Article

UAV-based high-throughput approach for fast growing *Cunninghamia lanceolata* (Lamb.) cultivar screening by machine learning

Xiaodan Zou[1], Anjie Liang[1], Bizhi Wu[2], Jun Su[2,3] *, Renhua Zheng[4] *, and Jian Li[1] *

[1] College of Forestry, Fujian Agriculture and Forestry University, Fuzhou 350002, China;

[2] Basic Forestry and Proteomics Research Center, College of Forestry, Fujian Provincial Key Laboratory of Haixia Applied Plant Systems Biology, Fujian Agriculture and Forestry University, Fuzhou 350002, China;

[3] Department of Molecular, Cell & Developmental Biology, University of California, Los Angeles, CA 90095, USA;

[4] Fujian Academy of Forestry, the Key Laboratory of Timber Forest Breeding and Cultivation for Mountainous Areas in Southern China, State Forestry Administration Engineering Research Center of Chinese Fir, the Key Laboratory of Forest Culture and Forest Product Processing Utilization of Fujian Province, Fuzhou, 350012, China.

* Correspondence: junsu@ucla.edu; zrh08@126.com; jianli@fafu.edu.cn

Abstract: Obtaining accurate measurements of tree height and diameter at breast height (DBH) in forests to evaluate the growth rate of cultivars is still a significant challenge, even when using light detection and ranging (LiDAR) and 3-D modeling. We propose an integrated pipeline methodology to measure the biomass of different tree cultivars in plantation forests with high crown density, which combines unmanned aerial vehicles (UAVs), hyperspectral image sensors, and data processing algorithms using machine learning. Using a planation of *Cunninghamia lanceolate*, commonly known as Chinese fir, in Fujian, China, images were collected using a hyperspectral camera and orthorectified in HiSpectral Stitcher. Vegetation indices and modeling were processed in Python using decision trees, random forests, support vector machine, and eXtreme Gradient Boosting (XGBoost) third-party libraries. The tree height and DBH of 2880 samples were measured manually and clustered into three groups—"fast", "median," and "normal" growth groups—and 19 vegetation indices from 12,000 pixels were abstracted as the input of features for the modeling. After modeling and cross-validation, the classifier generated by random forests had the best prediction accuracy compared to other algorithms (75%). This framework can be applied to other tree species to make management and business decisions.

Keywords: *Cunninghamia lanceolate*; UAVs; hyperspectral camera; machine learning; random forests; XGBoost

1. Introduction

Cunninghamia lanceolata (Lamb.) Hook, also known commonly as Chinese fir, is a primary species used for lumber production in southern China, especially between latitudes 20 and 34°N [1]. The species is highly popular due to its high timber quality, rapid growth rate, and straight and uniform stems [2-4]. Following the implementation of the Belt and Road Initiative （"B&R"） strategy by the Chinese government, the demand for high quality Chinese fir seedlings has expanded rapidly, especially in southeast Asia. This booming industry is believed to offer new opportunities for those living in poverty [4, 6].

Peer-reviewed version available at Forests 2019, 10, 815; doi:10.3390/f10090815

As with all trees, tree height and tree diameter at breast height (DBH) are two critical measurements for evaluating the quality of Chinese fir and planning forest management. However, these variables may change with plantation conditions [5, 6]. Several models have been established by ecologies to make these predictions, but these models were generated from limited data and in specific conditions (e.g., locations). Hence, more accurate and general approaches are needed [6, 7].

High-throughput phenotyping that uses satellites and aircraft tends to have a low image resolution and is associated with high costs [8]. However, with recent improvement in unmanned aerial vehicle (UAV) technology, field-based phenotyping of forests has become possible. This approach is also precise and can be conducted at a competitive cost. The UAV approach includes remote sensors, which can adapt to the objectives and more accurately collect measurements [9]. The RGB camera, multispectral sensors, and hyperspectral sensors are frequently used to determine image traits of trees from the canopy. At the same time, light detection and ranging (LiDAR) technology is applied to measure the tree height and DBH [10]. However, it is extremely hard to conduct measurements in areas that have a high crown density with LiDAR [9]. It is also difficult to conduct 3-D modeling without the support of accurate mapping level data [11, 12]. Unfortunately, a high crown density and a lack of mapping level data support happen in most forestry field trials.

Current models for predicting the biomass of Chinese fir are quite simple due to inadequate volumes of data [6, 7]. By taking advantage of unmanned aerial vehicles (UAVs) and remote sensors, it will be much easier to develop better quality and larger datasets from forests. These datasets can provide a better way to develop more complex models with advanced algorisms [13]. Machine learning allows for the classification and prediction of high volumes of data and has gained popularity in recent years. Scikit-learn, a free software machine learning library for the Python programming language, has been used in the forestry field to monitor vegetation levels, assess pest damage in the canopy, and for early detection and quantification of verticillium wilt. The prediction accuracy of this model has been as high as 90% [13-15]. The eXtreme Gradient Boosting (XGBoost) algorithm has been used in models related to pathogen damage and has had a predicting accuracy of >95% [12]. This combination of UAVs, remote sensing imagery, and artificial intelligence still needs to be trained on relevant features to specific datasets, data correlations, and validation processes [16].

Among all machine learning algorithms, decision trees are simpler to understand and interpret than association rules or logistic regressions. Decision trees require only a simple data preparation stage to handle categorical data, and are applied in many medical consulting programs (e.g., providing telehealth service) [17, 18]. The random forests (RF) algorithm was developed to fix the overfitting issue of decision trees, which generate classifiers by multiple decision trees and their mean regression. It has been shown that RF has a good prediction power in clinical diagnoses (e.g., glaucoma) [19]. In machine learning, the support-vector machines (SVMs) algorithm has been widely applied in the biological sciences, due to its powerful function in manipulating massive image data [20]. XGBoost was developed based on the gradient boosting decision tree. The main purpose of gradient boosting is that each model is based on the gradient descent direction of the loss function of the last established model. In addition, the XGBoost algorithm prevents overfitting and uses random forests to sample data columns. The XGBoost algorithm supports parallel computing and implements the processing of missing values, which in the case of missing values, automatically learns the iterative manner of the base classifier [21].

In this study, we selected a typical Chinese fir plantation that was measured to have a high crown density but insufficient mapping level data to support 3-D modeling. We sought to determine better connections between image traits and the biomass of each cultivar. To do this, we described a pipeline which includes UAV-based hyperspectral data collection, image processing, data combination, preprocessing and splitting, and classifier development and evaluation using multiple machine learning algorithms. The entire pipeline was tested using a case study comparing the growth rate (whole tree height and DBH) of eight commercial Chinese fir cultivars, for which there are already manually collected data measurements for three years. To develop the learning (growth rate prediction) model, we considered four machine learning algorithms: decision trees, RF, SVMs,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 July 2019 doi:10.20944/preprints201907.0158.v1

Peer-reviewed version available at Forests 2019, 10, 815; doi:10.3390/f10090815

and XGBoost. We repeatedly composed a learning model using training datasets and evaluated it with a validation dataset. The model which showed the best validation accuracy was chosen as the best learning model. All the image processing and the data manipulation was conducted using open source software or Python, with the goal of building a user-friendly system for people without a technology background.

2. Material and methods

2.1. Study area and experimental design

A 26-acre field was selected in Jiangle, Fujian province, in southeast China (26.6952°N, 117.4344°E). Plants had been generated in tissue culture to maintain the traits of the original variety (i.e., growth rate) (Figure 1). In 2007, the field site was divided into 3 rows: Row1 (up-row), Row2 (middle-row) and Row2 (bottom-row), with 8 sub-plots for each one. A total of 8 Chinese fir cultivars were planted in each row. Each row contained a single replicate of each cultivar. In Row 1, cultivars were in the order of: C1, C2, C3, C4, C5, C6, C7, and C8. In Row 2, cultivars were in the order of C2, C5, C8, C7, C3, C6, C4, and C1. In Row 3, cultivars were in the order of C6, C4, C3, C2, C8, C1, C5, and C7. To minimize the experimental deviation, all trees received the same management measures (fertilizers, farming, and spraying). Randomly, 120 individuals were selected in each sub-plot, and their whole tree height and diameter at breast height (DBH) over 1.3 m were measured manually each year from 2016 to 2018.

2.2. UAV-based hyperspectral image data collection

In this study, a DJI M600 Pro was used as a flight platform and equipped with a 176-band hyperspectral camera (Gaiasky-mini2-VN, Zuolihanguang, Beijing, China) with a wavelength range of 400 to 1000 nm. The resolution for this hyperspectral camera was 960 × 1057 pixels, and it yielded a 4.5 cm spatial resolution at a flight altitude of 90 m.

Prior to measurement, the exposure time was calibrated in direct sunlight by placing a standard whiteboard with a reflectivity of 100% perpendicular to the lens. Two dark background images and one white frame image were then used for lens calibration and reflectance calibration. The dark background images were collected by attaching the lens cap, and one dark background image was obtained by increasing the exposure time by 0.1 s. Cloths of 20%, 40%, and 60% reflectivity were placed in the field and their gray values in the images were later utilized for atmospheric correction.

Hyperspectral data were harvested on 18 April 2019, and were repeated 3 times from 9:00 to 10:00 am, 1:00 to 2:00 pm, and 4:00 to 5:00 pm, respectively, which corresponded to different light conditions.

2.3. Image processing

2.3.1. Digital Surface Models (DSM) generation and Region of Interests (ROI) selection

The entire data correction process included lens correction, reflectance correction, and atmospheric correction. HiSpectral Stitcher (Beta13) was used for orthomosaic and DSM generation using global positioning system (GPS) (Suppl. file1). For background removal and noise reduction in the corrected hyperspectral images, ROIs containing only Chinese fir and no soil, as well as ROIs without Chinese fir and with soil, were selected. A reference spectral library was generated from the average spectra of these ROIs. With these reference spectra, the original spectra were classified using the spectral angle classification method to remove the background and eliminate soil noise.

2.3.2. Parcel detection, random sampling, and dimensionality reduction

After removing the background, the planting area for each cultivar was marked. Areas along edges where adjacent cultivars touched were removed during labeling. In the selected ROIs, 300 points were randomly selected without repetition. Pixels identified as background and noise were

Peer-reviewed version available at Forests 2019, 10, 815; doi:10.3390/f10090815

removed during the preprocessing steps and points with a value of zero were excluded from the sampling process. The hyperspectral data contained information from 176 bands, which had a higher sensitivity than the analysis requirements. Therefore, we reduced the dimensionality to 22 spectra by averaging every 8 adjacent bands.

2.3.3. Calculation of vegetation indices

The wavelengths of visible and near-infrared radiation have been widely used to measure vegetation cover, growth vigor, or biomass [16], and have been used to generate qualitative and quantitative vegetation indices. We calculated 41 different vegetation indices (Figure 3) described by ENVI software (Version 5.5.2, 2019 Harris Geospatial Solutions, Inc., US) based on our dataset which was collected with the Gaiasky-mini2-VN hyperspectral camera (Suppl. file2 and Suppl. file3).

2.4. Data processing and modeling

2.4.1. Manual measurement data processing

K-mean clustering was conducted to classify the manually collected data from 2016–2018 by using height and DBH of the 8 cultivars (Figure. 2a). Three clusters were enough for this classification, because the within group sum of squares decreased dramatically for K values less than 3 (Figure 2).

2.4.2. Trimming the vegetation indices data

For machine learning, the quality of the training data is crucial for the accuracy of the training and prediction. In our case, first we plotted mean values of all 41 vegetation indices against the 3 clusters to understand the data distribution (Figure 3). This showed that most of the vegetation index values of the Fast group were centralized around the means, but that the data distribution of the Median and Normal groups was more normal. Considering the growth rate of these 3 groups of cultivars, we kept only 19 of all 41 vegetation indices ($p < 0.05$ between groups, one-way ANOVA), which were: ARVI, EVI, GARI, OSAVI, VARI, MACARI, MCARI2, MRENDVI, MRESR, MTVI, MTVI2, TCARI, TVI, PSRI, ARI1, ARI2, CRI2, SIPI, and RGRI (described by the ENVI software user manual).

2.5. Modeling by machine learning

2.5.1. Classification algorithms and evaluation metrics

A sample of 500 pixels was randomly selected from each grid. All the vegetation indices for each grid were used as the features, and the 3 classes (Fast, Median, and Normal) were set as labels. The dataset was divided into two sections: 75% for training and 25% to test the set using the train_test_split library from sklearn in Python. Four supervised machine learning algorithms were applied to our dataset: decision tree (DT), conditional random forest (CRF), support vector machine (SVM), and XGBoost. All four models were utilized by applying the corresponding libraries in Python.

To evaluate the accuracy of our classifiers, we conducted predictions on random selected test samples 1000 times and plotted the accuracy (Number of correct predictions) / (Total number of predictions made) for each time. The cross-validation score (CVS) was calculated utilizing the classifiers generated by the decision tree, random forest, SVM, and XGBoost. For XGBoost, Logarithmic Loss (Log Loss) was calculated, which works by penalizing false classifications to evaluate the prediction accuracy.

3. Results

We compared the height and DBH of eight Chinese Fir cultivars (C1–C8) beginning in 2007. A total of 120 out of 650 individuals from each repeat of every cultivar were randomly marked for

Peer-reviewed version available at Forests 2019, 10, 815; doi:10.3390/f10090815

future measurement. From 2007 to 2016, we focused on the establishment of trees and monitored the survival rate (not shown). Beginning in 2016, tree height and DBH were measured manually until 2018. No changes were detected in height and DBH over the course of these three years. Thus, we averaged plant height and DBH for each plant for the three years. According to the classification (Figure 2), we assigned cultivar C5 to the Normal group, cultivars C6, C7, and C8 to the Median group, and cultivars C1, C2, C3, and C4 to the Fast group. These groups represented the different growth rates (tree height and DBH) of Chinese fir as shown in previous field tests (Figure 4).

A total of 19 features from 12,000 pixels were abstracted as the input of this workflow. After repeating 1000 times with default paraments for the prediction with test data sets, the decision tree or random forest classifier had a 32% chance of making an accuracy score of 0.6, 30% chance of 0.7, 17% chance of 0.4, and a 14% chance of 0.9. Both had an average accuracy of 0.5 (Figure 5a and Figure 5b). With the optimal gamma (0.03) for the support vector machine classification, the prediction accuracy was always 0.5 (Figure 5c). To generate the classifier, XGBoost was set to a max depth value of 5 and the "objective" paraments were set to "binary: logistic." The new classifier had a degressive log loss value from 0.5 to 0.001 after 30 rounds (Figure S1), and a detection accuracy of 0.5, 0.65, and 0.95 for all three groups, fast, median and normal, respectively, with an average accuracy of 0.5 (Figure 5d).

The cross validation of each classifier was evaluated to further improve the model. A maximum depth was set for 1–19 and cross validation accuracy (CVA) was expected to be 0.67 for the decision tree classification (Figure 5e) and 0.75 for the random forest (Figure 5g). Due to the importance of the gamma value, we plotted the CVA at different gamma values for SVMs (Figure 5g). This showed a cross validation accuracy of 0.71 when the gamma value was 0.03. For boosting, XGBoost prediction accuracy improved to 0.67 with a gamma of 0 (Figure 5h). The model developed by DT, RF, SVMs, and XGBoost was improved by cross validation, and the model generated by the RF algorithm was modified to have the best prediction accuracy of 0.75.

4. Discussion

In the pre-experiments, we selected several sampling fields, and tried to measure the tree height and DBH using LiDAR and 3-D modeling by high resolution RGB imagery as previously described [10, 11]. Although our manually collected measurements found difference in these variables between cultivars, this was not detected from the data measured by LiDAR and 3-D modeling. This was problematic for the screening process, mainly because the spectral difference between cultivars ranges from 900 to 1000 nm, which is beyond RGB detection range (Figure 6a, b). After conducting 3-D modeling, an inaccurate measurement was made for subplot 3 (shown as a blank) due to high crown density and insufficient geographical information (Figure 6c).

It has been well documented that spectral traits of canopies are closely related to biomass, as well as some vegetation indices [22, 23]. Therefore, it is still possible to evaluate the biomass of the forest with the canopy images. From raw hyperspectral data collected by UAVs, we found the same pattern between the biomass and 19 vegetation indices of different cultivars. This enhanced the possibility of developing a highly accurate classifier. Furthermore, we evaluated the importance of these 19 vegetation indices during development of the model by RF, and found that nine (EVI, RGRI, MTVI, CRI2, PSR1, ARVI, OSAVI, TCARI, and MCARI2) had a larger F-score (>20) than the other 10 vegetation indices (Figure S2). This suggests that these nine vegetation indices are enough for the prediction of the growth rate of Chinese fir.

High resolution RGB imagery is quick and cost effective, and can be utilized to measure some characteristics of crops. However, the shortcomings of spectral information limits the usage of RGB cameras [24, 25]. Although we obtained enough data from our hyperspectral sensor for modeling, efficiency still needs to be improved to reduce the time associated with the sampling process. In this study, each group of eight adjacent bands was combined without sacrificing the data integrity, which implies we can narrow down 176 bands to less than 22 (Figure 3). This would allow us to customize multispectral sensing and accelerate the sampling process to make the pipeline more efficient [26]. We show that the XGBoost algorithm is a better tool for modeling in our case (Figure

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 July 2019 doi:10.20944/preprints201907.0158.v1

Peer-reviewed version available at Forests 2019, 10, 815; doi:10.3390/f10090815

5), but the arguments need to be optimized due to the environmental variables to further improve the pipeline [27].

Reliable and timely information of forest areas is crucial for governments, the commercial sector, and scientists, in order to make decisions on policies, investments, and research. Although tremendous efforts have been made to access that information, it is still difficult to gather customized information for specific cases, which has contributed to a slow-growing forestry industry [28, 29]. In this study, we collaborated with the government and private industry to collect information on forest land. During our investigation, we were amazed by the popularity of UAVs in the forest area with a coverage of the village level. Thus, we identified a strategy based on "Internet of Things," which connects all UAVs via the Internet, named "Internet of UAVs (IoU)", to share all useful information (e.g., growth conditions of different forests, real time weather information, and nutrient condition of ROI) with a wide range of users. Our largest concern for IoU is that most investigators using UAVs are not well-trained to quantify or conduct quality control on massive datasets and images. Thus, we aimed to develop a user-friendly software to serve those investigators with a single click.

5. Conclusions

To understand the connection between canopy spectral features collected by UAVs and the growth rate of Chinese fir, 19 vegetation indices were calculated by randomly selecting 500 pixels from every ROI and assigned to three categories according to the manually measured data. By comparing the prediction accuracy of the machine learning prediction models that were generated by decision tree (DT), random forest (RF), support vector machine (SVM), and XGBoost algorithms, we found that the model generated by the random forest had the highest prediction accuracy (0.75). This provides a novel approach to evaluate the growth rate of Chinese fir in high canopy density area.

Supplementary Materials: Suppl. file1: Digital Surface Models (DSM) of sample filed. Suppl. file2: Values of 41 different vegetation indices (VI) in this research. Suppl. file3: Calculation scripts of vegetation indices (VI).

Suppl. Figure S1 and S2.

Acknowledgments: This work was supported by the National Natural Science Foundation of China (31700582) and was sponsored by the People's Republic of China Scholarship Council.

Author Contributions: Investigation, X.Z., A.L., and B.W.; software, B.W.; writing, original draft, J.S.; review and editing: R.Z. and J.L.; project administration: R.Z. and J.L.; funding acquisition: J.S., R.Z. and J.L.

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

Peer-reviewed version available at *Forests* **2019**, *10*, 815; doi:10.3390/f10090815

Reference

1. Shi, J.;Zhen, Y.; and Zheng, R. Proteome profiling of early seed development in *Cunninghamia lanceolata* (Lamb.) Hook. J. Exp. Bot. **2010**, 61, 2367-2381.
2. Zheng, W.; Chen, J.; Hao, Z.; and Shi, J. Comparative analysis of the chloroplast genomic information of *Cunninghamia lanceolata* (Lamb.) Hook with sibling species from the Genera Cryptomeria D. Don, Taiwania Hayata, and Calocedrus Kurz. Int. J. Mol. Sci. **2016**, 17, 1084.
3. Guan, F.Y.; Tang, X.L.; Fan, S.H.;Zhao, J.C.; Peng, C. Changes in soil carbon and nitrogen stocks followed the conversion from secondary forest to Chinese fir and Moso bamboo plantations. Catena. **2015**,133,455–460.
4. Zhao, M.F.; Xiang, W.H.; Peng, C.H.; Tian, D.L. Simulating age-related changes in carbon storage and allocation in a Chinese Fir plantation growing in southern China using the 3-PG model. Forest Ecology and Management. **2009**, 257, 1520–1531.
5. Ablo, P.I.H.;Mathieu, J.; Claude, N.; Laurent S-A, Quentin P. Improving the robustness of biomass functions: from empirical to functional approaches. Annals of Forest Science. **2015**, 72, 795–810.
6. Guangyi, M.; Yujun, S.; Saeed, S. Models for Predicting the Biomass of *Cunninghamia lanceolata* Trees and Stands in Southeastern China. PLoS ONE. **2017**, 12(1): e0169747. doi:10.1371/journal.pone.0169747.
7. Hao, X.; Yujun, S.; Xinjie, W.; Jin, W.; Yao, F. Linear Mixed-Effects Models to Describe Individual Tree Crown Width for China-Fir in Fujian Province, Southeast China. PLoS ONE. **2015**, 10(4): e0122257. doi:10.1371/journal.pone.0122257.
8. White, J. F.; Andrade-Sanchez, P.; Gore, M. A.; Bronson, K. F.; Coffelt, T. A.; Conley, M. M.; et al.. Field-based phenomics for plant genetics research. Field Crop Res. **2012**, 133, 101–112.
9. Yang, G.; Liu, J.;Zhao, C.; Li, Z.;Huang, Y.; Yu, H.; Xu, B.; Yang, X.; Zhu, D.; Zhang, X.; Zhang, R.;Feng, H.; Zhao, X.; Li, Z.; Li, H.; Yang, H. Unmanned Aerial Vehicle Remote Sensing for Field-Based Crop Phenotyping: Current Status and Perspectives. Front. Plant Sci. **2017**, 8, 1111.
10. Benjamin, B.; Alvaro, L.; Harm, M.; Bartholomeus.; Lammert, K. Comparing RIEGL RiCOPTER UAV LiDAR Derived Canopy Height and DBH with Terrestrial LiDAR. Sensors (Basel). **2017**, 17, 2371.
11. Torres-Sánchez, J.; López-Granados, F.; Serrano, N.; Arquero, O.; Peña, J.M. High Throughput 3-D Monitoring of Agricultural-Tree Plantations with Unmanned Aerial Vehicle (UAV) Technology. PLoS ONE. **2015**, 10(6): e0130479. doi:10.1371/journal.pone.0130479
12. Sandino, J.; Pegg, G.; Gonzalez, F.; Smith, G. 2018. Aerial Mapping of Forests Affected by Pathogens Using UAVs, Hyperspectral Sensors, and Artificial Intelligence. Sensors. **2018**, 18, 944.
13. Aasen, H.; Burkart, A.; Bolten, A.; Bareth, G. Generating 3D hyperspectral information with lightweight UAV snapshot cameras for vegetation monitoring: From camera calibration to quality assurance. ISPRS J. Photogram. Remote Sens. **2015**, 108, 245–259.
14. Calderon, R.; Navas-Cortes, J.; Lucena, C.; Zarco-Tejada, P. High-resolution airborne hyperspectral and thermal imagery for early detection of Verticillium wilt of olive using fluorescence, temperature and narrow-band spectral indices. Remote Sens. Environ. **2013**, 139, 231–245.
15. Nasi, R.; Honkavaara, E.; Lyytikainen-Saarenmaa, P.; Blomqvist, M.; Litkey, P.; Hakala, T.; Viljanen, N.; Kantola, T.; Tanhuanpaa, T.; Holopainen, M. Using UAV-Based photogrammetry and hyperspectral imaging for mapping bark beetle damage at tree-level. Remote Sens. **2015**, 7, 15467–15493.
16. Pause, M.; Schweitzer, C.; Rosenthal, M.; Keuck, V.; Bumberger, J.; Dietrich, P.; Heurich, M.; Jung, A.; Lausch, A. In situ/remote sensing integration to assess forest health–A review. Remote Sens. 2016, 8, 471.
17. Quinlan, J.R. Induction of decision trees. Mach Learn. **1986**, 1, 81–106.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 July 2019 doi:10.20944/preprints201907.0158.v1

Peer-reviewed version available at *Forests* **2019**, *10*, 815; doi:10.3390/f10090815

18. Ching-Chin, C.; Yu-Jen, C.; Hsiao, B. Decision tree–based classifier in providing telehealth service. BMC Medical Informatics and Decision Making. **2019**, 19, 104.
19. Kim, S.J.; Cho, K.J.; Oh, S. Development of machine learning models for diagnosis of glaucoma. PLoS ONE. **2017**, 12(5): e0177726. doi:10.1371/journal.pone.0177726.
20. Bilwaj, G.; Christos, Davatzikos. Analytic estimation of statistical significance maps for support vector machine based multi-variate image analysis and classification. Neuroimage. **2013**, 78, 270–283.
21. Chen, T.; Guestrin, C. "XGBoost: A scalable tree boosting system", Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, pp. 785-794, 2016.
22. Wang, Y.; Wang, D.; Shi, P.; Omasa, K. Estimating rice chlorophyll content and leaf nitrogen concentration with a digital still color camera under natural light. Plant Method. **2014**, 10, 36.
23. Juan, J. Q. V.; Zhang, C.Y.; Jamin, A.; Smitchger, R. J.; McGee and Sindhuja, S. Phenotyping of Plant Biomass and Performance Traits Using Remote Sensing Techniques in Pea (*Pisum sativum*, L.). Sensors. 2019, 19, 2031.
24. Duan, T.; Chapman, S.C.; Guo, Y.; Zheng, B. Dynamic monitoring of NDVI in wheat agronomy and breeding trials using an unmanned aerial vehicle. Field Crops Res. **2018**, 210, 71–80.
25. Madec, S.; Jin, X.; Lu, H.; Solan, B.; Liu, S.; Duyme, F.; Heritier, E.; Baret, F. Ear density estimation from high resolution RGB imagery using deep learning technique. Agric. For. Meteorol. **2019**, 264, 225–234.
26. Maes, W.H.; Steppe, K. Perspectives for Remote Sensing with Unmanned Aerial Vehicles in Precision Agriculture. Trends Plant Sci. In Press.
27. Zhong, J.C.; Sun, Y.S.; Peng, W.; Xie, M.Z.; Yang, J.H.; Tang, X.W. XGBFEMF: An XGBoost-Based Framework for Essential Protein Prediction. IEEE Transactions on NanoBioscience. **2018**, 17, 243 – 250.
28. MacDicken, K.G. Global Forest Resources Assessment 2015: What, why and how? For. Ecol. Manage. **2015**, 352, 3–8.
29. Rodney, J. K.; Gregory, A. R.; Frédéric, A.; Joberto, V. D. F.; Alan, G.; Erik, L. Dynamics of global forest area: Results from the FAO Global Forest Resources Assessment 2015. For. Ecol. Manage. **2015**, 352, 9–20.

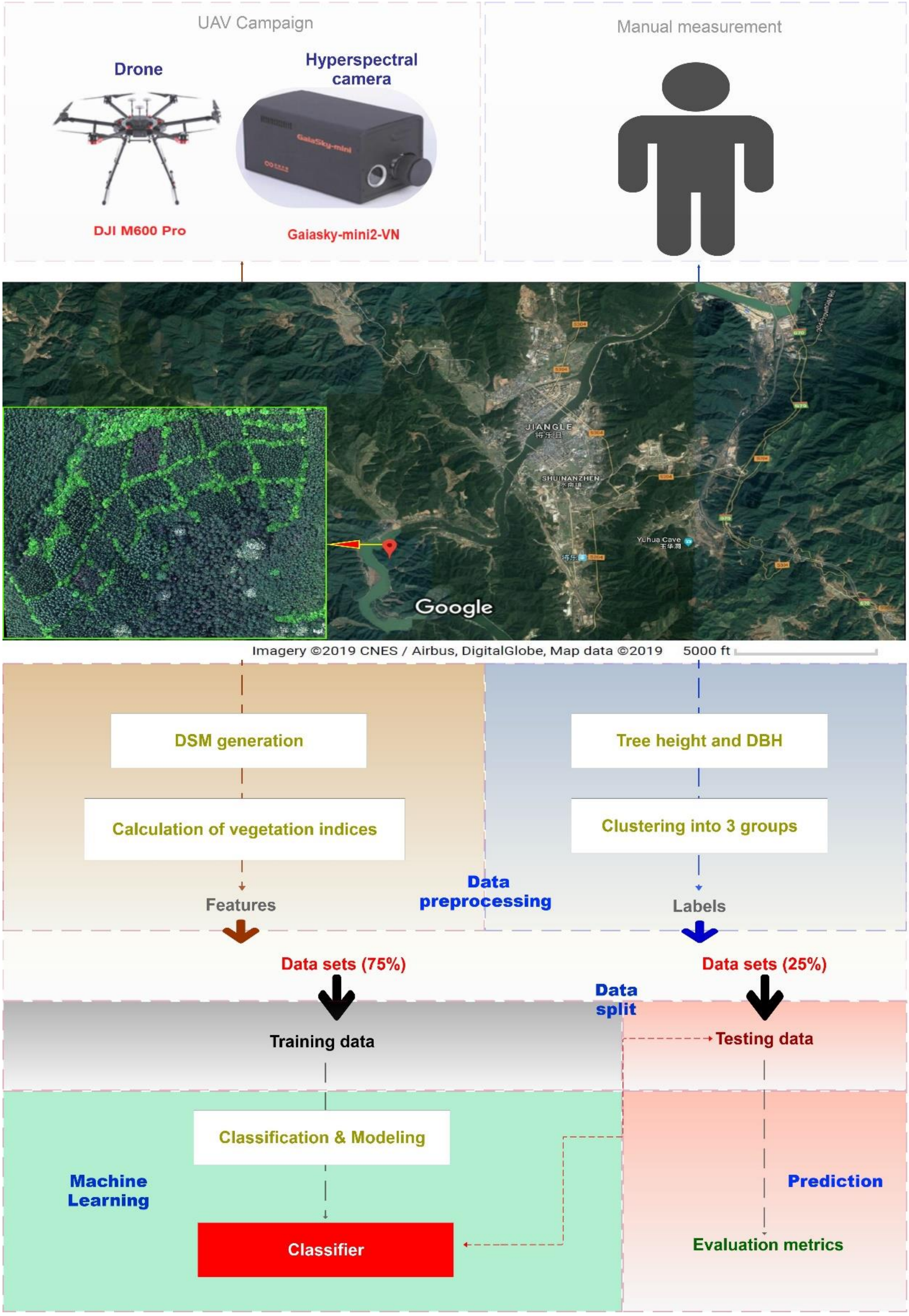

Figure 1. Pipeline process for screening fast growing Chinese fir.

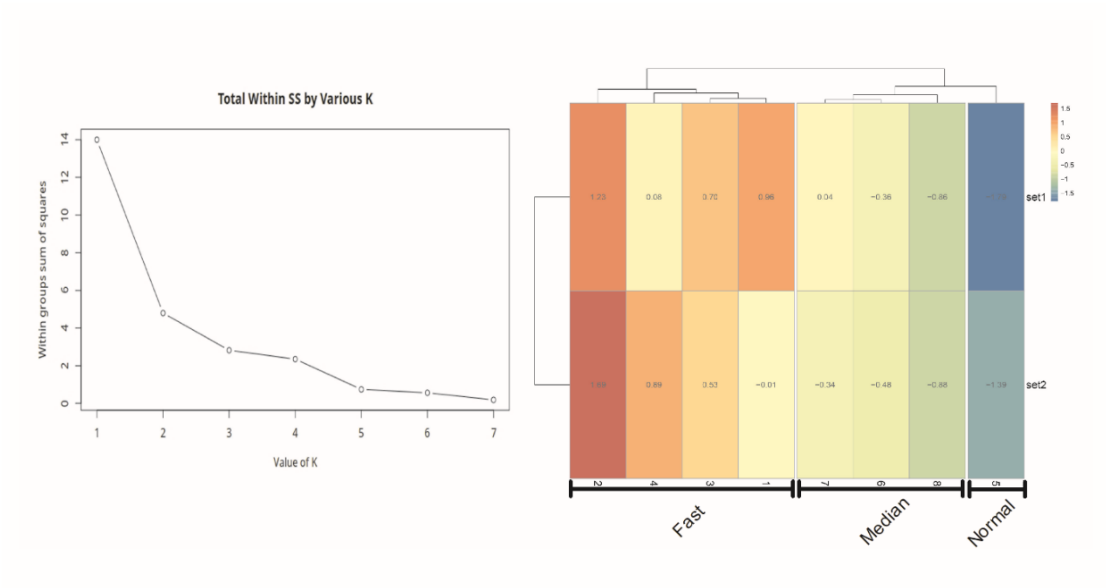

Figure 2. Clustering results of manually collected data on tree height and DBH of Chinese fir in Jiangle, China. (Left) represents K-mean clustering results plotted by the number of clusters and the validity; (Right) shows the heatmap plot against first 2 principal components (set1 & set2) of each cluster (Fast, Median, and Normal).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 July 2019 doi:10.20944/preprints201907.0158.v1

Peer-reviewed version available at Forests 2019, 10, 815; doi:10.3390/f10090815

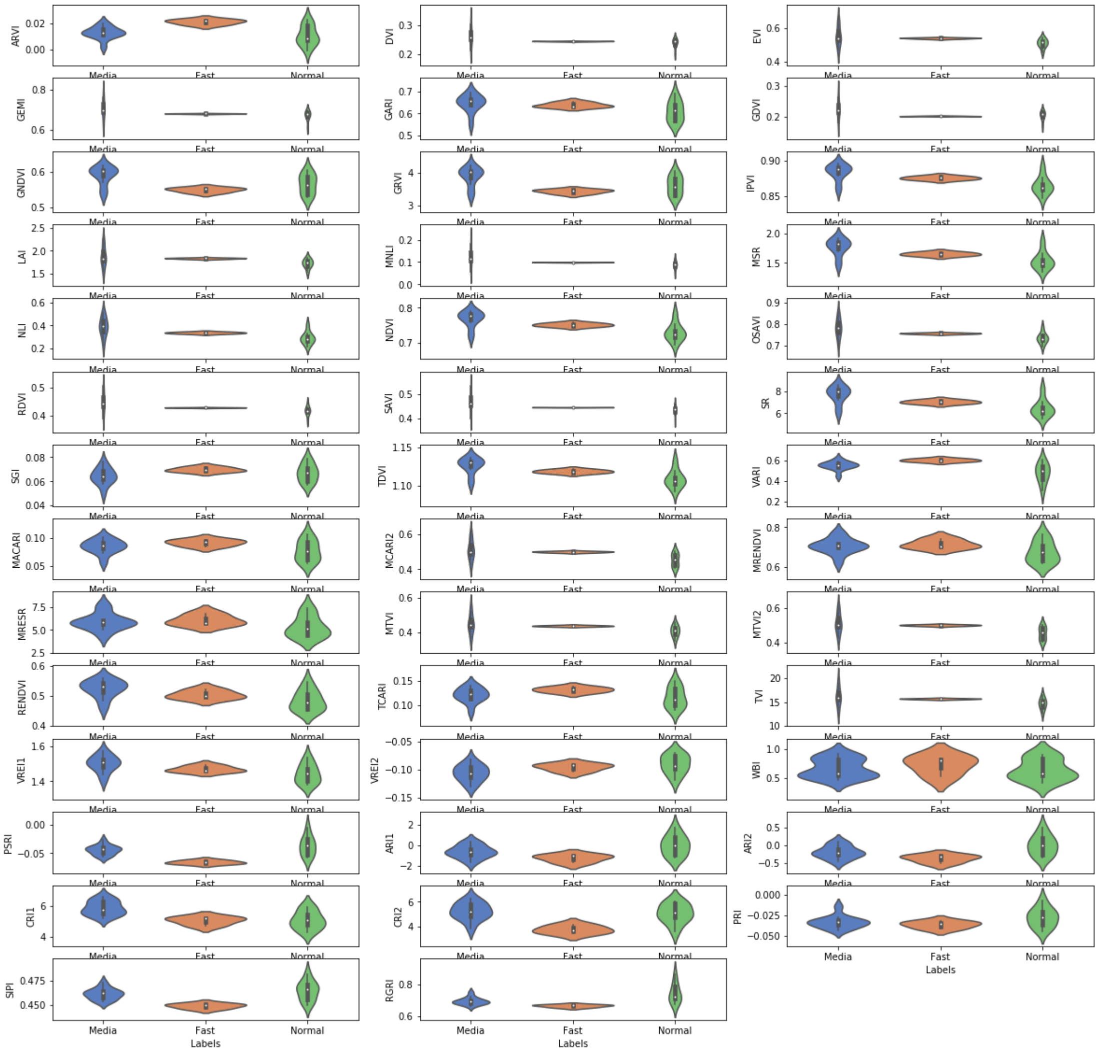

Figure 3. Data distribution of 41 vegetation indices of 3 clusters. Each column represents the data distribution of the represented group, the mean was represented by the white spot inside the column.

Peer-reviewed version available at *Forests* **2019**, *10*, 815; doi:10.3390/f10090815

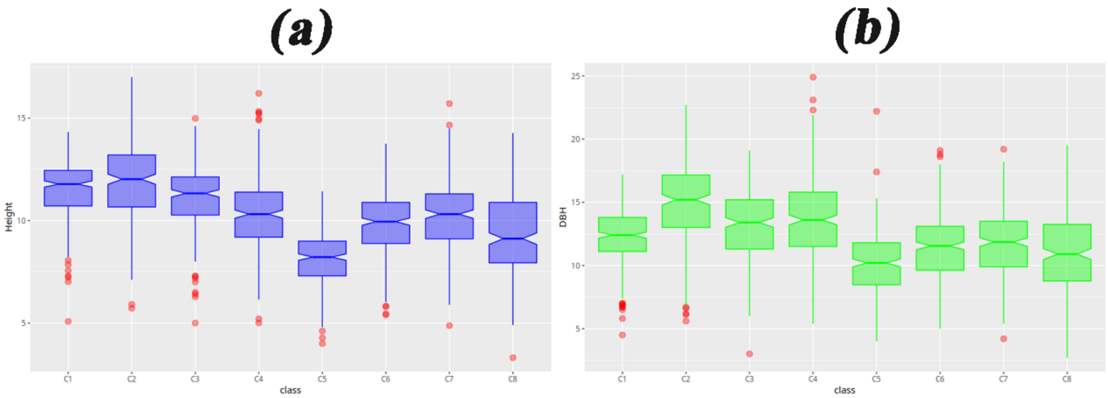

Figure 4. The manually collected data for tree height and DBH of each Chinese fir cultivar. Each box represents mean and 95% data distribution, where the bar represents standard error for the represented cultivar and red spots represent outlier values. (a) Whole tree height and (b) DBH.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 July 2019 doi:10.20944/preprints201907.0158.v1

Peer-reviewed version available at Forests 2019, 10, 815; doi:10.3390/f10090815

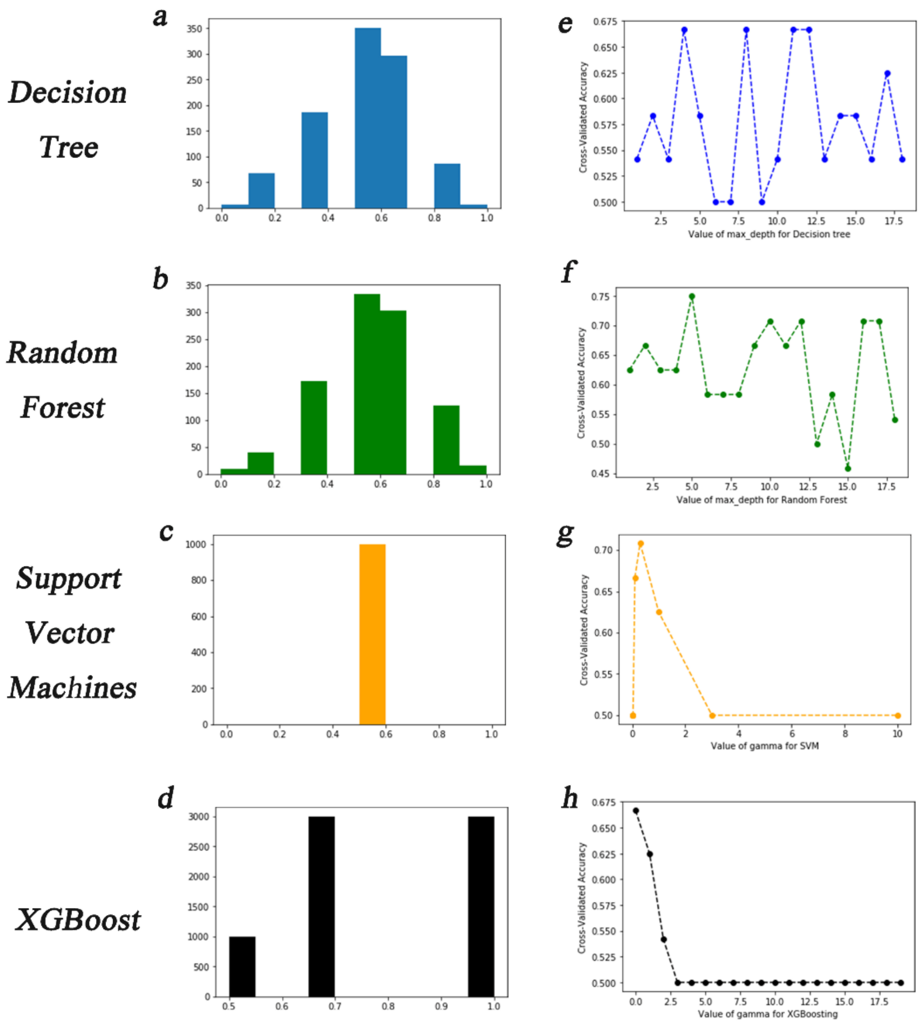

Figure 5. Evaluation results of different algorithms. Prediction accuracy of randomly selected test samples (1000 replicates) was calculated for decision tree (a), random forest (b), support vector machines (c), and XGBoost (d). The cross-validated accuracy was calculated with the values of max_depth from 1 to 18 for decision tree (e) and random forest (f), for gamma values from 1 to 10 for the support vector machines (g), and gamma values from 1 to 18 for XGBoost (h).

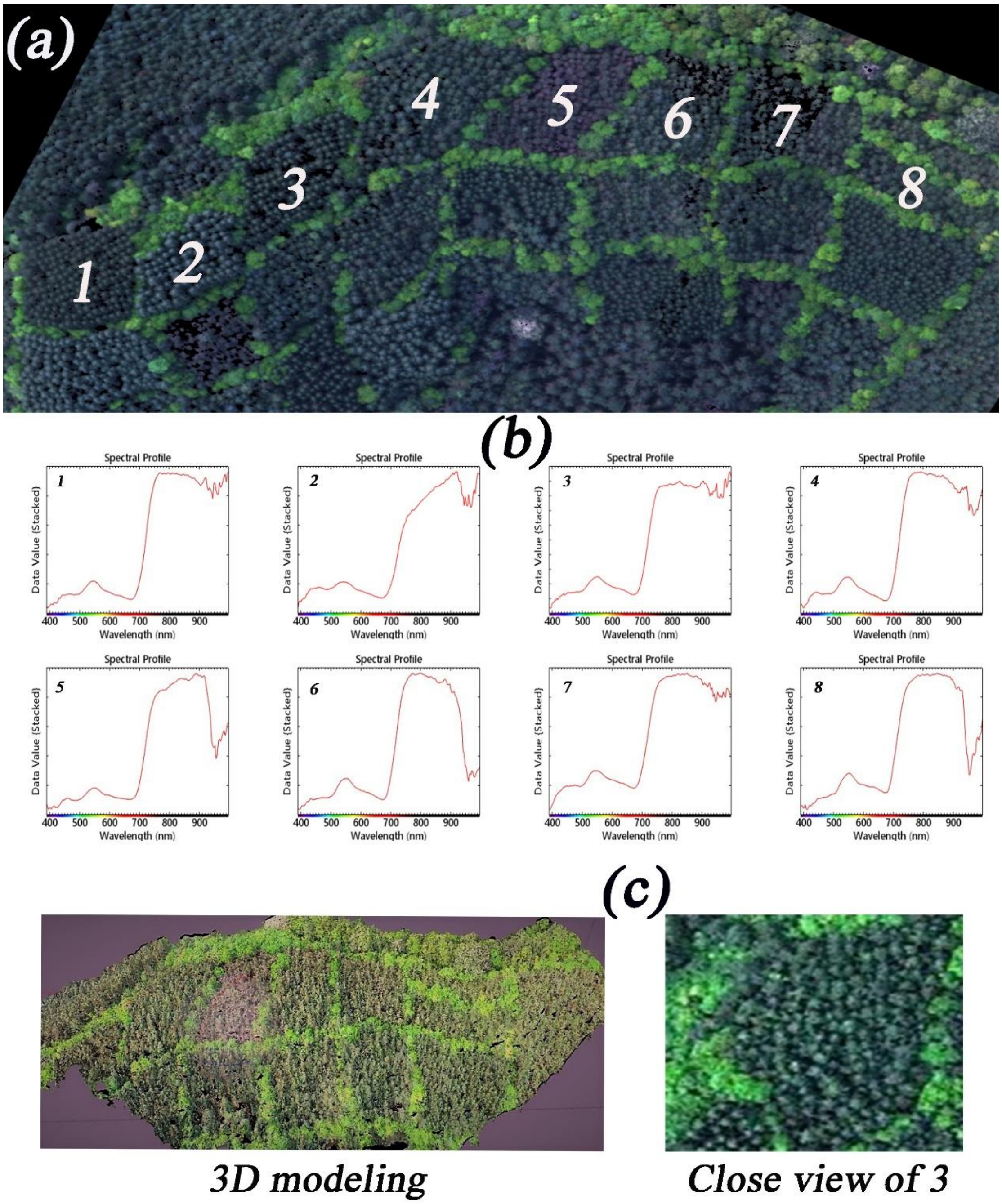

Figure 6. Limitations on measuring the real tree height and DBH of Chinese fir directly by RGB camera. Pre-experiment results of different locations (a); the variation of spectral profiles for 8 Chinese fir cultivars (b) range from 900–1000 nm; (c) 3-D modeling results of 8 Chinese fir cultivars (left) and a real color close view of subplot 3 (right).