*Article*

# Monitoring the Severity of *Pantana phyllostachysae* Chao Infestation in Moso Bamboo Forests Based on UAV Multi-Spectral Remote Sensing Feature Selection

Zhanghua Xu [1,2,3,4,5,*], Qi Zhang [1,6], Songyang Xiang [1,6], Yifan Li [1,2], Xuying Huang [1,7], Yiwei Zhang [1,2], Xin Zhou [1,2], Zenglu Li [5,8], Xiong Yao [5,9], Qiaosi Li [10] and Xiaoyu Guo [5]

1   Academy of Geography and Ecological Environment, Fuzhou University, Fuzhou 350108, China; n195527042@fzu.edu.cn (Q.Z.); 205527048@fzu.edu.cn (S.X.); 200620005@fzu.edu.cn (Y.L.); dg1827013@smail.nju.edu.cn (X.H.); n190627069@fzu.edu.cn (Y.Z.); n190620019@fzu.edu.cn (X.Z.)
2   College of Environment and Safety Engineering, Fuzhou University, Fuzhou 350108, China
3   Key Laboratory of Spatial Data Mining & Information Sharing, Ministry of Education, Fuzhou 350108, China
4   Postdoctoral Research Station of Information and Communication Engineering, Fuzhou University, Fuzhou 350108, China
5   Fujian Provincial Key Laboratory of Resources and Environment Monitoring & Sustainable Management and Utilisation, Sanming 365004, China; sukd200720@segi.edu.my (Z.L.); fjyx@fjut.edu.cn (X.Y.); 20161138@fjsmu.edu.cn (X.G.)
6   The Academy of Digital China, Fuzhou University, Fuzhou 350108, China
7   International Institute for Earth System Science, Nanjing University, Nanjing 210023, China
8   Faculty of Education, SEGi University, Damansara 47810, Malaysia
9   College of Architecture and Planning, Fujian University of Technology, Fuzhou 350118, China
10   Department of Earth Sciences, The University of Hong Kong, Hong Kong 999077, China; qiaosili@hku.hk
*   Correspondence: fzucar@fzu.edu.cn

**Abstract:** In recent years, the rapid development of unmanned aerial vehicle (UAV) remote sensing technology has provided a new means to efficiently monitor forest resources and effectively prevent and control pests and diseases. This study aims to develop a detection model to study the damage caused to Moso bamboo forests by *Pantana phyllostachysae* Chao (PPC), a major leaf-eating pest, at 5 cm resolution. Damage sensitive features were extracted from multispectral images acquired by UAVs and used to train detection models based on support vector machines (SVM), random forests (RF), and extreme gradient boosting tree (XGBoost) machine learning algorithms. The overall detection accuracy (OA) and Kappa coefficient of SVM, RF, and XGBoost were 81.95%, 0.733, 85.71%, 0.805, and 86.47%, 0.811, respectively. Meanwhile, the detection accuracies of SVM, RF, and XGBoost were 78.26%, 76.19%, and 80.95% for healthy, 75.00%, 83.87%, and 79.17% for mild damage, 83.33%, 86.49%, and 85.00% for moderate damage, and 82.5%, 90.91%, and 93.75% for severe damage Moso bamboo, respectively. Overall, XGBoost exhibited the best detection performance, followed by RF and SVM. Thus, the study findings provide a technical reference for the regional monitoring and control of PPC in Moso bamboo.

**Keywords:** UAV multispectral remote sensing; Moso bamboo forest; *Pantana phyllostachysae* Chao; feature selection; detection model

## 1. Introduction

Moso bamboo is the largest and most widely distributed bamboo species in China with a high economic value. It has various ecological functions, including conserving water, maintaining soil and water quality, as well as the balance between carbon and oxygen. However, serious forest pests and diseases have been threatening the ecological health of the Moso bamboo forest; chief among them is *Pantana phyllostachysae* Chao (PPC), which can cause waterlogging inside the bamboo nodes or even kill the plants when the infestation is severe, thus, seriously restricting the healthy development of bamboo plants. Traditional

field survey methods are generally used to control the spread of pests by determining the spatial location of pest occurrence and the degree of damage [1]. However, the emergence of pests is affected by a variety of factors, making it impossible to obtain comprehensive and accurate pest information using traditional methods [2]. Therefore, to overcome these limitations, it is necessary to directly develop a fast and accurate method to detect damage caused by PPC.

At present, remote sensing technology is widely used for forest pest detection and has greatly reduced the labour and workload requirements for exploration of pests [3,4]. The "ground" to "space" and microscopic to macroscopic paradigm has been applied to the remote sensing monitoring of forest pests and diseases [5,6]. Using these models, progress has been made in understanding the response mechanism of these plants to the damage caused by PPC. In fact, changes in leaf loss, chlorophyll content, water level, and spectral reflectance of Moso bamboo leaves, in response to the damage caused by PPC, has been demonstrated both at the leaf level and the remote sensing image scale, facilitating the analysis of the remote sensing response mechanism of these forests to the damage caused by PPC [7–9]. In regions with large spatial heterogeneity, the direct use of single-point measurements or multi-point sampling, averaged to represent the value at the image pixel scale, is common; however, this method can cause large uncertainties. In general, it is difficult to correlate the results of studies based on specific scales to those at other scales. Indeed, blindly applying the data obtained in leaf scale studies to low-resolution satellite remote sensing may result in significant errors. Therefore, it is difficult to draw the objective conclusions based on scale transformation from point to surface [10].

Currently, unmanned aerial vehicle (UAV) remote sensing technology is increasingly being used in forest pest monitoring research as it can efficiently obtain high spatial and temporal resolution remote sensing images with outstanding structure and texture information. As such, this method better compensates for the drawbacks of satellite-based remote sensing and provides a real-time and accurate "ground-aerial-space" integrated platform for pest monitoring. Numerous studies have successfully detected forest pests and diseases using indicators such as original wavelengths, vegetation indexes, and texture characteristics [11–25]. However, pests and diseases can affect the structure and spectral characteristics of the host canopy [26]. It is therefore essential to fully exploit remote sensing features that are sensitive to pest response, and select effective classification models when monitoring pests in forests. Using Pearson correlation analysis and stepwise discriminant analysis methods, Liu et al. [27] selected features sensitive to pine forest insect damage and developed a diagnostic model for assessing the damage level in pine forests using multiple linear regression (MLR). Meanwhile, Iordache et al. [28] combined Pearson correlation analysis with intra-class and inter-class distance methods to filter pest indicators and established an identification model based on radio frequency for pine wood nematode infestation. Moreover, five yellow pest identification models were developed for betel leaf based on vegetation indices extracted from high-resolution UAV multispectral images, of which those based on back propagation neural network and support vector machine (SVM) algorithms exhibited superior detection [29]. Deng et al. [30] examined four models for detecting diseased citrus plants based on different feature combinations and reported that the performance of the model based on the XGBoost algorithm was superior to the other models. Furthermore, although existing studies have investigated the damage response mechanism of PPC at ground and satellite remote sensing scales, none have assessed its damage mechanism at the UAV scale.

In view of this, in the current study, UAV multispectral images were used as the data source to obtain the original spectra of Moso bamboo canopy at different damage levels, analyse the spectral variations, and select the parameters sensitive to the damage response of Moso bamboo forest to PPC infestation. Thereafter, the optimal feature subsets were screened by the RF-recursive feature elimination (RFE) algorithm and SVM, RF and XGBoost detection models were established. Finally, the detection effect of each model was

evaluated. This study aimed to provide a reference for the application of UAV multispectral data for Moso bamboo forest pest detection.

## 2. Materials and Methods

### 2.1. Experimental Data and Pre-Processing

The test area was located in Shunchang County, Nanping City, Fujian Province. The geographical coordinates of the county are 117°30′–118°14′ E and 26°39′–27°12′ N. The landform is mainly mountainous and hilly, with a mild climate, and a clear distinction between wet and dry seasons. Shunchang County was selected as the first batch of "the hometown of bamboo" in China and has completed the country's first bamboo forest carbon sink transaction. By the end of 2018, the county had a forest area of more than 160,000 hectares, including approximately 440,00 hectares of bamboo. According to statistics, all forest types in Shunchang County suffer from pest infestation throughout the year. As one of the main insect pests of Moso bamboo, PPC seriously restricts the sustainable development of the county's Moso bamboo economy.

On 14 May 2021, the research team travelled to Dagan Town in Shunchang County to carry out field research. A typical Moso bamboo plantation in Shakeng Village with an area of approximately 21 hectares was selected as the test area. The DJ-Innovations (DJI) Elf Phantom 4 Real-Time Kinematic (RTK) platform with Complementary Metal-Oxide Semiconductor (CMOS) multispectral sensor was used to collect aerial imagery. The device has one visible and five multispectral light sensors, which can detect five wavebands, including blue, green, red, red-edge, and near-infrared light (Table 1). Before the flight, the grey plate matching the sensor was used for correction. To obtain the UAV data, the flight height, course overlap degree, and side overlap degree were set to 93 m, 80%, and 70%, respectively. The composition of the tree species and the level of damage in the test area were then investigated at the ground level. In the test area, a total of 441 Moso bamboo canopy positions were measured using Global Navigation Satellite System (GNSS) receivers with centimetre-level positioning accuracy. Under pest damage stress, the external characteristics and internal physiological state of plant leaves become altered. The colour of the affected leaves becomes yellow or even scorched and has diseased spots or nicks. The water loss and photosynthetic capacity of the leaves become reduced following pest infestation [7,8]. As such, the process of determining pest level in a forest is highly complicated as various factors must be considered, including the colour of canopy leaves, disease spots, leaf integrity, etc. [5,9,31]. Hence, "The General Principles of Investigates on Main Forestry Pest" (LY/T 2011–2012), which considers these multiple variables, is used as a reference to determine the level of damage to the Moso bamboo canopy. Canopy photos were taken on site for a further review with relevant experts to determine the damage level. The damage levels of the Moso bamboo canopy were classified by leaf damage percentage: healthy was assigned 0%, mild damage was 0–20%, moderate damage was 20–50%, and severe damage was >50%. Among the 441 Moso bamboo canopy samples collected, 16.33% were classified as healthy, 22.45% as mild damage, 28.57% as moderate damage, and 32.65% as severe damage. These data were used as the foundation for the subsequent labelling and extraction of image samples and spectral texture features.

**Table 1.** Band parameters of multispectral sensor.

| Bands | Wavelength Range/nm | Centre Wavelength/nm | Band Width/nm |
|---|---|---|---|
| Blue band (B) | 434–466 | 450 | 32 |
| Green band (G) | 544–576 | 560 | 32 |
| Red band (R) | 634–666 | 650 | 32 |
| Red-edge band (RE) | 714–746 | 730 | 32 |
| Near-infrared band (NIR) | 814–866 | 840 | 52 |

The DJI Terra 3.0.1 software was used to pre-process the UAV images for radiometric calibration, image stitching, and orthorectification. Finally, the standard orthophoto product with a spatial resolution of approximately 5 cm was obtained with UTM/WGS84 projection coordinates. As the UAV images have a high resolution and strong ability to recognise background information, the pre-processed images (Figure 1c) also included those of non-bamboo forest areas, such as bare ground and shadows. To exclude these areas, thresholds were applied to the normalised difference vegetation index (NDVI) and red-edge (RE) bands. The test was repeated several times using the stepwise method, and the best extraction results were obtained when NDVI < 0.4845 and RE < 0.0574. In addition, to remove broadleaf forests in the test area individual image bands, vegetation indexes and texture features were used in combination with logistic regression models (LR) to extract moso bamboo forests, where the vegetation indices and texture quantities used are ratio vegetation index (RVI), normalised vegetation index (NDVI), difference vegetation index (DVI), normalised differential greenness index (NDGI), and normalised red-edge vegetation index (NDRE), as well as mean, variance, homogeneity, contrast, dissimilarity, entropy, entropy, second moment, and correlation. The overall accuracy (OA) of LR was 97.34%, and the results are shown in Figure 1d.
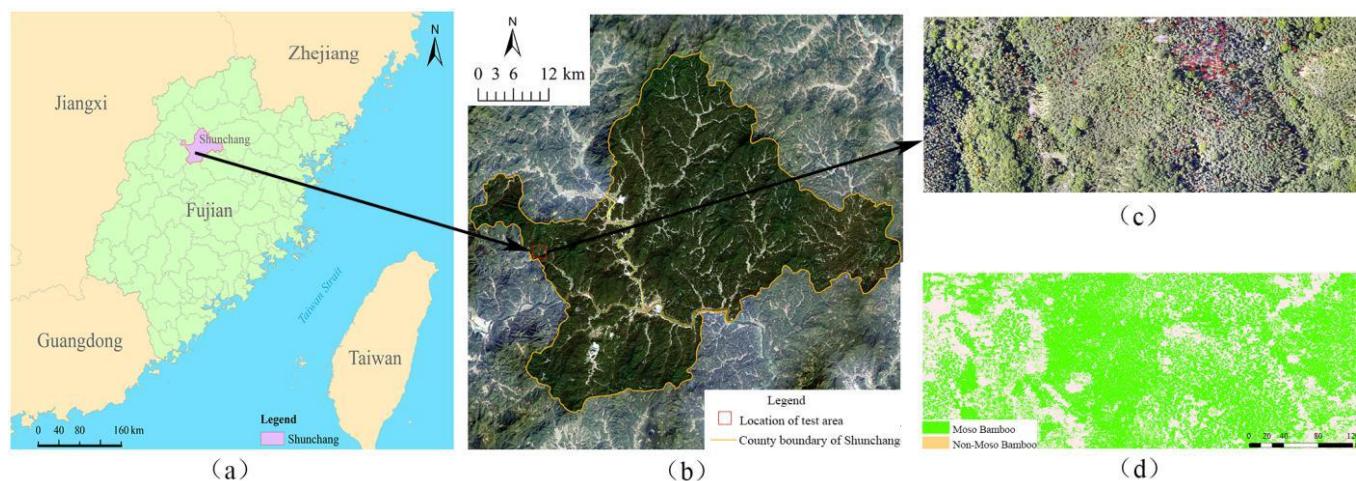


**Figure 1.** Overview of the test area and data processing: (**a**) geographic location of Shunchang County; (**b**) aerial image of Shunchang County; (**c**) unmanned aerial vehicle (UAV) orthophoto image of the test area; (**d**) image obtained after extracting the thematic information of Moso bamboo forest from UAV images.

### 2.2. UAV Remote Sensing Feature Selection

This study mined the remote sensing response characteristics of the bamboo forest at the canopy scale, on the basis of high-resolution UAV multispectral image data, to assess the damage caused by PPC. The pre-selected indicators included the individual image bands and 18 vegetation indexes, namely, RVI, renormalised difference vegetation index (RDVI), transformed normalised difference vegetation index (TNDVI), enhanced vegetation index 2 (EVI2), enhanced chlorophyll absorption rate index (MCARI), green normalised vegetation index (GNDVI), modified ratio vegetation index (MSR), optimisation soil-adjusted vegetation index (OSAVI), normalised vegetation index (NDVI), soil-adjusted vegetation index (SAVI), modified soil-adjusted vegetation index (MSAVI), green band chlorophyll index (RedGreen), enhanced red-edge ratio vegetation index (MSR$_{RE}$), normalised red-edge vegetation index (NDRE), leaf chlorophyll index (LCI), transformed chlorophyll uptake rate index (TCARI), RE band chlorophyll index (CIrededge) and feature spectral index of CSI. Additionally, eight texture features were included, namely, mean, variance, homogeneity, contrast, dissimilarity, entropy, second moment, correlation.

Not all of the above-mentioned feature indicators contribute to detection of the damage level caused by PPC; therefore, feature selection is needed to improve the classification

and detection accuracy of the model. One of the most commonly used methods, employed in previous feature optimisation studies, is recursive feature elimination based on SVM (SVM-RFE) [32,33]. Meanwhile, an increasing number of studies have combined random forest algorithms with RFE for feature optimisation [34–36]. Meanwhile, related studies have reported that RF-RFE is a mature feature selection method that offers superior results to SVM-RFE [37]. The RF-RFE selects features primarily by the results obtained from training the classifier. The basic process is as follows: (1) RF is used to rank the importance of features; (2) the features with low importance are removed before each classification, and iteration is stopped when the feature set is empty [38]. Since the random decision tree generation process uses the self-sampling method (Bootstrap), not all samples are used in the generation process of each tree—the unused samples are referred to as out_of_bag (OOB)—through which the accuracy of the tree can be evaluated. RF focuses on OOB estimation of the prediction error on a "random" dataset by assigning a random number to each feature in turn; each feature is then assigned a score accordingly; the larger the score, the more important the feature. RF uses OOB estimation to calculate the relevance of the features of the model, and this measure of feature importance makes the complexity of the computation unincreased. In addition, RF accounts for the influence of each feature on the classification. To further filter the number of feature parameters significant to the severity of the damage, the RF-RFE algorithm, based on cross-validation, is used to cross-validate different feature combinations on the basis of RF-RFE to determine the final optimal feature subset. In this study, the construction and optimisation of the detection model for the damage level caused by PPC is carried out on the basis of simplifying the features by RF importance ranking and RFE backward iteration.

### 2.3. Construction and Optimisation of Pest Detection Models

In this study, three models, SVM, RF, and XGBoost, were selected to detect and compare the damage levels of PPC; in each model, the feature subset screened by RF-RFE was considered as the independent variables and the damage level of PPC was the dependent variable.

#### 2.3.1. Support Vector Machine (SVM)

The core idea of SVM is to map the data to a high-dimensional space to find the optimal classification hyperplane by minimising the upper limit of the classification error. For linearly separable data, SVM focuses on identifying the separating hyperplane $\omega^T x + b$ [39]. Its objective function is given by Equation (1):

$$argmax_{\omega,b}\left\{ min_n\left( label \cdot \left(\omega^T x + b\right)\right) \cdot \frac{1}{\|\omega\|}\right\} = \pi r^2 \tag{1}$$

where $\omega$ is the normal vector which determines the direction of the hyperplane, *label* is the category label, and $label \bullet \left(\omega^T x + b\right) \geq 0$ is the constraint. By introducing Lagrange multipliers, Equation (1) can be transformed into Equation (2):

$$\begin{cases} max_\alpha \left( \sum\limits_{i=1}^{m} \alpha - \frac{1}{2} \sum\limits_{i,j-1}^{m} label^{(i)} \cdot label^{(i)} \cdot \alpha_i \alpha_j <x^{(i)}, x^{(j)}> \right) \\ C \geq \alpha \geq 0, \sum\limits_{i=1}^{m} \alpha_i label^{(i)} = 0 \end{cases} \tag{2}$$

where $\alpha$ is the displacement that determines the distance of the hyperplane from the origin and $C$ is the relaxation variable. The separation hyperplane can be obtained by solving $\alpha$ and $\omega$.

The SVM algorithm deals with nonlinear problems by setting a suitable kernel function which essentially calculates the similarity between the samples and landmark points to define new features and thus train complex nonlinear decision boundaries. SVM has four different kernel functions: linear, polynomial, hyperbolic tangent (sigmoid), and

Gaussian radial basis function (RBF). As different kernel functions are used to identify the hyperplanes under different data distributions, the settings of the kernel functions, and their corresponding penalty coefficients C and gamma parameters, affect the accuracy of model classification.

### 2.3.2. Random Forest (RF)

RF is a machine learning algorithm that contains multiple decision trees [40]. The basic concept is to use the self-help sampling method (bootstrap) to randomly select *k* samples from the original training set which are then added back as a new training set. The classification decision trees are then constructed to generate a random forest composed of *k* classification trees, and finally the attribution of new samples is decided by majority voting based on the results of each classification tree. The essence of the algorithm is to improve the decision trees by combining multiple decision trees together, with the creation of each tree depending on an independently drawn sample set.

### 2.3.3. Extreme Gradient Boosting Tree (XGBoost)

XGBoost is a boosted tree-based machine learning algorithm used for studying gradient-boosted decision tree algorithms [41]. The premise of this algorithm is to perform a second-order Taylor expansion of the loss function, add a regularization term to the objective function to control the complexity of the model, and continuously perform feature splitting to develop a tree to fit the residuals of the last prediction during the training process. *K* trees are obtained after training, and according to the characteristics of this sample, there will be a corresponding prediction score for each tree; the sum of the corresponding scores of each tree is the prediction value of this sample.

XGBoost considers the regularization term, and its objective function is defined as follows:

$$obj = \sum_{i=1}^{m} l(y_i, \hat{y}_i) + \sum_{k=1}^{k} \Omega(f(k)) \tag{3}$$

$$\Omega(f(k)) = \gamma^T + \frac{1}{2}\lambda \sum_{j=1}^{T} \omega_j^2 \tag{4}$$

where $\sum_{i=1}^{m} l(y_i, \hat{y}_i)$ represents the loss function, $\sum_{i=1}^{m} \sum_{k=1}^{k} \Omega(f(k))$ represents the regularization term, $\hat{y}_i$ is the predicted output, $y_i$ is the label value, $f(k)$ is the k-th tree model, $T$ represents the weight value, $\gamma$ is the leaf tree penalty regularization term with pruning effect, and $\lambda$ is the leaf weight penalty regularisation term to prevent overfitting.

### 2.4. Test Effect Evaluation

To ensure the reliability of the evaluation results, improve the stability and generalisation ability of the models, and attenuate the influence of sample data on the evaluation results, 441 sets of Moso bamboo canopy spectral sample data were divided into modelling and validation sets in the ratio 7:3, and 309 training samples were used for model construction while 132 validation samples were used for model evaluation. The OA and Kappa coefficient were used to evaluate the effectiveness of the three models for detecting the damage level of PPC. Data analysis in this work was conducted with Python (anaconda 4.9.2), jupyterlab 2.2.6, and machine learning library scikit-learn 0.23.2, under the Windows 10 operating system.

## 3. Results

### 3.1. UAV Multispectral Characterisation of PPC Damage in Moso Bamboo Forests

Changes in the internal structure and external morphological characteristics of Moso bamboo caused by PPC will present as changes in the spectral reflectance of the affected Moso bamboo at both the visible and near-infrared wavelengths. Hence, the reflectance of extracted individual image bands was normalised, and the spectral variation of the Moso

bamboo canopy under different damage levels was analysed based on the normalised individual image bands, as shown in Figure 2, where the horizontal coordinates 1, 2, 3, and 4 represent healthy, mild damage, moderate damage, and severe damage, respectively. No obvious change was observed in B-band reflectance under different severities, and the R-band reflectance was higher for moderate and severe damage than for healthy and mild damage. Moreover, the reflectance of R-band under severe damage was particularly visible, likely due to the image resolution being higher and the canopy of severely damaged, or even dead, Moso bamboo was sparser; therefore, the reflectance was influenced by the background such as branches and soil, thus not showing a proper monotonic trend. Unlike the B and R bands, the reflectance of the G, RE, and NIR bands tended to decrease as the damage level increased. As the host nutrient deficiency caused by PPC tends to obliviate the "green peak" and "red valley" of the spectrum, the reflectance of G, RE and NIR bands tends to decrease with an increasing damage level.
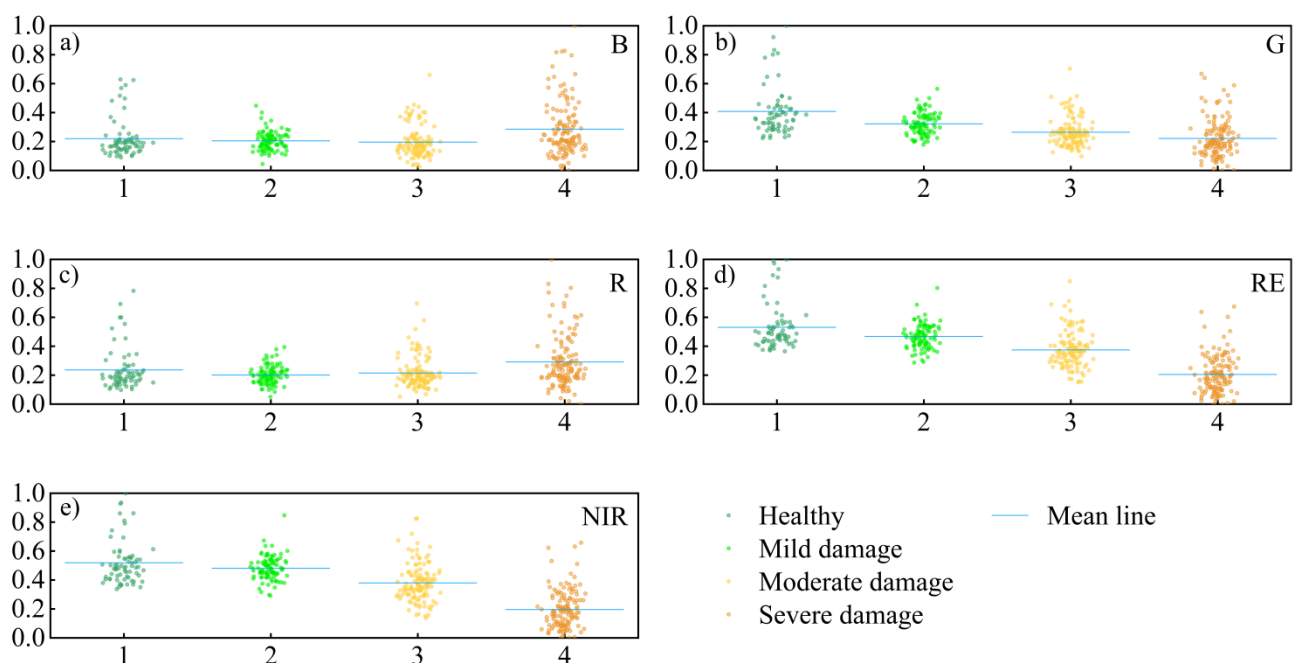


**Figure 2.** Variation characteristics of each index according to damage severity. (**a**) B, (**b**) G, (**c**) R, (**d**) RE and (**e**) NIR.

### 3.2. Feature Optimisation and Analysis Based on RF-RFE

To determine the optimal subset of features for the model, RF-RFE was used to screen 31 features (Figure 3). From the figure, it can be seen that classification accuracy reaches a maximum at ten features and tends to decline as more features are added. Therefore, the first ten features were selected as the optimal subset. The importance of the features in the optimal feature subset in descending order of importance are RedGreen, CSI, NDVI, MSR, TNDVI, RVI, correlation, MCARI, GNDVI, and CIrededge (Figure 4).
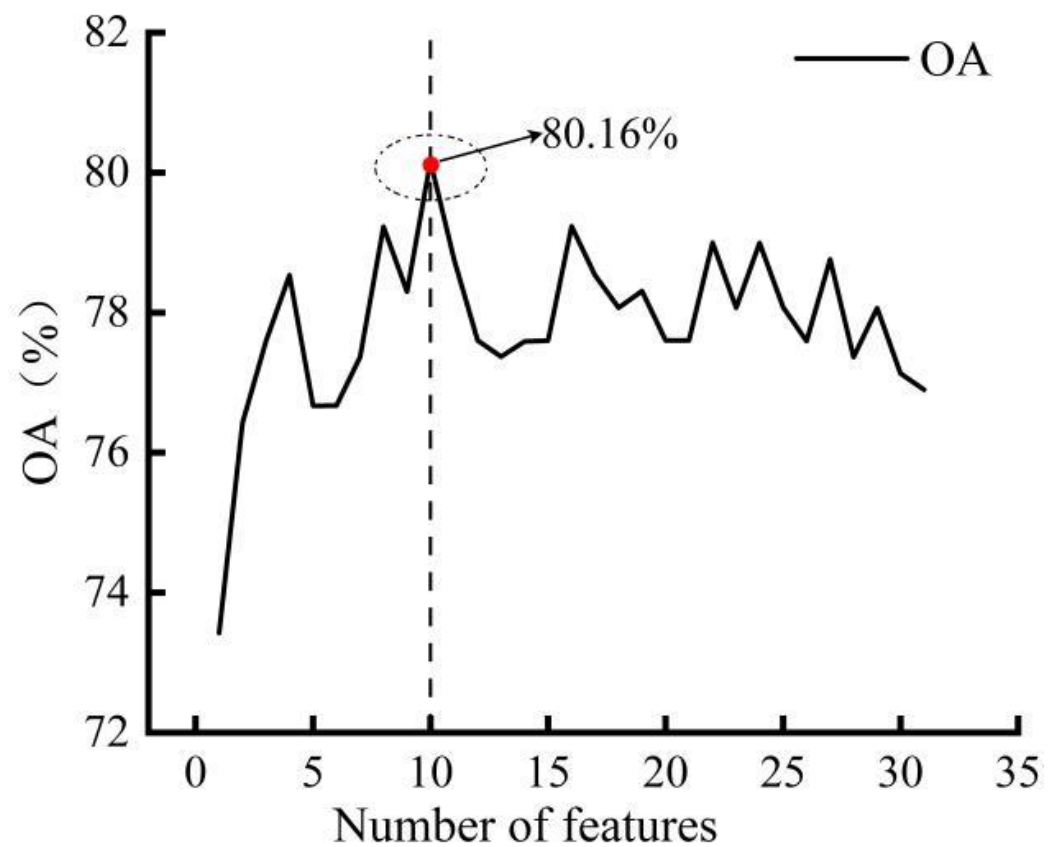
**Figure 3.** RF-RFE feature selection. The red dot inside the circle represents the overall accuracy (OA) corresponding to the optimal number of feature dimensions, i.e., 80.16%.
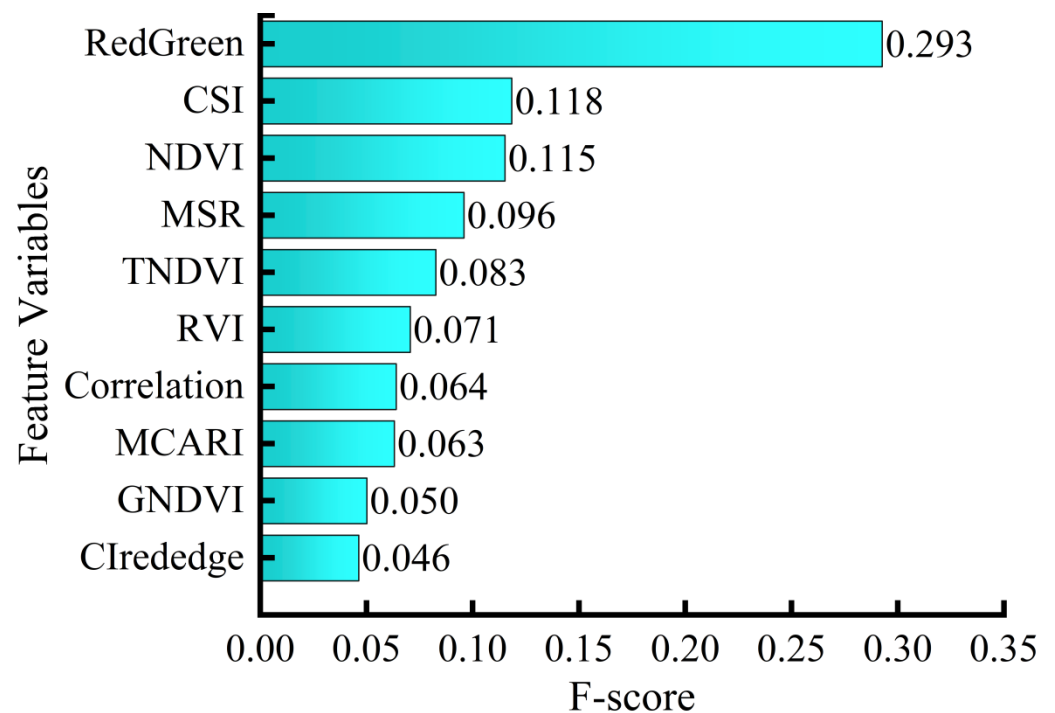


**Figure 4.** Importance scores of characteristic variables.

The values of each optimal feature were normalised and plotted as a scatter plot (Figure 5), where the horizontal coordinates 1, 2, 3, and 4 represent healthy, mild, moderate,

and severe damage, respectively. As seen in the figure, the vegetation indices under different severities exhibit some variation. As the damage level increased, RedGreen and CSI showed an increasing trend, while NDVI, MSR, TNDVI, RVI, correlation, MCARI, GNDVI, and CIrededge exhibited a decreasing trend. Correlation, which is a texture feature, showed a tendency to initially decrease, subsequently increase, and then decrease as the damage level increased. This indicates that the features selected by the RF-RFE feature selection algorithm showed a clear pattern of damage response, and large differences occurred among different damage levels of Moso bamboo canopies.
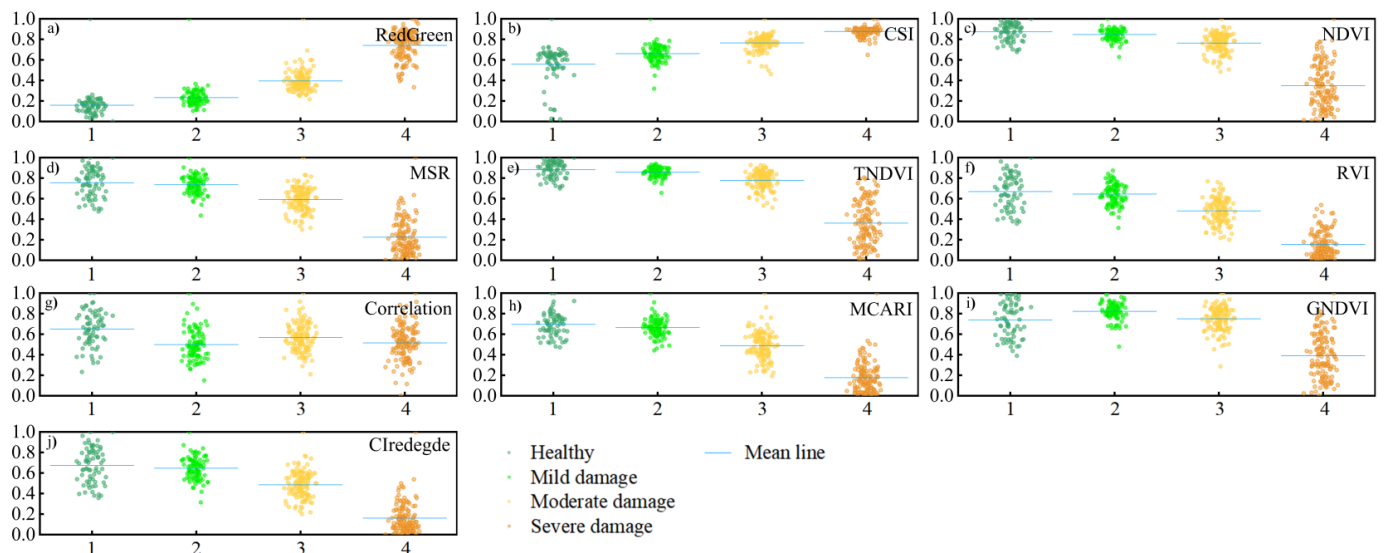


**Figure 5.** Variation characteristics of the selected features according to damage severity. (**a**) RedGreen, (**b**) CSI, (**c**) NDVI, (**d**) MSR, (**e**) TNDVI, (**f**) RVI, (**g**) Correlation, (**h**) MCARI, (**i**) GNDVI, (**J**) CIrededge.

### 3.3. Damage Detection Model and Effect Evaluation of PPC in Moso Bamboo Forest

3.3.1. Establishment of a Damage Detection Model for PPC Infestation in Moso Bamboo Forests

Based on the sample data of the experimental group, ten features, including RedGreen, CSI, NDVI, MSR, TNDVI, RVI, correlation, MCARI, GNDVI, and CIrededge, were used as independent variables, and healthy, mild damage, moderate damage, and severe damage were used as dependent variables to train a model for detecting the damage level of PPC using the SVM, RF, and XGBoost algorithms.

The learning curve and the grid search were used to tune the parameters and improve the accuracy of the models. The main tuning parameters of SVM are C, kernel, degree, and gamma. Here, C is the penalty factor or the tolerance for error; the higher the C value, the less likely it is to overfit; however, the computation will be slower. The gamma parameter comes with the RBF function as kernel; this parameter implicitly determines the distribution of the data after mapping to the new feature space. Degree is the dimensionality of the polynomial function. The main parameters regulated by random forest (RF) are n_estimators, random_state, and max_depth. The n_estimators are the number of trees, that is, the number of base evaluators; if n_estimators are too small, it will result in underfitting, and vice versa. The random_state is the seed used in any class or function with randomness to control the random pattern. The max_depth is the maximum depth of the tree which reflects the complexity of a single tree. The main parameters regulated by XGBoost are n_estimators, random_state, max_depth, gamma, and eta. n_estimators, random_state, and max_depth are the same as those in RF, while gamma is the minimum value of the loss function at which the node can be split and eta is the weight or learning rate of the model generated by each iteration.

To ensure the stability of the models, the training sets were subjected to five-fold cross-validation. The optimisation of the main parameters of each model and the accuracies of the five-fold cross-validation are listed in Table 2.

**Table 2.** Parameters of the SVM, RF, and XGBoost models.

| Classifier | Parameters | Five-Fold Cross-Validation Accuracy (%) |
|---|---|---|
| Support Vector Machine, SVM | kernel = 'poly', C = 1, gamma = 5, degree = 2 | 82.34 |
| Random Forests, RF | n_estimators = 50, random_state = 80, max_depth = 9 | 85.41 |
| Extreme Gradient Boosting, XGBoost | n_estimators = 137, random_state = 100, max_depth = 2, gamma = 0.1, eta = 0.05 | 86.63 |

3.3.2. Evaluation of the Detection Effect of PPC in Moso Bamboo Forest

The final classification results of each model after tuning and enhancing the parameters and effects, respectively, of the detection model using the learning curve and the grid search, are shown in Figure 6. The results show that the degree of Moso bamboo damage predictive ability is good for all three models. Among them, SVM exhibited slight confounding for the prediction of healthy and mild damage, while RF and XGBoost had similar prediction results. Furthermore, the prediction effect of XGBoost was slightly superior to that of RF for moderate and severe damage.
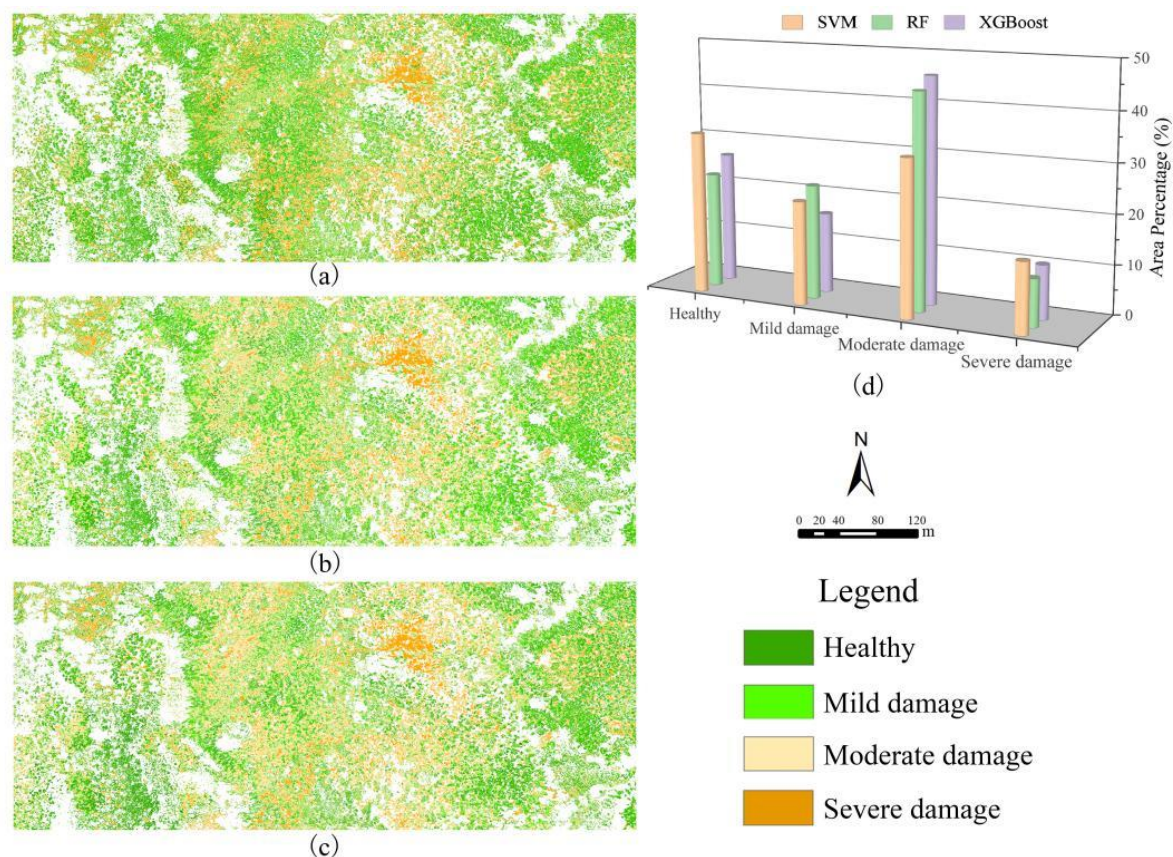


**Figure 6.** Prediction of the damage level of PPC based on three machine learning algorithms: (**a**) prediction results of SVM classification algorithm; (**b**) prediction results of RF classification algorithm; (**c**) prediction results of XGBoost classification algorithm; (**d**) the ratio of the area occupied by each damage level according to the three models.

Next, the sample data of the validation set were substituted into the three models and the OA and Kappa coefficient values for detecting the damage level of PPC were calculated (Table 3). The results show that the OA and Kappa coefficient values for all three models were above 80%, 0.700, respectively, and that the OA, Kappa coefficient value of the XGBoost algorithm were the highest at 82.54%, 0.811, respectively, compared with those of the SVM and RF models. This shows that the XGBoost pest detection model is more accurate than the SVM and RF models. The OA of the XGBoost pest detection model was higher by 4.52% and 0.76%, and the Kappa coefficient was higher by 0.078 and 0.006, compared to those of the SVM and RF models, respectively.

**Table 3.** Evaluation of the detection performance for SVM, RF, and XGBoost models.

| Damage Levels | SVM | | RF | | XGBoost | |
|---|---|---|---|---|---|---|
| | OA (%) | Kappa Coefficient | OA (%) | Kappa Coefficient | OA (%) | Kappa Coefficient |
| Healthy | 78.26 | 0.876 | 76.19 | 0.722 | 80.95 | 0.776 |
| Mild damage | 75.00 | 0.565 | 83.87 | 0.788 | 79.17 | 0.746 |
| Moderate damage | 83.33 | 0.712 | 86.49 | 0.809 | 85.00 | 0.788 |
| Severe damage | 82.50 | 0.922 | 90.91 | 0.866 | 93.75 | 0.900 |
| Total | 81.95 | 0.733 | 85.71 | 0.805 | 86.47 | 0.811 |

The accuracy of the three models in detecting the degree of damage (healthy, mild, moderate, and severe damage) was then analysed. The prediction accuracy of SVM and XGBoost for mild damage was lower than that for the other degrees, at 75.00% and 79.17%, while the Kappa coefficients were 0.565 and 0.746, respectively. As for RF, the prediction accuracy and Kappa coefficient were lower for healthy plants compared to the other levels, at 76.19% and 0.722, respectively. The detection performance of the three models was better for moderate and severe damage, with an accuracy between 80–93.75% and a Kappa coefficient between 0.7–0.922. Overall, XGBoost had the best detection performance for the four degrees of damage, with the lowest detection accuracy being close to 80% and the lowest Kappa coefficient being 0.746.

Based on the results of the above analysis, it can be seen that although all three models can effectively identify the severity of damage in Moso bamboo forests, the performance of the XGBoost-based detection model was superior compared to that of the other two models, and hence, can be used to fully exploit the damage information of PPC.

## 4. Discussion

In recent years, studies have demonstrated the use of medium-resolution multispectral satellite images to monitor forest pest disturbances at the regional scale [42,43], as well as the applicability of satellite remote sensing imagery to monitor moth infestation in Congo bamboo, as well as other forest pests and diseases [9,44,45]. Although satellite remote sensing images have the advantages of large coverage and low cost, they are susceptible to a relatively long revisit period, and low spatial resolution, among other factors. Even the relatively short revisit period for Sentinel 2 data is one week, which is not conducive to timely detection of early infestation. Moreover, the low spatial resolution can result in the presence of mixed pixels, making it difficult to achieve real-time and accurate monitoring, which to some extent restricts their application for accurate monitoring of forest pest and disease stress. UAV remote sensing technology can efficiently and rapidly acquire remote sensing images with high spatial and temporal resolution; the types of data acquired are also abundant and includes visible, multispectral, and hyperspectral data, which can better solve the drawbacks of satellite remote sensing and can be used for pest monitoring at the canopy scale [46–49]. In our previous studies, four dimensions of pest response, including leaf loss, greenness, humidity, and original wavelength, were clarified at the ground level [7–9]. In this study, spectrum-derived indicators that are sensitive to the damage

caused by PPC were selected and screened based on UAV multispectral data. The results were consistent with those of the previous studies. In addition, the leaf feeding pattern of PPC is from the top of the Moso bamboo canopy downward, and therefore, canopy leaf loss becomes more pronounced as the infestation increases, leading to variability in the canopy structure according to the degree of damage. Moreover, the structural variation of the Moso bamboo canopy was not considered in the previous studies. In this study, texture features were also incorporated into the analysis and detection of the damage level to the Moso bamboo canopy at the canopy scale. From the index filtering results, it can be seen that the selected texture features does not account for a large proportion, which may be related to the window size set when extracting texture features, which will be further investigated in a future study.

Since numerous spectral indices and texture feature variables were used to analyse the models used in this study, feature redundancy was inevitable. Therefore, it is necessary to select a suitable method to screen for the initial selection of indicators. In this study, the RF algorithm combined with the RFE method was used for feature optimisation. To further screen the number of features significant to each damage level, the idea of cross-validation was also introduced into the RFE algorithm, and the RF-RFE algorithm based on cross-validation was used to assess different feature combinations on the basis of RF-REF and determine the final optimal feature subset. However, as a black box model, it is difficult to explain the relationship between each preferred characteristic indicator and the occurrence of the pest; therefore, a variance analysis of the selected indicators was performed to explain the sensitivity of each preferred indicator to the damage level of PPC. In this study, the process of "feature selection of RF-RFE followed by variance analysis" allows for the objective selection and interpretation of the best feature indicators and their usability, respectively.

Although all of the detection models proposed in this study performed well with respect to detecting the damage level of PPC, certain limitations remain. For instance, the accuracy of the SVM, RF, and XGBoost models in detecting the level of damage (healthy, mild damage, moderate damage, and severe damage) all varied to some extent. The detection accuracies of the three models for moderate and severe damage were correspondingly greater than those for healthy and mild damage, and the average detection accuracies were 78.47%, 79.35%, 84.94%, and 89.05%, for healthy, mild damage, moderate damage, and severe damage, respectively. It can be seen that the detection accuracy of each model for healthy and mild damage has certain shortcomings, of which the SVM model performed particularly well. In general, the damage process of a single Moso bamboo plant is as follows: first, the uppermost leaves are destroyed by PPC; after the top leaves are eaten, the pest will move downwards and harm the leaves in the middle and lower part of the canopy [50]. In this study, the spectrum of the top leaves of the canopy did not change significantly between the initial damage phase (mild damage) and healthy Moso bamboo, as evidenced by the results of the variability analysis of the selected features. This may lead to errors in differentiating between healthy and mild damage plants when judging and sampling, thereby affecting the detection accuracy of both, and resulting in missing the areas of early mild infestation. Pest infestation can increase the frequency and intensity of forest disturbances and therefore require effective methods for accurate monitoring and mapping of damage levels [51]. Traditional methods of ground surveys are not sufficient to determine the damage level caused by pests to the host. The proposed framework for detecting the damage level of PPC using data mining and multi-model classification comparison achieves the identification and evaluation of the damage level of PPC at the canopy scale with positive results. However, this process requires manual mapping of the ROI of the canopy, and the subsequent extraction of spectral texture values of the canopy, which may lead to a reduction in detection accuracy; moreover, this process cannot be used to map the extent of damage at the individual plant level. The framework can be further improved by using better artificial intelligence algorithms to automatically extract and identify the canopy and damage level of individual trees. In addition, owing to the

limitation of the spectral resolution of the UAV data itself, the extracted spectral-derived indicators cannot characterise dimensions such as water content or humidity, leading to inadequate mining of damage response features, which then affects the accuracy of damage class classification. High-resolution hyperspectral images can provide higher spectral resolution and more detailed texture features to identify subtle changes in the tree canopy, which can subsequently be used in Moso bamboo single-plant pest detection studies by combining Light Detection and Ranging (LiDAR) data with hyperspectral image data.

## 5. Conclusions

PPC is a major leaf-eating pest in bamboo forests. Studying the response mechanism of the pest at the UAV remote sensing scale is necessary for the detection of the damage caused by PPC in Moso bamboo forests as it can provide a reference for intelligent monitoring of forest resources and accurate prevention and control of pests. In this study, multispectral UAV images acquired by DJI Elf Phantom 4 RTK small multirotor multispectral image acquisition platform were used to extract and analyse the features sensitive to the damage response of Moso bamboo, and establish the damage level detection models based on SVM, RF, and XGBoost by combining machine learning algorithms. The results showed the following:

1. The spectra of G, RE, and NIR bands of the Moso bamboo canopy differed significantly according to the degree of damage, and their values showed a decreasing trend with the increase in damage class.
2. The ten features selected using the RF-RFE algorithm, including nine vegetation indices and one texture feature, were ranked in descending order of importance as RedGreen, CSI, NDVI, MSR, TNDVI, RVI, correlation, MCARI, GNDVI, and CIred-edge. Each of the selected features showed relatively clear pest response patterns, and large differences were observed between the different damage classes of Moso bamboo canopies. The selected texture feature was also shown to play an important role in the detection of damage classes at the UAV scale.
3. All three models were able to detect the damage level of PPC, and XGBoost showed the best detection performance; its OA and Kappa coefficient were 86.47%, 0.811, respectively. The RF model, with an OA and Kappa coefficient value of 85.71%, 0.805, respectively, was ranked second, and SVM, with an OA and Kappa coefficient of 81.95%, 0.733, respectively, was ranked third.

## References

1.  Hall, R.J.; Castilla, G.; White, J.C.; Cooke, B.J.; Skakun, R.S. Remote sensing of forest pest damage: A review and lessons learned from a Canadian perspective. *Can. Entomol.* **2016**, *148*, S296–S356. [CrossRef]
2.  Park, J.; Sim, W.; Lee, J. Detection of trees with pine wilt disease using object-based classification method. *J. For. Environ. Sci.* **2016**, *32*, 384–391. [CrossRef]
3.  Havašová, M.; Ferenčík, J.; Jakuš, R. Interactions between windthrow, bark beetles and forest management in the Tatra national parks. *For. Ecol. Manag.* **2017**, *391*, 349–361. [CrossRef]
4.  Royle, D.D.; Lathrop, R.G. Monitoring hemlock forest health in New Jersey using landsat TM data and change detection techniques. *For. Sci.* **1997**, *43*, 327–335.
5.  Xu, Z.H.; Huang, X.Y.; Lin, L.; Wang, Q.F.; Liu, J.; Chen, C.C.; Yu, K.Y.; Zhou, H.K.; Zhang, H.F. Dendrolimus Punctatus Walker Damage Detection Based on Fisher Discriminant Analysis and Random Forest. *Spectrosc. Spect. Anal.* **2018**, *38*, 2888–2896.
6.  Xi, G.L. *Early Hyperspectral Identification of Erannis Jacobsoni Djak Disaster*; Inner Mongolia Normal University: Hohhot, China, 2020.
7.  Huang, X.Y.; Xu, Z.H.; Lin, L.; Shi, W.C.; Yu, K.Y.; Liu, J.; Chen, C.C.; Zhou, H.K. *Pantana Phyllostachysae* Chao Damage Detection Based on Physical and Chemical Parameters of Moso Bamboo Leaves. *Spectrosc. Spect. Anal.* **2019**, *39*, 857–864.
8.  Huang, X.Y.; Xu, Z.H.; Wang, X.P.; Yang, X.; Ju, W.M.; Hu, X.Y.; Li, K.; Chen, Y.Z. Spectral Characteristics of Moso Bamboo Leaves Damaged by *Pantana Phyllostachysae* Chao and Monitoring of Pest Rating. *Spectrosc. Spect. Anal.* **2021**, *41*, 1253–1259.
9.  Huang, X.Y. *Remote Sensing Response Mechanism of Phyllostachys pubescena Forests Damaged by Pantana phyllostchysae Chao*; Fuzhou University: Fuzhou, China, 2018.
10. Hao, D.L.; Xiao, Q.; Wen, J.G.; You, D.Q.; Wu, X.D.; Lin, X.W.; Wu, S.B. Advances in upscaling methods of quantitative remote sensing. *Natl. Remote Sens. Bull.* **2018**, *22*, 408–423.
11. Jordan, C.F. Derivation of leaf area index from quality of light on the forest floor. *Ecology* **1969**, *50*, 663–666. [CrossRef]
12. Roujean, J.; Breon, F. Estimating PAR absorbed by vegetation from bidirectional reflectance measurements. *Remote Sens. Environ.* **1995**, *51*, 375–384. [CrossRef]
13. Fu, Y.Z.; Wang, X.Q.; Jiang, H. The Correlation between LAI and Vegetation Index of Masson Pine. *Remote Sens. Land Resour.* **2010**, *3*, 41–46.
14. Jiang, Z.Y.; Huete, A.R.; Didan, K.; Miura, T. Development of a Two-Band Enhanced Vegetation Index without a Blue Band. *Remote Sens. Environ.* **2008**, *112*, 3833–3845. [CrossRef]
15. Daughtry, C.S.T.; Walthall, C.L.; Kim, M.S.; Colstoun, E.B.; McMurtrey, J.E. Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance. *Remote Sens. Environ.* **2000**, *74*, 229–239. [CrossRef]
16. Abdulridha, J.; Ampatzidis, Y.; Roberts, P.; Kakarla, S.C. Detecting powdery mildew disease in squash at different stages using UAV-based hyperspectral imaging and artificial intelligence. *Biosyst. Eng.* **2020**, *197*, 135–148. [CrossRef]
17. Verrelst, J.; Schaepman, M.E.; Koetz, B.; Kneubuhler, M. Angular sensitivity analysis of vegetation indices derived from CHRIS/PROBA data. *Remote Sens. Environ.* **2008**, *112*, 2341–2353. [CrossRef]
18. Steven, M.D. The Sensitivity of the OSAVI Vegetation Index to Observational Parameters. *Remote Sens. Environ.* **1998**, *63*, 49–60. [CrossRef]
19. Tanaka, S.; Kawamura, K.; Maki, M.; Muramoto, Y.; Yoshida, K.; Akiyama, T. Spectral Index for Quantifying Leaf Area Index of Winter Wheat by Field Hyperspectral Measurements: A Case Study in Gifu Prefecture, Central Japan. *Remote Sens.* **2015**, *7*, 5329–5346. [CrossRef]
20. Huete, A.R. A soil-adjusted vegetation index (SAVI). *Remote Sens. Environ.* **1988**, *25*, 295–309. [CrossRef]
21. Qi, J.; Chehbouni, A.; Huete, A.R.; Kerr, Y.H.; Sorooshian, S. A Modified soil adjusted vegetation index. *Remote Sens. Environ.* **1994**, *48*, 119–126. [CrossRef]
22. Wu, C.Y.; Niu, Z.; Tang, Q.; Huang, W.J. Estimating chlorophyll content from hyperspectral vegetation indices: Modeling and validation. *Agric. For. Meteorol.* **2008**, *148*, 1230–1241. [CrossRef]
23. Raymond, H.E.; Daughtry, C.S.T.; Eitel, J.U.H.; Dan, S.L. Remote Sensing Leaf Chlorophyll Content Using a Visible Band Index. *Agron. J.* **2011**, *103*, 1090–1099.
24. Penuelas, J.; Frederic, B.; Filella, I. Semiempirical indexes to assess carotenoids chlorophyll A ratio from leaf spectral reflectance. *Photosynthetica* **1995**, *31*, 221–230.
25. Haboudane, D.; Miller, J.R.; Tremblay, N.; Zarco-Tejada, P.J.; Dextraze, L. Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture. *Remote Sens. Environ.* **2002**, *81*, 416–426. [CrossRef]
26. Lin, Q.N.; Huang, H.G.; Wang, J.X.; Huang, K.; Liu, Y.Y. Detection of Pine Shoot Beetle (PSB) Stress on Pine Forests at Individual Tree Level using UAV-Based Hyperspectral Imagery and Lidar. *Remote Sens.* **2019**, *11*, 2540–2560. [CrossRef]
27. Liu, M.Y.; Zhang, Z.H.; Liu, X.L.; Yao, J.; Du, T.; Ma, Y.Q.; Shi, L. Discriminant Analysis of the Damage Degree Caused by Pine Shoot Beetle to Yunnan Pine Using UAV-Based Hyperspectral Images. *Forests* **2020**, *11*, 1258–1280. [CrossRef]
28. Iordache, M.D.; Mantas, V.; Baltazar, E.; Pauly, K.; Lewyckyj, N. A Machine Learning Approach to Detecting Pine Wilt Disease Using Airborne Spectral Imagery. *Remote Sens.* **2020**, *12*, 2280–2303. [CrossRef]
29. Lei, S.H.; Luo, J.B.; Tao, X.J.; Qiu, Z.X. Remote Sensing Detecting of Yellow Leaf Disease of Arecanut Based on UAV Multisource Sensors. *Remote Sens.* **2021**, *13*, 4562–4584. [CrossRef]

30. Deng, X.L.; Zeng, G.L.; Zhu, Z.H.; Huang, Z.X.; Yang, J.C.; Tong, Z.J.; Yin, X.B.; Wang, T.W.; Lan, Y.B. Classification and feature band extraction of diseased citrus plants based on UAV hyperspectral remote sensing. *J. South. China Agric. Univ.* **2020**, *41*, 100–108.

31. Zheng, B.J.; Chen, Y.Z.; Li, K.; Wang, X.Q.; Xu, Z.H.; Huang, X.Y.; Hu, X.Y. Detection of Pest Degree of Phyllostachys Chinese With Hyperspectral Data. *Spectrosc. Spect. Anal.* **2021**, *41*, 3200–3207.

32. MAO, Y.; Pi, D.Y.; Liu, Y.M.; Sun, Y.X. Accelerated Recursive Feature Elimination Based on Support Vector Machine for Key Variable Identification. *Chin. J. Chem. Eng.* **2006**, *14*, 65–72. [CrossRef]

33. Zhao, J.; Li, Z.M.; Lu, L.Q.; Jia, P.; Yang, H.B.; Lan, Y.B. Weed Identification in Maize Field Based on Multi-Spectral Remote Sensing of Unmanned Aerial Vehicle. *Sci. Agric. Sin.* **2020**, *53*, 1545–1555.

34. Liu, X.X.; Wang, L.; Xu, S.H.; Liang, Y. A remote sensing feature selection method of forest biomass estimation based on RF-RFE. *Sci. Surv. Mapp.* **2017**, *42*, 100–105.

35. Attarzadeh, R.; Amini, J. Towards an object-based multi-scale soil moisture product using coupled Sentinel-1 and Sentinel-2 data. *Remote Sens. Lett.* **2019**, *10*, 619–628. [CrossRef]

36. Zhao, Y.C.; Zhang, Y.; Wang, H.Y.; Du, X.; Li, Q.Z.; Zhu, J. Intraday Variation Mapping of Population Age Structure via Urban-Functional-Region-Based Scaling. *Remote Sens.* **2021**, *13*, 805. [CrossRef]

37. Granitto, P.M.; Furlanello, C.; Biasioli, F.; Gasperi, F. Recursive feature elimination with random forest for PTR-MS analysis of agroindustrial products. *Chemometr. Intell. Lab. Syst.* **2006**, *83*, 83–90. [CrossRef]

38. Zhou, X.C.; Zheng, L.; Huang, H.Y. Classification of Forest Stand Based on Multi-Feature Optimization of UAV Visible Light Remote Sensing. *Sci. Silv. Sin.* **2021**, *57*, 24–36.

39. Xie, Z.L.; Chen, Y.L.; Lu, D.S.; Li, G.Y.; Chen, E.X. Classification of Land Cover, Forest, and Tree Species Classes with ZiYuan-3 Multispectral and Stereo Data. *Remote Sens.* **2019**, *11*, 164–191. [CrossRef]

40. Zhang, L.; Gong, Z.N.; Wang, Q.W.; Jin, D.D.; Wang, X. Wetland mapping of Yellow River Delta wetlands based on multi-feature optimization of Sentinel-2 images. *Natl. Remote Sens. Bull.* **2019**, *23*, 313–326.

41. Xu, Y.; Zhen, J.N.; Jiang, X.P.; Wang, J.J. Mangrove species classification with UAV-based remote sensing data and XGBoost. *Natl. Remote Sens. Bull.* **2021**, *25*, 737–752.

42. Rullan-Silva, C.D.; Olthoff, A.E.; Mata, D.; Pajares-Alonso, J.A. Remote monitoring of forest insect defoliation. A Review. *For. Syst.* **2013**, *22*, 377–391. [CrossRef]

43. Senf, C.; Seidl, R.; Hostert, P. Remote sensing of forest insect disturbances: Current state and future directions. *Int. J. Appl. Earth Obs. Geoinf.* **2017**, *60*, 49–60. [CrossRef] [PubMed]

44. Beurs, K.M.; Townsend, P.A. Estimating the effect of gypsy moth defoliation using MODIS. *Remote Sens. Environ.* **2008**, *112*, 3983–3990.

45. Bhattarai, R.; Rahimzadeh-Bajgiran, P.; Weiskittel, A.; Meneghini, A.; MacLean, D.A. Spruce budworm tree host species distribution and abundance mapping using multi-temporal Sentinel-1 and Sentinel-2 satellite imagery. *ISPRS J. Photogramm. Remote Sens.* **2021**, *172*, 28–40. [CrossRef]

46. Huang, H.S.; Deng, J.Z.; Lan, Y.B.; Yang, A.; Deng, X.L.; Zhang, L.; Wen, S.; Jiang, Y.; Suo, G.Y.; Chen, P.C. A two-stage classification approach for the detection of spider mite- infested cotton using UAV multispectral imagery. *Remote Sens. Lett.* **2018**, *9*, 933–941. [CrossRef]

47. Marston, Z.P.D.; Cira, T.M.; Hodgson, E.W.; Knight, J.F.; MacRae, I.V.; Koch, R.L. Detection of Stress Induced by Soybean Aphid (Hemiptera: Aphididae) Using Multispectral Imagery from Unmanned Aerial Vehicles. *J. Econ. Entomol.* **2019**, *113*, 779–786. [CrossRef] [PubMed]

48. Albetis, J.; Duthoit, S.; Guttler, F.; Jacquin, A.; Goulard, M.; Poilve, H.; Feret, J.B.; Dedieu, G. Detection of Flavescence. *Remote Sens.* **2017**, *9*, 308–328. [CrossRef]

49. Yu, R.; Ren, L.L.; Luo, Y.Q. Early detection of pine wilt disease in Pinus tabuliformis in North China using a field portable spectrometer and UAV-based hyperspectral imagery. *For. Ecosyst.* **2021**, *8*, 44–63. [CrossRef]

50. Huang, X.Y.; Xu, Z.H.; Yang, X.; Shi, J.M.; Hu, X.Y.; Ju, W.M. Monitoring the Severity of *Pantana phyllostachysae* Chao on Bamboo Using Leaf Hyperspectral Data. *Remote Sens.* **2021**, *13*, 4146–4164. [CrossRef]

51. Zhang, N.; Zhang, X.L.; Guijun, Y.; Zhu, C.H.; Huo, L.N.; Feng, H.K. Assessment of defoliation during the Dendrolimus tabulaeformis Tsai et Liu disaster outbreak using UAV-based hyperspectral images. *Remote Sens. Environ.* **2018**, *217*, 323–339. [CrossRef]