

Article

Maximizing the Diversity of Ensemble Random Forests for Tree Genera Classification Using High Density LiDAR Data

Connie Ko ^{1,*}, Gunho Sohn ², Tarmo K. Remmel ¹ and John R. Miller ²

¹ Department of Geography, York University, 4700 Keele Street, Ross North 430, Toronto, ON M3J 1P3, Canada; remmelt@yorku.ca

² Department of Earth and Space Science and Engineering, York University, 4700 Keele Street, Petrie Building 149, Toronto, ON M3J 1P3, Canada; gsohn@yorku.ca (G.S.); jrmiller@yorku.ca (J.R.M.)

* Correspondence: cko@yorku.ca; Tel.: +1-416-736-2100; Fax: +1-416-736-5988

Academic Editors: Ioannis Gitas, Clement Atzberger and Prasad S. Thenkabail

Received: 20 June 2016; Accepted: 3 August 2016; Published: 8 August 2016

Abstract: Recent research into improving the effectiveness of forest inventory management using airborne LiDAR data has focused on developing advanced theories in data analytics. Furthermore, supervised learning as a predictive model for classifying tree genera (and species, where possible) has been gaining popularity in order to minimize this labor-intensive task. However, bottlenecks remain that hinder the immediate adoption of supervised learning methods. With supervised classification, training samples are required for learning the parameters that govern the performance of a classifier, yet the selection of training data is often subjective and the quality of such samples is critically important. For LiDAR scanning in forest environments, the quantification of data quality is somewhat abstract, normally referring to some metric related to the completeness of individual tree crowns; however, this is not an issue that has received much attention in the literature. Intuitively the choice of training samples having varying quality will affect classification accuracy. In this paper a Diversity Index (*DI*) is proposed that characterizes the diversity of data quality (Q_i) among selected training samples required for constructing a classification model of tree genera. The training sample is diversified in terms of data quality as opposed to the number of samples per class. The diversified training sample allows the classifier to better learn the positive and negative instances and; therefore; has a higher classification accuracy in discriminating the “unknown” class samples from the “known” samples. Our algorithm is implemented within the Random Forests base classifiers with six derived geometric features from LiDAR data. The training sample contains three tree genera (pine; poplar; and maple) and the validation samples contains four labels (pine; poplar; maple; and “unknown”). Classification accuracy improved from 72.8%; when training samples were selected randomly (with stratified sample size); to 93.8%; when samples were selected with additional criteria; and from 88.4% to 93.8% when an ensemble method was used.

Keywords: tree genera classification; LiDAR; random forests; ensemble classification; diversity maximization

1. Introduction

The use of LiDAR (Light Detection and Ranging) data for forestry applications has advanced in many ways in recent years, described, for example, by [1] with a series of methodologies along with case studies. In particular, research into obtaining tree species/genera information from LiDAR has been increasing and classification results have been reported by several researchers including [2–13] and continues to gain greater attention. However, there are two common challenges in using LiDAR to classify tree species/genera that have received little attention in the literature. The first challenge

is the existence of tree genera classes in the forest that have not been field validated. In supervised classification, this is a problem which arises when the validation data have more classes than the training data and it is a fundamental problem for sampling a subset of a population during training. In this case, normally an “unknown class” will be assigned to the extra class as has been described in [14,15]. The second challenge is the acquisition of inconsistent data from specific 3D tree objects, due to occlusion of tree parts due to varying LiDAR scan angles and canopy architecture. Moreover, when trees grow closely together, branches (and therefore LiDAR points) are often co-mingled among multiple trees and their membership to a specific tree canopy cannot be determined. Single trees having multiple tops or leaders can also be confused quite easily as being multiple trees, and variability in growing environments or tree age can lead trees of a common species to appear vastly different in LiDAR point clouds. This results in a large variation in training and validation samples, in terms of appearance and quality. This paper addresses these two issues by proposing a workflow that takes advantage of the variability observed within a common genera.

The use of ensemble methods is proposed to address this problem. Ensemble classification is the training of multiple classifiers to solve a common labeling problem by different communities. This classification strategy is also called committee-based learning, multiple classifier systems, or a mixture of experts [16–18]. The criterion of a good ensemble system is that it should provide an increase in classification accuracy. There are numerous ways of combining classifiers as suggested by previous studies [19] with a wide variety of applications contributed from text categorization [20] and hand-written word recognition [21]. In particular, a few examples of related studies in remote sensing include [17,22–24].

There are three popular techniques for reducing a multi-class problem into a series of binary classification problems [25], namely: “one-versus-one” (OVO) [26], “one-versus-all” (OVA) [27], and Error Correcting Output Codes (ECOC). OVO consists of binary classifiers that differentiate between two classes (for example pine vs. maple), and therefore for a k class problem, $\binom{k}{2}$ (or “ k combination 2”, calculates the number of combinations needed to separate k classes with only two outcomes for each classifier, namely “positive” and “negative”) classifiers are needed for ensemble implementation. On the other hand OVA consists of binary classifiers that discriminate between particular classes from the rest of the classes (for example, pine vs. non-pine), and therefore k classifiers are needed to solve a k -class problem. ECOC is a method where a code word is generated from a series of binary classifiers for each class label. For any validation sample, a code word from that sample will be generated, and compared to the code word generated from the training samples. The class with the minimum Hamming distance will be labeled [28]. To compare the performance between OVO and OVA, studies such as [25] show that although there is no significant difference found between OVA and OVO, both strategies outperform the original classifier. In another study, the authors of [29] showed that OVA attains more accurate classification when comparing: (i) a concept-adapting very fast decision tree (CVFDT), a single multi-classifier; (ii) a weighted classifier ensemble (WCE), and streaming ensemble algorithm (SEA), both of which are ensembles of multi-class classifiers; and (iii) an ultrafast forest tree (UFFT), an OVO method. In [30] it was shown that OVO outperformed OVA, in contrast to [27] where it was suggested that OVA performed as well as OVO. These studies suggest that neither OVO, nor OVA, consistently outperform the other. Instead, these studies indicate that the decomposition of a multiclass problem into series of binary classification problems is an efficient approach that often outperforms the original multi-class classifier. In fact, OVO and OVA are popular methods for combining Support Vector Machine (SVM) classifiers [30–33].

As mentioned earlier, it is not always easy to obtain a completely isolated single tree; as a result, the “data quality” of segmented LiDAR trees is not uniform. Consequently, when training data are randomly selected from these segmented LiDAR trees, the problem of imbalanced training, with respect to data quality, arises. Normally, the problem of data distribution imbalance refers to the differences in the number of samples for each class [34,35], which is a common problem in real

world data simply because some phenomena occur infrequently in nature. Classifiers generated from an imbalanced training sample will be specialized in classifying the majority classes, and therefore will bias the results towards the majority class. At the data level solution, there are three common techniques to overcome this problem: (1) oversampling the minority class; (2) under sampling the majority class; or (3) a combination of both methods [34–38]. The goal of these strategies is to diversify the sampling distribution in terms of the number of samples per class and these studies show that the distribution of the training sample is important in affecting the classification accuracy.

Diversification is an important concept in ensemble learning; when the classifiers to be combined have substantial diversity, the predictive power tends to be higher [39]. During the construction of the Random Forests classifier, the classifier already incorporates the concept of diversification in the following ways. According to [40], Random Forests diversify the number of training samples by bootstrap aggregating (bagging) in two ways. First, the same number of samples are drawn from the minority and majority class, balancing the minority/majority sample. Second, a heavier penalty is placed on misclassifying the minority class by giving it a higher weight, and the final label takes the majority voting of the individual classification trees. Also, during the construction of an individual classification tree, a subset of features is selected for defining the splitting node. The subset of features is drawn randomly and this randomization promotes diversification in classification. The authors of [41] further increased the diversity of Random Forests by splitting the training samples into smaller subspaces, and showed improved classification accuracy in their medical datasets over regular Random Forests classification. However, improving the performance of supervised learning through balancing the per-class distribution of training samples can be achieved only if the quality of individual samples exhibits relatively similar condition across and within classes. As discussed earlier, the segmentation quality of LiDAR tree samples is irregular, and its regularization is a non-trivial task, especially using 3D LiDAR point clouds captured in forest environments.

There are two goals for this paper: First, to address and quantify the quality of variability observed in the data that have been field validated. From those results, the relationship between the diversity present in the training samples are examined in relation to the ultimate classification accuracy. Second, training data are selected that contain the most diversity, and classification is performed using a series of binary classifiers with a scheme that has been designed to generate an “unknown” class. This will address the problem of having more classes in the validation data than the training data.

2. Study Area and Data Acquisition

The LiDAR data were collected on 7 August 2009, over field study sites located about 75 km east of Sault Ste. Marie, Ontario, Canada. A Riegl LMS-Q560 scanner on multiple flight passes was flown at altitudes between 122 and 250 m above ground level; the combined pulse density of the dataset is about 40 pulses per m². Since each pulse generates up to five returns, the point density can be as high as 200 points per m². The data were acquired at a lower altitude for the power-line corridor site to obtain a higher point density data for the purpose of power line risk management, whereas the other forested sites were acquired at a higher altitude. Field surveys were conducted in the summers of 2009 and 2011 at eight field sites (Table 1), selected to capture the diversity of environmental conditions, and are named: *Poplar*₁, *Poplar*₂, *Maple*₁, *Maple*₂, *Maple*₃, *Pine*₁, *Pine*₂, and *Corridor*. The poplar sites (*Poplar*₁ and *Poplar*₂) contain mostly poplar trees, which are located in the northern part of the study area. *Maple*₁, *Maple*₂, and *Maple*₃ are dominated by maple, and share very similar characteristics. All the maple-dominated sites have a closed canopy, and are growing with other deciduous species such as birch and oak; the understory growth is vigorous.

The Corridor site is the most complex site, and was selected because it includes trees that are difficult to identify from LiDAR data. The trees in the corridor sites are growing very close together, such that individual crown isolation from the LiDAR data for this site is difficult and often the LiDAR derived tree crowns that were isolated into individual trees, contain points from neighboring trees. Furthermore, some LiDAR tree crowns collected from this site contain partial occlusion by shadows,

and also the site is characterized by vigorous understory growth. The growing conditions between the two sides of the transmission corridor are also different due to the differences in topography, resulting in differences in sunlight penetration to the vegetated area, as well as in the abundance of understory growth. One side of the site has abundant understory growth whereas the other side has very little.

Table 1. Description of trees in each field site.

Site	No. of Trees	Genera	Structure
<i>Poplar</i> ₁	6	Poplar (6)	M, SS
<i>Poplar</i> ₂	40	Poplar (40)	D, MS
<i>Maple</i> ₁	20	Birch (8), Maple (12)	D, MS
<i>Maple</i> ₂	20	Birch (1), Maple (16), Oak (3)	D, MS
<i>Maple</i> ₃	6	Maple (6)	D, MS
<i>Corridor</i>	48	Pine (20), Poplar (14), Birch (2), Spruce (10), Larch (2)	D, MS
<i>Pine</i> ₁	40	Pine (40)	O
<i>Pine</i> ₂	6	Pine (6)	M, SS

M = Moderate, D = Dense, O = Open, SS = Single Stratum, MS = Multiple Strata; Bracketed number in the genera column indicates the number of trees belonging to the genera.

The pine-dominated sites (*Pine*₁ and *Pine*₂) were selected to represent an open canopy area. *Pine*₁ is dominated by red and white pine and *Pine*₂ is dominated by white pine stands. Of the 186 trees sampled, 160 of them belong to *Pinus* (pine), *Populus* (poplar) or *Acer* (maple), which are the three main genera (class labels) of classification for this paper. The rest of the trees (26 trees) will be treated as other category, and will be classified as “unknown”, which contain 11 *Betula* (birch), 3 *Quercus* (oak), 10 *Picea* (spruce), 2 *Larix* (larch) (Table 1). Within the tree genera collected in the field sites, the identified species are white birch (*Betula papyrifera* Marsh.), maple (*Acer saccharum* Marsh.), red oak (*Quercus rubra* L.), jack pine (*Pinus banksiana* Lamb.), poplar (*Populus temuloides*), white pine (*Pinus strobus* L.), white spruce (*Picea glauca* (Moench) Voss), and larch (*Larix laricina*).

3. Methods

3.1. Overview of the Methodology

In ensemble learning, there are two types of models. The first is the parallel ensemble model where base learners are used to make decisions in parallel (e.g., bagging methods). The second is the sequential ensemble model where base learners are used to make decisions sequentially (e.g., boosting methods) [18]. One of the most important objectives for combining classifiers is to reduce the overall error, that is, to increase the overall classification accuracy as compared to a single classifier [24,42–45]. In parallel models, [18] showed mathematically that errors reduce exponentially to ensemble size T by the Hoeffding inequality using the following example: Let h_i be the base classifier (binary) $\{+1, -1\}$ with error ϵ such that

$$P(h_i(x) \neq f(x)) = \epsilon \quad (1)$$

where $f(x)$ is the field validated data, and by combining T classifiers, one can form a regular ensemble classifier $H(x)$:

$$H(x) = \text{sign}\left(\sum_{i=1}^T h_i(x)\right) \quad (2)$$

$$\text{where } \text{sign}(x) = \begin{cases} -1 & \text{for } x < 0 \\ 0 & \text{for } x = 0 \\ 1 & \text{for } x > 0 \end{cases}.$$

Assume that if the final decision is made by a majority voting scheme, $H(x)$ will misclassify if at least half of the base classifiers make an error. Therefore, the generalization error can be written as:

$$P(H(x) \neq f(x)) = \sum_{k=0}^{T/2} \binom{T}{k} (1-\epsilon)^k \leq \exp(-\frac{1}{2}T(2\epsilon-1)^2) \quad (3)$$

This shows that the generalization error reduces exponentially as T becomes larger, such that as $T \rightarrow \infty$, $\epsilon \rightarrow 0$.

In sequential models, the overall errors are reduced in a residual decreasing manner. An efficient sequential model has base classifiers ordered in a way that the subsequent base classifier is able to correct the mistake of a prior base classifier. Also, base classifiers will have to be sufficiently diverse such that the same mistake will not be made in the succeeding level.

For this paper, classification is performed with two models, but because the parallel model shows better classification accuracy, our discussions will be restricted to the parallel model. Since one of the goals for this paper is to design a classification scheme that is capable of classifying the negative samples (the non-pine, non-poplar, and non-maple), the OVA decomposition is more suitable for this task. We use the OVA decomposition within a parallel model for classifying pine, poplar, maple, and “unknown” with Random Forests as our base classifiers. The process comprises three main components: (1) selection of training and validation samples; (2) using Random Forests as base classifiers; and (3) running the parallel ensemble model. Random Forest classification is discussed first although it is the second step in the overall methodology, primarily because the other components rely on an understanding of Random Forests classification.

3.2. Random Forests Classification

In this paper, each “LiDAR tree” refers to an individual, isolated tree that contains LiDAR points that are segmented from the LiDAR point cloud scene whereas classification tree refers to the decision tree constructed by the Random Forests method. As alluded to earlier, increased diversity is important for improving classification performance; Random Forests has already applied this concept by balancing the number of training samples per class, and had included randomization in selecting training samples, and features for splitting node definitions. We will further increase diversity by having as much variety in terms of data quality in training samples as possible. Random Forests is being used in two different areas in this paper. The first is for quantifying the quality of each LiDAR tree; this information is then used for the final selection of training data. An experiment is also performed to show the change in classification accuracy due to the increase in diversity included in the training sample. Since this part of the analysis is for selecting training samples, where training samples contain no “unknown” class (as training samples had been field validated), the Random Forests classifier is used as a multi-class classifier. The second way that Random Forests classifier is used is for constructing the binary base classifiers where classification will be performed on the validation data. There are three base classifiers: h_p (classifier that will produce class labels of pine (p) and non-pine (p')); h_o (classifier that will produce class labels of poplar (o) and non-poplar (o')); and h_m (classifier that will produce class labels of maple (m) and non-maple (m')). In this paper, we refer to Random Forests that is being treated as a multi-class classifier as Regular Random Forests (R-RF) in order to differentiate it from Ensemble Random Forests (E-RF).

The Random Forests algorithm itself is an ensemble classifier, where the final classification labels are obtained by combining multiple classification trees for categorical data, or regression trees for continuous data trained from a subset of the data [46,47]. Random Forests uses approximately 63% of the data for training (in-bag data) and therefore uses approximately 37% of the data (out-of-bag data) for validation. The value 37% comes from the approximation of the value $\frac{1}{e}$ where a number of observations N is drawn with replacement and therefore $\frac{1}{e}$ of data are being omitted from the draw and therefore it is called “out-of-bag” (OOB). The main input variables for Random Forests relevant

for this paper are: (1) training sample labeled with known genera, and a description of the geometric classification features; the detail for deriving these features can be found in [48]; (2) the number of feature variables randomly sampled at each split ($M_{try} = 2$); (3) the number of trees generated within each iteration ($N_{tree} = 1000$); and (4) minimum size of terminal node ($K_{node} = 1$). These values were set according to the suggestion from [49] where N_{tree} is a large number; $M_{try} \approx$ square root of the number of features and $nodesize = 1$ is the default for classification trees. Random Forests produce the following output relevant for this paper: (1) a classification scheme generated using in-bag training data; (2) the mis-classification rate as a percentage using OOB data; and (3) average vote calculated for each class in each LiDAR tree, where the final prediction of the OOB data are made by the maximum average vote. The classification algorithm was implemented within the *randomForest* package for R [50].

3.3. Goal 1: Maximizing Quality Diversity in the Training Data

In this part of the experiment, the quality of each LiDAR tree is quantified as has been validated in the field as belonging to pine, poplar, and maple. One can only quantify LiDAR trees belonging to these genera because a subset of these trees eventually became our training data, where our validation data will contain the “unknown” class. A sensitivity analysis is also performed on how data quality affects classification accuracy, since it was assumed that a larger range of quality in training sample would yield higher classification rate. With the lack of field validated data, it was decided to include all the field validated LiDAR trees for running the experiment excluding the “unknown” instances for testing diversity against classification accuracy. As mentioned before, Random Forests randomly select approximately 37% of the data (OOB data) as validation. A class prediction is made for each LiDAR tree that is selected as OOB data, from which an OOB error is calculated. One of our research goals is to propose a quality index measuring a diversity of training data quality used for learning Random Forests. Through maximizing the diversity of training data quality, we aim to improve classification accuracy obtained by Random Forests classification.

Suppose that we have a training dataset, T , which contains k LiDAR trees, that is $T = \{t_i \mid t_1, t_2, \dots, t_k\}$, where $k = 160$ tree samples in our case. Given a set of T , OOB data are generated by randomly selecting N_{tree} numbers of its subsamples, $OOB = \{OOB_j \mid OOB_1, OOB_2, \dots, OOB_{N_{tree}}\}$, each of which is used for producing a set of decision trees, $h = \{h_j \mid h_1, h_2, \dots, h_{N_{tree}}\}$. Then, we can measure NC_i and ND_i for any given t_i by introducing two indicator functions $I_C(t_i)$ and $I_D(t_i)$ as follows:

$$NC_i = \sum_{j=1}^{N_{tree}} I_C(t_i \in T; OOB_j, h_j); I_C(t_i) = \begin{cases} 1, & \text{if } h_j(t_i) = f(t_i) \\ 0, & \text{if } h_j(t_i) \neq f(t_i) \end{cases} \quad (4)$$

$$ND_i = \sum_{j=1}^{N_{tree}} I_D(t_i \in T; OOB_j); I_D(t_i) = \begin{cases} 1, & \text{if } t_i \in OOB_j \\ 0, & \text{if } t_i \notin OOB_j \end{cases} \quad (5)$$

where NC_i indicates the number of times a LiDAR tree (t_i) is predicted correctly, while ND_i represents the number of times a LiDAR tree (t_i) is selected as OOB data; h_j is the base classifier, and $f(t_i)$ is the field validated data. Next, we compute a data quality Q_i , which can be described as:

$$Q_i = \frac{NC_i}{ND_i} \quad (6)$$

Q_i from Equation (6) is a measurement being made of the i th LiDAR tree sample, and is normalized by the denominator ND_i . If Q_i is small, it means that the particular LiDAR tree cannot be correctly predicted easily, and vice versa. Q_i has a minimum value of 0 where the particular tree can never be classified correctly and a maximum value of 1 where the particular tree can always be classified correctly. This ratio will represent data quality for the rest of the paper. The data quality measure described in Equation (6) is calculated over the entire training set T to produce $\{Q_i\}$.

To obtain classification accuracy of validation data, the OOB error provided by Random Forests classifier was not used. Instead, 25% of the total data were partitioned for training and 75% of the data for testing (and to obtain the classification accuracy). A previous study [48] shows that a 25%:75% ratio for training and validation is optimal for this dataset and therefore this same ratio will be used to perform the following experiments.

Next, we plot a frequency distribution over $\{Q_i\}$ computed by Equation (6). Then, we had quantized $\{Q_i\}$ into a bins (where $a = 10$), from the entire training set T . Subsequently, from this distribution, we select 25% of the data (N) optimally with maximized diversity where each bin contains approximately $\frac{N}{a}$ LiDAR tree samples. We can only approximate $\frac{N}{a}$ for two reasons: (1) $\frac{N}{a} \in R$, but the number of samples in each bin is an integer; (2) in the case where the number of samples per bin is less than $\frac{N}{a}$, the sample gets allotted to another bin such that t_n remains 25% of the number of total samples. We then characterize the distribution of data quality by calculating our diversification index, DI , using Equation (7):

$$DI = 1 - \frac{\sum_{i=1}^a \left| t_i - \frac{t_n}{a} \right|}{2(t_n - \frac{t_n}{a})} \quad (7)$$

where a = number of bins, N = number of training data, N_i = number of training data in the i th bin, such that $0 \leq DI \leq 1$.

When the distribution of the dataset across bins is uniform, each bin contains $(\frac{N}{a})$ samples, $DI = 1$, indicating the most diversified distribution of Q_i contained in the training data. When selected N is contained in a single bin, $DI = 0$, indicating the least diversity in the distribution of Q_i . To test the relationship between DI , and classification accuracy, we restrict our range of Q_i ten times, each time the range begins from 0.0 and ends at a threshold increasing from 0.1 through to 1.0 in 0.1 unit increments; within each of these ranges we calculate: (1) DI ; (2) the classification accuracy for R-RF without “unknown” samples in the validation data; (3) classification accuracy for R-RF with “unknown” samples in the validation data; (4) classification accuracy for E-RF to be discussed in the next section, without “unknown” samples in the validation data; and (5) classification accuracy for E-RF to be discussed in the next section, with “unknown” samples in the validation data.

3.4. Goal 2: Ensemble Random Forests (E-RF)

The optimized set of training samples (47 LiDAR trees, 10 bins) from the previous section are used for training the base classifiers, the remaining 139 trees are used as validation data. The selected training data are used for training the three base classifiers, h_p , h_o , and h_m with maximized diversity, allowing the base classifiers to learn as much variability as possible.

The parallel model is summarized in Table 2, where h_p , h_o , and h_m are the base classifiers. In the model, the base classifier simultaneously classifies pine (non-pine), poplar (non-poplar), and maple (non-maple) at the same time. In cases where there are no conflicts in decisions made among the base classifiers (i.e., Cases 1, 2, 3, 8), the final decision is made by the classifier voted for by a positive case (Cases 1, 2, and 3). Where all three classifiers vote negatively (Case 8), the tree will be labeled as “unknown”. In the cases where classifiers are conflicted (Cases 4–7), the final decision is made by the classifier that has the majority positive vote. A majority vote is calculated by Random Forests over T base classifiers described in Equation (8) [50]. The classification accuracy of this model, ensemble Random Forests (E-RF) is compared with regular Random Forests (R-RF).

$$y^* = \operatorname{argmax}_{y \in L} \frac{1}{T} \sum_{i=1}^T p_i(y|X) \quad (8)$$

where $X \subset \mathbb{R}^f$ is the features selected for classification; $p_i(y|X)$ is the binary indicator variable for voting the L instances with given X ; y is the predicted class label such that $y \in L$, and y^* is the final prediction for a particular base classifier.

Table 2. Summary of the parallel ensemble model.

Case #	Decision Made by h_p	Decision Made by h_o	Decision Made by h_m	Final Decision
1	p	o'	m'	p
2	p'	o	m'	o
3	p'	o'	m	m
4	p	o	m'	If max vote (p) > max vote(o), p; else, o
5	p'	o	m	
6	p	o'	m	If max vote(o) > max vote(m), o; else, m
7	p	o	m	
8	p'	o'	m'	If max vote(p) > max vote(m), p; else, m
				Label with max(max vote)
				u

p = "pine"; p' = "non-pine"; o = "poplar"; p' = "non-poplar"; m = "maple"; m' = "non-maple"; u = "unknown".

4. Results

4.1. Quantifying Data Quality and Its Effect on Classification Accuracy

To quantify the quality of the LiDAR data for individual trees, Q_i for each LiDAR tree is calculated with Equation (6). Figure 1a shows the frequency distribution of Q_i over 10 bins for 160 trees, and Figure 1b shows the frequency distribution of Q_i for selected training samples ($N = 47$) with maximum diversity. In Figure 1a, it is shown that most of the LiDAR tree data collected are considered easily identifiable, with $0.8 \leq Q_i \leq 1.0$, meaning that these trees have a chance in excess of 80% to be classified correctly. To visualize the differences between the calculated ratios, six examples of LiDAR trees with lower Q_i are shown in Figure 2a and six examples with higher Q_i are shown in Figure 2b. From Figure 2b, it is observed visually that these LiDAR trees have less occlusion, fewer overlapping tree crown points with adjacent trees and are larger in size (more data points per tree) and are therefore more easily classified.

To study the relationship between diversity contained in the training data indicated by DI and classification accuracy, we plot the classification accuracy over different values of DI . To separate the performance of the classifiers with the presence of "unknown" data, we included two sets of curves. Figure 3 displays the result of the analysis performed with and without the "unknown" samples in the validation data, and is performed by using R-RF (regular Random Forests) and E-RF (parallel ensemble Random Forests). In Figure 3, results show that classification accuracies increase as DI increases, with two methods of classification, E-RF and R-RF. Additionally, we calculated the Pearson product-moment correlation coefficients for R-RF without an "unknown" class in the validation data, R-RF with an "unknown" class in the validation data, E-RF without an "unknown" class in the validation data, and E-RF with an "unknown" class in the validation data, as 0.94, 0.93, 0.98, and 0.83, respectively.

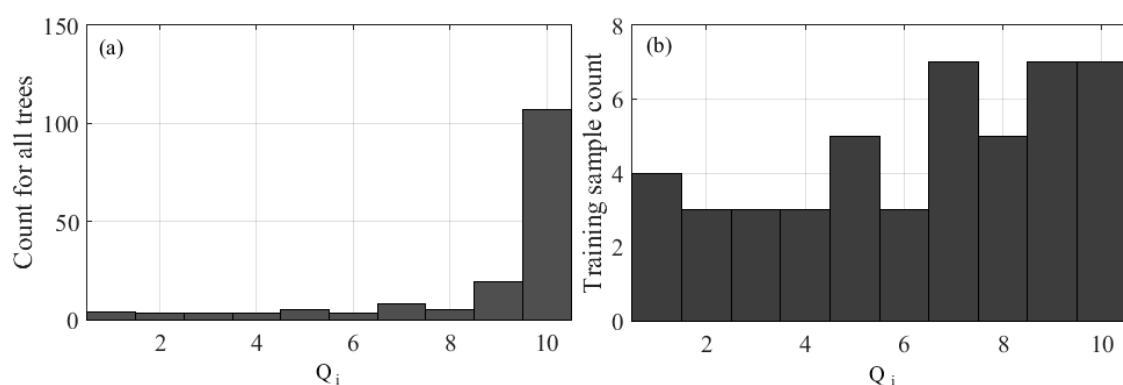


Figure 1. (a) Frequency distribution of Q_i for all 160 trees; (b) Frequency distribution of Q_i for selected 47 training sample trees.

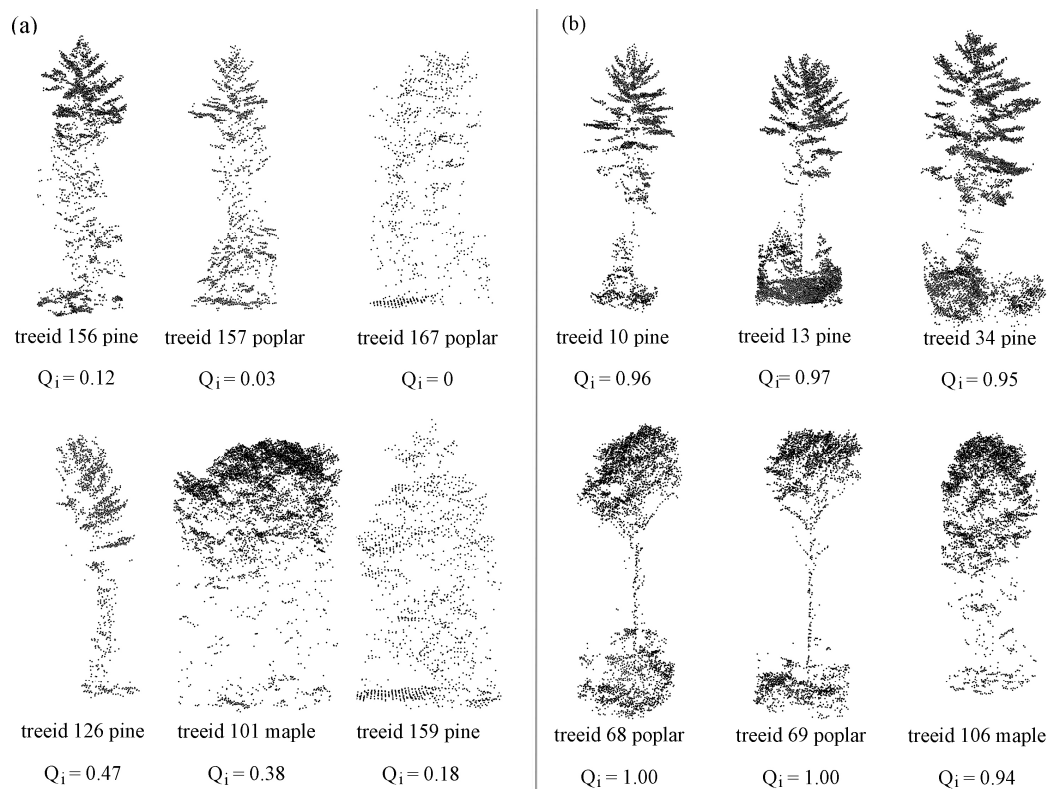


Figure 2. (a) Example LiDAR trees that have $Q_i < 0.5$; and (b) example LiDAR trees that have $Q_i > 0.9$.

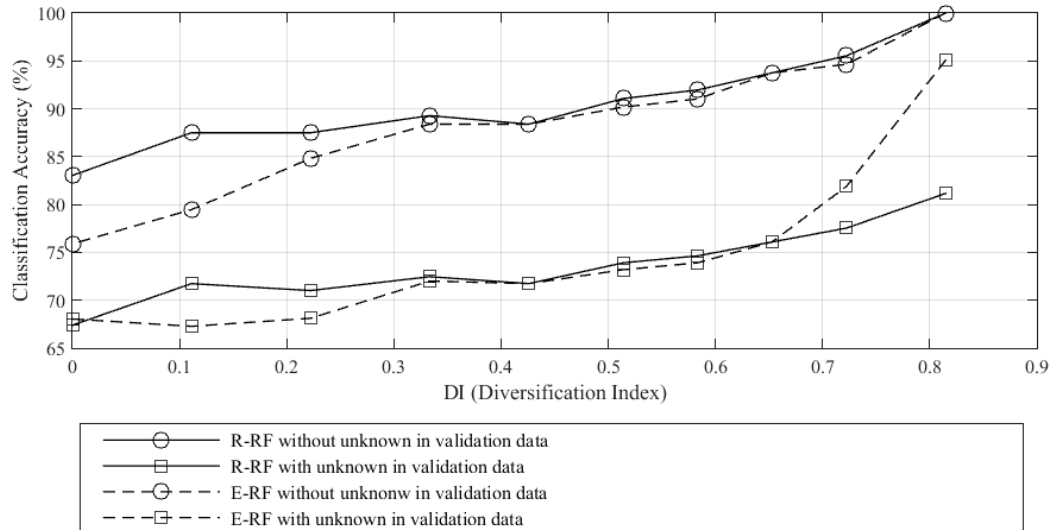


Figure 3. Classification accuracy changes relative to the number of bins being included in the training sample.

To supplement the importance of including a diversified set of Q_i in the training data, we perform E-RF with unknown samples in the validation 40 times where the training data are randomly selected. This test is performed under the condition where we have kept the variables the same as in Figure 3, such as the percentage of training data and validation data, variables used in E-RF including N_{tree} , M_{try} , and K_{node} . The classification results are shown in Figure 4. Within the 40 randomly selected samples, the classification accuracy ranges from 60% to 88%, caused by the difference in training

sample selection. This indicates the importance of selecting a representative training sample set in order to achieve accurate classification.

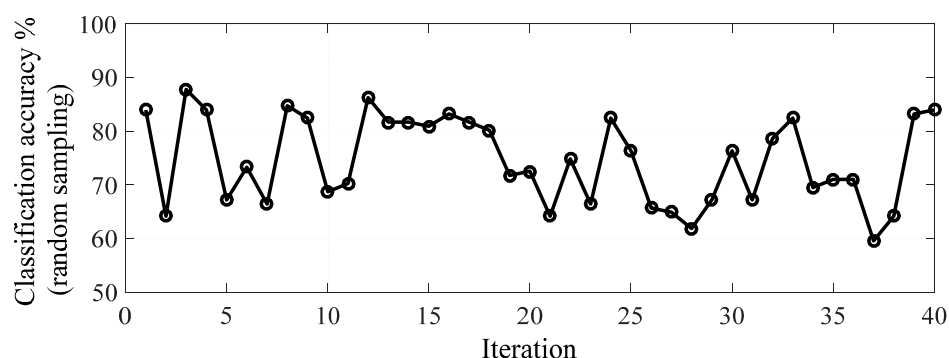


Figure 4. Classification accuracy (%) performed using Random Forests (Ensemble method with “unknown” class in validation data) with 40 different training samples set selected randomly.

4.2. Classification for Random Forests and Ensembled Random Forests

The classification results for R-RF and E-RF is summarized in Tables 3–5. The confusion matrix obtained from E-RF using random sampling is shown in Table 3 and the confusion matrix obtained from E-RF using diversified training sample is shown in Table 4. The confusion matrix obtained from R-RF using diversified training sample is presented in Table 5.

Table 3. Confusion matrix obtained from E-RF using random sampling; values are averaged over 20 different random samples.

		Measured				
		<i>Pine</i>	<i>Poplar</i>	<i>Maple</i>	<i>Unknown</i>	Commission Error
predicted	<i>pine</i>	37.45	3.85	0.05	6.75	0.22
	<i>poplar</i>	4.05	35.70	0.20	2.45	0.16
	<i>maple</i>	1.15	0.15	23.90	12.55	0.37
	<i>unknown</i>	4.15	1.80	0.55	4.25	0.60
Omission error		0.20	0.14	0.03	0.84	
Overall Accuracy: 72.8%						

Table 4. Confusion matrix obtained from E-RF using diversified sampling, where values are averaged over running the E-RF 20 times.

		Measured				
		<i>Pine</i>	<i>Poplar</i>	<i>Maple</i>	<i>Unknown</i>	Commission Error
predicted	<i>pine</i>	47.00	0.00	0.20	3.40	0.07
	<i>poplar</i>	0.00	41.00	0.00	2.95	0.07
	<i>maple</i>	0.00	0.00	23.80	2.00	0.08
	<i>unknown</i>	0.00	0.00	0.00	18.65	0.00
Omission error		0.00	0.00	0.01	0.31	
Overall Accuracy: 93.8%						

Table 5. Confusion matrix obtained from R-RF classification using diversified sampling, where values are averaged over running the ensemble 20 times.

		Measured				
		<i>Pine</i>	<i>Poplar</i>	<i>Maple</i>	<i>Unknown</i>	Commission Error
predicted	<i>pine</i>	39.60	0.00	0.00	0.00	0.00
	<i>poplar</i>	7.40	25.00	2.55	4.20	0.36
	<i>maple</i>	0.00	0.00	38.45	1.00	0.03
	<i>unknown</i>	0.00	1.00	0.00	19.8	0.05
Omission error		0.16	0.04	0.06	0.21	
Overall Accuracy: 88.4%						

By comparing Tables 3 and 4, the overall accuracy improved from $\bar{x} = 72.8\%$, $s = 0.026$ to $\bar{x} = 93.8\%$, $s = 0.027$ when the training samples were maximally diverse. By comparing Table 4 to Table 5, the overall accuracy improved from $\bar{x} = 88.4\%$, $s = 0.017$ to $\bar{x} = 93.8\%$ when E-RF is performed as opposed to R-RF (\bar{x} represents the mean and s represents the standard deviation).

We observe that the error of the “unknown” class results from the difficulty in differentiating genera that appear similar within a LiDAR point cloud. It is found that birch and oak trees are normally mistaken as maple trees and spruces and larch trees are normally mistaken as pine trees. Figure 5 shows an example of each mistaken pair.

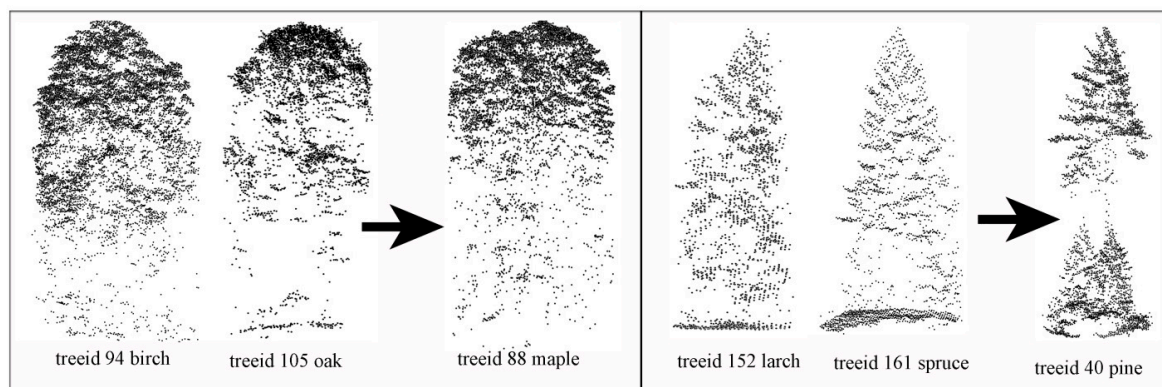


Figure 5. Example misclassification from birch and oak to maple; larch and spruce to pine.

5. Discussion

The majority of the data collected for this research have similar properties; by randomly selecting the training data we ensure that the training data will have a similar statistical distribution as does the population. Given that the majority of the training data is similar, any validation data that deviates from the training data will be misclassified. To solve this problem, the quality of collected samples is quantified by computing Q_i , and the frequency distribution examined (Figure 1a) so that a diversified training sample is obtained (Figure 1b). When training samples are selected randomly from the distribution as shown in Figure 1a, there is a higher chance of having empty bins in the training data due to lower frequencies in the lower Q_i bins, resulting in lower diversity in the training data. As a result, when base classifiers are trained based on randomly selected samples (with lower diversity), more emphasis will be placed on the LiDAR trees that have a high Q_i . We propose using a distribution such as in Figure 1b for training sample selection in order to maximize diversity where training samples contain trees that have both low, and high Q_i , resulting in improved classification accuracy with such a selection. The classification accuracy assessment is compared through two comparative experiments: in the first experiment the classification results are compared from E-RF random sampling versus E-RF with diversified training samples (Figure 1b) by comparing results shown in Tables 3 and 4; in the

second experiment, the classification results are compared between R-RF and E-RF, both using the diversified training samples by comparing results shown in Tables 4 and 5.

In Figure 3, the smallest value of DI is zero when all training samples are drawn from one class; this is the least diversified example. Alternatively, the largest value, $DI = 0.82$ is the most diversified training sample attainable (Figure 1b), where the samples were not able to distribute uniformly across all calculated Q_i bins. In theory, $DI = 1.0$ if all the samples are distributed uniformly across all bins. When DI increases, the classification accuracies also increase (both with and without unknown samples in the validation data). This shows that diversity in the training data improves classification accuracy, regardless of the choice of algorithm or the presence of an “unknown” class in the validation set. This result indicates that taking DI into consideration during training is advisable. We believe that the variations in classification accuracy observed in Figure 4 are attributed to differences arising from the selection of training samples. We further observe that when unknown samples are included in the validation data, E-RF is outperformed by R-RF by 4.3% and 13.9% when $DI = 0.72$ and 0.82 respectively. This refers to training data which include values of Q_i from 0.0–0.2 and 0.0–0.1, respectively. These results imply the importance of including low Q_i samples in the training data. Given that E-RF outperformed R-RF implies that the higher DI included in the training data are more effective than the R-RF and in the future, E-RF should be a preferred choice for classification, especially with maximized DI . Conversely, if DI is small, R-RF and E-RF exhibit similar performance with the presence of unknown samples in the validation set. Further, R-RF outperformed E-RF in the case where no unknown class is present in the data set and when DI is small (<0.33). E-RF shows a lower classification accuracy compared to R-RF by 8.0% and 7.1% when $DI = 0$ and 0.11 , respectively, and when validation data are absent; in these cases E-RF should be avoided. However, this is normally not a real life situation where the validation data contain only classes obtained from the training data. Furthermore, it is hard to guarantee that the field-validated data will contain all tree species in the forest plot given the complexity of the natural environment.

Computed Shannon entropy for the Q_i ranges and the plot of classification accuracies over different entropy values present corresponding results. However, since DI values are bounded between 0 and 1, their interpretation is easier and we propose the use of DI over entropy.

By comparing Tables 3 and 4, the commission and omission errors were reduced to 60% and 53% respectively for the “unknown” class when the training sample is diversified. This is because the “unknown” classes are labeled based on the classification accuracies of negative samples from the base classifiers. When the training samples have low diversity, it indicates that most of the training samples appear to be similar and therefore the distributions of the features related to the specific class are narrow. As a result, any validation data that deviate from the training data will be classified as negative (high omission error, 84% in Table 3, reduced to 31% in Table 4). Conversely, when the training samples contains high diversity, the distributions of values for the features that train the base classifiers are broader allowing a broader definition of a specific class, reducing the chance of classifying negative samples as positive (commission error of “unknown” class reduced from 60% to 0%).

Since the training sample contains only pine, poplar, and maple, by default, the class labels generated from Random Forests will be the three mentioned classes. However, in order to be able to compare with ensemble methods, an additional condition is included so that “unknown” classes can be identified with Random Forests alone. As mentioned, for each validation sample in Random Forests, the final class label is assigned by the class with the majority vote stated in Equation (8). The additional condition is that if the majority vote calculated for the particular tree is less than 67%, it should be classified as “unknown” instead of one of the three classes. The additional condition ensures that the “unknown” class is assigned when the majority vote calculated for the particular tree is $<67\%$. This threshold represents at least a two-thirds vote among the three classes in order to be valid. Table 5 shows the confusion matrix obtained from Random Forests using diversified samples, where values are averaged by running the ensemble 20 times. Comparison of Table 5 with Table 4 shows that E-RF classification yields an overall accuracy of 93.8% whereas R-RF yields an overall accuracy of 88.4%.

The omission error for R-RF is less than E-RF but the commission error for classifying poplar for R-RF is higher. With E-RF, the omission error for pine, poplar, and maple is also lower.

There are several reasons proposed for this improvement over the traditional multi-class classification: (1) E-RF are able to generate an “unknown” class without a pre-defined threshold; (2) the method does not require the presence of an “unknown” class in the training data; (3) the implementation of binary classification is simple, so that in the future, if another class label is being collected in the field, an additional binary classifier can be built but all the previous training results can be re-used without any changes; and (4) the method of combination can be altered and additional rules can be implemented to improve the aggregation of information such as in Table 3. The methodology and discussion in this paper applies beyond the use of LiDAR for forest applications, as it is natural to obtain 3D data with varying quality. This study aids in making better choices related to the selection of training data, such that classification accuracy can also be boosted. Furthermore, the generation of an “unknown” class is practical for classification problems when there is a presence of classes in the testing data that were not included in the training data.

6. Conclusions

This study has two important conclusions, the first relates to the importance of diversifying the quality of training samples in order to improve classification accuracy. The quality of LiDAR tree samples can vary due to reasons such as shadow or occlusions from scanning angle effects, overlapping of adjacent trees or trees that appear to have multiple tree tops making tree crown isolation difficult. The problem of quality inconsistency, an inherent problem with 3D point cloud data, is an unavoidable problem that naturally occurs with all types of objects, not only trees. Intuitively we assume that a better sample quality should improve classification accuracy; our results show that high diversity in training samples helps with classifying negative instances and is especially useful in determining the “unknown” class.

This paper quantifies data quality, Q_i , a term that is normally prone to subjective definition, and shows that by increasing diversity in the training data, classification accuracy can be improved. Diversity is a concept being applied to improve learning capacity in machine learning and we have included additional diversity beyond what the Random Forests algorithm already provides. By computing the DI we also quantify the amount of diversification in the dataset, the diversified Q_i range (high DI) in the training data provide an improved accuracy. This paper was guided by the hypothesis that when the classifiers are trained with samples that contain as much variability as possible, there will be a better decision boundary between the positive and negative class labels. With a broader definition of each genus, results show that classification accuracy can be improved from 72.8% to 93.8%, both using E-RF, with the accuracy especially improved for the classification of negative labels. In the future, when we have increased number of samples, it would be possible to partition part of the data for the computation of Q_i and select training samples from the computation. While we are unable to do this in this experiment, we have shown that the increased DI would provide improved classification accuracy and provides a basis for future research in this direction.

The second conclusion we can draw from this paper is the effectiveness of E-RF compared to R-RF. In E-RF, we have designed the base classifiers such that k binary classification models are required for classifying k classes. By combining these base classifiers instead of using R-RF as a multi-class classifier, a better accuracy result can be obtained (improved from 88.4% to 93.8%). Although the classification accuracy can be improved by implementing E-RF as suggested in comparing Tables 3 and 4, nevertheless, the omission error for the “unknown” class is still the highest in Table 4 when compared to other classes. This is because of the confusion of the similarities shared among genera. Examples are birch and oak trees are normally mistaken as maple trees; spruces and larch trees are normally mistaken as pine trees (Figure 5). This problem can be resolved if new features for classification can be developed that are independent of geometry. This can increase the diversity in the

classification features and enable building base classifiers that are capable of processing different types of information.

In this paper, we have proposed a method to quantify data quality by computing Q_i . We have also examined the importance of having a diverse training data set in order to improve classification accuracy. Experimentation has shown that as DI increases, the classification accuracy increases. In our dataset, it is especially important to include the data with the lowest quality in the training data set in order to maximize classification accuracy; results show that classification accuracy improves from 72.8% (random sampling) to 93.8% (diversified sampling). To tackle the problem of having more classes in the validation data than training data, we proposed E-RF for the generation of an “unknown” class without implementing any threshold. Results show that E-RF (93.8%) outperformed R-RF (88.4%) when the DI was maximized.

Acknowledgments: This research was jointly funded by GeoDigital International Inc., the Ontario Centres for Excellence, and two Discovery Grants from the Natural Sciences and Engineering Research Council of Canada. We thank Richard Pollock, Konstantin Lisitsyn, Doug Parent, and Yulia Lazukova at GeoDigital International Inc. for their assistance in preparing the data. We also acknowledge the extensive field validation assistance provided to the primary three authors of this paper by Junji Zhang, Jili Li, and Yoonseok Jwa, (York University, Canada), and by Nakhyn Song (Inha University, South Korea) for acquiring field surveying data for this study.

Author Contributions: Connie Ko, Gunho Sohn, Tarmo K. Remmel, and John R. Miller wrote and edited the manuscript. Connie Ko designed and implemented the experiments with the advice of Gunho Sohn and Tarmo K. Remmel. All authors discussed the results, presented in the manuscript.

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

References

1. Vauhkonen, J.; Maltamo, M.; McRoberts, R.E.; Næsset, E. Introduction to forestry applications of airborne laser scanning. In *Applications of Airborne Laser Scanning: Concepts and Case Studies Forestry*; Maltamo, M., Næsset, E., Vauhkonen, J., Eds.; Springer: Dordrecht, The Netherlands, 2014; pp. 1–16.
2. Holmgren, J.; Persson, A. Identifying species of individual trees using airborne laser scanning. *Remote Sens. Environ.* **2004**, *90*, 415–423. [[CrossRef](#)]
3. Brandtberg, T. Classifying individual tree species under leaf-off and leaf-on conditions using airborne LiDAR. *ISPRS J. Photogramm. Remote Sens.* **2007**, *61*, 325–340. [[CrossRef](#)]
4. Holmgren, J.; Persson, Å.; Söderman, U. Species identification of individual trees by combining high resolution LiDAR data with multi-spectral images. *Int. J. Remote Sens.* **2008**, *29*, 1537–1552. [[CrossRef](#)]
5. Kato, A.; Moskal, L.M.; Schiess, P.; Swanson, M.E.; Calhoun, D.; Stuetzle, W. Capturing tree crown formation through implicit surface reconstruction using airborne LiDAR data. *Remote Sens. Environ.* **2009**, *113*, 1148–1162. [[CrossRef](#)]
6. Ørka, H.O.; Næsset, E.; Bollandsås, O.M. Utilizing Airborne Laser Intensity for Tree Species Classification. Available online: http://www.isprs.org/proceedings/XXXVI/3-W52/final_papers/Oerka_2007.pdf (accessed on 14 September 2007).
7. Ørka, H.O.; Næsset, E.; Bollandsås, O.M. Classifying species of individual trees by intensity and structure features derived from airborne laser scanner data. *Remote Sens. Environ.* **2009**, *113*, 1163–1174. [[CrossRef](#)]
8. Vauhkonen, J.; Tokola, T.; Maltamo, M.; Packalen, P. Effects of pulse density on predicting characteristics of individual trees of Scandinavian commercial species using alpha shape metrics based on ALS data. *Can. J. Remote Sens.* **2008**, *34*, S441–S459. [[CrossRef](#)]
9. Vauhkonen, J.; Tokola, T.; Packalen, P.; Maltamo, M. Identification of Scandinavian commercial species of individual trees from airborne laser scanning data using alpha shape metrics. *For. Sci.* **2009**, *55*, 37–47.
10. Vauhkonen, J.; Korpela, I.; Maltamo, M.; Tokola, T. Imputation of single-tree attributes using airborne laser scanning-based height, intensity, and alpha shape metrics. *Remote Sens. Environ.* **2010**, *114*, 1263–1276. [[CrossRef](#)]
11. Korpela, I.; Ørka, H.O.; Maltamo, M.; Tokola, T. Tree species classification using airborne LiDAR—Effects of stand and tree parameters, downsizing of training set, intensity normalization and sensor type. *Silva Fenn.* **2010**, *44*, 319–339. [[CrossRef](#)]

12. Kim, S.; Mcgaughey, R.J.; Andersen, H.; Schreuder, G. Tree species differentiation using intensity data derived from leaf-on and leaf-off airborne laser scanner data. *Remote Sens. Environ.* **2009**, *113*, 1575–1586. [[CrossRef](#)]
13. Kim, S.; Hinckley, T.; Briggs, D. Classifying individual tree genera using stepwise cluster analysis based on height and intensity metrics derived from airborne laser scanner data. *Remote Sens. Environ.* **2011**, *115*, 3329–3342. [[CrossRef](#)]
14. Mantero, P.; Moser, G.; Serpico, S.B. Partially supervised classification of remote sensing images through SVM-based probability density estimation. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 559–570. [[CrossRef](#)]
15. Muñoz-Marí, J.; Bruzzone, L.; Camps-Valls, G. A support vector domain description approach to supervised classification of remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 2683–2692. [[CrossRef](#)]
16. Ruta, D.; Gabrys, B. An overview of classifier fusion methods. *Comput. Inf. Syst.* **2000**, *7*, 1–10.
17. Samadzadegan, F.; Bigdeli, B.; Ramzi, P. A multiple classifier system for classification of LIDAR remote sensing data using multi-class SVM. *Multi. Classif. Syst.* **2010**, *5997*, 254–263.
18. Zhou, Z.H. *Ensemble Methods: Foundations and Algorithms* (Chapman & Hall/Crc Machine Learning & Pattern Recognition); CRC Press: Boca Raton, FL, USA, 2012.
19. Oza, N.C.; Tumer, K. Classifier ensembles: Select real-world applications. *Inf. Fusion* **2008**, *9*, 4–20. [[CrossRef](#)]
20. Bell, D.; Guan, J.W.; Bi, Y. On combining classifiers mass functions for text categorization. *IEEE Trans. Knowl. Data Eng.* **2005**, *17*, 1307–1319. [[CrossRef](#)]
21. Koerich, A.L.; Sabourin, R.; Suen, C.Y. Recognition and verification of unconstrained handwritten words. *IEEE Trans. Pattern Anal. Mach. Intell.* **2005**, *27*, 1509–1521. [[CrossRef](#)] [[PubMed](#)]
22. Lodha, S.K.; Kreps, E.J.; Helmbold, D.P.; Fitzpatrick, D. Aerial LiDAR data classification using support vector machines (SVM). In Proceedings of the IEEE International Symposium on 3D Data Processing, Visualization, and Transmission, Chapel Hill, NC, USA, 14–16 June 2006.
23. Kumar, S.; Ghosh, J.; Crawford, M.M. Hierarchical fusion of multiple classifiers for hyperspectral data analysis. *Pattern Anal. Appl.* **2002**, *5*, 210–220. [[CrossRef](#)]
24. Kavzoglu, T.; Colkesen, I. An assessment of the effectiveness of a rotation forest ensemble for land-use and land-cover mapping. *Int. J. Remote Sens.* **2013**, *34*, 4224–4241. [[CrossRef](#)]
25. Galar, M.; Fernández, A.; Barrenechea, E.; Bustince, H.; Herrera, F. An overview of ensemble methods for binary classifiers in multi-class problems: Experimental study on one-vs.-one and one-vs.-all schemes. *Pattern Recognit.* **2011**, *44*, 1761–1776. [[CrossRef](#)]
26. Hastie, T.; Tibshirani, R. Classification by pairwise coupling. *Ann. Stat.* **1998**, *26*, 451–471. [[CrossRef](#)]
27. Rifkin, R.; Klautau, A. In defense of one-vs.-all classification. *J. Mach. Learn. Res.* **2004**, *5*, 101–141.
28. Dietterich, T.G. An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting and randomization. *Mach. Learn.* **2000**, *40*, 139–158. [[CrossRef](#)]
29. Hashemi, S.; Yang, Y.; Mirzamomen, Z.; Kangavar, M. Adapted one-vs.-all decision trees for data stream classification. *IEEE Trans. Knowl. Data Eng.* **2009**, *21*, 624–637. [[CrossRef](#)]
30. Hsu, C.W.; Lin, C.J. A comparison of methods for multi-class support vector machines. *IEEE Trans. Neural Netw.* **2002**, *13*, 415–425. [[PubMed](#)]
31. Duan, K.B.; Rajapakse, J.C.; Nguyen, M.N. One-vs.-one and one-vs.-all multiclass SVM-RFE for gene selection in cancer classification. In *EvoBIO 2007*; Marchiori, E., More, J.H., Rajapakse, J.C., Eds.; Springer: Heidelberg, Germany, 2007; Volume 4447, pp. 47–56.
32. Milgram, J.; Cheriet, M.; Sabourin, R. One against “one” or “one against all”: Which one is better for handwriting recognition with SVMs? In Proceedings of 10th International Workshop on Frontiers in Handwriting Recognition, La Baule, France, 5 October 2006.
33. Yi, L.; Zheng, Y.F. One-against-all multi-class SVM classification using reliability measures. In Proceedings of the IEEE International Joint Conference on Neural Networks (IJCNN), Montréal, QC, Canada, 31 July–4 August 2005.
34. Chawla, N.V.; Japkowicz, N.; Kolcz, A. Editorial: Special issue on learning from imbalanced data sets. *ACM SIGKDD Explor. Newsl.* **2004**, *6*, 1–6. [[CrossRef](#)]
35. Fernández, A.; López, V.; Galar, M.; del Jesús, M.J.; Herrera, F. Analysing the classification of imbalanced data-sets with multiple classes: Binarization techniques and ad-hoc approaches. *Knowl. Based Syst.* **2013**, *42*, 97–110. [[CrossRef](#)]

36. Barandela, R.; Valdovinos, R.M.; Sánchez, J.S.; Ferri, F.J. The imbalanced training sample problem: Under or over sampling? In *Structural, Syntactic, and Statistical Pattern Recognition*; Fred, A., Caelli, T.M., Duin, R.P.W., Campilho, A.C., de Ridder, D., Eds.; Springer: Heidelberg, Germany, 2004; Volume 3138, pp. 806–814.
37. Weiss, G. Mining with rarity: A unifying framework. *ACM SIGKDD Explor. Newsl.* **2004**, *6*, 7–19. [[CrossRef](#)]
38. Japkowicz, N. *Learning from Imbalanced Data Sets: A Comparison of Various Strategies*; AAAI Technical Report, WS-00-05; AAAI: Palo Alto, CA, USA, 2000.
39. Kuncheva, L.; Whitaker, C. Measures of diversity in classifier ensembles. *Mach. Learn.* **2003**, *51*, 181–207. [[CrossRef](#)]
40. Chen, C.; Liaw, A.; Breiman, L. *Using Random Forest to Learn Imbalanced Data*; Technical Report 666; University of California: Berkeley, CA, USA, 2004.
41. Fawagreh, K.; Gaber, M.M.; Elyan, E. Diversified random forests using random subspaces. In *Intelligent Data Engineering and Automated Learning—IDEAL 2014*; Springer: Salamanca, Spain, 2014; pp. 85–92.
42. Ali, K.M.; Pazzani, M.J. Error reduction through learning multiple descriptions. *Mach. Learn.* **1996**, *24*, 173–202. [[CrossRef](#)]
43. Breiman, L. Arcing classifiers. *Ann. Stat.* **1998**, *26*, 801–824.
44. Dietterich, T.G.; Bakiri, G. Solving multiclass learning problems via error-correcting output codes. *J. Artif. Intell. Res.* **1995**, *2*, 263–286.
45. Bryll, R.; Gutierrez-Osuna, R.; Quek, R. Attribute bagging: Improving accuracy of classifier ensembles by using random feature subsets. *Pattern Recognit.* **2003**, *36*, 1291–1302. [[CrossRef](#)]
46. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
47. Liaw, A.; Wiener, M. Classification and regression by random Forest. *R. News.* **2002**, *2*, 18–22.
48. Ko, C.; Sohn, G.; Rimmel, T.K. Tree genera classification with geometric features from high-density airborne LiDAR. *Can. J. Remote Sens.* **2013**, *39*, S73–S85. [[CrossRef](#)]
49. R Development Core Team. The R Project for Statistical Computing. Available online: <http://www.R-project.org/> (accessed on 12 July 2016).
50. Schwing, A.; Zach, C.; Zheng, Y.; Pollefeys, M. Adaptive random forest—How many “experts” to ask before making a decision? In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Providence, RI, USA, 20–25 June 2011.



© 2016 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/>).