

Article

Classification of Tree Species Based on Point Cloud Projection Images with Depth Information

Zhongmou Fan *, Wenxuan Zhang, Ruiyang Zhang, Jinhuang Wei, Zhanyong Wang and Yunkai Ruan 

College of Transportation and Civil Engineering, Fujian Agriculture and Forestry University, Fuzhou 350100, China; 1221326001@fafu.edu.cn (W.Z.); 1221326005@fafu.edu.cn (R.Z.); wangzy1026@fafu.edu.cn (Z.W.); ruanyunkai@163.com (Y.R.)

* Correspondence: fujtfan2016@fafu.edu.cn

Abstract: To address the disorderliness issue of point cloud data when directly used for tree species classification, this study transformed point cloud data into projected images for classification. Building upon this foundation, the influence of incorporating multiple distinct projection perspectives, integrating depth information, and utilising various classification models on the classification of tree point cloud projected images was investigated. Nine tree species in Sanjiangkou Ecological Park, Fuzhou City, were selected as samples. In the single-direction projection classification, the X-direction projection exhibited the highest average accuracy of 80.56%. In the dual-direction projection classification, the XY-direction projection exhibited the highest accuracy of 84.76%, which increased to 87.14% after adding depth information. Four classification models (convolutional neural network, CNN; visual geometry group, VGG; ResNet; and densely connected convolutional networks, DenseNet) were used to classify the datasets, with average accuracies of 73.53%, 85.83%, 87%, and 86.79%, respectively. Utilising datasets with depth and multidirectional information can enhance the accuracy and robustness of image classification. Among the models, the CNN served as a baseline model, VGG accuracy was 12.3% higher than that of CNN, DenseNet had a smaller gap between the average accuracy and the optimal result, and ResNet performed the best in classification tasks.

Keywords: image recognition; deep learning; tree species classification; three-dimensional point cloud; projection image; convolutional neural network; residual neural network



Citation: Fan, Z.; Zhang, W.; Zhang, R.; Wei, J.; Wang, Z.; Ruan, Y. Classification of Tree Species Based on Point Cloud Projection Images with Depth Information. *Forests* **2023**, *14*, 2014. <https://doi.org/10.3390/f14102014>

Academic Editor: Temuulen Sankey

Received: 31 August 2023

Revised: 21 September 2023

Accepted: 6 October 2023

Published: 7 October 2023



Copyright: © 2023 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

Forest species classification [1] plays a vital role in forest resource monitoring [2], forest management [3], biodiversity assessment [4], and carbon storage [5], among others. Surveying tree species [6] relies on the visual inspection and measurement of individual trees through various parameters, such as tree height, canopy width, trunk diameter (diameter at breast height), morphological structure, leaf shape, and bark texture. Gathering this detailed information requires significant manpower, time, and prior knowledge, rendering it unsuitable for undertaking large-scale tree species surveys. With technological advancements, remote sensing has gradually been applied in tree species classification [7]. This approach initially extracts features from data that are then combined with traditional supervised classification methods, such as support vector machines [8], maximum likelihood [9], and random forest [10], to achieve tree species classification. However, due to spatial resolution constraints, early remote sensing images were only suitable for regional scale assessments alone and were incapable of achieving individual tree-level classification [11]. With the emergence of high-resolution remote sensing [12] and hyperspectral remote sensing [13,14], the resolution and accuracy of tree species classifications have significantly improved. However, there remain inherent limitations in passive remote sensing images, such as difficulties in acquiring information on tree species below the canopy, reliance on sunlight, and susceptibility to meteorological conditions and time factors [15].

In comparison, Light Detection and Ranging (LiDAR) [16–18], as a form of active remote sensing technology, possesses the ability to autonomously emit light sources and receive reflected signals, which are uniquely high in reflectivity when interacting with plant matter [19]. The quality of its signals is not affected by meteorological conditions or time. LiDAR allows for changes in the signal transmitter position to obtain information about forest trees from various angles. It exhibits higher spatial resolution for complex forest terrains and vegetation structures along with superior penetration capabilities, enabling it to capture information beneath the canopy. Furthermore, LiDAR can collect three-dimensional (3D) point cloud data under various environmental conditions, among many other advantages. Consequently, it has gradually become a research hotspot for the classification of tree species [20–22].

One challenge in the application of point cloud data for tree species classification lies in its unordered nature [23]. Given that each point in a point cloud dataset is independently collected in space, the arrangement of these points in the dataset is random and unrelated to their physical locations. This disorder implies that point cloud data cannot be directly applied to traditional machine learning methods. To address this issue, it is often necessary to transform unordered point cloud data into a format that can be processed by traditional machine learning classifiers. Currently, the most prevalent method is feature-based classification. This approach involves extracting a series of features from the point cloud data, which are then used as inputs for conventional machine learning classifiers for tree species identification. In this manner, the inherent disorder of point cloud data can be converted into ordered information suitable for machine learning classifiers through feature extraction and selection, thereby enabling effective identification of different tree species. Xiaoyi et al. [24] used an optimal feature parameter set based on point cloud distribution characteristics for tree species classification, achieving an average classification accuracy of 58.8%. Cao et al. [25] used full-waveform LiDAR data to achieve an overall classification accuracy of 68.6% for six subtropical forest tree species, including *Pinus massoniana* and *Cunninghamia lanceolata*. In the process of feature-based classification, it is evident that the selection of features typically relies heavily on deep prior knowledge. The accuracy of the classification results demonstrates a high sensitivity to the categories of selected features, which greatly limits the effectiveness of this method. Concurrently, although point cloud data provide comprehensive 3D spatial information about trees, feature-based classification methods often fail to effectively exploit and utilise the features and 3D information inherent within point cloud data.

Simultaneously, considering the strong correlation between a tree species and its morphological structure [26–28], image-based tree classification stands as one of the traditional methods for tree species classification [29]. However, due to its intensive demands for human labour and time, it is not suitable for current large-scale tree species surveys. The advent of LiDAR technology has addressed the previous difficulty in obtaining tree images. By merely segmenting individual trees from point cloud data and projecting them, it is possible to acquire images exhibiting the complete morphological structure of the trees. The introduction of these two-dimensional (2D) images also circumvents the disorderliness issue inherent in point cloud data. Therefore, tree species classification can be based on point cloud projection images. Hamid et al. [30] converted cloud data into a 2D projection image dataset for individual trees. They used a convolutional neural network (CNN) to classify the crowns of 124 conifers, achieving an average accuracy of 87% with limited tree features provided by canopy information. Mizoguchi et al. [31] converted point cloud data from the trunk sections of cedar and cypress trees into images and used the CNN method to classify the two types of trunks, achieving an average accuracy of 89%.

In recent years, image classification algorithms have made significant progress. The evolution of these algorithms is notable, transitioning from early machine learning feature extraction methods to today's advanced deep learning techniques. The mainstay of current image classification algorithms is the CNN [32]. CNNs are often used for feature extraction and dimension reduction, making them a crucial component of modern image classification.

Another significant development is the visual geometry group (VGG) [33], characterised by the use of small convolutional kernels. This technique enhances the effectiveness of the image classification process, especially in dealing with detailed and complex image content. The advent of the residual network (ResNet) [34,35] marked a considerable advancement in the field. ResNet addresses the vanishing gradient problem encountered in deep neural networks by introducing cross-layer residual connections. This innovation significantly enhances the learning capability of deep networks. Furthermore, the densely connected convolutional network (DenseNet) introduced dense connections, another leap forward in the evolution of image classification algorithms. These methods continuously explore the potential of image classification, offering innovative perspectives for the classification of individual tree species.

For the aforementioned reasons, this study focused on tree species classification based on projected images from point cloud data. The main research emphasis lies in exploring the impact of different projection directions, the various classification models, and the incorporation of colour information as a method to restore depth information lost during the transformation of 3D point cloud data into 2D projection images. The feasibility of these methods in resolving the issue of dimensional information loss during the projection process and in enhancing classification accuracy in the context of tree species identification using point cloud projected images was investigated with the aim of informing and benefiting future research.

2. Materials and Methods

2.1. Study Area

The study area was located in the Sanjiangkou Ecological Park in the south-eastern part of Fuzhou City, Fujian Province. Fuzhou City has a typical subtropical monsoon climate. In spring and autumn, temperatures average between 15 and 25 °C. The summers are hot and rainy with temperatures of around 33–37 °C, while winters are warm and humid with temperatures of around 6–10 °C. The city also sustains an average annual relative humidity of approximately 77% and an average annual precipitation of 1224.2 mm. Within this region lies a natural forest with a canopy closure of about 0.5. The main tree species include the council birch (*Betula*), mango (*Mangifera indica*), Tung (*Alstonia scholaris*), banyan (*Ficus religiosa*), and Chinese soapberry (*Sapindus saponaria*), as illustrated in Figure 1. A partial top-down view of the research area is shown in Figure 2. The 3D point cloud data of the research area were obtained using the SAL-1500 3D scanning system (South Group, Beijing, China) mounted on the SF1650 flight platform (DJI, Shenzhen, China) on 15 March 2022. Table 1 lists the main parameters of the 3D laser scanning system.

Table 1. SAL-1500 instrument parameters.

Parameter	SAL-1500
Measurement Rate	2,000,000 points per second
Scanning Speed	400 lines per second
Flight Altitude	200 m
System Relative Accuracy	20 mm
Field of View	20 mm

2.2. Data Processing

The experiment was conducted using the deep-learning framework PyTorch 1.8 coupled with CUDA 11.4. The workstation used for this research ran on Windows 10 Professional equipped with an Intel Core i7-13700F CPU, 32 GB of RAM, and an NVIDIA GeForce RTX 4080 (16 GB) GPU.

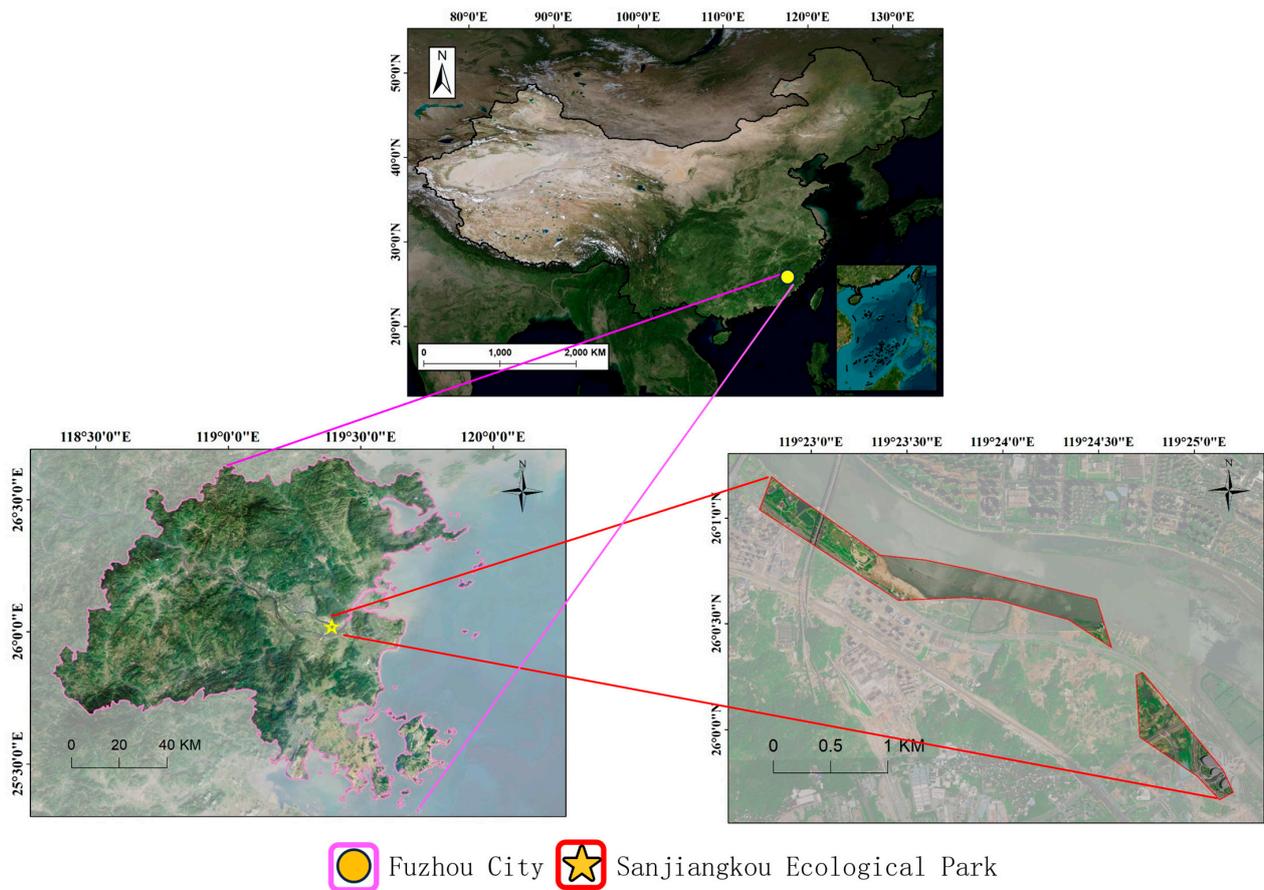


Figure 1. Spatial distribution of the study area.

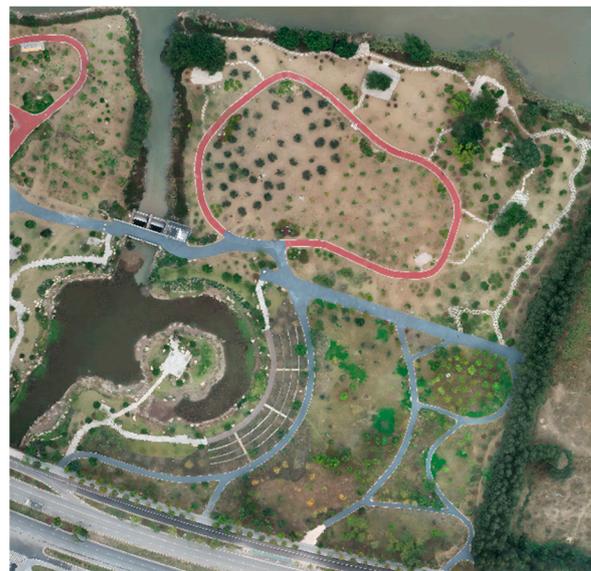


Figure 2. Partial top-down view of the research area.

Nine tree species; namely, council (*Ficus altissima*), birch (*Betula*), mango (*M. indica*), Tung (*A. scholaris*), banyan (*F. religiosa*), Chinese soapberry (*S. saponaria*), Simon poplar (*Populus simonii*), and camphor (*Cinnamomum cam*), were selected as classification samples in the study area. By comparing the coordinates of each tree species that were manually identified and labelled on-site during the collection of point cloud data, the point cloud data of different tree species were separately extracted. Subsequently, the point cloud data collected from airborne LiDAR were pre-processed in three steps. (1) Data calibration: first,

coordinate system conversion was performed, allowing for data analysis and processing in a uniform coordinate system to obtain precise raw data. (2) Data conversion: statistical analysis and smoothing filter methods were adopted to denoise the collected data and address the issues related to noise, overlapping points, and missing points. (3) Data cleaning: finally, outliers were removed based on various measures, such as distance, density, or fitting error, and the number of point clouds were reduced to a consistent quantity using a range of techniques, such as interpolation enhancement [36], jitter augmentation [37], and neighbourhood-based methods, which ultimately improved the data quality. The point cloud data utilised in this study were denoised using a radius outlier removal method with specific parameters set as a minimum neighbouring point count of six and a neighbourhood radius of one. Data augmentation was implemented through jitter interpolation, with specific parameters defined as a mean value of 0.0001 and a standard deviation of 0.01 for the normal distribution of the jitter.

Considering that some data-point clouds had too few points and provided insufficient feature information for image classification, a threshold of 512 points [38,39] was established for screening. Trees with point cloud counts <512 were removed. To avoid difficulties in reading folders named in Chinese during training, the nine tree folders were numerically renamed from zero to eight. A total of 557 files were collected. Overall, 80% of each type of tree projection image was used to train the classification model, whereas the remaining 20% was used to validate the training results. Ultimately, nine numerically named folders were obtained, each containing samples for the test and training sets, as listed in Table 2.

Table 2. Number of trees for each species.

Tree Species	Latin Names	Number of Trees	
		Train	Test
Council tree	<i>Ficus altissima</i>	66	19
Birch	<i>Betula</i>	40	10
Mango tree	<i>Mangifera indica</i>	64	16
Scholar tree	<i>Alstonia scholaris</i>	43	11
Bodhi tree	<i>Ficus religiosa</i>	43	11
Wingleaf soapberry	<i>Sapindus saponaria</i>	38	10
Terminalia neotaliala	<i>Terminalia neotaliala</i>	40	10
Simon poplar	<i>Populus simonii</i>	39	9
Camphor tree	<i>Cinnamomum cam-phora</i>	70	18
Total	—	443	114

2.3. Classification Methods

Considering the requirements of individual tree projection quality and dataset size, four models were selected for dataset classification: CNN, VGG, ResNet, and DenseNet. Figure 3 shows schematic diagrams of the models selected. Fundamental CNNs include convolutional, pooling, and fully connected layers [40]. Convolutional and fully connected layers output the classification results, enabling automatic feature learning and optimisation of feature extractors. A CNN can effectively handle large batches of data, mitigate overfitting issues caused by large data volumes, and perform admirably when dealing with data with a grid-like structure. In order to enhance the comparability between models, this study standardised the parameters of the classification models. The specific parameters are detailed in Table 3.

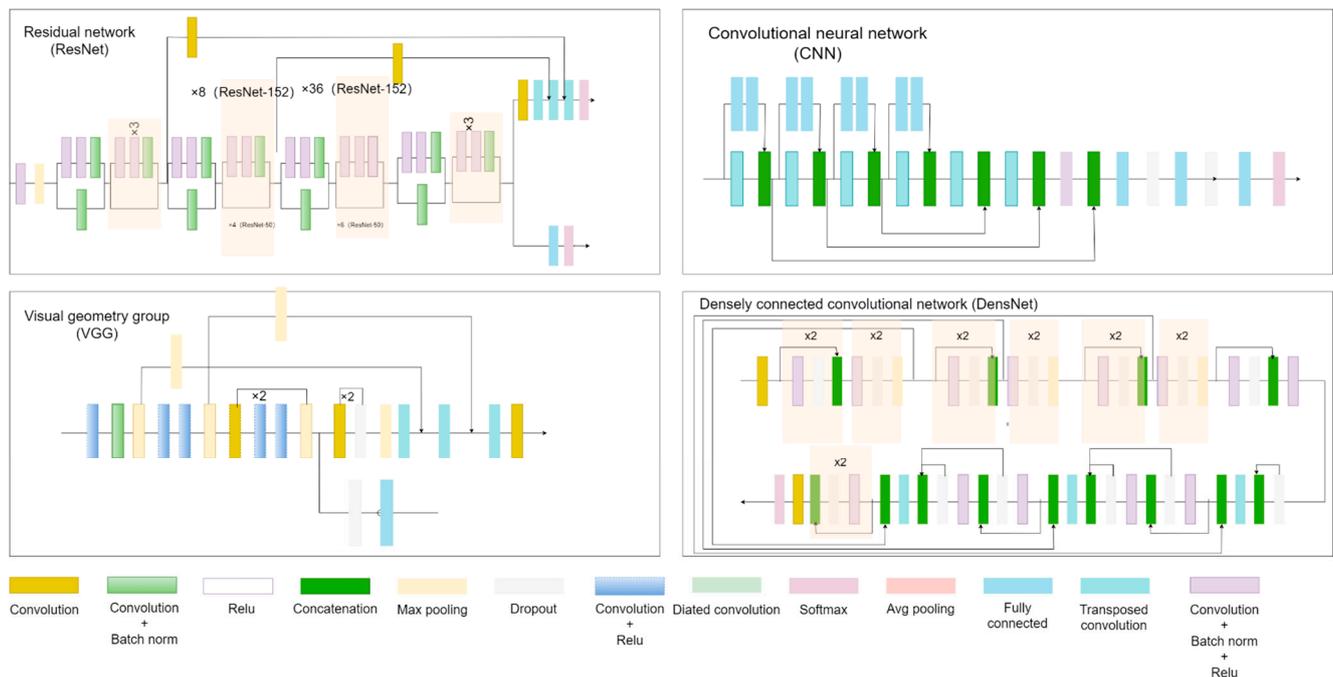


Figure 3. Architecture of image classification models.

Table 3. Parameter settings for the classification model.

Parameter	Value
Optimizer	Adam
Batch size	24
Epoch	300
Learning rate	0.0001

Figure 3 illustrates the key operations of the models used. These include convolution, which intermixes information from input pixels or nodes to learn data features, and batch normalisation, which improves network stability by standardising the output of previous activation layers, thereby reducing overfitting and boosting performance. The ReLU activation function is employed to pass non-negative values alone to the subsequent layer, enhancing the network's non-linearity. Feature maps from various layers are combined using concatenation, enabling the network to retain information from previous layers. Dimensionality reduction and overfitting prevention are achieved via max pooling and dropout, respectively. The latter disregards random neurons during training for this purpose. Dilated convolution expands the receptive field without resolution or coverage loss by applying filters over an area larger than their size because of the added gaps. Softmax, an activation function, is typically utilised in the final network layer for multi-class classification as it converts numbers into probabilities. Average pooling further reduces dimensionality by down-sampling an input using average values over a window defined by a filter. In the fully connected layer, each neuron connects to every neuron in the subsequent layer to learn more global patterns. Finally, transposed convolution, also known as deconvolution, is employed in various tasks, such as segmentation, to increase the input's spatial dimension.

We adopted a strategy of training a shallow, simple network (VGG11) and then reused the weights of VGG11 to initialise VGG13. This iterative training and initialisation process was repeated for VGG19, accelerating convergence during training and addressing issues, such as weight initialisation. VGG builds on a CNN by proposing a more refined learning structure based on depth. It uses multiple 3×3 convolutional and pooling layers for gradual feature extraction. As a simple and deep CNN structure, it enhances the image

processing performance primarily by increasing the network depth, resulting in higher accuracy and a more efficient learning process for complex feature representation. However, owing to its simple structure and excessive depth, it is susceptible to overfitting during training, and the training process takes a significant amount of time [41].

ResNet uses residual blocks to extend the depth of the model by incorporating residual units through a shortcut mechanism and replacing the fully connected layer with a global average pool layer. This network resolves the problem of the gradients becoming progressively smaller during the training of the dataset, which causes slow weight updates during backpropagation. Optimising the training effect of the neural network reduces the training difficulty, accelerates convergence, and improves the overall model performance [42]. Model performance refers to the ability of the neural network to accurately predict or classify new, unseen data based on the learned patterns from the training phase. Higher performance models have lower error rates and better generalisability to different datasets, making them more reliable and robust in various applications.

Although ResNet uses element-wise addition to connect each layer to the two preceding layers, DenseNet extends this principle by proposing that all layers are interconnected [43]. Each layer accepts all previous layers as additional inputs, thereby improving the gradient utilisation. DenseNet also introduces a parameter called “Growth Rate” that controls the growth of feature maps in each layer, enabling better control of complex networks and parameter quantity.

2.4. Point Cloud Projection Transformation

To transform the point cloud data into the 2D images required by the classifier, it is necessary to project the point cloud. Taking the x-axis (east–west direction) as an example, it was first necessary to normalise the point cloud file. In this process, the x-axis coordinates of each 3D point were disregarded, their values were changed to zero, and they were projected on to the 2D (Y–Z) plane. Consequently, the point cloud data were rendered as a scatterplot on the Y–Z plane, and the formula for the x-axis normalisation was as follows:

$$normalised_x = \frac{x - x_{min}}{(x_{max} - x_{min})} \times 255, \quad (1)$$

where X represents the x-axis coordinates in the point cloud data and x_{min} and x_{max} denote the minimum and maximum values of the x-axis coordinates, respectively. The X value after grayscale normalisation, denoted as $normalised_x$, was scaled between 0 and 255.

Partial projection results are illustrated in Figure 4.

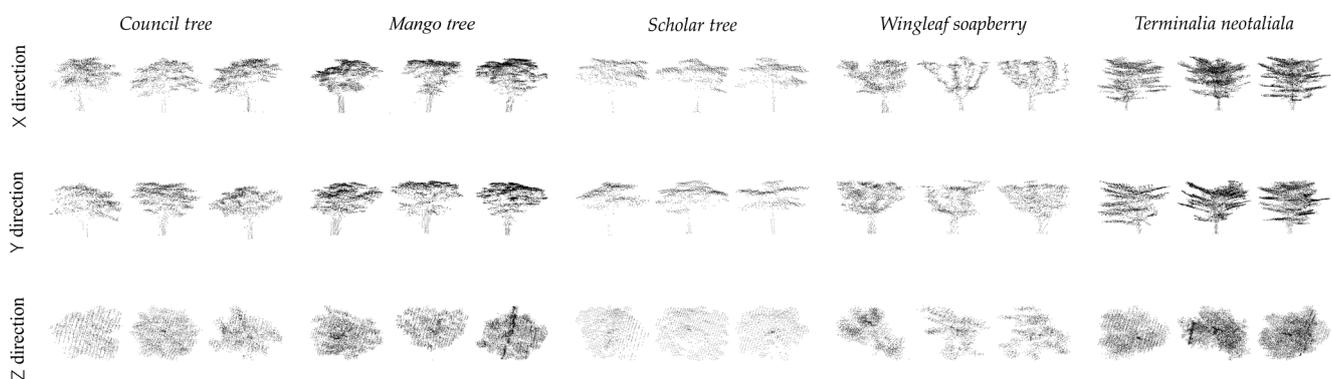


Figure 4. Schematic of projection images with different directions.

This study also incorporated depth information, converting it into colour depth values which were then assigned to the projected images. Therefore, it was necessary to normalise the coordinate information of the compressed dimension in the point cloud data, convert it into grayscale values, and use it to colour the originally colourless 2D point cloud

projection image. Taking the x-axis in Figure 5 as an example, when an x-axis projection was performed, all the x-axis coordinate values change to zero, resulting in a projection image without x-axis depth information. By converting the x-axis coordinate values into colour values to colour the projection image, a projection image was obtained with x-axis depth information.

The formula for colouring the point cloud is as follows:

$$colours[:,0] = normalised_x \tag{2}$$

In Equation (2), the colour is an RGB colour array containing all points. $colours[:,0]$ indicate setting the green channel value of all points as $normalised_x$. This implies that points with smaller x-axis coordinates have lower green-channel values, whereas points with larger x-axis coordinates have higher green-channel values. Meanwhile, the red and blue values were set to zero; therefore, all the points appeared green. As illustrated in Figure 6, contrast diagrams were developed with and without depth in the X and Y directions for four of the nine tree types. For instance, in the case of a mango tree projection with depth in the X-direction, the further away from the projection surface, the deeper the green dot colour, and the closer to the projection surface, the lighter the green dot colour.

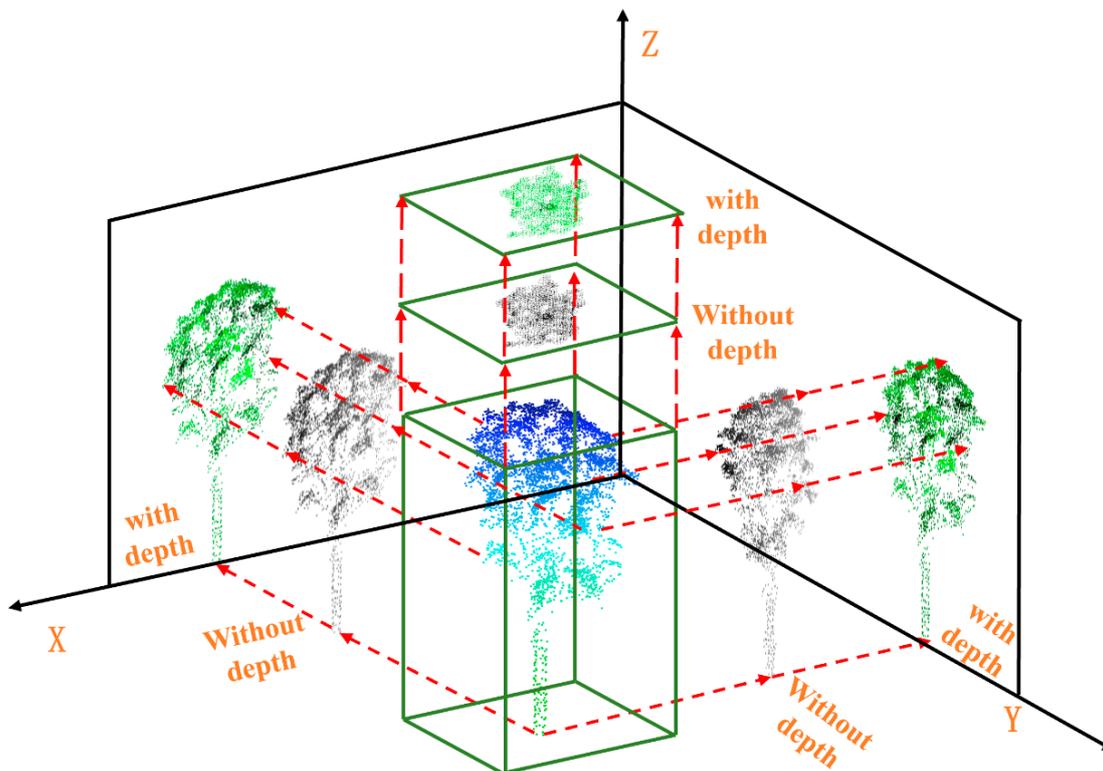


Figure 5. Schematic of projection with different directions and depth information.

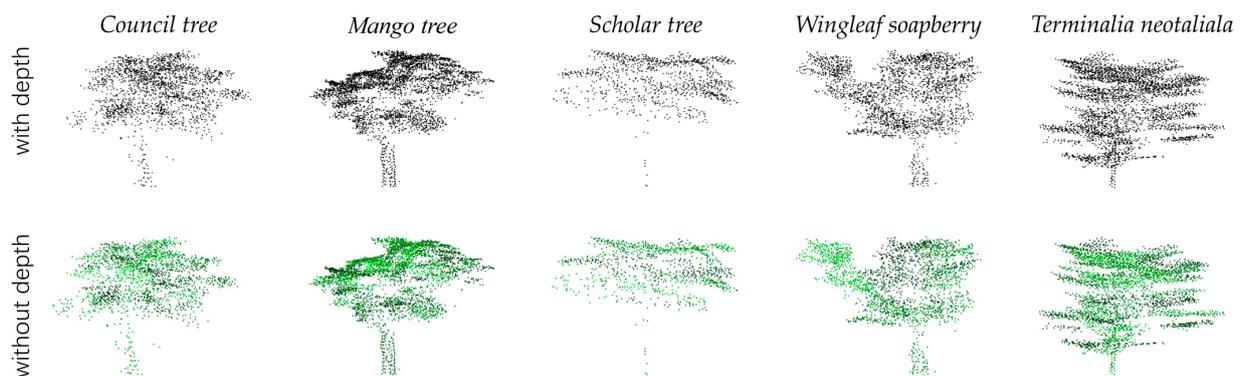


Figure 6. Schematic of projection images with different depths.

Considering the phototropism exhibited by trees during their growth, there are certain differences in the canopy structure in the north–south and east–west directions [44]. In this study, projection occurred from the X (east–west), Y (north–south), and $-Z$ (top–down) directions. Consequently, a single tree generated three different point cloud projection images, as illustrated in Figure 5.

The four classification models employed in this study all incorporated the normalisation of image data. Therefore, the images obtained from the projection could be directly inputted into the classification models, obviating the need for any further processing of the images. The entire data processing workflow is illustrated in Figure 7. The arrows indicate the sequence of the process.

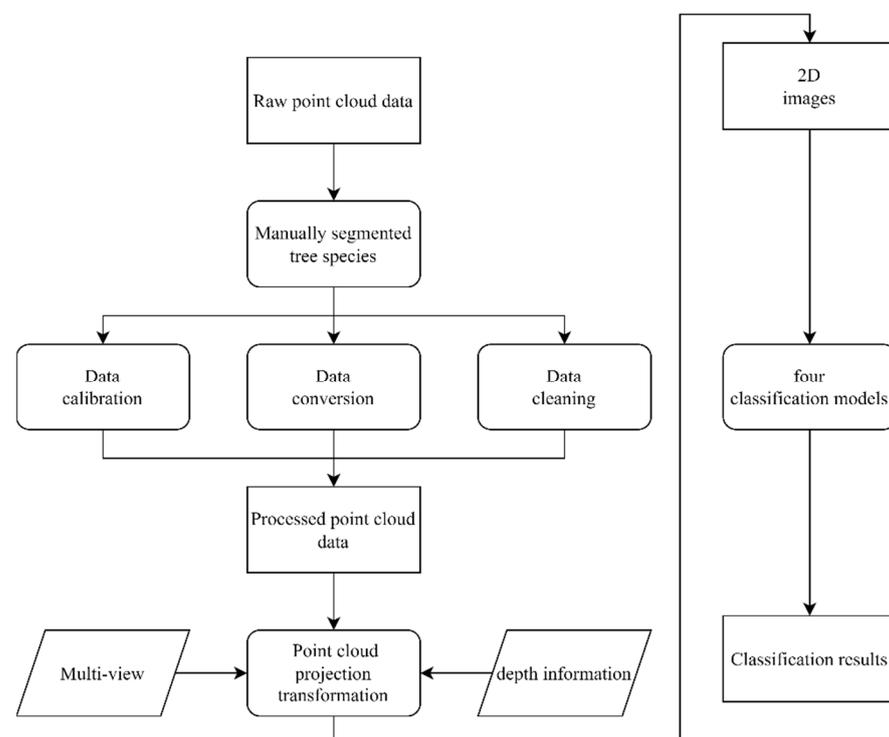


Figure 7. Workflow diagram.

2.5. Evaluation Metrics

The confusion matrix is a common tool in machine learning used to display the statistical information comparing model predictions to existing classifications [45]. It refers to the actual similarity scores generated using a benchmark. It reports four primary values: true positive (TP), true negative (TN), false positive (FP), and false negative (FN). TP

represents the number of instances that the machine learning model correctly predicts as belonging to a specific class, which is consistent with the actual results of the reference alignment. Conversely, TN represents the number of instances that the model correctly predicts as not belonging to a specific class, consistent with the actual reference alignment results. FP indicates the number of instances that the model predicts as belonging to a specific class; however, the reference alignment result suggests otherwise. FN denotes the number of instances that the model predicts as not belonging to a specific class; however, the actual result of the reference alignment indicates the opposite.

The evaluation metrics used in this study were precision and recall. Precision refers to the proportion of correctly predicted tree species quantity to the total number of prediction results, as formulated below:

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

Recall refers to the proportion of correctly predicted quantities to actual quantities, as formulated below:

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

The F-Score is the harmonic mean of precision and recall, offering a balance between the two. It is computed as follows:

$$F - Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (5)$$

3. Results

3.1. Exploring the Potential of the Bi-Directional Approach for Classification

To compare the classification effects, the X single-direction (without depth), Y single-direction (without depth), and Z single-direction (without depth) datasets were classified into the following four models: CNN, VGG, ResNet, and DenseNet. As shown in Table 4, the average precision and recall rates were 79.08% and 78.36% for the X-direction and 78.50% and 77.49% for the Y-direction, respectively. The training accuracy of the Z-direction projection dataset was extremely low, with average precision and recall rates of 57.73% and 56.36%, respectively. This may be because the tree crown obscured the trunk information during the Z-axis projection, resulting in a substantial loss of 3D information. Therefore, this study did not use the Z-direction in subsequent experiments.

Table 4. Comparison of the precision and recall of different methods for the X-, Y-, and Z-axes.

Classification Model	X		Y		Z		Average	
	Precision (%)	Recall (%)						
CNN	68.43	67.54	67.01	65.79	38.87	37.72	58.10	57.02
VGG	82.81	81.58	84.62	84.21	66.08	63.16	77.84	76.32
ResNet	84.98	83.33	83.82	84.21	57.85	58.77	75.55	75.44
DenseNet	86.01	85.96	83.88	82.46	68.13	65.79	79.34	78.07

The scale of the training set directly influences the quantity and quality of knowledge and patterns that the algorithm can learn. In other words, the larger the size of the training set, the richer the information it contains, consequently enhancing the performance of the algorithm's training results. In the next step, the projection images of the same tree were combined in the X- and Y-directions to create a training set, while keeping the original feature information unchanged, with the aim of improving the generalisation ability and accuracy of the model. This was done to expand the training set and sample size and avoid problems such as overfitting and underfitting.

3.2. Classification Results

Six datasets were created based on the tree projection images: X-direction without depth, Y-direction without depth, XY bi-direction without depth, X-direction with depth, Y-direction with depth, and XY bi-direction with depth. Four classification models (CNN, VGG, ResNet, and DenseNet) were employed to classify these datasets. The confusion matrix of some classification results is shown in Figure 8.

As shown in Table 5, the CNN performed poorly in image classification across various datasets, with average precision and recall rates of only 73.53% and 71.20%, respectively. The highest accuracy for the XY bi-directional projection was 79.29%. Considering that the CNN utilized in this study represents an early version of the CNN algorithm, it has a simple structure and fewer feature-extraction capabilities, and thus generally yields mediocre accuracy.

Table 5. Comparison of CNN precision and recall in different directions.

	Precision (%)	Recall (%)	F-Score (%)
X-direction without depth	68.43	67.54	67.98
Y-direction without depth	67.01	65.79	66.39
XY bidirectional without depth	77.99	76.32	77.16
X-direction with depth	74.94	71.05	72.94
Y-direction with depth	73.51	69.30	71.34
XY bidirectional with depth	79.29	77.19	78.23

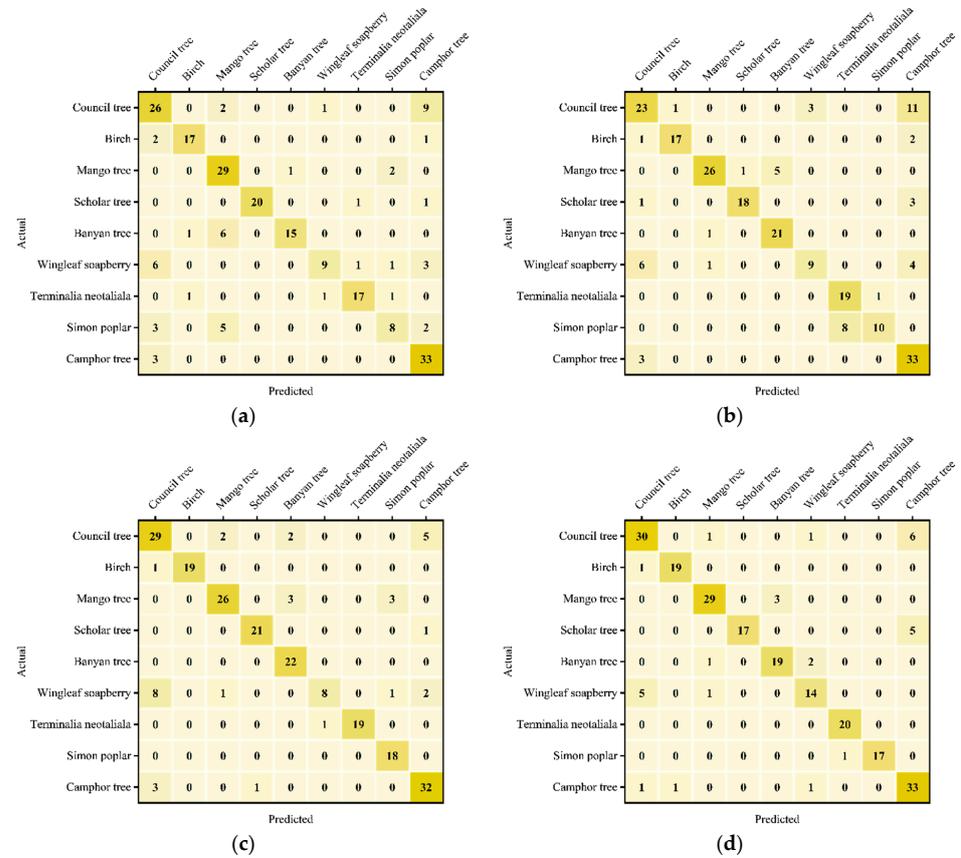


Figure 8. Cont.

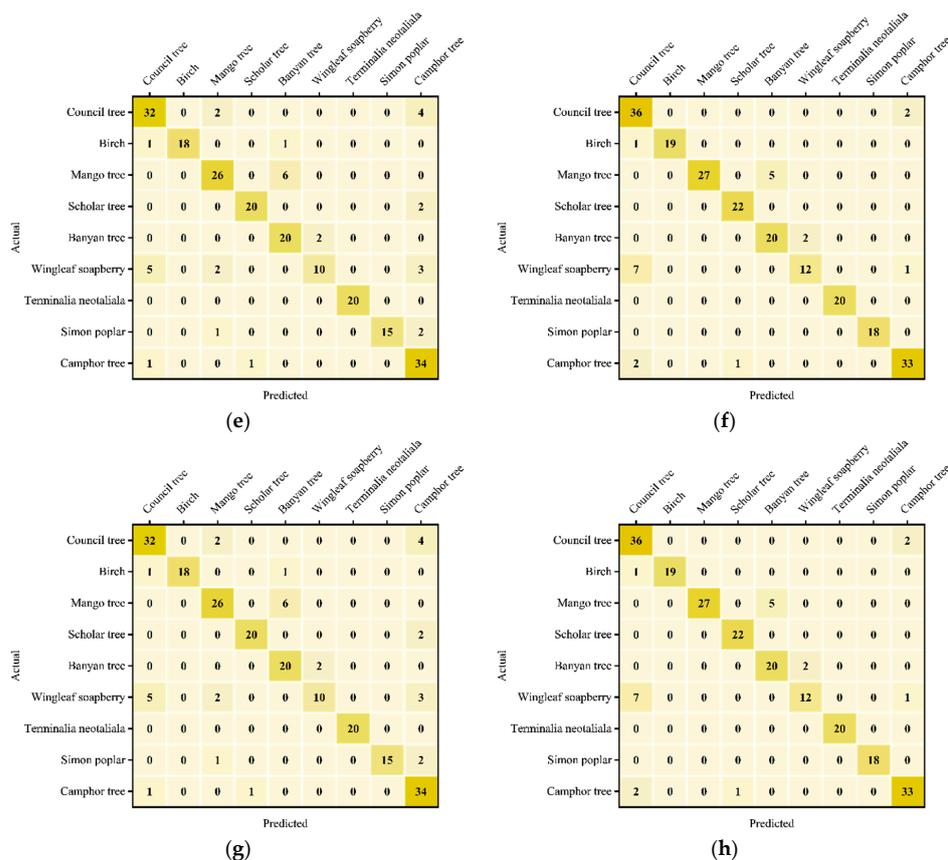


Figure 8. Confusion matrix of various classification models for projected image classification. (a) CNN_XY bi-directional without depth, (b) CNN_XY bi-directional with depth, (c) VGG_XY bi-directional without depth, (d) VGG_XY bi-directional with depth, (e) ResNet_XY bi-directional without depth, (f) ResNet_XY bi-directional with depth, (g) DenseNet_XY bi-directional without depth, (h) DenseNet_XY bi-directional with depth.

As presented in Table 6, the average precision and recall rates of the VGG image classification model were 85.83% and 85.09%, which represented improvements of 12.30% and 13.89% over the CNN classification model, respectively. The precision of the XY bi-directional projection reached 89.36%, which was 10.07% higher than that of the CNN. This may have been because VGG uses multiple convolutional and pooling layers to reduce the number of parameters, enabling the model to learn the rich features of the tree crown and trunk during training. Furthermore, by loading fewer fully connected layers, training becomes more stable and accurate.

Table 6. Comparison of VGG precision and recall in different directions.

	Precision (%)	Recall (%)	F-Score (%)
X-direction without depth	82.81	81.58	82.19
Y-direction without depth	84.62	84.21	84.41
XY bidirectional without depth	85.88	85.09	85.48
X-direction with depth	86.27	85.09	85.69
Y-direction with depth	86.04	85.96	86.00
XY bidirectional with depth	89.36	88.60	88.98

As shown in Table 7, the average precision and recall rates of the ResNet image classification model were 87% and 86.26%, respectively. This may have been due to the use of cross-layer connections in ResNet, which simplify the propagation of tree contour information in the network and deepen the learning of important tree features and solve the

problems of gradient vanishing and exploding in deep network training, thereby improving the precision of learning tree features and patterns. Consequently, both metrics were 1.17% higher than those of the VGG classification model. The precision of the XY bi-directional projection reached 91.46%, which was 2.1% higher than that of VGG.

Table 7. Comparison of ResNet precision and recall in different directions.

	Precision (%)	Recall (%)	F-Score (%)
X-direction without depth	84.98	83.33	84.15
Y-direction without depth	83.82	84.21	84.01
XY bidirectional without depth	86.46	85.53	85.99
X-direction with depth	87.89	86.84	87.36
Y-direction with depth	87.37	86.84	87.10
XY bidirectional with depth	91.46	90.79	91.12

As shown in Table 8, the average precision and recall rates of the DenseNet image classification model were 86.79% and 86.04%, which were 13.26% and 14.84% higher than those of the CNN classification model and 0.96% higher and 0.22% lower than those of the VGG classification model, respectively. The precision of the XY bi-directional projection reached 88.43%, which was 9.14% higher than that of the CNN, 0.93% lower than that of VGG, and 3.03% lower than that of the ResNet model. A possible reason is that DenseNet connects each layer to all subsequent layers in deep learning training. This more comprehensive connection method can cause the amount of tree feature transmission in the point cloud projection image to increase rapidly, thereby increasing the computation and storage costs. This may also cause overfitting and gradient vanishing problems during the tree data training. Therefore, the performance of the DenseNet model in tree classification was slightly lower than that of the ResNet model.

Table 8. Comparison of DenseNet precision and recall in different directions.

	Precision (%)	Recall (%)	F-Score (%)
X-direction without depth	86.01	85.96	85.98
Y-direction without depth	83.88	82.46	83.16
XY bidirectional without depth	88.70	87.72	88.21
X-direction with depth	87.89	86.84	87.36
Y-direction with depth	85.83	85.09	85.4
XY bidirectional with depth	88.43	88.16	88.29

As shown in Figure 9, the basic CNN model exhibited the lowest average accuracy among the four models. The VGG model, which increased the depth based on the CNN, had an average precision of 85.83%. By optimising the connection mode between the layers, the performance of ResNet was 1.17% higher than that of VGG. Owing to the large dataset, the accuracy of DenseNet was slightly inferior to that of ResNet. The bi-directional training results provided more 3D spatial relationships and the training result was 3.99% higher than that of the single direction. With the addition of depth information, which provided more 3D data, the average precisions of the models with and without depth were 81.72% and 84.86%, respectively. The depth-enabled direction was 3.14% higher than that without depth.

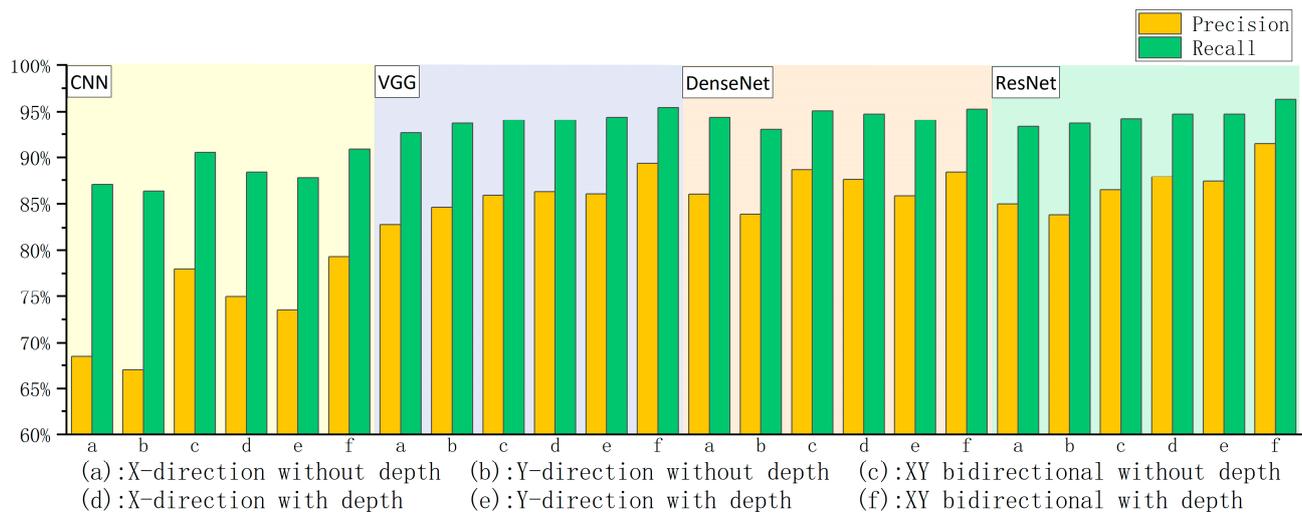


Figure 9. Comparison of precision and recall rates among the four classification models for different projected images.

4. Discussion

The results of this study show that the method of projecting point cloud data into 2D images can effectively address their issue of a lack of order. This approach transforms the original unordered points in point cloud data into pixels in a 2D image with explicit adjacency relationships and order, further employing existing machine learning techniques for classification. However, there is indeed a risk of information loss in this transformation process. To mitigate this risk, this study adopted several strategies. These strategies and their impact on the experimental results are discussed below.

The quality of information contained within the projection images derived from different point cloud projection directions can vary, resulting in different classification accuracies. From the X-, Y-, and Z-direction projection images selected for this study, the average classification accuracies obtained in the X- and Y-directions were superior to those in the Z-direction. Upon comparing the projection images, it was observed that only the tree crown information, including shape, area, and degree of closure, were obtained from the Z-direction projection. As the selected trees in this study were concentrated in one region, similar climatic conditions caused insufficient crown differentiation in some tree species, ultimately leading to lower information content in this direction. Therefore, the classification model had a relatively low accuracy in this direction, and as such, the Z-axis projection model was not discussed. In comparison to the Z-direction, the information content of the X- and Y-direction projection images was often richer.

Compared to single-directional projection images, XY dual-directional projection images provided classification models with shape contour information from different directions while effectively increasing the sample size. Consequently, dual-directional classification outperformed single-directional classification. This strategy not only expanded the training set and sample size compared to single-directional training but also added further information dimensions, thereby enhancing the model's generalisation ability and accuracy. This enabled the model to comprehensively understand the 3D coordinate information of point clouds, enabling a more accurate acquisition of precise positional information, thereby improving the classification accuracy and robustness. The average precision increased from 81.34% to 85.43%.

By supplementing part of the information that was compressed in the dimension during point cloud image projection using colouration methods, classification models could obtain more spatial information, aiding the model in better understanding the 3D information features and distance relationships between the points of the point cloud data. Therefore, the accuracy of the classification model significantly improved, with the average precision increasing from 81.72% to 84.86%. This result indicates that adding depth

information can beneficially impact point cloud projection image classification tasks and that utilising spatial information increases the accuracy of the classification models.

By comparing the training results of the four models, it was observed that the simpler CNN structure had lower accuracy; the VGG model had deeper training and, compared to the ordinary CNN model, its accuracy improved by nearly 12%, reaching 85.83%; ResNet possessed the advantages of VGG with lower time costs and resource occupation rates and its classification accuracy was the highest; DenseNet, while occupying more computational costs and storage space, had a precision that was slightly lower (3.03%) than that of ResNet.

In line with initial expectations, the strategies employed successfully mitigated the problem of dimensionality information loss during the projection process and enhanced classification accuracy. Compared to previous related studies [30,31], the results of the present study expanded the classification from two species to nine. While this undoubtedly increased the complexity of the classification task, a peak classification accuracy of 91.46% was still achieved. This result provides substantial evidence for the efficacy and feasibility of the research method displayed here and holds significant implications for advancing the study of tree species classification using point cloud projection images.

There exists a significant correlation between the morphological structure of trees and their corresponding species. Each tree has unique growth patterns and morphological features, which, in most instances, are directly associated with the species. For instance, some species might exhibit rapid vertical growth, resulting in slender, erect trunks, whereas others might lean towards lateral expansion, forming expansive canopies. These characteristics are intrinsic attributes of trees, manifesting as distinct dendritic structures. Moreover, these differences in dendritic structure are reflected within point cloud data. Specifically, by analysing and interpreting point cloud data, detailed 3D information about trees can be obtained, encompassing various aspects such as trunk thickness, leaf distribution, and canopy shape. This information can significantly aid in the accurate determination of a tree's species. In other words, to some extent, dendritic structure provides pivotal clues for species identification.

During the tree species classification process, the classification model's ability to learn these dendritic structural features specific to tree species can be enhanced. This can be achieved by incorporating additional viewing angle information, augmenting depth information, and adjusting the classification model's parameters and the model. These methods aim to fulfil the objective of effective tree species classification. However, certain misclassification issues persisted. For instance, council trees were misclassified as camphor trees, mango trees as bodhi trees, and Wingleaf soapberries as council trees. Upon manual comparison with the original point clouds, these misclassifications were attributed to specific individual morphologies or issues with point cloud quality. For example, council trees with poor growth may exhibit morphological similarities to camphor trees, mango trees with fewer lateral branches were erroneously identified as bodhi trees, and point clouds of inferior quality misclassified Wingleaf soapberries as council trees. These scenarios underscore that while the majority of trees manifest similar morphological structures during their growth, free growth in natural environments might result in distinct morphological deviations, leading machines to misidentify them as other species. Nevertheless, these discrepancies still fall within an acceptable margin of error.

These challenges suggest the potential for further improving the classification accuracy of point cloud projection images. However, given the complexity of these issues, it might be essential to delve into deeper-level features, optimise the model parameters of the classification methods, or employ superior techniques to observe the understory, such as integrating data from airborne laser scanning (ALS) [46], terrestrial laser scanning [47], and backpack laser scanning [48]. These strategies could potentially offer more effective solutions to these challenging classification problems.

There were some limitations to this study that need to be addressed in future investigations. First, the tree segmentation method required manual assistance and the sampling areas were relatively concentrated. In the future, we will explore the effects of

more projection angles on tree classification and investigate the optimal combination of projection angles to improve the speed and efficiency of classification and minimise redundant samples. Additionally, we will introduce classification models, such as multi-view CNN [49], which treat multiple views of an object as the same object for classification to avoid the problem of treating different projections of the same object as separate entities. For example, Silva et al. [50] achieved 95% accuracy in tree species classification by using microscopic images of three major anatomical parts of wood and combining them with the multi-view random forest model, which is different from the traditional approach of using cross-sectional images alone. In future research, we will explore the optimal combination of multi-view images and the multi-view classification model, along with the point cloud data that are currently in use, to further investigate the upper limit of the classification of tree species using multi-view projection images, which can be obtained quickly and conveniently. Furthermore, considering the differences in tree growth patterns between different terrains and regions, we will further validate the classification performance of trees in different areas and climates and on different mountain slopes (shady versus sunny) to enhance the reliability and generalisability of this method in practical applications. Finally, the point cloud data processed in this study only cover nine specific tree species. As a result, the applicability of the trained classification model is somewhat limited at the current stage, being effective only for the identification and classification of these nine tree types. To augment the model's versatility, robustness, and address the issue of parameter generalization, [51] future research directions will focus on collecting and processing point cloud data from a broader array of tree species across different geographical areas and time frames. This expansion will broaden the model's applicability, further elevating its comprehensiveness and effectiveness in practical forestry applications.

5. Conclusions

Four classification models were used to classify the six datasets. The results indicated that ResNet had the highest overall accuracy, with an average of 87%. The results of this study revealed that multi-directional datasets provide more complete contour features and spatial information for the classification model, and depth information helps compensate for the content lost when converting 3D information into two dimensions. These two methods can make better use of the spatial advantages of point cloud data, thereby improving the classification accuracy. Among the four classification models, the CNN had a relatively low accuracy as a baseline model. The VGG model exhibited a notable improvement over the CNN. DenseNet performed the best for image classification without depth but was not proficient in classifying images with depth. Finally, ResNet performed the best in all classification tasks.

In summary, this study validated the feasibility of using ALS point cloud data for tree species classification using point cloud projection images and improved classification accuracy by adding projection directions and supplementing depth information for projection images. In summary, this study validated the feasibility of using ALS point cloud data for tree species classification through point cloud projection images and enhanced classification accuracy by incorporating projection directions and integrating depth information. Currently, the acquisition and application of airborne point cloud data have reached a certain scale. However, addressing the disorderliness of point cloud data is pivotal to achieving tree species classification. Compared to traditional feature value classifications, this research reduces the prerequisite prior knowledge and various intricate extraction formulas. It efficiently and swiftly transforms point cloud data into two-dimensional images while ensuring limited information loss. This facilitates individuals with foundational knowledge in point cloud and image classification to accomplish tasks such as rare tree species identification, invasive tree species detection, and forest resource surveys using the tree species classification method based on point cloud projection images. This contributes significantly to forestry research.

Author Contributions: Conceptualisation, Z.F.; methodology, Z.F.; software, R.Z.; validation, W.Z. and R.Z.; formal analysis, Z.F.; investigation, J.W., W.Z. and R.Z.; resources, Y.R.; data curation, W.Z.; writing—original draft preparation, R.Z. and Z.F.; writing—review and editing, Z.W., J.W. and W.Z.; visualization, W.Z.; supervision, Z.F.; project administration, Z.F.; funding acquisition, Z.F. and Z.W. All 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, grant number 32101523 and 42007261, the Major Project Funding for Social Science Research Base in Fujian Province Social Science Planning, grant number FJ2020JDZ035, and the Natural Science Foundation of Fujian Province, grant number 2023J01080.

Data Availability Statement: Not applicable.

Acknowledgments: The authors would like to thank all their colleagues for the fruitful discussions on this work.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Fassnacht, F.E.; Latifi, H.; Stereńczak, K.; Modzelewska, A.; Lefsky, M.; Waser, L.T.; Straub, C.; Ghosh, A. Review of studies on tree species classification from remotely sensed data. *Remote Sens. Environ.* **2016**, *186*, 64–87.
- Berie, H.T.; Burud, I. Application of unmanned aerial vehicles in earth resources monitoring: Focus on evaluating potentials for forest monitoring in Ethiopia. *Eur. J. Remote Sens.* **2018**, *51*, 326–335. [\[CrossRef\]](#)
- Fontes, L.; Bontemps, J.D.; Bugmann, H.; Van Oijen, M.; Gracia, C.; Kramer, K.; Lindner, M.; Rötzer, T.; Skovsgaard, J.P. Models for supporting forest management in a changing environment. *For. Syst.* **2010**, *19*, 8–29. [\[CrossRef\]](#)
- Kacic, P.; Kuenzer, C. Forest Biodiversity Monitoring Based on Remotely Sensed Spectral Diversity—A Review. *Remote Sens.* **2022**, *14*, 5363. [\[CrossRef\]](#)
- Li, Y.; Brando, P.M.; Morton, D.C.; Lawrence, D.M.; Yang, H.; Randerson, J.T. Deforestation-induced climate change reduces carbon storage in remaining tropical forests. *Nat. Commun.* **2022**, *13*, 1964. [\[CrossRef\]](#)
- Lechner, A.M.; Foody, G.M.; Boyd, D.S. Applications in remote sensing to forest ecology and management. *One Earth* **2020**, *2*, 405–412. [\[CrossRef\]](#)
- Saatchi, S.; Buermann, W.; Ter Steege, H.; Mori, S.; Smith, T.B. Modeling distribution of Amazonian tree species and diversity using remote sensing measurements. *Remote Sens. Environ.* **2008**, *112*, 2000–2017. [\[CrossRef\]](#)
- Raczko, E.; Zagajewski, B. Comparison of support vector machine, random forest and neural network classifiers for tree species classification on airborne hyperspectral APEX images. *Eur. J. Remote Sens.* **2017**, *50*, 144–154. [\[CrossRef\]](#)
- Hagner, O.; Reese, H. A method for calibrated maximum likelihood classification of forest types. *Remote Sens. Environ.* **2007**, *110*, 438–444. [\[CrossRef\]](#)
- Immitzer, M.; Atzberger, C.; Koukal, T. Tree species classification with random forest using very high spatial resolution 8-band WorldView-2 satellite data. *Remote Sens.* **2012**, *4*, 2661–2693. [\[CrossRef\]](#)
- Immitzer, M.; Vuolo, F.; Atzberger, C. First experience with Sentinel-2 data for crop and tree species classifications in central Europe. *Remote Sens.* **2016**, *8*, 166. [\[CrossRef\]](#)
- Wan, H.; Tang, Y.; Jing, L.; Li, H.; Qiu, F.; Wu, W. Tree Species Classification of Forest Stands Using Multisource Remote Sensing Data. *Remote Sens.* **2021**, *13*, 144. [\[CrossRef\]](#)
- Zhang, B.; Zhao, L.; Zhang, X. Three-dimensional convolutional neural network model for tree species classification using airborne hyperspectral images. *Remote Sens. Environ.* **2020**, *247*, 111938. [\[CrossRef\]](#)
- Nezami, S.; Khoramshahi, E.; Nevalainen, O.; Pölonen, I.; Honkavaara, E. Tree species classification of drone hyperspectral and RGB imagery with deep learning convolutional neural networks. *Remote Sens.* **2020**, *12*, 1070. [\[CrossRef\]](#)
- Dhingra, S.; Kumar, D. A review of remotely sensed satellite image classification. *Int. J. Electr. Comput. Eng.* **2019**, *9*, 1720. [\[CrossRef\]](#)
- Si, H.; Qiu, J.; Li, Y. A review of point cloud registration algorithms for laser scanners: Applications in large-scale aircraft measurement. *Appl. Sci.* **2022**, *12*, 10247.
- Carmer, D.C.; Peterson, L.M. Laser radar in robotics. *Proc. IEEE* **1996**, *84*, 299–320. [\[CrossRef\]](#)
- Means, J.E.; Acker, S.A.; Fitt, B.J.; Renslow, M.; Emerson, L.; Hendrix, C.J. Predicting forest stand characteristics with airborne scanning lidar. *Photogramm. Eng. Remote Sens.* **2000**, *66*, 1367–1372.
- Puttonen, E.; Suomalainen, J.; Hakala, T.; Räikkönen, E.; Kaartinen, H.; Kaasalainen, S.; Litkey, P. Tree species classification from fused active hyperspectral reflectance and LIDAR measurements. *For. Ecol. Manag.* **2010**, *260*, 1843–1852. [\[CrossRef\]](#)
- Budei, B.C.; St-Onge, B.; Hopkinson, C.; Audet, F.A. Identifying the genus or species of individual trees using a three-wavelength airborne lidar system. *Remote Sens. Environ.* **2018**, *204*, 632–647.
- Hovi, A.; Korhonen, L.; Vauhkonen, J.; Korpela, I. LiDAR waveform features for tree species classification and their sensitivity to tree-and acquisition related parameters. *Remote Sens. Environ.* **2016**, *173*, 224–237.

22. Vaughn, N.R.; Moskal, L.M.; Turnblom, E.C. Tree species detection accuracies using discrete point lidar and airborne waveform lidar. *Remote Sens.* **2012**, *4*, 377–403. [[CrossRef](#)]
23. Qi, C.R.; Su, H.; Mo, K.; Guibas, L.J. Pointnet: Deep learning on point sets for 3d classification and segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 652–660.
24. Xiao, Y.L.; Ting, Y.; Lian, F.X. Effective Feature Extraction and Identification Method Based on Tree Laser Point Cloud. *Chin. J. Lasers* **2019**, *46*, 411–422.
25. Cao, L.; Coops, N.C.; Innes, J.L.; Dai, J.; Ruan, H.; She, G. Tree species classification in subtropical forests using small-footprint full-waveform LiDAR data. *Int. J. Appl. Earth Obs. Geoinf.* **2016**, *49*, 39–51.
26. Minore, D. Comparative Autecological Characteristics of Northwestern Tree Species: A Literature Review. University of Illinois at Urbana-Champaign: Champaign, IL, USA, 1979.
27. Rahman, M.A.; Stratopoulos, L.M.; Moser-Reischl, A.; Zölch, T.; Häberle, K.H.; Rötzer, T.; Pretzsch, H.; Pauleit, S. Traits of trees for cooling urban heat islands: A meta-analysis. *Build. Environ.* **2020**, *170*, 106606.
28. Poorter, L.; Bongers, L.; Bongers, F. Architecture of 54 moist-forest tree species: Traits, trade-offs, and functional groups. *Ecology* **2006**, *87*, 1289–1301. [[CrossRef](#)]
29. Ibrahim, I.; Khairuddin, A.S.M.; Abu Talip, M.S.; Arof, H.; Yusof, R. Tree species recognition system based on macroscopic image analysis. *Wood Sci. Technol.* **2017**, *51*, 431–444.
30. Hamraz, H.; Jacobs, N.B.; Contreras, M.A.; Clark, C.H. Deep learning for conifer/deciduous classification of airborne LiDAR 3D point clouds representing individual trees. *arXiv* **2018**, arXiv:abs/1802.08872.
31. Mizoguchi, T.; Ishii, A.; Nakamura, H. Individual tree species classification based on terrestrial laser scanning using curvature estimation and convolutional neural network. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2019**, *42*, 1077–1082. [[CrossRef](#)]
32. Li, S.; Song, W.; Fang, L.; Chen, Y.; Ghamisi, P.; Benediktsson, J.A. Deep learning for hyperspectral image classification: An overview. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 6690–6709. [[CrossRef](#)]
33. Chua, L.O.; Roska, T. The CNN paradigm. *IEEE Trans. Circuits Syst. I Fundam. Theory Appl.* **1993**, *40*, 147–156.
34. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:14091556.
35. Natesan, S.; Armenakis, C.; Vepakomma, U. Resnet-based tree species classification using uav images. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2019**, *42*, 475–481. [[CrossRef](#)]
36. 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.
37. Hasegawa, T.; Emaru, T.; Ravankar, A.A. Real-time interpolation method for sparse lidar point cloud using rgb camera. In Proceedings of the 2021 IEEE/SICE International Symposium on System Integration (SII), Fukushima, Japan, 11–14 January 2021.
38. Chen, J.; Chen, Y.; Liu, Z. Classification of Typical Tree Species in Laser Point Cloud Based on Deep Learning. *Remote Sens.* **2021**, *13*, 4750. [[CrossRef](#)]
39. Ren, N.; Fu, Z.; Zhou, D.; Kong, D.; Liu, H.; Tian, S. Jitter Decomposition by PointNet-Based Dual-Dirac Model. *IEEE Trans. Electromagn. Compat.* **2022**, *64*, 840–849.
40. Kattenborn, T.; Leitloff, J.; Schiefer, F.; Hinz, S. Review on Convolutional Neural Networks (CNN) in vegetation remote sensing. *ISPRS J. Photogramm. Remote Sens.* **2021**, *173*, 24–49.
41. Sengupta, A.; Ye, Y.; Wang, R.; Liu, C.; Roy, K. Going deeper in spiking neural networks: VGG and residual architectures. *Front. Neurosci.* **2019**, *13*, 95.
42. Duta, I.C.; Liu, L.; Zhu, F.; Shao, L. Improved residual networks for image and video recognition. In Proceedings of the 2020 25th International Conference on Pattern Recognition (ICPR), Milan, Italy, 10–15 January 2021.
43. Takahashi, N.; Mitsufuji, Y. D3net: Densely connected multidilated densenet for music source separation. *arXiv* **2020**, arXiv:201001733.
44. Pretzsch, H. Canopy space filling and tree crown morphology in mixed-species stands compared with monocultures. *For. Ecol. Manag.* **2014**, *327*, 251–264.
45. Swetapadma, A.; Yadav, A. A novel decision tree regression-based fault distance estimation scheme for transmission lines. *IEEE Trans. Power Deliv.* **2016**, *32*, 234–245. [[CrossRef](#)]
46. Michałowska, M.; Rapiński, J. A review of tree species classification based on airborne LiDAR data and applied classifiers. *Remote Sens.* **2021**, *13*, 353. [[CrossRef](#)]
47. Kuma, P.; McDonald, A.J.; Morgenstern, O.; Querel, R.; Silber, I.; Flynn, C.J. Ground-based lidar processing and simulator framework for comparing models and observations (ALCF 1.0). *Geosci. Model Dev.* **2021**, *14*, 43–72.
48. Su, Y.; Guo, Q.; Jin, S.; Guan, H.; Sun, X.; Ma, Q.; Hu, T.; Wang, R.; Li, Y. The development and evaluation of a backpack LiDAR system for accurate and efficient forest inventory. *IEEE Geosci. Remote Sens. Lett.* **2020**, *18*, 1660–1664.
49. Su, H.; Maji, S.; Kalogerakis, E.; Learned-Miller, E. Multi-view convolutional neural networks for 3D shape recognition. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 945–953.

50. Rosa da Silva, N.; Deklerck, V.; Baetens, J.M.; Van den Bulcke, J.; De Ridder, M.; Rousseau, M.; Bruno, O.M.; Beeckman, H.; Van Acker, J.; De Baets, B.; et al. Improved wood species identification based on multi-view imagery of the three anatomical planes. *Plant Methods* **2022**, *18*, 79. [[CrossRef](#)] [[PubMed](#)]
51. Kent, M.G.; Schiavon, S. Predicting window view preferences using the environmental information criteria. *LEUKOS* **2023**, *19*, 190–209. [[CrossRef](#)]

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