indices representing the physiological information of apples. In contrast, complex models, due to their stronger learning capabilities, may fit unnecessary noise.



Hence, vegetation indices are more effective on shallower networks. Compared with the model trained with RGB images, the accuracy of the subset of ResNet18 trained with vegetation indices increased by 3.59%, and the F1-score increased by 2.68%.

### 3.2.5. Impact of RF-ML Algorithm on Model Accuracy

The ReliefF algorithm controls the size of the dataset it adapts to by selecting the number of neighbors  $k$ . A larger value means that more distant neighbors can influence the weight results, which improves the robustness and accuracy of the algorithm, but it also means that the algorithm will be subject to more noise interference. The corresponding weights are calculated by the RF-ML suitable for multi-label classification tasks, and the influence of the calculated weights combined with the vegetation indices on the accuracy of the model when  $k = 5, 10, 20$  is compared. As shown in Table 11, when  $k = 5$ , the RF-ML algorithm performs best, and the subset accuracy on ResNet18 is 0.61% higher than the unweighting vegetation indices. This shows that when the number of adjacent samples is 5, the importance weight calculated by RF-ML can better fit the sample distribution of the dataset and help correct the focus of attention of the model. However, with the increase in the  $k$  value, the algorithm will produce overfitting, which will lead to a decrease in the accuracy of the model. Figure 6 shows the importance of the combination of the vegetation index calculated by RF-ML at  $k = 5$  weight.

**Table 11.** Classification results of the model guided by the RF-ML algorithm under different  $k$  values.

Value	Subsetacc/%	Accuracy/%	Recall/%	Precision/%	F1/%
$k = 5$	<b>88.14</b>	<b>81.13</b>	<b>81.84</b>	<b>82.40</b>	<b>82.22</b>
$k = 10$	85.64	79.06	80.08	80.34	80.21
$k = 20$	86.89	80.05	81.33	81.24	81.28

The bold content is the column optimal value.

### 3.2.6. Influence of Multi-Source Data Fusion Method on Model Accuracy

RGB images can make up for the missing texture features of pests and diseases in vegetation index features. Therefore, the method of RGB image fusion vegetation indices is adopted to further improve the precision of the model. Data-level fusion and feature-level fusion are, respectively, used in model training to explore appropriate multi-source data fusion methods. Data-level fusion refers to aligning the raw data of RGB images and vegetation indices and then concatenating them in the channel dimension to form 25-dimensional channel data for model prediction. Feature-level fusion, on the other hand, involves extracting features from RGB images and vegetation indices separately through a backbone network, concatenating these features in the channel dimension, and then sending them into a classifier for multi-label classification results. In the experiment, ResNet18 was adopted as the feature extraction network of AMMFNet. As shown in Table 12, the experimental results show that the data-level fusion is more suitable for the fusion of the RGB image and the vegetation indices. Because the vegetation indices are calculated by the meridian linearity and nonlinearity of the multispectral image, and the multispectral image is taken at the same time as the RGB image, the two positions are aligned and have a high degree of spatial consistency. Therefore, the fusion of RGB images and vegetation indices at the data level can ensure that the two data form a more consistent data expression form so as to promote the model to better capture the detailed features in the two. While the features extracted by the network are independent feature expressions of the two data types, they have become advanced semantics that can be classified. At this point, performing feature fusion between the two semantics is equivalent to introducing noise into the clear semantic information, which interferes with the original feature representation and results in classification performance that is even lower than that of single-source input.

**Table 12.** Classification results of the model under different fusion methods.

Method	Model	Subsetacc/%	Accuracy/%	Recall/%	Precision/%	F1/%
Data-level fusion	ResNet18	<b>90.02</b>	<b>83.60</b>	<b>84.51</b>	<b>84.90</b>	<b>84.71</b>
Feature-level fusion	ResNet18	85.64	79.06	80.08	80.34	80.21

The bold content is the column optimal value.

### 3.2.7. Influence of Noise on Model Performance

To further evaluate the performance of the proposed method, AMMFNet was tested using a validation set with noise added to test the recognition ability of the model in the actual situation when the image quality is impaired.

Specifically, Gaussian noise and salt and pepper noise are added to the RGB and multispectral images in the verification set. Gaussian noise is a noise whose probability density function obeys normal distribution, which can simulate the random interference that image sensors may receive. Salt and pepper noise is formed by black and white dots randomly appearing on the image and is often used to simulate possible errors during data transmission. The standard deviation of Gaussian noise is set to 0.3, and the percentage of salt and pepper noise is set to 0.1. The experimental results are shown in Table 13.

**Table 13.** Single-class apple disease and pest classification results of AMMFNet under noise interference.

Evaluation Metrics	Aphid	Alternaria	Mosaic	Brown Spot	Gray Spot
Accuracy/%	94.83	97.41	<b>99.98</b>	93.94	96.21
Recall/%	73.33	59.72	60.83	<b>91.75</b>	76.32
Precision/%	91.15	93.90	<b>94.68</b>	87.02	92.62

The bold content is the row optimal value.

Comparing the performance metrics on the original validation set (Table 8) and the noisy validation set (Table 13), although the addition of noise has had some impact on the model's performance, the changes in the various evaluation metrics of the model are relatively small. Especially, it still maintains a high accuracy rate. Therefore, the proposed model has good stability and robustness.

### 3.2.8. Main Contributions of This Work

To further illustrate the advancements made in this study, we provide a comprehensive comparison between the key contributions of our work and those reported in previous studies. Table 14 summarizes the main differences and advancements achieved in our study.

**Table 14.** Contribution comparison.

Contribution Aspect	This Work	Literature References
Methodology	Developed a multi-source image fusion approach combining RGB and multispectral images.	Utilized single-source RGB images or multispectral images for classification [4,17].
Feature selection	Implemented global feature selection using ReliefF algorithm to enhance multi-source data fusion.	Did not include feature selection steps or use artificial feature screening [5,17].

Table 14. Cont.

Contribution Aspect	This Work	Literature References
Attention mechanism	Applied channel attention to re-weight fused features, improving model performance by 2.72%. Evaluated the effectiveness of different input data combinations on various model architectures, demonstrating the superiority of the vegetation indices combination.	Used different attention mechanisms [5], which did not achieve comparable improvements.  Compared single-source input methods without exploring the impact of different input data combinations on model performance [4,5,17].
Input data combination	Achieved a subset accuracy of 92.92%, outperforming single-source methods by 8.93% for RGB and 10.9% for multispectral images.	Reported subset accuracies of 83.99% for RGB images and 82.02% for multispectral images [5].
Accuracy		

This comprehensive comparison highlights the advances and innovations introduced in our research, highlighting the effectiveness of our multi-source image fusion method and the significant improvements made in accuracy.

#### 4. Conclusions

A regional classification method of apple diseases and insect pests based on multi-source image fusion is proposed. Firstly, the RGB images and multispectral canopy images of an apple orchard are collected by UAV. After regional division and labeling, a multi-source apply canopy image dataset of diseases and pests is constructed. Secondly, a vegetation indices selection method based on salient attention is proposed, which calculates the vegetation index with a strong correlation between the physiological information of 22 apples and optimizes the selection of the vegetation indices. Finally, a multi-label classification model of the apple disease and pest areas based on RGB and multispectral image fusion is proposed, which can effectively combine the texture information of RGB images and the apple physiological information reflected by multispectral images. The experiments show that the subset accuracy of the model proposed in this paper, tested on the apple pest and disease canopy multi-source image dataset, is 8.93% and 10.9% higher than prediction methods based on single-source RGB images or multispectral images, reaching 93.92%. It is worth mentioning that the classification accuracy for individual apple pest and disease regions is higher than 97%. Therefore, this model meets the accuracy requirements of practical production, providing technical support for coarse-grained localization of apple pests and diseases and ensuring the healthy development of the apple industry. At the same time, in order to extend the proposed method to other plant species to improve the universality of disease and insect pest diagnosis, how to develop spectral feature extraction methods suitable for more plant species to avoid differences in spectral features between species, data quality and quantity, and the influence of changes in spectral characteristics under different conditions is an important direction for further research in the future.

**Author Contributions:** Conceptualization, H.L. (Hanye Liu) and H.L. (Hengzhao Li); methodology, L.S. and H.L. (Hanye Liu); software, L.S. and H.L. (Hengzhao Li); validation, L.S. and B.T.; formal analysis, H.L. (Hengzhao Li) and B.T.; investigation, B.T.; resources, B.L.; data curation, B.T. and L.S.; writing—original draft preparation, B.T.; writing—review and editing, H.L. (Hanye Liu) and H.Z.; visualization, B.T.; supervision, H.Z. and H.L. (Hanye Liu); project administration, B.L.; funding acquisition, B.L. All the authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Natural Science Foundation of China (No. 62376226), by Shaanxi's Key Research and Development Program (No. 2024NC-YBXM-191), by the Key Industry Chain of Shaanxi Provincial Key Research and Development Program (No. 2023-ZDLNY-63), by Xianyang's Key Research and Development Program (No. L2022-ZDYF-NY-019), by Technology Innovation Leading Program of Shaanxi (No. 2024QCY-KXJ-094), by Key technology projects of key agricultural industry chain in Xi'an (NO. 2024JH-NYZD-0027), and by National Natural Science Foundation Youth Project (NO. 62406254).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Some or all data, models, or codes that support the findings of this study are available from the corresponding author upon reasonable request. The data are not publicly available due to privacy.

**Conflicts of Interest:** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References

1. Guo, H.; Cao, Y.; Wang, C.; Rong, L.; Li, Y.; Wang, T.; Yang, F. Recognition and application of apple defoliation disease based on transfer learning. *Trans. CSAE* **2024**, *40*, 184–192.
2. Zhou, M.; Ou, Y.; Zhang, L.; Tong, G.; Wang, Y. Effect of dielectric properties on radio frequency heating uniformity of apple. *Trans. CSAE* **2019**, *35*, 273–279.
3. Qiu, M.; Liu, B.; Liu, Y.; Wang, K.; Pang, J.; Zhang, X. Simulation of first flowering date for apple and risk assessment of late frost in main producing areas of northern China. *Trans. CSAE* **2020**, *36*, 154–163.
4. Zhong, Y.; Zhao, M. Research on deep learning in apple leaf disease recognition. *Comput. Electron. Agric.* **2020**, *168*, 105146. [[CrossRef](#)]
5. Mahmud, M.S.; He, L.; Zahid, A.; Heinemann, P.; Choi, D.; Krawczyk, G.; Zhu, H. Detection and infected area segmentation of apple fire blight using image processing and deep transfer learning for site-specific management. *Comput. Electron. Agric.* **2023**, *209*, 107862. [[CrossRef](#)]
6. Liu, B.; Jia, R.; Zhu, X.; Yu, C.; Yao, Z.; Zhang, H.; He, D. A lightweight identification model for apple leaf diseases for mobile terminals. *Trans. CSAE* **2022**, *38*, 130–139.
7. Ahad, M.T.; Li, Y.; Song, B.; Buhuiyan, T. Comparison of CNN-based deep learning architectures for rice diseases classification. *Artif. Intell. Agric.* **2023**, *9*, 22–35. [[CrossRef](#)]
8. Liu, B.; Ding, Z.; Tian, L.; He, D.; Li, S.; Wang, H. Grape leaf disease identification using improved deep convolutional neural networks. *Front. Plant Sci.* **2020**, *11*, 1082. [[CrossRef](#)]
9. Xie, X.Y.; Ma, Y.; Liu, B.; He, J.; Li, S.; Wang, H. A deep-learning-based real-time detector for grape leaf diseases using improved convolutional neural networks. *Front. Plant Sci.* **2020**, *11*, 751. [[CrossRef](#)]
10. Zeng, T.; Li, S.; Song, Q.; Zhong, F.; Wei, X. Lightweight tomato real-time detection method based on improved YOLO and mobile deployment. *Comput. Electron. Agric.* **2023**, *205*, 107625. [[CrossRef](#)]
11. Tang, Z.; Lu, J.; Chen, Z.; Qi, F.; Zhang, L. Improved Pest-YOLO: Real-time pest detection based on efficient channel attention mechanism and transformer encoder. *Ecol. Inform.* **2023**, *78*, 102340. [[CrossRef](#)]
12. Tian, L.; Zhang, H.; Liu, B.; Qi, F.; Zhang, L. VMF-SSD: A novel V-space based multi-scale feature fusion SSD for apple leaf disease detection. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2022**, *20*, 2016–2028. [[CrossRef](#)] [[PubMed](#)]
13. Liu, B.; Ren, H.; Li, J.; Duan, N.; Yuan, A.; Zhang, H. RE-RCNN: A novel representation-enhanced RCNN model for early apple leaf disease detection. *ACM Trans. Sens. Netw.* **2023**, *1550–4867*. [[CrossRef](#)]
14. Cob-Parro, A.C.; Lalangui, Y.; Lazcano, R. Fostering Agricultural Transformation through AI: An Open-Source AI Architecture Exploiting the MLOps Paradigm. *Agronomy* **2024**, *14*, 259. [[CrossRef](#)]
15. Singha, A.; Moon, M.S.H.; Dipta, S.R. An End-to-End Deep Learning Method for Potato Blight Disease Classification Using CNN. In Proceedings of the 2023 International Conference on Computational Intelligence, Networks and Security (ICCINS), Mylavaram, India, 22–23 December 2023; Volume 12, pp. 1–6.
16. Mohite, J.; Sawant, S.; Agarrwal, R.; Pandit, A.; Pappula, S. Detection of crop water stress in maize using drone based hyperspectral imaging. In Proceedings of the IGARSS 2022—2022 IEEE International Geoscience and Remote Sensing Symposium, Kuala Lumpur, Malaysia, 17–22 July 2022; pp. 5957–5960.
17. Enciso, J.; Avila, C.A.; Jung, J.; Elsayed-Farag, S.; Chang, A.; Yeom, J.; Landivar, J.; Maeda, M.; Chavez, J.C. Validation of agronomic UAV and field measurements for tomato varieties. *Comput. Electron. Agric.* **2019**, *158*, 278–283. [[CrossRef](#)]
18. Theau, J.; Gavelle, E.; Menard, P. Crop scouting using UAV imagery: A case study for potatoes. *J. Unmanned Veh. Syst.* **2020**, *8*, 99–118. [[CrossRef](#)]

19. Brewer, K.; Clulow, A.; Sibanda, M.; Gokool, S.; Odindi, J.; Mutanga, O.; Naiken, V.; Chimonyo, V.G.P.; Mabhaudhi, T. Estimation of maize foliar temperature and stomatal conductance as indicators of water stress based on optical and thermal imagery acquired using an Unmanned Aerial Vehicle platform. *Drones* **2022**, *6*, 169. [[CrossRef](#)]
20. Jiang, Y.; Liu, B.; Zhang, C.; Zhao, D.; Chen, R.Q.; Xu, B.; Long, H.L.; Yang, G.J.; Yang, H. Monitoring the maturity of multi-variety corn using multi-spectral imagery from an unmanned aerial vehicle. *Trans. CSAE* **2023**, *39*, 84–91.
21. Fei, S.; Hassan, M.A.; Xiao, Y.; Su, X.; Chen, Z.; Cheng, Q.; Duan, F.; Chen, R.; Ma, Y. UAV-based multi-sensor data fusion and machine learning algorithm for yield prediction in wheat. *Precis. Agric.* **2023**, *24*, 187–212. [[CrossRef](#)]
22. Yokoyama, Y.; De Wit, A.; Matsui, T.; Tanaka, T.S.T. Predicting plant-level cabbage yield by assimilating UAV-derived LAI into a crop simulation model. *Precis. Agric.* **2023**, 1043–1048.
23. Šupčík, A.; Beranová, V. *Grape Yield Prediction Based on Vine Canopy Morphology Obtained by 3D Point Clouds from UAV Images*; Wageningen Academic: Wageningen, The Netherlands, 2023; pp. 619–625.
24. Yan, H.; Zhuo, Y.; Li, M.; Wang, Y.; Guo, H.; Wang, J.; Li, C.; Ding, F. Prediction of alfalfa yield based on machine learning and remote sensing of multi-spectral images from an unmanned aerial vehicle. *Trans. CSAE* **2022**, *38*, 64–71.
25. Kent, O.W.; Chun, T.W.; Choo, T.L.; Lai, W.K. Early symptom detection of basal stem rot disease in oil palm trees using a deep learning approach on UAV images. *Comput. Electron. Agric.* **2023**, *213*, 108192. [[CrossRef](#)]
26. Antolínez García, A.; Cáceres Campana, J.W. Identification of pathogens in corn using near-infrared UAV imagery and deep learning. *Precis. Agric.* **2023**, *24*, 783–806. [[CrossRef](#)]
27. Das, A.K.; Mathew, J.; Zhang, Z.; Friskop, A.; Huang, Y.; Flores, P.; Han, X. Corn goss's wilt disease assessment based on UAV imagery. In *Unmanned Aerial Systems in Precision Agriculture: Technological Progresses and Applications*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 123–136.
28. Zhao, J.; Jin, Y.; Ye, H.; Huang, W.; Dong, Y.; Fan, L.; Ma, H. Remote sensing monitoring of areca nut yellowing disease based on multi-spectral images from an unmanned aerial vehicle. *Trans. CSAE* **2020**, *36*, 54–61.
29. Wang, C.; Liu, Y.; Zhang, Z.; Han, L.; Li, Y.; Zhang, H.; Wongsuk, S.; Li, Y.; Wu, X.; He, X. Spray performance evaluation of a six-rotor unmanned aerial vehicle sprayer for pesticide application using an orchard operation mode in apple orchards. *Pest Manag. Sci.* **2022**, *78*, 2449–2466. [[CrossRef](#)] [[PubMed](#)]
30. Huang, J.; Luo, Y.; Quan, Q.; Wang, B.; Xue, X.; Zhang, Y. An autonomous task assignment and decision-making method for coverage path planning of multiple pesticide spraying UAVs. *Comput. Electron. Agric.* **2023**, *212*, 108128. [[CrossRef](#)]
31. Zeng, W.; Deng, J.; Gao, Q. Pest control of rice leaf folder with reduced pesticide application using the P20 type plant protection UAV. *Trans. CSAE* **2021**, *37*, 53–59.
32. Ye, H.; Huang, W.; Huang, S.; Cui, B.; Dong, Y.; Guo, A.; Ren, Y.; Jin, Y. Recognition of banana fusarium wilt based on UAV remote sensing. *Remote Sens.* **2020**, *12*, 938. [[CrossRef](#)]
33. Chen, T.; Yang, W.; Zhang, H.; Zhu, B.; Zeng, R.; Wang, X.; Wang, S.; Wang, L.; Qi, H.; Lan, Y.; Zhang, L. Early detection of bacterial wilt in peanut plants through leaf-level hyperspectral and unmanned aerial vehicle data. *Comput. Electron. Agric.* **2020**, *177*, 105708. [[CrossRef](#)]
34. Bhandari, M.; Ibrahim, A.M.; Xue, Q.; Jung, J.; Chang, A.; Rudd, J.C.; Maeda, M.; Rajan, N.; Neely, H.; Landivar, J. Assessing winter wheat foliage disease severity using aerial imagery acquired from small Unmanned Aerial Vehicle. *Comput. Electron. Agric.* **2020**, *176*, 105665. [[CrossRef](#)]
35. Musci, M.A.; Persello, C.; Lingua, A.M. UAV images and deep-learning algorithms for detecting flavescence doree disease in grapevine orchards. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2020**, *XLIII-B3-2020*, 1483–1489.
36. Wu, G.; Fang, Y.; Jiang, Q.; Cui, M.; Li, N.; Ou, Y.; Diao, Z.; Zhang, B. Early identification of strawberry leaves disease utilizing hyperspectral imaging combing with spectral features, multiple vegetation indices and textural features. *Comput. Electron. Agric.* **2023**, *204*, 107553. [[CrossRef](#)]
37. Huang, S.; Tang, L.; Hupy, J.P.; Wang, Y.; Shao, G. A commentary review on the use of normalized difference vegetation index (NDVI) in the era of popular remote sensing. *J. For. Res.* **2021**, *32*, 1–6. [[CrossRef](#)]
38. Guo, Z.; Xu, H.; Ma, J.; Ning, H.; Shen, J.; Zhang, Z. Construction of three-dimensional remote sensing ecological index (TRSEI) based on stereopair images: A case study of Miaodao Archipelago in China. *Ecol. Indic.* **2024**, *159*, 111737. [[CrossRef](#)]
39. Cardoso, L.A.S.; Farias, P.R.S.; Soares, J.A.C.; Caldeira, C.R.T.; de Oliveira, F.J. Use of a UAV for statistical-spectral analysis of vegetation indices in sugarcane plants in the Eastern Amazon. *Int. J. Environ. Sci. Technol.* **2024**, *21*, 6947–6964. [[CrossRef](#)]
40. Zhao, X.; Qi, J.; Xu, H.; Yu, Z.; Yuan, L.; Chen, Y.; Huang, H. Evaluating the potential of airborne hyperspectral LiDAR for assessing forest insects and diseases with 3D Radiative Transfer Modeling. *Remote Sens. Environ.* **2023**, *297*, 113759. [[CrossRef](#)]
41. Lanucara, S.; Praticò, S.; Pioggia, G.; Di Fazio, S.; Modica, G. Web-based spatial decision support system for precision agriculture: A tool for delineating dynamic management unit zones (MUZs). *Smart Agric. Technol.* **2024**, *8*, 100444. [[CrossRef](#)]
42. Fu, X.; Bao, Y.; Tubuxin, B.; Bao, Y. Bias Correction of Sentinel-2 MSI Vegetation Indices in a Desert Steppe with Original Assembled Field Online Multi-Angle Spectrometers. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 1–11.
43. Han, D.; Cai, H.; Zhang, L.; Wen, Y. Multi-sensor high spatial resolution leaf area index estimation by combining surface reflectance with vegetation indices for highly heterogeneous regions: A case study of the Chishui River Basin in southwest China. *Ecol. Inform.* **2024**, *80*, 102489. [[CrossRef](#)]
44. Sun, H. Crop vegetation indices. In *Encyclopedia of Smart Agriculture Technologies*; Springer International Publishing: Cham, Switzerland, 2023; pp. 1–7.

45. Marcello, J.; Eugenio, F.; Rodriguez-Esparragón, D.; Marqués, F. Assessment of forest degradation using multitemporal and multisensor very high resolution satellite imagery. In Proceedings of the International Geoscience and Remote Sensing Symposium, Pasadena, CA, USA, 16–21 July 2023; pp. 3233–3236.
46. Zhao, R.; Tang, W.; An, L.; Qiao, L.; Wang, N.; Sun, H.; Li, M.; Liu, G.; Liu, Y. Solar-induced chlorophyll fluorescence extraction based on heterogeneous light distribution for improving in-situ chlorophyll content estimation. *Comput. Electron. Agric.* **2023**, *215*, 108405. [[CrossRef](#)]
47. KADakci KOCA, T. A statistical approach to site-specific thresholding for burn severity maps using bi-temporal Landsat-8 images. *Earth Sci. Inform.* **2023**, *16*, 1313–1327. [[CrossRef](#)]
48. Fu, Z.; Zhang, J.; Jiang, J.; Zhang, Z.; Cao, Q.; Tian, Y.; Zhu, Y.; Cao, W.; Liu, X. Using the time series nitrogen diagnosis curve for precise nitrogen management in wheat and rice. *Field Crops Res.* **2024**, *307*, 109259. [[CrossRef](#)]
49. Gao, S.; Zhong, R.; Yan, K.; Ma, X.; Chen, X.; Pu, J.; Gao, S.; Qi, J.; Yin, G. Evaluating the saturation effect of vegetation indices in forests using 3D radiative transfer simulations and satellite observations. *Remote Sens. Environ.* **2023**, *295*, 113665.
50. Huang, X.; Lin, D.; Mao, X.; Zhao, Y. Multi-source data fusion for estimating maize leaf area index over the whole growing season under different mulching and irrigation conditions. *Field Crops Res.* **2023**, *303*, 109111. [[CrossRef](#)]
51. Sun, X.; Zhou, Y.; Jia, S.; Shao, H.; Liu, M.; Tao, S.; Dai, X. Impacts of mining on vegetation phenology and sensitivity assessment of spectral vegetation indices to mining activities in arid/semi-arid areas. *J. Environ. Manag.* **2024**, *356*, 120678. [[CrossRef](#)] [[PubMed](#)]
52. Hu, Y.; Han, C.; Li, W.; Hu, Q.; Wu, H. Experimental evaluation of SOFC fuel adaptability and power generation performance based on MSR. *Fuel Process. Technol.* **2023**, *250*, 107919. [[CrossRef](#)]
53. Jemaa, H.; Bouachir, W.; Leblom, B.; LaRocque, A.; Haddadi, A.; Bouguila, N. UAV-based computer vision system for orchard apple tree detection and health assessment. *Remote Sens.* **2023**, *15*, 3558. [[CrossRef](#)]
54. Trubin, A.; Kozhoridze, G.; Zabihi, K.; Modlinger, R.; Singh, V.V.; Surový, P.; Jakuš, R. Detection of green attack and bark beetle susceptibility in Norway Spruce: Utilizing PlanetScope Multispectral Imagery for Tri-Stage spectral separability analysis. *For. Ecol. Manag.* **2024**, *560*, 121838. [[CrossRef](#)]
55. Kesselring, J.; Morsdorf, F.; Kükenbrink, D.; Gastellu-Etchegorry, J.P.; Damm, A. Diversity of 3D APAR and LAI dynamics in broadleaf and coniferous forests: Implications for the interpretation of remote sensing-based products. *Remote Sens. Environ.* **2024**, *306*, 114116. [[CrossRef](#)]
56. Spolaör, N.; Cherman, E.A.; Monard, M.C.; Lee, H.D. ReliefF for multi-label feature selection. In Proceedings of the Brazilian Conference on Intelligent Systems, Fortaleza, Brazil, 19–24 October 2013; pp. 6–11.
57. Kononenko, I.; Šikonja, M.R. Non-myopic feature quality evaluation with (R) ReliefF. In *Computational Methods of Feature Selection*; Chapman and Hall/CRC: Boca Raton, FL, USA, 2007; pp. 185–208.
58. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
59. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. CBAM: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision, Munich, Germany, 8–14 September 2018; pp. 3–19.
60. Wang, F.; Jiang, M.; Qian, C.; Yang, S.; Li, C.; Zhang, H.; Wang, X.; Tang, X. Residual attention network for image classification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 3156–3164.
61. Brock, A.; De, S.; Smith, S.L.; Simonyan, K. High-performance large-scale image recognition without normalization. In Proceedings of the International Conference on Machine Learning, Virtual, 21–24 July 2021; PMLR: London, UK, 2021; pp. 1059–1071.
62. Liu, S.; Zhang, L.; Yang, X.; Su, H.; Zhu, J. Query2label: A simple transformer way to multi-label classification. *arXiv* **2021**, arXiv:2107.10834.

**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.