

Article

Evaluating Variable Selection and Machine Learning Algorithms for Estimating Forest Heights by Combining Lidar and Hyperspectral Data

Sanjiwana Arjasakusuma ^{1,*} , Sandiaga Swahyu Kusuma ¹ and Stuart Phinn ² 

¹ Department of Geographic Information Science, Faculty of Geography, Gadjah Mada University, Bulaksumur, Yogyakarta 55281, Indonesia; sandiaga.kusuma@mail.ugm.ac.id

² Remote Sensing Research Centre, School of Earth and Environmental Sciences, The University of Queensland, Brisbane, QLD 4072, Australia; s.phinn@uq.edu.au

* Correspondence: sanjiwana.arjasakusuma@ugm.ac.id

Received: 16 July 2020; Accepted: 19 August 2020; Published: 24 August 2020



Abstract: Machine learning has been employed for various mapping and modeling tasks using input variables from different sources of remote sensing data. For feature selection involving high- spatial and spectral dimensionality data, various methods have been developed and incorporated into the machine learning framework to ensure an efficient and optimal computational process. This research aims to assess the accuracy of various feature selection and machine learning methods for estimating forest height using AISA (airborne imaging spectrometer for applications) hyperspectral bands (479 bands) and airborne light detection and ranging (lidar) height metrics (36 metrics), alone and combined. Feature selection and dimensionality reduction using Boruta (BO), principal component analysis (PCA), simulated annealing (SA), and genetic algorithm (GA) in combination with machine learning algorithms such as multivariate adaptive regression spline (MARS), extra trees (ET), support vector regression (SVR) with radial basis function, and extreme gradient boosting (XGB) with trees (XGBtree and XGBdart) and linear (XGBlin) classifiers were evaluated. The results demonstrated that the combinations of BO-XGBdart and BO-SVR delivered the best model performance for estimating tropical forest height by combining lidar and hyperspectral data, with $R^2 = 0.53$ and RMSE = 1.7 m (18.4% of nRMSE and 0.046 m of bias) for BO-XGBdart and $R^2 = 0.51$ and RMSE = 1.8 m (15.8% of nRMSE and −0.244 m of bias) for BO-SVR. Our study also demonstrated the effectiveness of BO for variables selection; it could reduce 95% of the data to select the 29 most important variables from the initial 516 variables from lidar metrics and hyperspectral data.

Keywords: forest height; feature selection; dimensionality reduction; lidar; AISA hyperspectral

1. Introduction

Forest structural properties are critical information sources for measuring and monitoring aboveground biomass (AGB), which is used to predict the amount of carbon stock in forest stands [1–3]. Accurate biomass data are needed in order to monitor the progress toward the 5.2% reduction in carbon emission compared with 1990 levels, agreed by 37 nations in the Kyoto Protocol of the United Nations Framework Convention on Climate Change (UNFCCC) [4]. In addition, forest structural properties are also important to infer forest conditions, as well as to assess the habitat and biodiversity within forest structures, especially for canopy-dwelling organisms [5–7].

The UNFCCC has set up three levels of accuracy for mapping carbon emissions, from national to regional and global scales, where remote sensing has been recognized as one of the promising technologies for mapping these features. Carbon measurement was traditionally carried out by field

plot sampling, with high accuracy; however, this method was expensive and time-consuming [1,8,9]. Therefore, the development of remote sensing technologies that can map broader areas from regional (1000's km^2) to global (10^8 s km^2) scales at high levels of detail has been widely explored for mapping and modeling forest structural properties. However, mapping forest structural properties using remote sensing is still a challenging task, especially when dealing with complex forest environments; thus, further assessment related to the uncertainties of using remote sensing methods for mapping forest structural properties is needed [10].

1.1. Remote Sensing for Forest Structural Property Modeling

Forest structural variables are a complex set of properties that portray the quantities and spatial distribution of forest components, including leaves and branches [11]. These variables, including tree height, diameter at breast height (DBH), basal area, and AGB, have been modeled from multi/hyperspectral data [12,13], radar intensity backscattering [14,15], and height metrics from both large- and small-footprint lidar data [2,16,17]. Owing to the inability of remote sensing methods to directly measure the forest structural properties, they are typically measured by linking the field plot inventory data with remote sensing metrics through empirical modeling using single or multivariate analysis.

Univariate analysis employing single variables from remote sensing metrics has several limitations. For example, the strong relationship between broadband multispectral vegetation indices and the leaf area index tends to lower the prediction accuracy for measuring high-density complex forest areas [18], while radar backscatter tends to saturate for biomass levels greater than 100 mg/ha [15] and is also affected by precipitation levels; the saturation of the radar backscatter and the influence of precipitation levels increase the soil and vegetation moisture level and reduce the dynamic range of radar backscatter, hence lowering the sensitivity of the model [14,19,20]. The most appropriate remote sensing technology for measuring forest structural properties is the use of lidar data, which can provide accurate and precise measurement of ground and objects elevation, enabling the measurement of aboveground object properties closely related to forest structural properties. Hyde et al. [1] found that the performance of lidar data can be enhanced by combining lidar sensors with other sensors.

Multi-sensor modeling, such as by combining lidar and other remote sensing sensors, can increase the modeling performance of forest structural properties, mainly tropical forest biomass, and can fulfill the standard accuracy for the monitoring, reporting, and verification (MRV) activities; however, the performance and accuracy of the multi-sensor fusion can be varied [21,22]. The improvement when combining various multi-sensor remote sensing data comes from the unique features or metrics that each datum can generate, which are able to increase the accuracy of modeling forest structural properties. Zolkos, Goetz and Dubayah [21] pointed out that lidar corresponded strongly with the vertical profiles of trees. Passive optical sensors provided information regarding the species and types of forests, and radar provided the variability of tree foliage and branches through the backscattering values. Meanwhile, Lu, Chen, Wang, Liu, Li and Moran [22] listed the possible features of spatial, spectral, and height that can be generated separately from different sources of remote sensing (RS) data from active and passive sensors and this has been used in forest structural property modeling. Therefore, it is essential to explore the possible optimization of the multisensory fusion model for forest structure mapping [22].

1.2. Objectives

Previous research has indicated that combining remote sensing data, particularly those of passive and active sensors, can improve the accuracy of forest structural parameter mapping and modeling. More variables are being used in the modeling, which can increase the computation time and model complexity and affect the classifier accuracy [23,24]. In addition, the curse of dimensionality affects most machine learning algorithms, so that optimization is necessary in order to select the significant variables [25]. Therefore, this study explores the best combination between variable selection strategies

and machine learning algorithms for modeling using high-dimensionality input data. Several variable selection strategies, such as Boruta (BO), principal component analysis (PCA), simulated annealing (SA), and genetic algorithms (GAs), were combined with machine learning algorithms such as extra trees (ET), support vector machine (SVM) with radial basis function kernel, and a family of extreme gradient boosting (XGB) algorithms, including linear booster XGB (XGBlin), tree booster XGB (XGBtree), and regression tree booster XGB (XGBdart). The forest height was modeled using variables collected from the AISA (airborne imaging spectrometer for applications) hyperspectral data (479 bands) and airborne lidar elevation statistical metrics (37 variables).

2. Materials and Methods

2.1. Study Site

The study was conducted at Robson Creek, Far North Queensland Rainforest, Queensland, Australia, which is part of the Wet Tropics bioregion located along the coastline and adjacent to the Great Barrier Reef. The Robson Creek area is a tropical rainforest dominated by dense notophyll vine forests, which hold the highest biomass in Australia [26]. The study area is located in the northeast of Atherton, on the western slopes of the Lamb Range in Danbulla Park, Queensland, Australia. It has an average elevation of 700 m, with a 2300 mm rainfall intensity and an average temperature of 19 °C. The forest in this area is dominated by simple notophyll vine forests, with an average height of 26 to 40 m. A 500 × 500 m grid area with average biomass values of 418.5 mg/ha [27], located at 145.232° east and 17.12° south within Robson Creek, was used as the study site in this area (Figure 1).

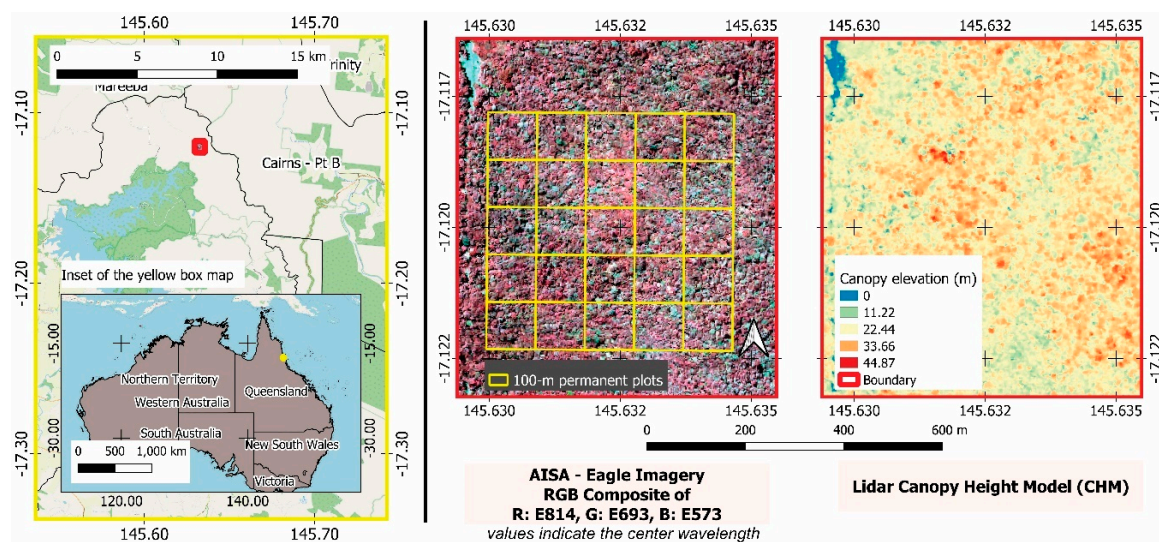


Figure 1. Study site area of Robson Creek, Far North Queensland, Australia, with the (right figure) showing the canopy height model from airborne lidar and the middle figure showing the AISA (Airborne Imaging Spectrometer for Applications) false color composites and the 100-m plots, while the (left figures) show the inset of the study area over broader scales.

2.2. Data

2.2.1. Field Datasets

Field data were acquired from the study of Bradford, Metcalfe, Ford, Liddell, Green and Mckeown [27], which conducted a census of the trees with diameter at breast height (DBH) > 10 cm within the Danbulla National Park in the northwest of Atherton, Far North Queensland, Australia. The Robson Creek study site is a 25-ha permanent plot consisting of 100 × 100 m plots and 20 × 20 m quadrat-plots. This area is managed by the Terrestrial Ecosystem Research Network and CSIRO Tropical Forest Research. From December 2009 to December 2012, this area was surveyed to collect

the vegetation species data and the tree structures configuration, including the height of the trees (m), particularly for trees with a DBH of above 10 cm.

In the field survey, approximately 23,000 individual trees, belonging to 209 species, were mapped and measured using a differential GPS system, Trimble Pro XRT, equipped with OmniSTAR DGPS signal with 2.3 m and 1.8 m standard deviation accuracy. The vegetation species in this area are dominated by *Litsea lefeana*, *Cardwellia sublimis*, *Findersia bourjotiana*, *Elaeocarpus largiflorens*, and *Alphitonia whitei*, which account for 28% of the species in this area. The vegetation height varied from 2 to 120 m, with an average height of 18.4 m and an average DBH of 20.9 cm (Table 1). On the field plots, there were no signs of man-made disturbance, including the signs for silvicultural treatments with the last selective loggings that were conducted between 1960 and 1969, as described in the study by Bradford, Metcalfe, Ford, Liddell, Green and Mckeown [27].

Table 1. Field inventory data statistics for the Robson Creek study site (Bradford et al. [27]).

Parameters	Statistical Analysis					
	Min	Mean	Max	Standard Deviation	5th Percentile	95th Percentile
Height (H) (m)	2.0	18.4	120.0	6.6	10	30
Diameter at Breast Height (DBH) (cm)	10.0	20.9	152.5	13.1	10.4	48.5
Aboveground Biomass (AGB) (mg/ha)	294.81	402.53	540.75	56.59	311.529	534.85
Number of Trees (count/ha)	715	935.2	1074	75.78	743.5	1065

2.2.2. Hyperspectral Data

The SPECIM AISA hyperspectral system consists of two hyperspectral sensors, Eagle and Hawk, mounted under the wing of the ARA's (Aircraft Research Association) ECO-Dimona research aircraft. These data were acquired on 13 and 14 September 2012. The Eagle sensor has 252 spectral bands, from 400.7 to 999.2 nm, with 2.2 to 2.45 nm of full-width half maximum (FWHM), while the Hawk sensor has 227 bands, from 993.1 to 2497.4 nm, with 6.22 to 6.32 nm of FWHM (Table 2), with a swath width of 296 pixels. Examples of the spectral responses of vegetation in the study areas from Eagle and Hawk sensors are presented in Figure 2. Images were acquired at an altitude of 500 m above ground level, resulting in 30 cm to 1 m spatial resolution, and the sensors were operated in a push-broom mode. The data were then atmospherically corrected to reflectance values using the FLAASH (Fast Line-of-Sight Atmospheric Analysis of Spectral Hypercubes) module in ENVI 4.8. The data were corrected for cross-track anomaly to remove limb or edge brightening effects, which change the value for pixels located across the flight track [28].

Table 2. Hyperspectral metrics variable acronyms with values representing the center of wavelength.

Sensors	Wavelength	Acronyms	Number of Bands	Full-Width Half Maximum (FWHM)
Eagle (E)	400.7–450 nm	E ₄₀₁ –E ₄₅₀	23	2.2 to 2.45 nm
	452.6–501 nm	E ₄₅₃ –E ₅₀₁	22	
	503.3–580.4 nm	E ₅₀ –E ₅₈₀	34	
	582.8–599.4 nm	E ₅₈₃ –E ₅₉₉	8	
	601.9–680.7 nm	E ₆₀₂ –E ₆₈₁	34	
	683–749.8 nm	E ₆₈₃ –E ₇₅₀	29	
	752.4–999.2 nm	E ₇₅₂ –E ₉₉₉	102	
Hawk	993.1–1396.7 nm	H ₉₉₃ –H ₁₃₉₇	61	6.22 to 6.32 nm
	1403–2497.37 nm	H ₁₄₀₃ –H ₂₄₉₇	166	
Total bands				479 bands

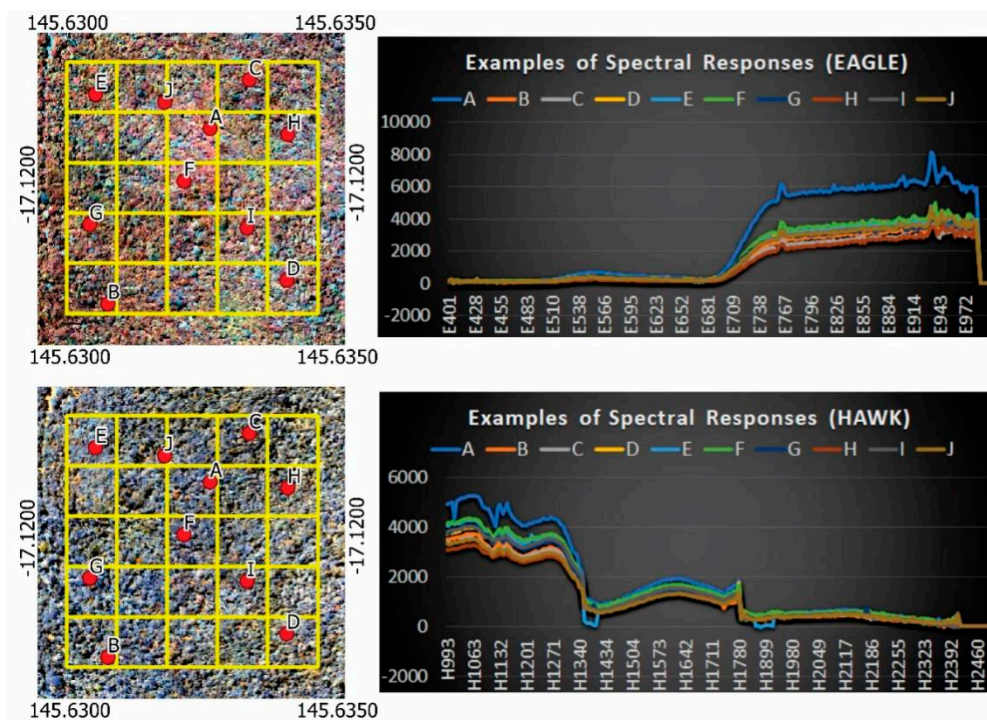


Figure 2. Examples of spectral responses from Eagle (top figure) and Hawk (bottom figure) sensors at 10 random points at the study area.

2.2.3. Lidar Metrics

Small-footprint discrete-return lidar data were used in this study. The data were generated from full-waveform lidar using a Riegl Q560 instrument mounted under the wing of ARA's ECO-Dimona aircraft and were taken at the same time as the hyperspectral measurements. The lidar sensor recorded at 300 m above ground level with a velocity of 40 m/s and created a 0.30 m point distribution along-track and across-track, with a diameter of <0.15 m for each pulse, in order to produce lidar point clouds. The lidar flight was closer to the ground to ensure a higher penetration rate and higher first/last return, useful for forest structural property assessment [29,30].

The lidar processing was started by generating the height normalized lidar. To create the height-normalized lidar, the raw lidar data were normalized by using the triangulated irregular network (TIN) interpolation surface by using the ground returns. Lidar sensor was able to record the multiple returns from a pulse that hits multiple objects, with the first return being the highest object, and the ground return can be identified from the high intensity of the last return when large lidar footprint was used [31], although not all last returns can be classified as ground returns. Another method to detect ground return can be identified from the lidar return, which only has a single return, that typically happened when the pulse hits a solid object or a surface. The process of height normalization was conducted in the R environment by using the "LidR" package [32]. A comparison of non-normalized lidar and height normalized lidar is presented in Figure 3.

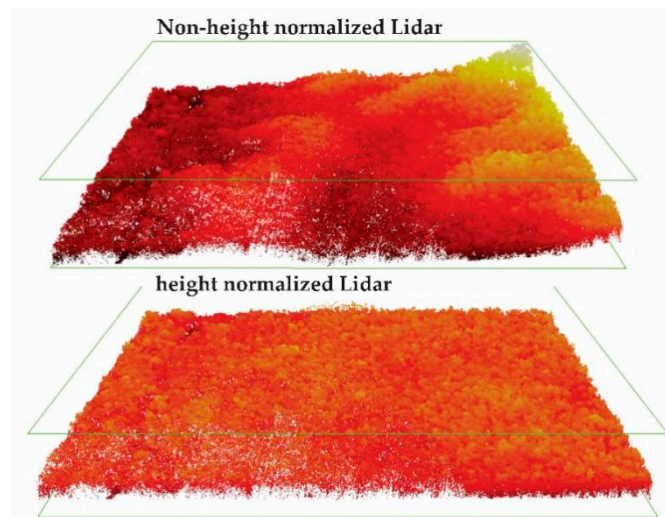


Figure 3. The differences between the non-height normalized lidar (**top figure**) and the height normalized lidar (**bottom figure**); color represents elevation.

To derive the lidar height statistical metrics, the “standard_z” function in the “LidR” package [32] was used to derive the standard lidar metrics calculated using the elevation values. This process generated 36 raster metrics, which represented the statistical properties of the point clouds within a $5 \times 5 \text{ m}^2$ bin. Additional canopy height metrics (CHMs) were generated using the pit-free algorithm developed by Khosravipour et al. [33] to account for the canopy height irregularities and remove pits from the final CHM data. Therefore, 37 lidar variables in total were used for the forest height modeling (Table 3).

Table 3. Lidar metrics variable acronyms.

Acronyms	Statistical Metric Type	Number of Variables
LD01–LD06	Maximum, mean, standard deviation, skewness, kurtosis, entropy	6
LD07–LD08	Height above mean and number of returns from the 2nd and above 2nd return, divided by the 1st return	2
LD09–LD27	Height percentile from 5% to 95% (5% interval)	11
LD28–LD36	Cumulative height based on nine breaks from minimum to maximum	9
CHM	Pit-free canopy height metrics	1
Total lidar statistical variables		37

3. Methodology

3.1. Feature Selection and Dimensionality Reduction

3.1.1. Random Forest Implementation in Boruta

Boruta was developed by Kursa and Rudnicki [34], designed as a wrapper around the random forest (RF) algorithm. The original RF algorithm can determine significant variables by considering the decrease in accuracy when permutation of the input variables is conducted. However, in BO, the important variables are determined by introducing the shadow variables generated by randomly shuffling the attribute values to be used in the RF classification alongside the original variables; then, an iterative search for a set of original variables that outperform the shadow variables is performed by referring to the Z-score as a measure for accuracy loss. The pseudocode for BO can be found in Paja et al. [35] study.

Boruta has been implemented in various types of machine learning modeling for variable selection, including gully erosion modeling [36], peat thickness modeling [37], and digital soil mapping and modeling [38,39]. It has also been applied for hyperspectral band selection [36] to reduce the number

of bands used in the analysis. In this study, Boruta was run with the number of iterations and number of trees set as 200 and 1000, respectively.

3.1.2. Principal Component Analysis

PCA is a classic method for compressing the data dimensionality by rotating the data according to the direction of the majority of the data distribution on n -dimensional spaces; the first principal component (PC) is created, while the second PC is directed to be perpendicular to the first, and so on, to create the PC bands with numbers equal to the number of input bands [40]. The PC bands are highly uncorrelated with each other and hold the most variance contained in the original data, especially for the first PC bands. Thus, PCA analysis is beneficial for reducing the dimensionality from the highly correlated data while retaining the majority of information stored at the original data. The analysis can also be extended to reduce the noises in the original data by transforming/rotating the PCA bands back to the original data. PCA has been applied mainly for classification using hyperspectral data, and it resulted in unstable accuracy when high-dimensionality data with high correlation between variables were used; this decrease in accuracy when using high-dimensionality data is commonly referred to as the “Hughes” phenomenon [41].

In our study, PCA was applied to several datasets to measure the performance of the PC bands when applied in modeling using machine learning algorithms. Here, PCA was applied to the hyperspectral data from each sensor (Eagle and Hawk sensors) and the lidar statistical metrics. The first ten PC bands from each electromagnetic spectrum and lidar metric were extracted. Therefore, the total number of PC variables to be used as the input for the machine learning algorithms was 30. The selection of a fixed number of PC bands from each dataset was performed due to the difficulty of automatically assessing the meaningful components of the PC bands without visual assessment [42]. The processing was conducted by using the “rasterPCA” function in the RSToolbox package in R [43]. The output was standardized (scaled and centered) so that determining the variable importance from the PC bands could be identified from the PC’s factor loadings.

3.1.3. Simulated Annealing

SA is a non-deterministic feature selection strategy that uses a probabilistic function [44]. This variable selection strategy, which is used for multivariate optimization, mimics the statistical mechanics process of achieving thermal equilibrium in the solidification of metal or glass [45]. The probability function in SA is constructed from a search process that allows trial subsets that worsen the objective function by controlling the range of temperature variables and the cooling scheme, mimicking the physical process of annealing [46]. The pseudocode for SA can be found in Gheyas and Smith [47] study. In this study, SA was implemented using an RF wrapper in the Caret package [48]. A three-fold cross-validation was run, with 1000 iterations to generate the variable selection and R^2 with the fitness measure. We also set the improvement values with 50 iterations, meaning that if there was no improvement at the 50th iteration after the best combination was found, the algorithm would reset the search grid using the latest best combination. This was conducted in order to avoid becoming stuck in local maxima.

3.1.4. Genetic Algorithm

GA is also a non-deterministic strategy for variable selection. This strategy mimics the natural evolution stages to find the best combination of variables (or chromosome) through the following processes: (1) selection based on the fitness criterion to generate parent candidate from the random initial combination, (2) combination of genes from two parents, and (3) the mutation of genes within the parents’ chromosome (or the variable combination) by randomly flipping the gene [49,50]. The pseudocode for GA can be found in the studies of Chau et al. [51] and Gheyas and Smith [47]. These three processes are iteratively conducted until the stopping criterion is met; that is, no significant improvement in the fitness level is found or the number of designated iterations has been achieved.

In this study, similar to SA, the GA was implemented using an RF wrapper in Caret, with a three-fold cross-validation run for 1000 iterations and the coefficient of determination (R^2) as the fitness measure. In addition, the same setting for improvement as used in SA was used to find the local best combination while avoiding becoming trapped in the local minimum/maxima.

3.2. Machine Learning Algorithm

3.2.1. Multivariate Adaptive Regression Spline

Multivariate adaptive regression spline (MARS) is a method developed by Friedman [52] and explained in detail by Friedman and Roosen [53]. It involves using a spline basis function to construct an accurate non-linear model from high-dimensionality data. Unlike the standard spline function, the non-linear model was constructed to resemble the general approximation of the data distribution by dividing the predictors into several linear functions based on the detected joints or knots. The pseudocode for the MARS algorithm can be found in the study by Alkaim and Al-Janabi [54]. MARS has been used in various types of multivariate modeling and has been reported to outperform other modeling algorithms, such as the generalized additive model for species distribution [55], Cubist for AGB modeling [56], maximum likelihood, and parallelepiped for land cover classification [57]; however, the accuracy of MARS may be dependent on the pixel size of the data used in the modeling [58]. This study used the MARS implementation in the “earth” package, called the “Caret” train function [59].

3.2.2. Extra Trees

RF, developed by Breiman [60], is a bagging ensemble method that has gained popularity in the study of remote sensing for regression and classification in recent years. This is due to the low number of parameters used in the analysis and its comparable performance with other machine learning methods such as SVM [61]. In addition, the algorithm is also less prone to overfitting, due to the parallel process in creating the ensemble of classifiers, and computationally less intensive [62]. RF combines multiple tree classifiers from the random subset of data and determines the final value from the majority vote (classification) or average (regression) [63].

This study used the extension of the original RF implementation, so-called the extremely randomized trees or extratrees (ET), in the “extraTrees” package [64]. This algorithm was developed by Geurts et al. [65], who modified the strategy for splitting the nodes by using a random subset of the best feature and the corresponding attributes and trained each tree classifier using the whole dataset instead of the random subset of data in the original RF. Geurts, Ernst and Wehenkel [65] study also listed the details and pseudocode for the ET algorithm. Therefore, the ET model will be less likely to overfit, and the randomization in the node splitting process will reduce the variance better than the original RF and minimize bias [65]. In addition, the algorithm also has also been reported to be faster in training and prediction processes, with accuracy comparable to that of the original bagging RF [66–68]. Moreover, ET used for land cover classification in remote sensing has been reported to provide better accuracy than the original RF and SVM [69].

3.2.3. Extreme Gradient Boosting

Extreme gradient boosting (XGB) is also an ensemble method based on the boosting strategy rather than the bagging strategy in RF. The boosting strategy works by assessing the loss function of the previously constructed ensemble and assigning a weight to the previous error of classification when constructing and refining a new set of classifiers [70]. This study used the linear, tree, and DART implementation of XGB in the “xgboost” package [71], which is based on the gradient boosting algorithm by Friedman [72]. The details, steps, and pseudocodes of the XGB algorithms were presented at Chen and Guestrin [73]. The XGB algorithm is capable of performing classification and regression tasks in several applications, including remote sensing. The growing popularity of this method is

due to its better or comparable accuracy and stability relative to other algorithms, such as RF and SVM for land cover classification in urban areas; SVM for modeling global solar radiation; SVM, RF, and Gaussian process regression for mangrove AGB modeling, and ET, SVM, and RF for high-resolution vegetation mapping [74–77].

3.2.4. Support Vector Regression

SVM, developed by Cortes and Vapnik [78], is a machine learning algorithm that transforms input data into a high-dimensional space to enable optimal linear separation between the different clusters of data. With the additional settings of soft margin and kernel, SVMs separate data by a linear surface model or hyperplane and project the data into a higher-dimensional space to accommodate the linear separation of the class [79]. The details and pseudocodes for SVM can be found in the report by Alloghani et al. [80]. This method can be used for classification and regression tasks and, when used for the latter, the method is also known as support vector regression (SVR) [81]. Both SVR and SVM have been widely applied for many remote sensing studies, particularly to perform modeling from high-dimensionality data [82]. In addition, SVM and SVR are easy to use, robust, and less prone to noise from the input data [83]. In this study, we employed the SVR with the radial basis function kernel in the R package of “kernlab” [84]. The radial basis function kernel is the most commonly used function in the modeling of vegetation biophysical properties [77].

3.3. Model Construction

In this study, each feature reduction and selection analysis method described above was combined with machine learning. Since there were four feature reduction and selection analysis methods and six machine learning algorithms, 24 combinations were tested in this study. In addition, three scenarios of input data were used: hyperspectral bands (HS), lidar CHM and height metrics (LD), and hyperspectral and lidar data (ALL). Therefore, 72 combinations of input data, feature reduction and selection analysis methods, and machine learning algorithms were tested in this study.

In the model construction, the input data were split into training and testing data in 70:30 proportion. Each model was trained using five-fold cross-validation, with R^2 as the model performance metric. Hyperparameter optimization was conducted via a random search with 50 iterations to find the best parameter for each machine learning algorithm. The random search can find the best optimization for the tuning parameter with a relatively faster computation time [85].

3.4. Model Validation

Error estimation of the models was performed to determine which models had better prediction levels. Root mean squared error (RMSE) analysis was used to determine the magnitude of the margin of error between predicted and actual field plot data. The RMSE value was calculated by taking the square root from the quadrat difference of the predicted P_i and the field data O_i , as shown in Equation (1). Additional metrics of bias and normalized RMSE (nRMSE) were also added to evaluate the error of the output models (Equations (2) and (3)), where bias was used to determine whether the output model is over- or underestimating the actual values [86] and nRMSE is a scaled RMSE value (can be represented by percentage) calculated by using the maximum and minimum values of population, where smaller values indicate the better fit of the model to the actual values.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (P_i - O_i)^2}{n}} \quad (1)$$

$$nRMSE (\%) = \frac{RMSE}{(Max(O_i) - Min(O_i))} \times 100 \quad (2)$$

$$Bias = \frac{\sum_{i=1}^n (P_i - O_i)}{\sum_{i=1}^n O_i} \quad (3)$$

where n is the number of observations; O_i is the observed/true value, which is the field data measurement; P_i is the model predicted regarding forest structure attributes from the remote sensing model. The RMSE, nRMSE, and bias were calculated for the output models by using metrics [87] and the hydroGOF package [88]. The error assessment was conducted by using 30% of the total field data (36 points), while the remaining 70% of the data (85 points) was used to construct the model.

This study can be summarized into the following workflow (Figure 4).

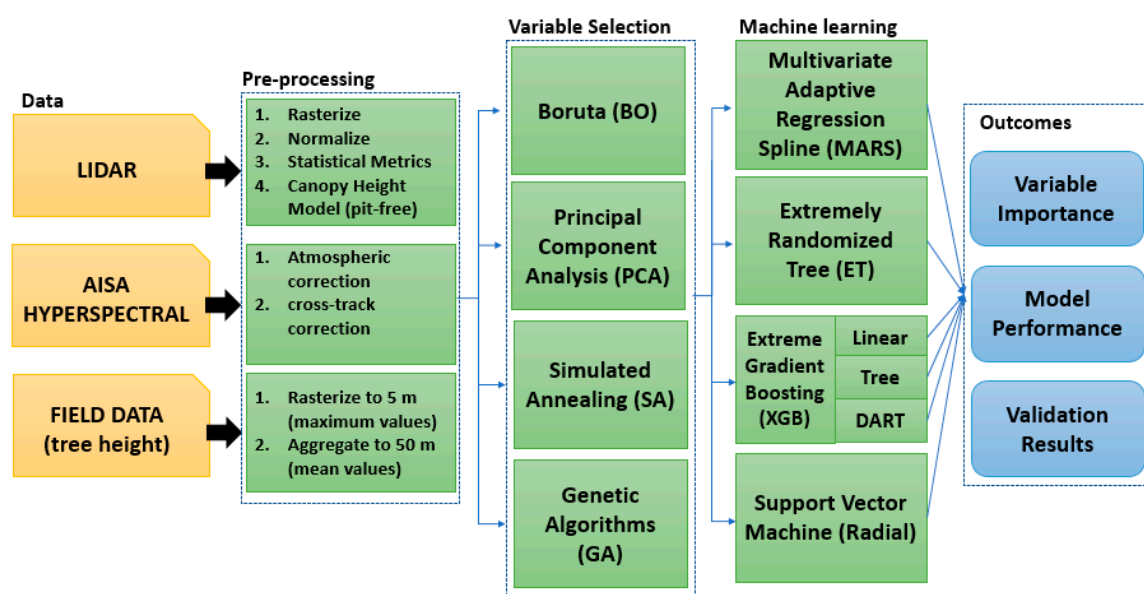


Figure 4. Research workflow.

4. Results

4.1. Feature Selection and Dimensionality Reduction

Feature selection analyses and dimensionality reduction were conducted using Boruta (BO), Simulated Annealing (SA), Genetic Algorithm (GA), and Principal Component Analysis (PCA). These methods were used to select the significant variables from AISA hyperspectral bands (479 variables), lidar statistical metrics (37 variables), and the combination of both datasets (516 variables) for modeling the forest height data. Each of the feature selection and dimensionality reduction frameworks was applied to three datasets: hyperspectral bands, lidar height metrics, and a combination of hyperspectral and lidar metrics. The feature selection identified different numbers of variables, ranging from 4 to 133 variables (Table 4), which were deemed significant for modeling forest height.

Table 4. Number of significant variables identified from the feature selection analysis.

Variables	All Data			Hyperspectral			Lidar		
	BO	GA	SA	BO	GA	SA	BO	GA	SA
Eagle	4	85	48	2	66	66			
Hawk	1	44	46	2	42	62			
Lidar	24	4	10				27	9	14
Total	29	133	104	4	108	128	27	9	14
% from initial datasets	5.62	25.78	20.16	0.84	22.55	26.72	72.97	24.32	37.84

From Table 4, GA and SA generally identified more significant variables than BO; they identified a range of 20.16% to 37.84% of the total number of variables, while BO identified a range of 0.84% to 72.97%. The higher identified significant variables in BO were obtained when it was applied to select the significant lidar variables. Boruta identified the significant variables by measuring the number

of times the original variables performed better than the randomly generated shadow variables and considered the variable to be significant when the importance frequency for the variable reached >95% quantile of confidence. Therefore, it seemed that most of the lidar variables performed better than the shadow variables, hence more detected significant variables. Details of the selected variables from BO can be seen in Figure 5.

Figure 5 shows the ranking of importance considering each selected variable, where a higher value indicates better performance. Based on the feature selection results from BO, lidar showed overall better performance than the hyperspectral bands as more lidar variables were detected as significant. The detection of fewer variables from hyperspectral metrics indicates the poor performance of most of the original spectral bands against the randomly generated “shadow” variables. In addition, BO also identified CHM metrics as the most significant variables, both for the combined variables and lidar metrics, while the wavelength of 2410 nm in the Hawk (H_{2410}) sensor was considered the most significant variable among the hyperspectral sensors (both Hawk and Eagle). However, in terms of the number of identified variables from hyperspectral bands, Eagle sensors performed better than Hawk, with more detected variables coming from the Eagle sensors.

Unfortunately, there was no ranking for variables generated from GA and SA analyses since the outputs of the identified variables were considered as a combination that generated the best performance, measured using the fitness function of RMSE and R^2 after 1000 iterations. However, the identified combination of the significant variables differed considerably from the BO results. For instance, GA and SA identified more hyperspectral bands than the lidar metrics, although most of the identified variables were from the Eagle sensors, similar to the results of BO analysis. Moreover, CHMs were not identified as significant variables in GA and SA, contrary to the BO results.

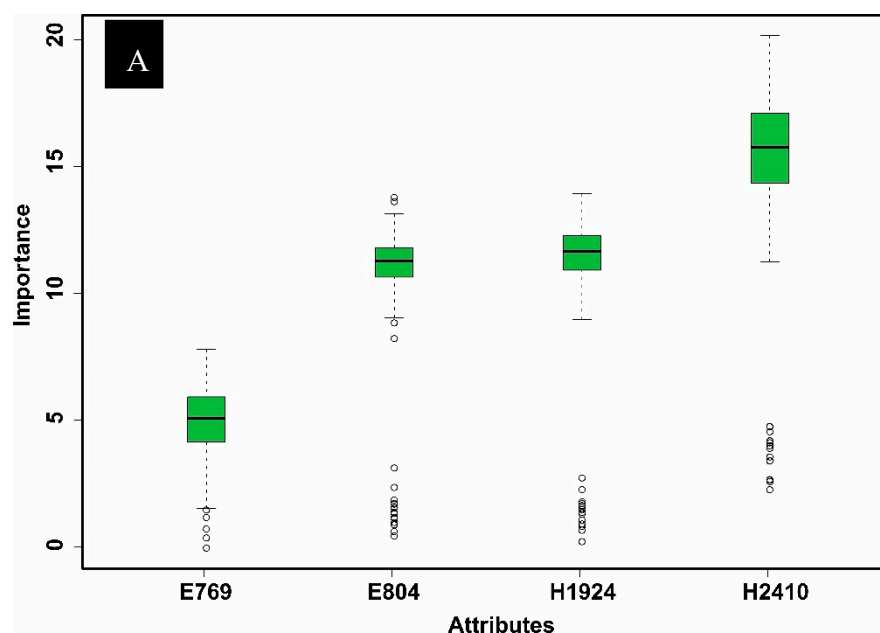


Figure 5. Cont.

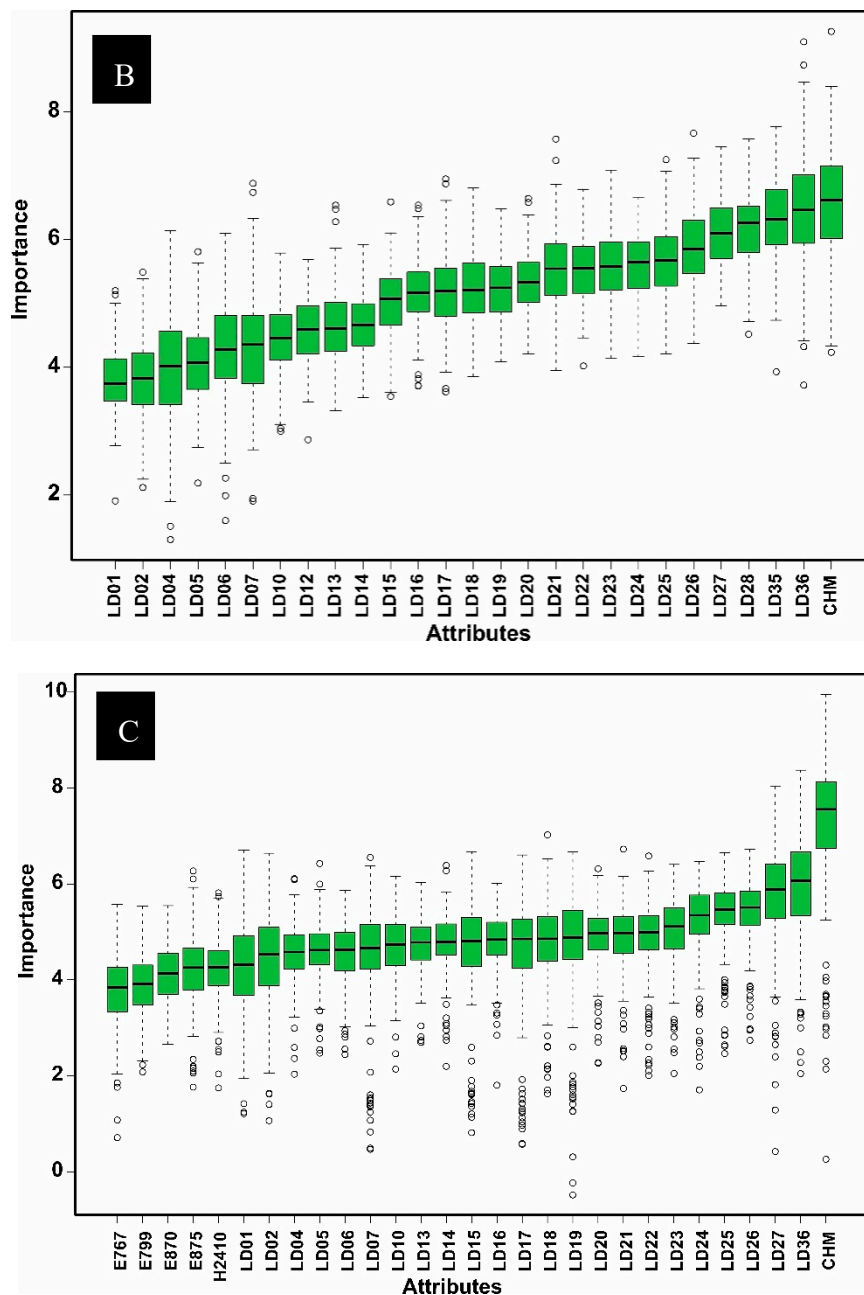


Figure 5. Selected variables by using Boruta. (A) Hyperspectral datasets (E₇₆₉, E₈₀₄, H₁₉₂₄, and H₂₄₁₀), (B) lidar height metrics (LD1, LD2, LD4–LD7, LD10, LD12–LD28, LD35–LD36, and CHM), and (C) hyperspectral and lidar metrics (E₇₆₇, E₇₉₉, E₈₇₀, E₈₇₅, H₂₄₁₀, LD1, LD2, LD4–LD7, LD10, LD13–LD27, and LD36), based on the number of variables with better performance against the shadow variables (y -axis; importance).

Different results were presented using PCA analysis since PCA did not select the best variables but rather compressed the data into a PC and retained most of the information or variance in the data. Here, we used the first 10 PCs from Eagle, Hawk, and lidar height metrics. The first 10 components of Eagle, Hawk, and lidar data stored a cumulative variance of 99.60%, 99.46%, and 96.11%, respectively. Most of the variance was stored in the first to third PCs, which could make up to 81% to 90% of the total variance. To understand the contribution of each variable to the PC, the loading factors of the first 10 components are presented in Figure 6.

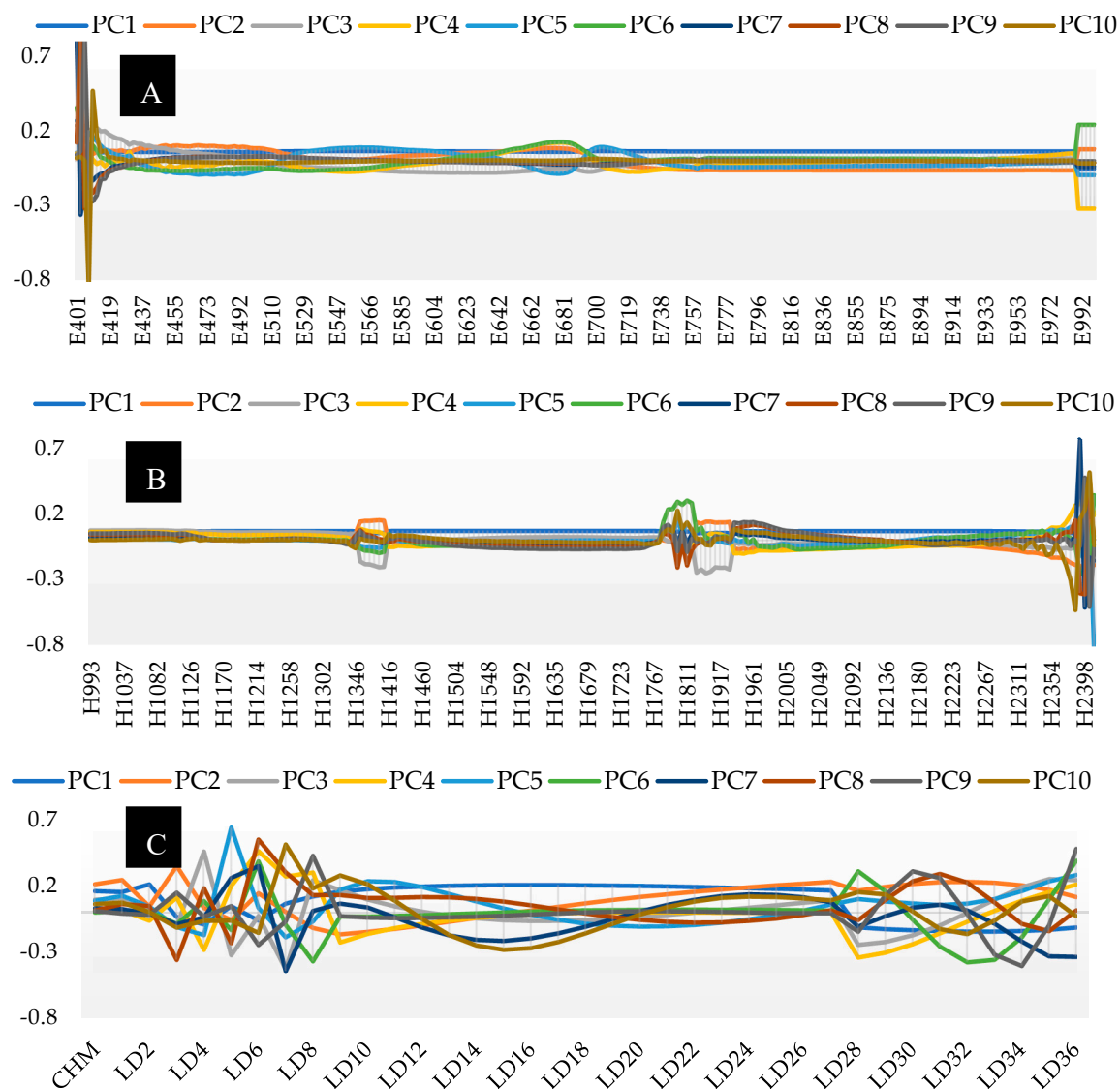


Figure 6. Factor loading (standardized) from each dataset, showing the linear contribution of each input variable to the PC bands: (A) Eagle sensor, (B) Hawk sensor, and (C) Lidar metrics.

The factor loading analysis showed that some hyperspectral bands and lidar height metrics had more contributions than other variables within each dataset. In the Eagle dataset, higher contributions can be found from the blue (E₄₀₁ to E₄₁₉) and near-infrared spectra (E₉₇₂ to E₉₉₉), while Hawk sensors showed high factor loadings at short infrared spectra (H₁₇₆₇ to H₁₉₆₁ and H₂₃₅₄ to H₂₄₁₀). In lidar height metrics, the variables from LD1 to LD9 and LD27 to LD36 showed high factor loadings, although, contrary to results from BO, the CHM metric did not show a high contribution to the PCs.

The results of the variable selection showed diverse combinations of the significant variables, although some agreement of the significant variables was also found. All the detected variables were then employed as the input variables for the machine learning modeling so that the accuracy of the modeling while using the set of identified variables could be determined.

4.2. Model Performance

The performance of the models generated using the machine learning algorithms in the Caret package was assessed by using the R^2 values to determine the best possible parameter combination. In this study, the R^2 and RMSE values using training and testing data were calculated and presented to evaluate the overall models' performance.

Figure 7 shows the performance of models of the forest heights generated using 85 datasets (in a plot size of $50 \times 50 \text{ m}^2$) from the selected hyperspectral metrics. In general, the model generated using the hyperspectral metrics had an R^2 of 0.127 to 0.432 and an RMSE of 1.78 m to 2.961 m. The best model performance was obtained using the dataset from the combination of GA and XGBtree ($R^2 = 0.432$ and $\text{RMSE} = 1.78 \text{ m}$) and GA and XGBdart ($R^2 = 0.424$ and $\text{RMSE} = 1.79 \text{ m}$); however, the validation process using the testing data yielded lower R^2 but stronger RMSE values, i.e., GA-XGBtree validation performances using testing data are R^2 of 0.15, RMSE of 1.37 m, nRMSE = 19.60%, and bias of -0.277 , suggesting a lower model fit but with lower error and an overall underestimation when being validated by using testing data. Some generated models resulted in overfitting, with lower R^2 or increased RMSE, when tested using the testing data ($n = 36$), such as in the combination of PCA with MARS (increased error of RMSE from 1.956 m (model) to 2.54 m (validation) with nRMSE of 26.7 % and bias of 0.763 (overestimation)) and SVR (increased error from 1.942 m (model) to 2.07 m (validation) with 18.7 % nRMSE and 0.252 (overestimation)), suggesting the increased error level from the generated model when tested with data not used in the input process.

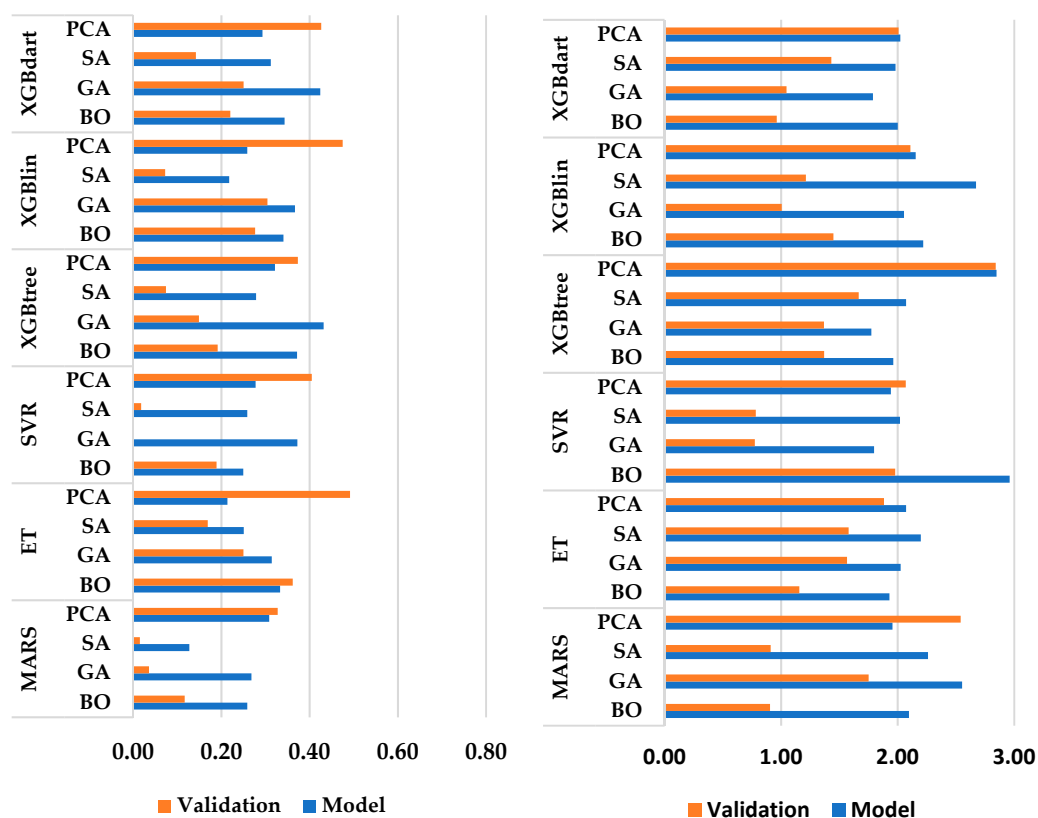


Figure 7. Machine learning model performance of R^2 (left) and RMSE (right) using input dataset from selected hyperspectral band.

In addition, although the best model performance was acquired by GA-XGBtree, the best validation performance by using hyperspectral data alone was displayed by using the combination of BO and ET, with an R^2 of 0.36, RMSE of 1.16 m, nRMSE of 16.6 %, and bias of 0.086 m. The complete table of nRMSE and bias values can be found in the Supplementary Materials.

By comparing Figures 7 and 8 above, some notable improvements can be found in the models' performance using selected lidar metrics. The models generated by selected lidar metrics yielded performances ranging from 0.27 to 0.50 (R^2) and 1.82 m to 3.62 m (RMSE). The best performance by selected lidar metrics was generated from the combination of PCA data and SVR ($R^2 = 0.50$ and $\text{RMSE} = 2.2 \text{ m}$), while the second-best model was generated from the GA and XGBtree combination ($R^2 = 0.45$ and $\text{RMSE} = 1.87 \text{ m}$). Although the PCA and SVR combination yielded the best performance

when lidar metrics were used, the case of overfitting with PCA data can still be seen, as indicated by the increased RMSE and lower R^2 when the model was tested with testing data ($R^2 = 0.31$, $RMSE = 2.2$ m, $nRMSE = 20.70\%$, $bias = 0.012$, $n = 36$). In addition, the GA and XGBtree combination also generated a lower R^2 value ($0.16 < 0.44$) in the validation process but a higher validation RMSE value than the RMSE value in the model performance (0.86 m > 1.86 m), with $nRMSE$ of 19% and bias of -0.0461 , suggesting that the model generated values were slightly lower than the actual field data. The best model performance by using lidar only data, however, came from the combination of SA-ET, with an R^2 of 0.33 , $RMSE$ of 0.94 , $nRMSE$ of 17.3% , and bias of -0.004 m, and GA-MARS, with an R^2 of 0.38 , $RMSE$ of 1.09 , $nRMSE$ of 16.7% , and bias of -0.4 m, showing the general underestimation of the GA-MARS model.

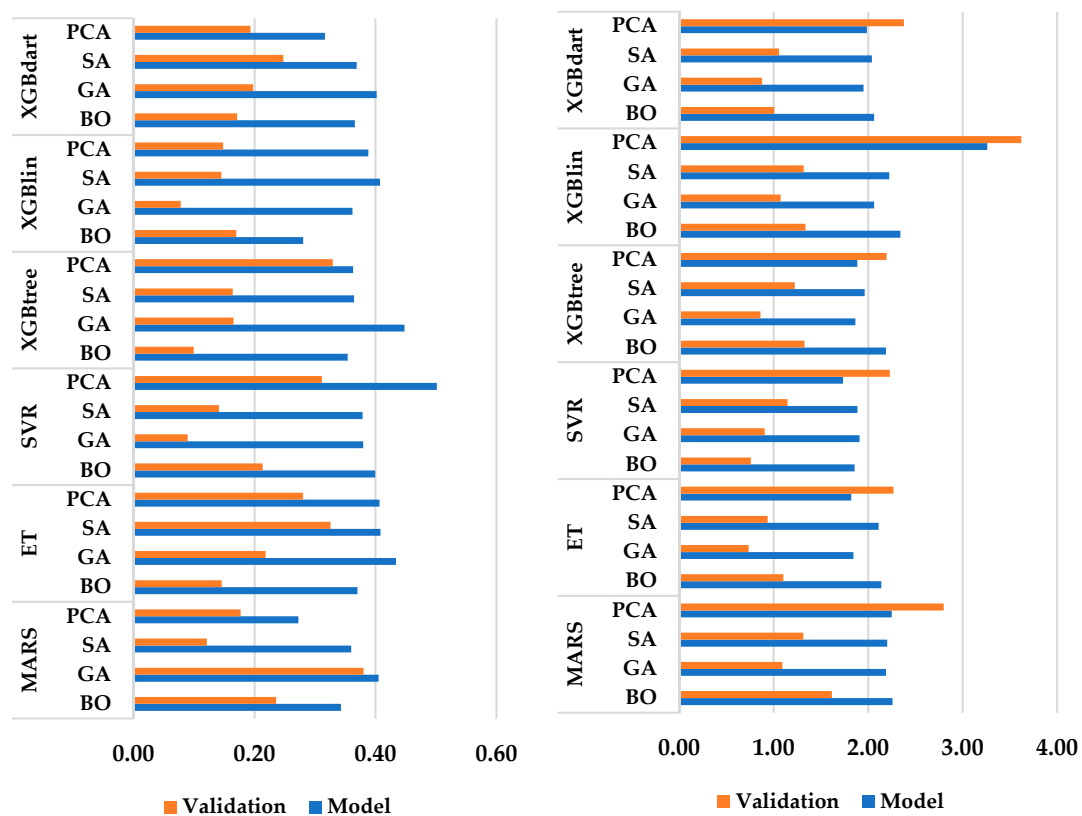


Figure 8. Machine learning model performance of R^2 (left) and RMSE (right) using input dataset from selected lidar metrics.

Additional improvements and the best performance among other models using lidar and hyperspectral only as the input data in the model could be obtained using the data combination of lidar and hyperspectral metrics, as summarized in Figure 9. The R^2 values ranged from 0.25 to 0.53 and the RMSE from 0.47 m to 2.3 m. Overall, the BO variable selection performed better at this section of the analysis, whereby the best and second-best performances were acquired by combining the BO-selected variables with XGBdart ($R^2 = 0.53$ and $RMSE = 1.7$ m) and SVR ($R^2 = 0.51$ and $RMSE = 1.8$ m), respectively. Both methods yielded lower validation R^2 ($0.43 < 0.53$ in XGBdart) but higher validation RMSE values (1.06 m > 1.7 m in XGBdart), with an $nRMSE$ of 15.8% , which was the lowest $nRMSE$ value among other models, and a bias of 0.046 , suggesting slight overestimation of values from BO-XGBdart model. Complete bias and $nRMSE$ values can be found in the Supplementary Materials.

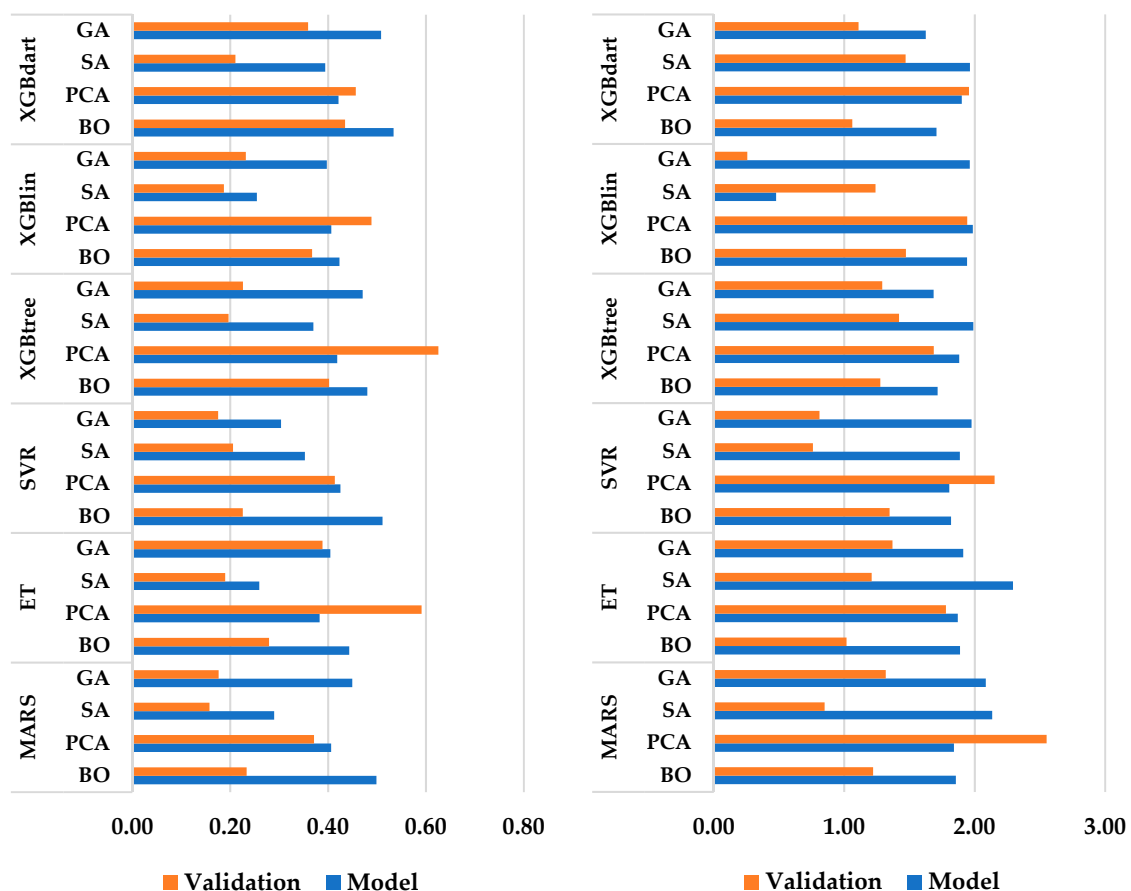


Figure 9. Machine learning model performance of R^2 (left) and RMSE (right) using input dataset from selected combinations of lidar and hyperspectral data.

5. Discussion

5.1. Results Overview

This study explored the combination of different input datasets, variable selection methods, and machine learning algorithms to model the forest heights at Robson Creek, Far North Queensland Rainforest, Queensland, Australia, using high-spatial-resolution and high-dimensionality airborne hyperspectral and lidar data. Overall, the performance measured in the analyses can be summarized under three headings: the performance of different input datasets, different variable selection methods, and different machine learning algorithms.

A summary of the performance of the input datasets can be seen in the boxplots in Figure 10. The model could be improved by combining lidar and hyperspectral data to obtain an average model performance of 0.41 (R^2) and 1.9 m (RMSE), compared with when hyperspectral or lidar metrics were solely used. However, some models generated using lidar metrics could produce similar or better performance compared with the combination/all datasets. Among the datasets, the lidar metrics yielded the best average RMSE values (Figure 10D). In addition, the improvement detected when incorporating hyperspectral data for modeling forest height was not statistically significant, as some of the models using lidar data could generate almost similar or better performance with a more efficient computation time compared with models using the combination dataset; this finding is similar to that of Swatantran et al. [89], who used lidar and AVIRIS hyperspectral data to model forest biomass.

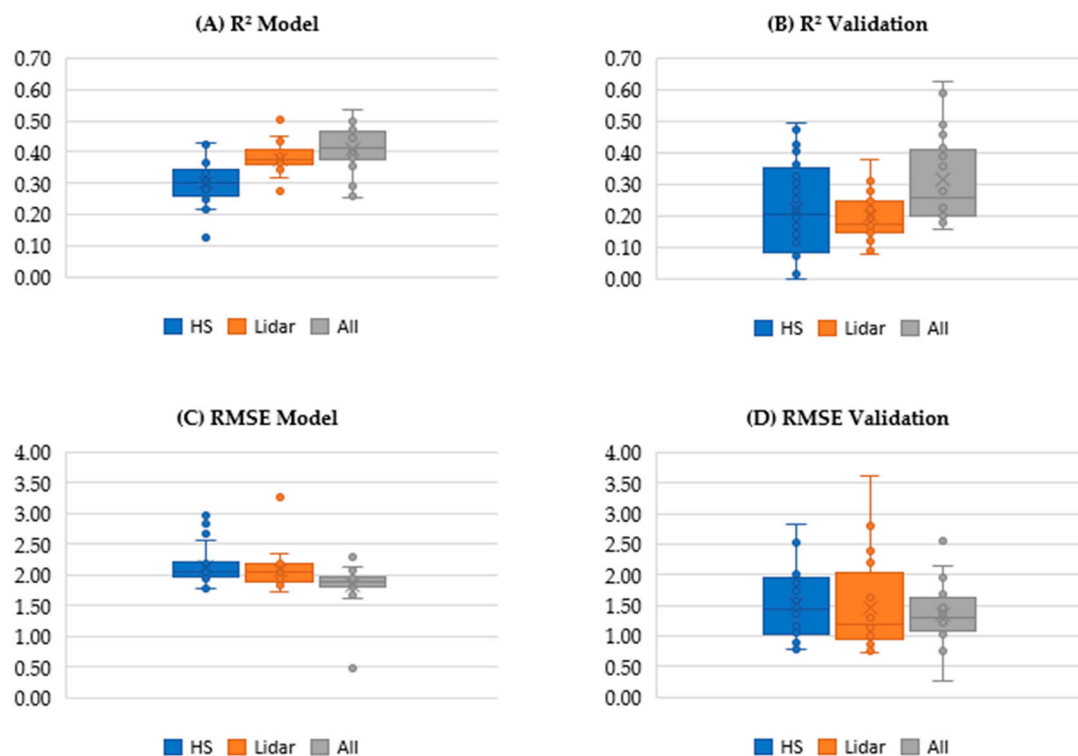


Figure 10. Performance of different input datasets when employed for modeling forest heights.

The better performance of lidar over hyperspectral is expected since the lidar measurement corresponds directly to the height structure of the vegetation as compared to the spectral values presented in the hyperspectral bands. Other studies have indicated the ability of lidar data to model tree height (especially tall top layer canopy) with high accuracy [90]. The maps of the best validation performance models using lidar (SA-ET), hyperspectral (BO-ET), and the combined datasets (BO-XGBdart) can be found in Figure 11. From the maps, it can be seen that hyperspectral-sensor was not able to portray the taller and shorter tree areas, as depicted on the lidar and lidar-hyperspectral combination based maps (highlighted with green circles in Figure 11). Figure 11 also showed the lower capability of hyperspectral bands alone when being used to model the vegetation structure in the study area. Unlike lidar sensors, hyperspectral sensors were rarely used for modeling vegetation structural properties, especially height, without the combination of other datasets. There was one study by Cho et al. [91] which showed the capability of this sensor for modeling DBH and tree densities better than height, which will be useful for biomass estimation, as demonstrated by the study by Laurin et al. [92]. Nevertheless, hyperspectral sensors can still be used to map different vegetation species to complete the 3D canopy information from lidar [25,93,94]. However, it should be noted that there was a time difference in the field data collection for some plots taken at the earlier census campaign (2009) and the flight observation (2012), which may affect the accuracy of models and the generated maps at some locations in the study area.

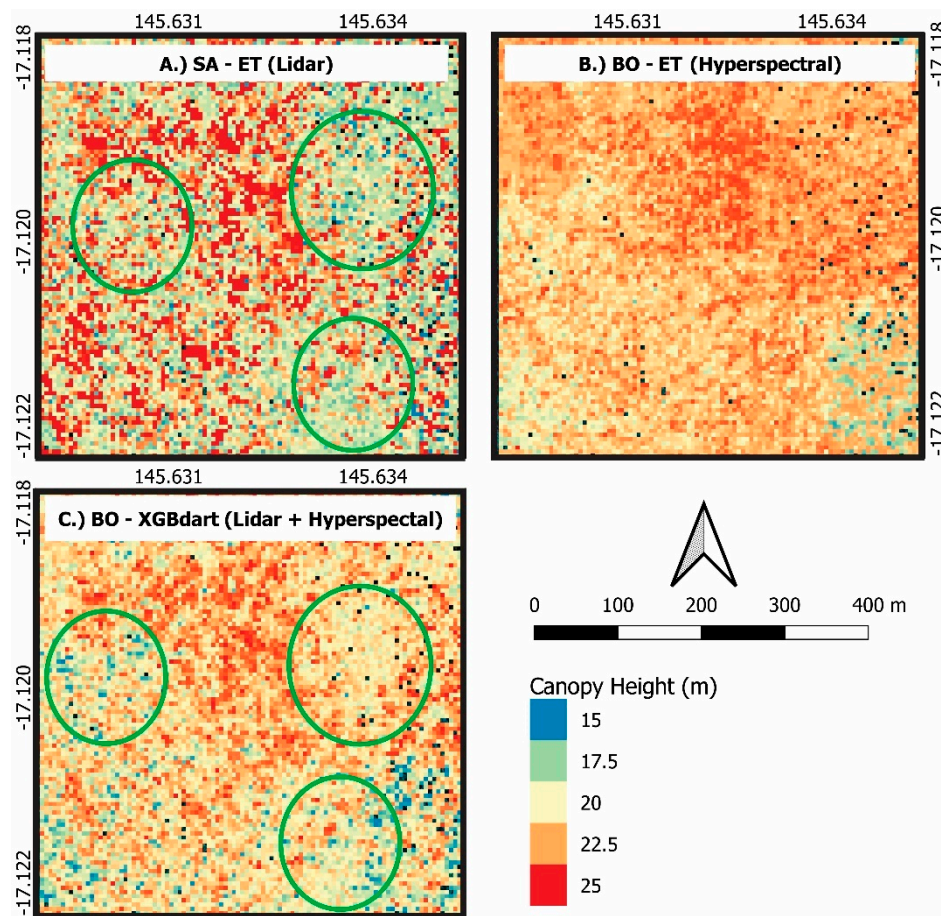


Figure 11. The height maps produced by using the best model found by using different datasets. (A) Lidar metrics only (SA-ET), (B) hyperspectral bands only (BO-ET), and (C) combination of lidar and hyperspectral (BO-XGBdart). Green circles showed the regions of lower tree height which went undetected by the hyperspectral model.

The important spectrum corresponding to the forest structural properties identified in Cho, Skidmore and Sobhan [91] study was similar to the important spectra identified in our study, especially the result from BO. Cho, Skidmore and Sobhan [91] study identified a red-edge spectrum (756 to 820 nm), near infrared spectrum (1172 to 1301 nm), and shortwave infrared spectrum (SWIR) (1953 to 1972 nm and 2221 to 2420 nm), whereas BO analysis in our study found the important spectra of 769 nm, 804 nm, 1924 nm, and 2410 nm as the important hyperspectral bands for modeling forest height.

A summary of different variable selection methods is presented in Figure 12. The average values displayed in the figure indicate that there were no significant differences in the output models generated by the input variables from different variable selection methods. However, BO yielded models with R^2 of above 0.4 (Figure 12A). Besides BO, GA could select variables that can be used to produce models with lower RMSE (Figure 12C,D). However, GA (and SA) selected more variables (>100 variables) than BO, especially when the input was high-dimensionality data, which subsequently increased the computation time in the modeling steps, with no significant improvement in modeling performance. As suggested by Latifi, Fassnacht and Koch [25], a model with 9 to 12 predictors is enough to generate an accurate model for forest structure, which will also reduce the computation time. In addition, the computation time from conducting the GA in the Caret package was notably longer than the computation time in BO or SA, a problem which has also been reported in other studies [95].

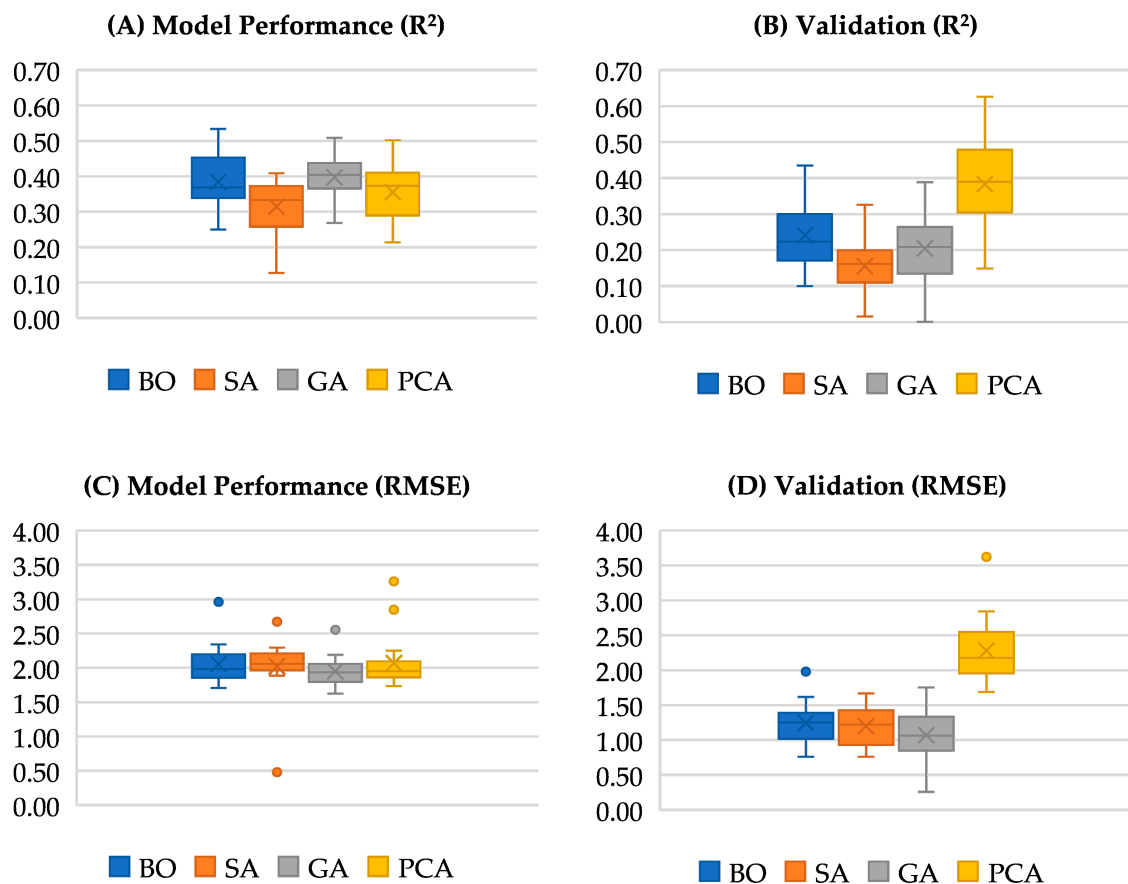


Figure 12. Performance of variables selected from different variable selection algorithms when used for modeling forest height.

The next part of this discussion section discusses the performance of different machine learning methods. Figure 13 displays a summary of the modeling and validation performance of the 72 generated models. Although the differences were not significant, the three methods generated slightly better models: ET, XGBtree, and SVR. Moreover, XGBtree, in particular, generated a relatively stable result, as displayed by the model and validation statistics, which may be due to the normalization of the loss function in XGBtree, which can reduce the likelihood of overfitting [96]. The better performance of the tree-based model of machine learning in our study is similar to the results obtained by Kattenborn et al. [97] when comparing random forests with generalized boosting regression model (GBM), generalized additive model (GAM), and boosted GAM for modeling forest biomass from the combination of Hyperion, Worldview-2, and Tandem-X.

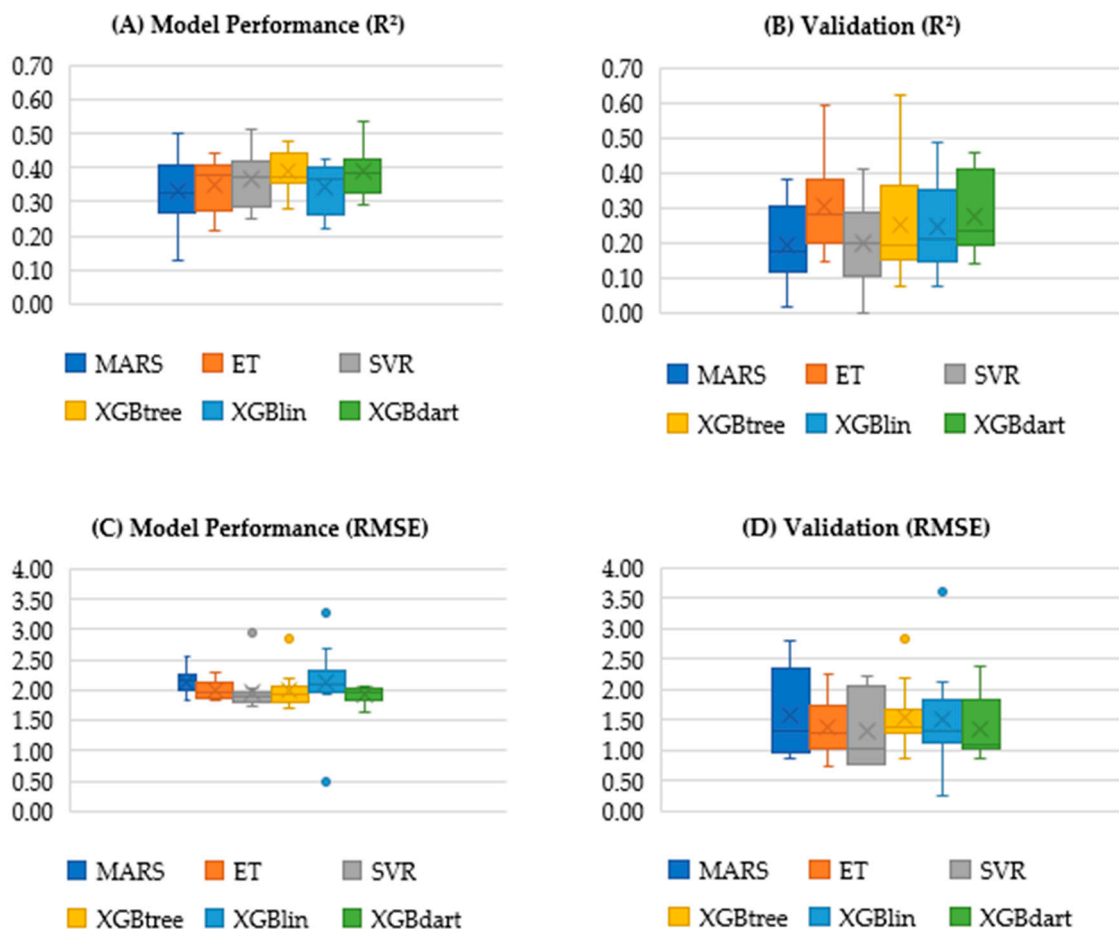


Figure 13. Machine learning model and validation performance using different input datasets.

5.2. Future Improvement

Our study reviewed the combination of various feature selection and extraction methods and machine learning algorithms for forest structural property modeling by using high spatial and spectral dimension remote sensing data of airborne lidar and hyperspectral sensors. However, the rapid development of the feature selection methods still leaves room for improvement in this study. For instance, the exploration of various nature-inspired optimization algorithms besides GA, such as particle swarm optimization (PSO) and other variations with the ability to solve the non-deterministic polynomial computational problem [98], to be used with high-dimensionality data is needed. A study by Hamedianfar et al. [99] has reported the superiority of combining PSO and XGB for land use land cover classification.

In addition, our study gave an overview of the performance of different combinations of feature selection and machine learning methods for forest structural assessment. The existence of the recent satellite-based hyperspectral sensors such as *PRecursore IperSpettrale della Missione Applicativa* (PRISMA) [100–102] and satellite-based laser altimetry of NASA's Global Ecosystem Dynamics Investigation (GEDI) [103] will trigger more studies by utilizing the combination of multi-sensor data. GEDI data, for instance, has been combined with radar backscatter intensity for forest structural assessment [104]. With the growing satellite data from optical (multispectral and hyperspectral) passive sensors and active sensors from radar and lidar, optimization of the processing workflow in the form of variable selection and machine learning methods is needed when a combination of different datasets is performed.

6. Conclusions

This study explored the combination of lidar and hyperspectral data using different combinations of variable selection methods and machine learning algorithms to model forest heights in Robson Creek, Australia. In total, 72 machine learning models were generated from different input variables using different variable selection methods, and a wide range of model performance was obtained. Generally, no noticeable performance differences were found when different variable selection and machine learning algorithms were combined. However, BO and GA showed some peculiarities; they generated input variables that performed better during the modeling process. Boruta in particular could distinguish and remove almost 95% of the original data to identify 29 variables that produced models with similar or better performance compared with other models. The combination of Boruta and XGdart could generate the best performance model in our study, with $R^2 = 0.53$ and RMSE = 1.7 m, with the Boruta and SVR combination yielding the second-best performance model, with $R^2 = 0.51$ and RMSE = 1.8 m. In addition, our study also found overall better performance of XGBdart, XGBtree, and SVR when they were combined with various datasets for estimating forest height. Our study highlighted the optimization results by using different strategies of variable selection and machine learning methods, which would be beneficial for the upcoming studies utilizing the combination of multi-sensor data, such as satellite-based hyperspectral, i.e., Hyperion and PRISMA, and satellites' laser altimetry data, i.e., ICESat GLAS and GEDI.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2220-9964/9/9/507/s1>. The supplementary material contains (1). Table S1. The Normalized RMSE (nrmse) and Bias for every model, and (2). Table S2. Summaries of average nRMSE and bias values for machine learning algorithms, variable selection methods, and different input datasets.

Author Contributions: Sanjiwana Arjasakusuma and Sandiaga Swahyu Kusuma performed the data analysis and prepared the manuscript; Stuart Phinn contributed to providing feedback for the research analysis and manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: The authors would like to acknowledge the data support given by Matt Bradford from Tropical Forest Research Centre, CSIRO, Rebecca Trevithick (DSITIA) and John Armstrong for providing the image and field datasets, and TERN/Auscover geoportal (<http://www.auscover.org.au/>) for collecting and providing lidar and AISA hyperspectral data. The authors also would like to thank the School of Earth and Environmental Sciences, the University of Queensland, Australia, for providing the computation infrastructures to perform this research. In addition, the authors also would like to acknowledge the language editing program provided by the Research and Publishing Agency (BPP), Universitas Gadjah Mada.

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

References

- Hyde, P.; Dubayah, R.; Walker, W.; Blair, J.B.; Hofton, M.; Hunsaker, C. Mapping forest structure for wildlife habitat analysis using multi-sensor (LiDAR, SAR/InSAR, ETM+, Quickbird) synergy. *Remote Sens. Environ.* **2006**, *102*, 63–73. [\[CrossRef\]](#)
- Drake, J.B.; Dubayah, R.O.; Clark, D.B.; Knox, R.G.; Blair, J.B.; Hofton, M.a.; Chazdon, R.L.; Weishampel, J.F.; Prince, S. Estimation of tropical forest structural characteristics using large-footprint lidar. *Remote Sens. Environ.* **2002**, *79*, 305–319. [\[CrossRef\]](#)
- Clark, D.B.; Clark, D.A. Landscape-scale variation in forest structure and biomass in a tropical rain forest. *For. Ecol. Manag.* **2000**, *137*, 185–198. [\[CrossRef\]](#)
- Venter, O.; Koh, L.P. Reducing emissions from deforestation and forest degradation (REDD+): Game changer or just another quick fix? *Ann. N. Y. Acad. Sci.* **2012**, *1249*, 137–150. [\[CrossRef\]](#)
- Pommerening, a. Approaches to quantifying forest structures. *Forestry* **2002**, *75*, 305–324. [\[CrossRef\]](#)
- Ingram, J.C.; Dawson, T.P.; Whittaker, R.J. Mapping tropical forest structure in southeastern Madagascar using remote sensing and artificial neural networks. *Remote Sens. Environ.* **2005**, *94*, 491–507. [\[CrossRef\]](#)

7. Lesak, A.a.; Radeloff, V.C.; Hawbaker, T.J.; Pidgeon, A.M.; Gobakken, T.; Contrucci, K. Modeling forest songbird species richness using LiDAR-derived measures of forest structure. *Remote Sens. Environ.* **2011**, *115*, 2823–2835. [\[CrossRef\]](#)
8. Kokaly, R.F.; Clark, R.N. Spectroscopic Determination of Leaf Biochemistry Using Band-Depth Analysis of Absorption Features and Stepwise Multiple Linear Regression. *Remote Sens. Environ.* **1999**, *67*, 267–287. [\[CrossRef\]](#)
9. Stojanova, D.; Panov, P.; Gjorgjioski, V.; Kobler, A.; Džeroski, S. Estimating vegetation height and canopy cover from remotely sensed data with machine learning. *Ecol. Inform.* **2010**, *5*, 256–266. [\[CrossRef\]](#)
10. Lu, D. The potential and challenge of remote sensing - based biomass estimation. *Int. J. Remote Sens.* **2007**, *27*, 1297–1328. [\[CrossRef\]](#)
11. Nadkarni, N.M.; McIntosh, A.C.S.; Cushing, J.B. A framework to categorize forest structure concepts. *For. Ecol. Manag.* **2008**, *256*, 872–882. [\[CrossRef\]](#)
12. Myeong, S.; Nowak, D.J.; Duggin, M.J. A temporal analysis of urban forest carbon storage using remote sensing. *Remote Sens. Environ.* **2006**, *101*, 277–282. [\[CrossRef\]](#)
13. Heiskanen, J. Estimating aboveground tree biomass and leaf area index in a mountain birch forest using ASTER satellite data. *Int. J. Remote Sens.* **2006**, *27*, 1135–1158. [\[CrossRef\]](#)
14. Rignot, E.; Way, J.; Williams, C.; Viereck, L. Radar estimates of aboveground biomass in boreal forests of interior Alaska. *IEEE Trans. Geosci. Remote Sens.* **1994**, *32*, 1117–1124. [\[CrossRef\]](#)
15. Carreiras, J.M.B.; Vasconcelos, M.J.; Lucas, R.M. Understanding the relationship between aboveground biomass and ALOS PALSAR data in the forests of Guinea-Bissau (West Africa). *Remote Sens. Environ.* **2012**, *121*, 426–442. [\[CrossRef\]](#)
16. Lefsky, M.a.; Cohen, W.B.; Acker, S.a.; Parker, G.G.; Spies, T.a.; Harding, D. Lidar Remote Sensing of the Canopy Structure and Biophysical Properties of Douglas-Fir Western Hemlock Forests. *Remote Sens. Environ.* **1999**, *70*, 339–361. [\[CrossRef\]](#)
17. Asner, G.P.; Mascaro, J.; Muller-Landau, H.C.; Vieilledent, G.; Vaudry, R.; Rasamoelina, M.; Hall, J.S.; van Breugel, M. A universal airborne LiDAR approach for tropical forest carbon mapping. *Oecologia* **2012**, *168*, 1147–1160. [\[CrossRef\]](#)
18. Clark, M.L.; Roberts, D.a.; Ewel, J.J.; Clark, D.B. Estimation of tropical rain forest aboveground biomass with small-footprint lidar and hyperspectral sensors. *Remote Sens. Environ.* **2011**, *115*, 2931–2942. [\[CrossRef\]](#)
19. Lucas, R.; Armston, J.; Fairfax, R.; Fensham, R.; Accad, A.; Carreiras, J.; Kelley, J.; Bunting, P.; Clewley, D.; Bray, S.; et al. An Evaluation of the ALOS PALSAR L-Band Backscatter—Above Ground Biomass Relationship Queensland, Australia: Impacts of Surface Moisture Condition and Vegetation Structure. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2010**, *3*, 576–593. [\[CrossRef\]](#)
20. Morel, A.C.; Saatchi, S.S.; Malhi, Y.; Berry, N.J.; Banin, L.; Burslem, D.; Nilus, R.; Ong, R.C. Estimating aboveground biomass in forest and oil palm plantation in Sabah, Malaysian Borneo using ALOS PALSAR data. *For. Ecol. Manag.* **2011**, *262*, 1786–1798. [\[CrossRef\]](#)
21. Zolkos, S.; Goetz, S.; Dubayah, R. A meta-analysis of terrestrial aboveground biomass estimation using lidar remote sensing. *Remote Sens. Environ.* **2013**, *128*, 289–298. [\[CrossRef\]](#)
22. Lu, D.; Chen, Q.; Wang, G.; Liu, L.; Li, G.; Moran, E. A survey of remote sensing-based aboveground biomass estimation methods in forest ecosystems. *Int. J. Digit. Earth* **2016**, *9*, 63–105. [\[CrossRef\]](#)
23. Yang, J.; Honavar, V. Feature subset selection using a genetic algorithm. In *Feature Extraction, Construction and Selection*; Springer: Boston, MA, USA, 1998; pp. 117–136.
24. Bhanu, B.; Lin, Y. Genetic algorithm based feature selection for target detection in SAR images. *Image Vis. Comput.* **2003**, *21*, 591–608. [\[CrossRef\]](#)
25. Latifi, H.; Fassnacht, F.; Koch, B. Forest structure modeling with combined airborne hyperspectral and LiDAR data. *Remote Sens. Environ.* **2012**, *121*, 10–25. [\[CrossRef\]](#)
26. Johansen, K.; Trevithick, R.; Bradford, M.; Hacker, J.; McGrath, A.; Lieff, W. Australian examples of field and airborne AusCover campaigns. In *AusCover Good Practice Guidelines: A Technical Handbook Supporting Calibration and Validation Activities of Remotely Sensed Data Products, Version 1.1*. TERN AusCover; TERN AusCover; The University of Queensland: St Lucia, Australia, 2015; Volume Version 1.1.
27. Bradford, M.; Metcalfe, D.; Ford, A.; Liddell, M.; Green, P.; McKeown, A. Floristics, stand structure and above ground biomass of a 25 ha rainforest plot in the Wet Tropics of Australia. *J. Trop. For. Sci.* **2014**, *26*, 543–553.

28. Aktaruzzaman, M. Simulation and Correction of Spectral Smile Effect and its Influence on Hyperspectral Mapping. Master's Thesis, ITC Faculty Geo-Information Science and Earth Observation, Enschede, The Netherlands, 2008.
29. Goodwin, N.R.; Coops, N.C.; Culvenor, D.S. Assessment of forest structure with airborne LiDAR and the effects of platform altitude. *Remote Sens. Environ.* **2006**, *103*, 140–152. [\[CrossRef\]](#)
30. Takahashi, T.; Awaya, Y.; Hirata, Y.; Furuya, N.; Sakai, T.; Sakai, A. Effects of flight altitude on LiDAR-derived tree heights in mountainous forests with poor laser penetration rates. *Photogramm. J. Finl.* **2008**, *21*, 86–96.
31. Dubayah, R.O.; Drake, J.B. Lidar remote sensing for forestry. *J. For.* **2000**, *98*, 44–46.
32. Roussel, J.-R.; Auty, D.; De Boissieu, F.; Meador, A. *lidR: Airborne LiDAR Data Manipulation and Visualization for Forestry Applications*; R Package Version 1; 2018. Available online: <https://cran.r-project.org/web/packages/lidR/index.html> (accessed on 22 August 2020).
33. Khosravipour, A.; Skidmore, A.K.; Isenburg, M.; Wang, T.; Hussin, Y.A. Generating pit-free canopy height models from airborne lidar. *Photogramm. Eng. Remote Sens.* **2014**, *80*, 863–872. [\[CrossRef\]](#)
34. Kursa, M.B.; Rudnicki, W.R. Feature selection with the Boruta package. *Stat. Softw.* **2010**, *36*, 1–13.
35. Paja, W.; Pancerz, K.; Grochowalski, P. Generational feature elimination and some other ranking feature selection methods. In *Advances in Feature Selection for Data and Pattern Recognition*; Springer: Cham, Switzerland, 2018; pp. 97–112.
36. Amiri, M.; Pourghasemi, H.R.; Ghanbarian, G.A.; Afzali, S.F.J.G. Assessment of the importance of gully erosion effective factors using Boruta algorithm and its spatial modeling and mapping using three machine learning algorithms. *Geoderma* **2019**, *340*, 55–69. [\[CrossRef\]](#)
37. Minasny, B.; Setiawan, B.I.; Saptomo, S.K.; McBratney, A.B. Open digital mapping as a cost-effective method for mapping peat thickness and assessing the carbon stock of tropical peatlands. *Geoderma* **2018**, *313*, 25–40.
38. Keskin, H.; Grunwald, S.; Harris, W.G. Digital mapping of soil carbon fractions with machine learning. *Geoderma* **2019**, *339*, 40–58. [\[CrossRef\]](#)
39. Xu, Y.; Smith, S.E.; Grunwald, S.; Abd-Elrahman, A.; Wani, S.P. Incorporation of satellite remote sensing pan-sharpened imagery into digital soil prediction and mapping models to characterize soil property variability in small agricultural fields. *ISPRS J. Photogramm. Remote Sens.* **2017**, *123*, 1–19. [\[CrossRef\]](#)
40. Jensen, J. Chapter 8. Image Enhancement. In *Introductory Digital Image Processing: A Remote Sensing Perspective*, 5th ed.; Pearson: Glenview, IL, USA, 2015; pp. 301–322.
41. Ma, W.; Gong, C.; Hu, Y.; Meng, P.; Xu, F. The Hughes phenomenon in hyperspectral classification based on the ground spectrum of grasslands in the region around Qinghai Lake. In *Proceedings of the International Symposium on Photoelectronic Detection and Imaging 2013: Imaging Spectrometer Technologies and Applications*, Beijing, China, 25–27 June 2013; p. 89101.
42. Asner, G.P.; Knapp, D.E.; Boardman, J.; Green, R.O.; Kennedy-Bowdoin, T.; Eastwood, M.; Martin, R.E.; Anderson, C.; Field, C.B. Carnegie Airborne Observatory-2: Increasing science data dimensionality via high-fidelity multi-sensor fusion. *Remote Sens. Environ.* **2012**, *124*, 454–465. [\[CrossRef\]](#)
43. Leutner, B.; Horning, N.; Leutner, M.B. Package 'RStoolbox'; Version 0.1; 2017. Available online: <https://cran.r-project.org/web/packages/RStoolbox/index.html> (accessed on 22 August 2020).
44. Meiri, R.; Zahavi, J. Using simulated annealing to optimize the feature selection problem in marketing applications. *Eur. J. Oper. Res.* **2006**, *171*, 842–858. [\[CrossRef\]](#)
45. Kirkpatrick, S.; Gelatt, C.D.; Vecchi, M.P. Optimization by simulated annealing. *Science* **1983**, *220*, 671–680. [\[CrossRef\]](#)
46. Brusco, M.J. A comparison of simulated annealing algorithms for variable selection in principal component analysis and discriminant analysis. *Comput. Stat. Data Anal.* **2014**, *77*, 38–53. [\[CrossRef\]](#)
47. Gheyas, I.A.; Smith, L.S. Feature subset selection in large dimensionality domains. *Pattern Recognit.* **2010**, *43*, 5–13. [\[CrossRef\]](#)
48. Kuhn, M.; Wing, J.; Weston, S.; Williams, A.; Keefer, C.; Engelhardt, A.; Cooper, T.; Mayer, Z.; Kenkel, B.; R Core Team; et al. *The Caret Package*; Vienna, Austria, 2012. Available online: <https://cran.r-project.org/package=caret> (accessed on 22 August 2020).
49. Tan, F.; Fu, X.; Zhang, Y.; Bourgeois, A.G. A genetic algorithm-based method for feature subset selection. *Soft Comput.* **2008**, *12*, 111–120. [\[CrossRef\]](#)
50. Oreski, S.; Oreski, G. Genetic algorithm-based heuristic for feature selection in credit risk assessment. *Expert Syst. Appl.* **2014**, *41*, 2052–2064. [\[CrossRef\]](#)

51. Chau, A.L.; Li, X.; Yu, W. Support vector machine classification for large datasets using decision tree and Fisher linear discriminant. *Future Gener. Comput. Syst.* **2014**, *36*, 57–65. [CrossRef]
52. Friedman, J.H. Multivariate adaptive regression splines. *Ann. Stat.* **1991**, *19*, 1–67. [CrossRef]
53. Friedman, J.H.; Roosen, C.B. *An Introduction to Multivariate Adaptive Regression Splines*; Sage Publications Sage CA: Thousand Oaks, CA, USA, 1995.
54. Alkaim, A.F.; Al-Janabi, S. Multi objectives optimization to gas flaring reduction from oil production. In Proceedings of the International conference on big data and networks technologies, Leuven, Belgium, 29 April–2 May 2019; pp. 117–139.
55. Leathwick, J.; Elith, J.; Hastie, T. Comparative performance of generalized additive models and multivariate adaptive regression splines for statistical modelling of species distributions. *Ecol. Model.* **2006**, *199*, 188–196. [CrossRef]
56. Filippi, A.M.; Güneralp, İ.; Randall, J. Hyperspectral remote sensing of aboveground biomass on a river meander bend using multivariate adaptive regression splines and stochastic gradient boosting. *Remote Sens. Lett.* **2014**, *5*, 432–441. [CrossRef]
57. Quirós, E.; Felicísimo, Á.M.; Cuartero, A. Testing multivariate adaptive regression splines (MARS) as a method of land cover classification of TERRA-ASTER satellite images. *Sensors* **2009**, *9*, 9011–9028. [CrossRef]
58. Güneralp, İ.; Filippi, A.M.; Randall, J. Estimation of floodplain aboveground biomass using multispectral remote sensing and nonparametric modeling. *Int. J. Appl. Earth Obs. Geoinf.* **2014**, *33*, 119–126. [CrossRef]
59. Milborrow, S. *Package 'Earth'*, R Package Version 5.1.2; 2020. Available online: <https://cran.r-project.org/web/packages/earth/index.html> (accessed on 22 August 2020).
60. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
61. Pal, M. Random forest classifier for remote sensing classification. *Int. J. Remote Sens.* **2005**, *26*, 217–222. [CrossRef]
62. Gislason, P.O.; Benediktsson, J.A.; Sveinsson, J.R. Random forests for land cover classification. *Pattern Recognit. Lett.* **2006**, *27*, 294–300. [CrossRef]
63. Liaw, A.; Wiener, M. Classification and regression by randomForest. *R News* **2002**, *2*, 18–22.
64. Simm, J.; Abril, I. *Extratrees: Extremely Randomized Trees (ExtraTrees) Method for Classification and Regression*; R Package Version 1.0.5; 2014. Available online: <https://cran.r-project.org/web/packages/extraTrees/extraTrees.pdf> (accessed on 22 August 2020).
65. Geurts, P.; Ernst, D.; Wehenkel, L. Extremely randomized trees. *Mach. Learn.* **2006**, *63*, 3–42. [CrossRef]
66. Ahmad, M.W.; Mourshed, M.; Rezgui, Y. Tree-based ensemble methods for predicting PV power generation and their comparison with support vector regression. *Energy* **2018**, *164*, 465–474. [CrossRef]
67. Ahmad, M.W.; Reynolds, J.; Rezgui, Y. Predictive modelling for solar thermal energy systems: A comparison of support vector regression, random forest, extra trees and regression trees. *J. Clean. Prod.* **2018**, *203*, 810–821. [CrossRef]
68. Zdravevski, E.; Lameski, P.; Kulakov, A.; Trajkovik, V. Performance comparison of random forests and extremely randomized trees. In Proceedings of the 13th Conference for Informatics and Information Technology (CIIT 2016), Faculty of Computer Science and Engineering (FCSE) and Computer Society of Macedonia, Struga, Macedonia, 22–24 April 2016.
69. Barrett, B.; Nitze, I.; Green, S.; Cawkwell, F. Assessment of multi-temporal, multi-sensor radar and ancillary spatial data for grasslands monitoring in Ireland using machine learning approaches. *Remote Sens. Environ.* **2014**, *152*, 109–124. [CrossRef]
70. Lawrence, R.; Bunn, A.; Powell, S.; Zambon, M. Classification of remotely sensed imagery using stochastic gradient boosting as a refinement of classification tree analysis. *Remote Sens. Environ.* **2004**, *90*, 331–336. [CrossRef]
71. Chen, T.; He, T.; Benesty, M.; Khotilovich, V.; Tang, Y. *Xgboost: Extreme Gradient Boosting*; R Package Version 1.1.1.1; 2020. Available online: <https://cran.r-project.org/web/packages/xgboost/index.html> (accessed on 22 August 2020).
72. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [CrossRef]
73. Chen, T.; Guestrin, C. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.

74. Georganos, S.; Grippa, T.; Vanhuysse, S.; Lennert, M.; Shimoni, M.; Wolff, E. Very high resolution object-based land use–land cover urban classification using extreme gradient boosting. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 607–611. [CrossRef]
75. Fan, J.; Wang, X.; Wu, L.; Zhou, H.; Zhang, F.; Yu, X.; Lu, X.; Xiang, Y. Comparison of Support Vector Machine and Extreme Gradient Boosting for predicting daily global solar radiation using temperature and precipitation in humid subtropical climates: A case study in China. *Energy Convers. Manag.* **2018**, *164*, 102–111. [CrossRef]
76. Zhang, H.; Eziz, A.; Xiao, J.; Tao, S.; Wang, S.; Tang, Z.; Zhu, J.; Fang, J. High-Resolution Vegetation Mapping Using eXtreme Gradient Boosting Based on Extensive Features. *Remote Sens.* **2019**, *11*, 1505. [CrossRef]
77. Pham, T.D.; Le, N.N.; Ha, N.T.; Nguyen, L.V.; Xia, J.; Yokoya, N.; To, T.T.; Trinh, H.X.; Kieu, L.Q.; Takeuchi, W. Estimating Mangrove Above-Ground Biomass Using Extreme Gradient Boosting Decision Trees Algorithm with Fused Sentinel-2 and ALOS-2 PALSAR-2 Data in Can Gio Biosphere Reserve, Vietnam. *Remote Sens.* **2020**, *12*, 777. [CrossRef]
78. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [CrossRef]
79. Noble, W.S. What is a support vector machine? *Nat. Biotechnol.* **2006**, *24*, 1565–1567. [CrossRef] [PubMed]
80. Alloghani, M.; Aljaaf, A.; Hussain, A.; Baker, T.; Mustafina, J.; Al-Jumeily, D.; Khalaf, M. Implementation of machine learning algorithms to create diabetic patient re-admission profiles. *Med Inform. Decis. Mak.* **2019**, *19*, 253. [CrossRef]
81. Drucker, H.; Burges, C.J.; Kaufman, L.; Smola, A.J.; Vapnik, V. *Support Vector Regression Machines*; MIT Press: Cambridge, MA, USA, 1997; pp. 155–161.
82. Gualtieri, J.A.; Crompt, R.F. Support vector machines for hyperspectral remote sensing classification. In Proceedings of the 27th AIPR Workshop: Advances in Computer-Assisted Recognition, Washington, DC, USA, 14–16 October 1998; pp. 221–232.
83. Pasolli, L.; Notarnicola, C.; Bruzzone, L. Estimating soil moisture with the support vector regression technique. *IEEE Geosci. Remote Sens. Lett.* **2011**, *8*, 1080–1084. [CrossRef]
84. Karatzoglou, A.; Smola, A.; Hornik, K. *Package ‘Kernlab’*; R Package Version 0.9-29; 2019. Available online: <https://cran.r-project.org/web/packages/kernlab/index.html> (accessed on 22 August 2020).
85. Bergstra, J.; Bengio, Y. Random search for hyper-parameter optimization. *J. Mach. Learn. Res.* **2012**, *13*, 281–305.
86. Walther, B.A.; Moore, J.L. The concepts of bias, precision and accuracy, and their use in testing the performance of species richness estimators, with a literature review of estimator performance. *Ecography* **2005**, *28*, 815–829. [CrossRef]
87. Hamner, B.; Frasco, M. *Metrics: Evaluation Metrics for Machine Learning*; R Package Version 0.1.4; 2018. Available online: <https://cran.r-project.org/web/packages/Metrics/index.html> (accessed on 22 August 2020).
88. Zambrano-Bigiarini, M. *Package ‘hydroGOF’: Goodness-of-Fit Functions for Comparison of Simulated and Observed Hydrological Time Series*; R Package Version 0.4; 2020. Available online: <https://cran.r-project.org/web/packages/hydroGOF/index.html> (accessed on 22 August 2020).
89. Swatantran, A.; Dubayah, R.; Roberts, D.; Hofton, M.; Blair, J.B. Mapping biomass and stress in the Sierra Nevada using lidar and hyperspectral data fusion. *Remote Sens. Environ.* **2011**, *115*, 2917–2930. [CrossRef]
90. Suárez, J.C.; Ontiveros, C.; Smith, S.; Snape, S. Use of airborne LiDAR and aerial photography in the estimation of individual tree heights in forestry. *Comput. Geosci.* **2005**, *31*, 253–262. [CrossRef]
91. Cho, M.A.; Skidmore, A.K.; Sobhan, I. Mapping beech (*Fagus sylvatica* L.) forest structure with airborne hyperspectral imagery. *Int. J. Appl. Earth Obs. Geoinf.* **2009**, *11*, 201–211. [CrossRef]
92. Laurin, G.V.; Chen, Q.; Lindsell, J.A.; Coomes, D.A.; Del Frate, F.; Guerriero, L.; Pirotti, F.; Valentini, R. Above ground biomass estimation in an African tropical forest with lidar and hyperspectral data. *ISPRS J. Photogramm. Remote Sens.* **2014**, *89*, 49–58. [CrossRef]
93. Sankey, T.; Donager, J.; McVay, J.; Sankey, J.B. UAV lidar and hyperspectral fusion for forest monitoring in the southwestern USA. *Remote Sens. Environ.* **2017**, *195*, 30–43. [CrossRef]
94. Anderson, J.; Plourde, L.; Martin, M.; Braswell, B.; Smith, M.; Dubayah, R.; Hofton, M.; Blair, J. Integrating waveform lidar with hyperspectral imagery for inventory of a northern temperate forest. *Remote Sens. Environ.* **2008**, *112*, 1856–1870. [CrossRef]
95. Xu, T.; Wei, H.; Hu, G. Study on continuous network design problem using simulated annealing and genetic algorithm. *Expert Syst. Appl.* **2009**, *36*, 1322–1328. [CrossRef]

96. Chang, Y.-C.; Chang, K.-H.; Wu, G.-J. Application of eXtreme gradient boosting trees in the construction of credit risk assessment models for financial institutions. *J. Appl. Soft Comput.* **2018**, *73*, 914–920. [[CrossRef](#)]
97. Kattenborn, T.; Maack, J.; Faßnacht, F.; Enßle, F.; Ermert, J.; Koch, B. Mapping forest biomass from space—Fusion of hyperspectral EO1-hyperion data and Tandem-X and WorldView-2 canopy height models. *Int. J. Appl. Earth Obs. Geoinf.* **2015**, *35*, 359–367. [[CrossRef](#)]
98. Brezočnik, L.; Fister, I.; Podgorelec, V. Swarm intelligence algorithms for feature selection: A review. *Appl. Sci.* **2018**, *8*, 1521. [[CrossRef](#)]
99. Hamedianfar, A.; Gibril, M.B.A.; Hosseinpour, M.; Pellikka, P.K. Synergistic use of particle swarm optimization, artificial neural network, and extreme gradient boosting algorithms for urban LULC mapping from WorldView-3 images. *Geocarto Int.* **2020**, 1–19. [[CrossRef](#)]
100. Loizzo, R.; Guarini, R.; Longo, F.; Scopa, T.; Formaro, R.; Facchinetti, C.; Varacalli, G. PRISMA: The Italian hyperspectral mission. In Proceedings of the IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 175–178.
101. Lopinto, E.; Ananasso, C. The Prisma hyperspectral mission. In Proceedings of the 33rd EARSeL Symposium, Towards Horizon 2020: Earth Observation and Social Perspectives, Matera, Italy, 3–7 June 2013.
102. Pignatti, S.; Palombo, A.; Pascucci, S.; Romano, F.; Santini, F.; Simoniello, T.; Umberto, A.; Vincenzo, C.; Acito, N.; Diani, M. The PRISMA hyperspectral mission: Science activities and opportunities for agriculture and land monitoring. In Proceedings of the 2013 IEEE International Geoscience and Remote Sensing Symposium-IGARSS, Melbourne, Australia, 21–26 July 2013; pp. 4558–4561.
103. Magruder, L.; Neuenschwander, A.; Neumann, T.; Kurtz, N.; Duncanson, L.; Dubayah, R. NASA's ICESat-2 and GEDI missions for land and vegetation applications. In Proceedings of the 21st EGU General Assembly, Vienna, Austria, 7–12 April 2019.
104. Duncanson, L.; Neuenschwander, A.; Hancock, S.; Thomas, N.; Fatoyinbo, T.; Simard, M.; Silva, C.A.; Armston, J.; Luthcke, S.B.; Hofton, M. Biomass estimation from simulated GEDI, ICESat-2 and NISAR across environmental gradients in Sonoma County, California. *Remote Sens. Environ.* **2020**, *242*, 111779. [[CrossRef](#)]



© 2020 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 (<http://creativecommons.org/licenses/by/4.0/>).