



# Audrey P. Riddell<sup>1</sup>, Stephen A. Fitzgerald<sup>2</sup>, Chu Qi<sup>2</sup> and Bogdan M. Strimbu<sup>2,\*</sup>

- <sup>1</sup> Division of Environmental Planning, California Department of Transportation, 100 S Main, Los Angeles, CA 90012, USA; audrey.riddell@dot.ca.gov
- <sup>2</sup> Department of Forest Engineering, Resources, and Management, Oregon State University, 280 Peavy Hall, Corvallis, OR 97331, USA; stephen.fitzgerald@oregonstate.edu (S.A.F.); chu.qi@oregonstate.edu (C.Q.)
- \* Correspondence: bogdan.strimbu@oregonstate.edu; Tel.: +1-541-737-1604

Received: 22 August 2020; Accepted: 10 October 2020; Published: 13 October 2020



Abstract: Forest species classifications are becoming increasingly automated as advances are made in machine learning. Complex algorithms can reach high accuracies, but are not always suitable for small-scale classifications, which may benefit from simpler conventional methods. The goal of this classification was to identify contiguous stands of ponderosa pine (Pinus ponderosa Douglas ex Lawson) against a mix of forest and non-forest background in the southern Willamette Valley, Oregon. The study area is approximately 816,600 ha, considerably larger than most study areas used for presenting techniques for tree species classification. To achieve the objective, we used two classification procedures, one parametric and one non-parametric. For the parametric method, we selected the maximum likelihood (ML) algorithm, whereas for the non-parametric method we chose the random forest (RF) algorithm. To identify ponderosa pine, we used 1 m spatial resolution red-green-blue-infrared (RGBI) aerial images supplied by the U.S. National Agriculture Imagery Program (NAIP) and 1 m spatial resolution canopy height models (CHMs) provided by the Oregon Department of Geology and Mineral Industries (DOGAMI). We tested four data variations for each method: Aerial imagery, CHM-masked aerial imagery, aerial imagery with an additional CHM band, and CHM-masked aerial imagery with a CHM band. The parametric classifications of aerial imagery alone reached an average kappa coefficient of 0.29, which increased to 0.51 when masked with CHM data. The incorporation of CHM data as a fifth band resulted in a similar improvement in kappa (0.47), but the most effective parametric method was the incorporation of CHM data as both a fifth band and a post-classification mask, resulting in a kappa coefficient of 0.89. The non-parametric classification of aerial imagery achieved a mean validation kappa coefficient of 0.85 collectively and 0.90 individually, which only increased by approximately 0.01 or less when the CHM masks were applied. The addition of the CHM band increased the kappa value to 0.91 for both individual and collective tile classifications. The highest kappa of all methods was achieved through five-band non-parametric classification with the addition of the CHM band (0.94) for both collective and individual classifications. Our results suggest that parametric methods, when enhanced with a CHM mask, could be suitable for large-area, small-scale classifications based on RGBI imagery, but a non-parametric classification of fused spectral and height data will generally achieve the highest accuracy for large, unbalanced datasets.

**Keywords:** tree species classification; binary maps; maximum likelihood; random forest; canopy height model



#### 1. Introduction

Remote sensing has become an indispensable tool in forest management, due to the spatial scales of forested land and management timelines. The use of satellite or aerial imagery in land cover analysis enables land managers and planners to make landscape level decisions quickly and accurately. This is crucial for managers who work with large tracts of land, or researchers performing small-scale analyses. Advances in remote sensing introduce increasingly fine-resolution datasets, both spatially and spectrally. This allows more precise measures of variations across the landscape, such as forest species composition. Fortunately, many of these datasets are freely available for landscape analyses.

The relatively recent availability of sub-meter spatial resolution introduced new opportunities for precise measurements and predictions of land cover, allowing managers to identify land classes at a very fine scale. With the introduction of such high spatial and spectral resolution imagery, forest vegetation classification can be performed at species-level precision [1-3]. Many recent studies conducted on species-level classification of forested landscapes make use of hyperspectral imagery with considerable success [4–7]. Fassnacht et al. (2016) found that hyperspectral sensors were used most frequently in tree species classification [8]. Hyperspectral imagery provides an abundance of valuable information when distinguishing between vegetation classes but can be cost-prohibitive due to the lack of publicly available hyperspectral datasets. It can also come at the cost of spatial resolution, making fine-scale feature recognition difficult, such as the publicly available hyperspectral Hyperion dataset, which has a 30m spatial resolution. In temperate forests, multispectral datasets with NIR bands often provide sufficient spectral data to distinguish between vegetation classes [3], especially when combined with light detection and ranging (LIDAR) data [9]. The combination of lidar data with hyperspectral or multispectral data has been nearly universal in recent tree species classifications [8]. Species-level classifications greatly benefit from the addition of lidar-derived height data [4,6,10]. These data can be integrated at several different stages of the classification process [11,12].

Most recent studies have made use of non-parametric algorithms for tree species classification, particularly support vector machine and random forest (RF) [5,8,9]. These algorithms do not assume normality within the dataset and allow a smoother integration of non-spectral data. While non-parametric algorithms generally produce the highest reported accuracies in tree species classifications, parametric algorithms like maximum likelihood (ML) are only surpassed in accuracy by a few percentage points in certain cases [13,14]. In some studies, a simple classifier is preferable, due to its low processing power requirements, user friendliness, and ease of access in many remote sensing software. Parametric classifiers are also suitable when ancillary data, such as spatial segmentation data, are not integrated into the algorithm [15]. ML can even surpass non-parametric algorithms in highly heterogeneous regions when classifying aerial imagery alone [14]. This is particularly true in very large study areas [13].

As data quality progresses and improves, so do the classification techniques performed upon them. Many of the advances of the past few decades are concentrated in machine learning, the ultimate objective of which is to achieve or surpass the pattern recognition capacity of a human [16]. Machine learning is a catch-all term that describes many different advanced algorithms used to perform difficult classifications or complex regressions [17]. The complexity of machine learning algorithms covers a broad spectrum, from basic decision trees to "black box" deep neural networks. The success of advanced models is undeniable, as is evident in the abundance of recent publications showcasing the use of machine learning algorithms in land cover classifications [5,13,18]. However, the complexity that makes these algorithms so effective also creates an obstacle to those with limited programming experience. High-level research certainly merits the use of such advanced techniques, but for a user with limited exposure to machine learning, can a more simplistic classification algorithm with extensive human training and oversight match the accuracy of an automated machine learning classification?

#### 1.1. Unbalanced Data in Species Classification

When large swaths of land are captured by a sensor, the rich spectral information used for land-cover classification can be challenging for classes that covers only a fraction of the image. Therefore, we considered an unbalanced image one that contain classes with a number of pixels three orders of magnitude smaller than the rest of the classes [19]. The presence of unbalanced data in species classification may be unavoidable for species that have a limited or unknown range within a large landscape. This is not an unusual occurrence for special status plant populations, as many species of high ecological value have rapidly declined in population size and spread. Additionally, as climate change affects species distribution, a larger study area might allow researchers to monitor the migration of the species into previously unpopulated areas, further contributing to an unbalanced distribution of the target class and background matrix. Several challenges arise when attempting to classify small, widely dispersed populations of a single species. There are two general methods of handling unbalanced classifications. The imbalance can be accommodated at the data level or the algorithm level, as reviewed by Ali et al. [20]. This decision will affect the way that training and validation is sampled across classes, and may affect the algorithm selected. The creation of training and validation data is another difficulty, particularly if the species' distribution is sporadic and difficult to predict. In this study, the target species is not necessarily limited to its natural preference of elevation and soil types, because the majority of Willamette Valley ponderosa pine is currently found in plantations, which may or may not occur within its historical range. These plantations can be found in forested, agricultural, or even developed environments, which make it difficult to fully eliminate regions in the search for training data. Class separability is another challenge in training a model with unbalanced class distribution, especially in species classifications. If the target species is not spectrally distinct from the background class(es) or otherwise clearly identifiable from auxiliary data such as height, it will be especially difficult to properly train the algorithm. Finally, the handling of accuracy measures must be tailored to the unbalanced classification. Most traditional algorithms are accuracy-centric, which can be artificially inflated by the correct classification of the non-target class. Metrics such as omission and commission error provide more insight into class-specific accuracy. For a holistic view of the classification accuracy, Cohen's kappa coefficient, as discussed in Section 2.7, is better suited than overall percent accuracy, as it represents the proportion of agreement between class accuracies, allowing the minority class greater representation in the metric.

#### 1.2. Parametric and Non-Parametric Classification of Unbalanced Data

Many factors influence the suitability of parametric and non-parametric algorithms for vegetation classification. Species-level studies vary in study area, quantity and precision of classes, model complexity, time constraints, and desired product. The decision to use a parametric or non-parametric algorithm can be influenced by all of these factors. For species classifications that target one class of vegetation, unbalanced data, as discussed above, can introduce another challenge in selecting an appropriate algorithm.

Species classifications require algorithm training, and for large study areas that are broken into dozens or even hundreds of rasters, this can be time consuming. When building an unbalanced classification, analysts may attempt to create a representative sample of class training data in which the target species is known to occur in greater quantities, to classify a large assortment of rasters with unknown target species abundance. For datasets that are unequally distributed across the study area, creating a representative set of training data can be particularly time consuming. Non-parametric algorithms, such as decision trees or machine learning classifiers, are often selected for such tasks, as they are better equipped to handle unbalanced data through ensemble mechanisms such as bagging or boosting [21]. The RF algorithm is among the most popular of these non-parametric classification methods, such as ML. ML classifications also require a training dataset, but commonly the dataset is limited by its spatial relationship to the raster. Unlike RF, a generic training dataset is unsuitable

for different spatial classifications when running an ML algorithm. Thus, a drawback of ML is the labor required to train the algorithm for each image. However, the level of detail required by ML can be one of its primary benefits when classifying an unbalanced dataset. The spectral signature of each class is represented more accurately for each raster, and the amount of variability within its spectral space is reduced. Significant variation can be found in a large aerial imagery dataset, due to atmospheric influences and flight timing. Parametric classifications may reduce the impacts of these variations by using training data exclusive to each individual raster being classified, while non-parametric classifications may require some accepted reduction in training data accuracy, or additional pre-processing, such as atmospheric or light correction.

## 1.3. Ponderosa Pine in the Willamette Valley

The Willamette Valley is located in northwestern Oregon (Figure 1), a region composed of agriculture, developed urban areas, and temperate forest. Douglas-fir (*Pseudotsuga menziesii*) dominates the majority of surrounding forests of the Willamette Valley, often growing with western hemlock (*Tsuga heterophylla*), western redcedar (*Thuja plicata*), and deciduous trees such as bigleaf maple (*Acer macrophylla*) and red alder (*Alnus rubra*). While these species are among the most common today, this was not the case prior to European settlement of the region.



**Figure 1.** The Willamette Valley and the footprints of the rasters used for identification of Willamette Valley Ponderosa Pine (WVPP) stands.

Historically, indigenous communities burned large areas of the valley to maintain an abundance of edible plant species and to maintain visibility for hunting. The routine burning minimized the growth of trees and hardwood shrubs, creating an abundance of open prairies. Surveyors noted that Willamette Valley Ponderosa Pine (WVPP) typically grew within white oak (*Quercus garryana*) woodlands, shallow hills and ridges, areas prone to flooding, and in open savannahs with white oak and Douglas-fir [22,23]. These forest types were most likely maintained by the routine burning by indigenous tribes [24]. However, the settlement of the Willamette Valley in the 1880s resulted in an almost complete conversion of prairie and open woodlands to agriculture and Douglas-fir forest by inhibiting natural and human-induced forest fires. Most WVPP stands were harvested for building materials as development expanded, resulting in just a few scattered stands of native pine. These stands are often found on three soil types: Poorly drained clay in flat areas or foothills; shallow, rocky clay in foothills; and in well-drained sandy areas in the Willamette River floodplain [25]. Natural regeneration of WVPP has been limited primarily by low light conditions in historically suitable locations [23].

In 1996, a group of concerned landowners and researchers formed the Willamette Valley Ponderosa Pine Conservation Association (WVPPCA) with the intention of maintaining the genetic diversity of WVPP and increasing the ease of acquiring and growing the pine locally [25]. The WVPPCA has located over 900 native WVPP stands, with 150 individual trees serving as seed stock sources to preserve and propagate. However, the extent of younger WVPP plantations, most having been planted after the advent of the WVPPCA, is still unknown. Methods explored by this study may be used to quantify the current extent of WVPP. To identify the most suitable methods for area prediction, we compare the accuracy and the trade-offs between parametric and nonparametric classifications in the context of a small-scale land assessment with open-source data.

## 2. Materials and Methods

### 2.1. Study Area

The Willamette Valley is an alluvial region in northwestern Oregon of about 14,457 km<sup>2</sup> [26]. It is bounded by the Coast Range to the west and the Cascade Range to the east, encompassing the Columbia River to the north and ending around the Klamath Mountains to the south. The true ecoregion includes a small portion of southwestern Washington, but for the purpose of this research, the northern boundary was clipped to the Oregon-Washington border. The shapefile delineating this region was acquired through the Oregon Spatial Data Library [27]. The southern half of the Willamette Valley level III ecoregion of Oregon, which encompasses the study area [28,29], is approximately 8166 km<sup>2</sup>.

The center of the valley is primarily developed urban and agricultural land, and encompasses several large cities, including Portland, Salem, and Eugene. Nutrient-rich soil and relatively flat topography (sea level to 122 m) makes the Willamette Valley a significant agricultural area. The Valley contains several prominent rivers, including the Columbia, Willamette, Tualatin, McKenzie, Santiam, and others. The perimeter of the valley rises into the foothills of the Coast and Cascade Ranges, where dense forests take over the landscape. Most WVPP has been planted at the interface of rural urban areas and commercial forestland.

## 2.2. Data

## 2.2.1. Aerial Imagery

High-resolution red-green-blue-infrared (RGBI) aerial imagery was acquired from the National Agriculture Imagery Program (NAIP), which is flown every two to three years in all continental states. The data were collected using Leica Geosystem's ADSH100/SH100 digital sensors, flown on a Cessna 441 at a height of about 8400 m. The sensors recorded imagery at 40cm or 80cm GSD in four wavelengths, namely red, green, blue, and near-infrared. Images were captured continuously at

three angles: Nadir, forward (26°), and backward (19°). Flight lines were oriented north/south with 30% sidelap.

The data were processed to contain four RBGI bands with a ground pixel resolution of 1 m. The imagery was delivered as georeferenced digital ortho-quarter quad tiles (DOQQ), each tile covering one USGS topographic quadrangle. The Willamette Valley ecoregion contained 507 orthophotos collected in June and August of 2016 with no cloud cover. Two hundred and thirty-one of these raster images fell within the focus area of the southern Willamette Valley. However, only a small fraction contains Ponderosa pine that could be reliable identified from the images. Therefore, we have focused the study on five DOQQ tiles, which covered an area of 69,722 hectares.

The training data was partially created through known locations of WVPP, but most sites were identified by visually scanning aerial imagery for WVPP-specific characteristics, such as color, crown structure, and planting patterns. This process was facilitated by a false color composite of the NAIP imagery and very-high-resolution satellite reference imagery. Across the used dataset, 12.8 ha of WVPP training regions were identified and used. An additional two or three classes were used in parametric classification to hone the pine training areas, namely non-pine coniferous forest and low vegetation classes were used for all classifications, and a deciduous class was used in riparian-dominated regions. For non-parametric classifications, only two classes were used (pine and non-pine). The training data for the additional classes were created using visual assessment of the NAIP imagery, as they were less spectrally precise than the WVPP regions.

#### 2.2.2. Lidar

Lidar-derived highest hit digital surface models (DSMs) and bare earth digital elevation models (DEMs) of the focus area were acquired from the State of Oregon Department of Geology and Mineral Industries. The majority of data used was collected in the 2009, with some additional data from the 2012, 2015, and 2016 [30].

Watershed Sciences, Inc. collected and processed the lidar data for the 2009 projects. The data were flown using a Leica ALS50 Phase II laser system mounted in a Cessna Caravan 208. Data were collected at near nadir, and flight lines overlap 60% to 75% laterally. The average pulse density in 2009 was 8.14 pulses per m<sup>2</sup> with an average ground density of 1.36 returns per m<sup>2</sup>. The project average relative accuracy was 0.050m, and root mean square error for the calculated absolute accuracy was 0.04 m. The DSMs were created using highest hit lidar data, and the DEMs were created using triangular irregular networks (TIN) processing of ground point returns using TerraScan and Microstation.

#### 2.2.3. Species Cover Maps

Estimated species cover datasets have been created for western Oregon and surrounding regions in the Pacific Northwest, such as the gradient nearest neighbor (GNN) species map created by the Landscape Ecology, Modeling, Mapping, and Analysis (LEMMA) group. LEMMA's species map provides a continuous raster image of estimated plant community composition over large landscapes [31]. Likelihood of species composition is calculated by pixel proximity to the nearest of approximately 25,000 plots, and the climate, soil, geographic, and topographical conditions of the pixel. The raster image has a spatial resolution of 30m. The accuracy of the model is highly variable, with a reported kappa of 0.00–0.80 for species presence or absence. Each pixel represents an "ecological system" containing multiple associated species, usually a mix of tree and understory vegetation.

The inherently mixed species pixels of this dataset limit its usefulness in training or validating a single-species classification map, especially a species such as WVPP, which is present in small quantities within a very fragmented matrix of agricultural, forested, and developed land covers, and often occurs in very small stands that may be lost in the 30 m resolution of the GNN map. However, in future studies of unbalanced species mapping, a GNN species map may be very useful in providing a preliminary filter that can mask out unsuitable vegetation communities in large study areas.

#### 2.3. Aerial Imagery Pre-Processing

Initial attempts to run an ML classification on the original NAIP aerial imagery resulted in a high level of noise and class blending. The high spatial resolution of the imagery led to wide spectral variation within each target class, causing substantial overlap between the vegetation classes. To decrease spectral variability within the classes, a smoothing filter was applied to each image before classification. This filter came from the Orfeo Toolbox "Mean-Shift Smoothing" algorithm, which draws spatial and spectral information from neighboring cells within given radii to produce an image that preserves integral edges [26]. Although this algorithm is typically used for image segmentation, full segmentation was not performed for this project because of the scale and complexity of the dataset. The addition of object data information would contain errors and would not have provided meaningful separability to the pine and non-pine forest classes.

## 2.4. Enhancing Spectral Data

#### 2.4.1. Canopy Height Model (CHM) Band Incorporation

The original NAIP imagery data had limited class separability potential. Although the spatial resolution was very high—so high that the smoothing filter was needed—the spectral resolution was limited to four bands. Many tree species classification studies make use of multispectral or hyperspectral data with much higher spectral resolution [5–7,18]. Studies that have similar spectral limitations usually enhance the imagery with spatial or structural data [9]. As noted by Fassnacht et al. (2016), the majority of recent tree species classification publications have incorporated lidar data into the classification at the pixel level [8]. For this study, we created canopy height models (CHMs) by subtracting the DEMs from the DSMs, which provided a normalized representation of the true feature heights in the dataset. Fassnacht et al. (2016) claim that CHMs can be of limited use when fused with spectral data pre-classification, a finding supported by Ghosh et al. (2014), who did not notice any significant improvement in classification accuracy when fusing lidar-derived height data with hyperspectral data [7]. However, we chose to test this claim in the context of height data fusion with low spectral resolution data. The CHM was added to the NAIP imagery as a fifth band, incorporating the lidar data at the decision level, as described by Shen [11].

### 2.4.2. CHM Mask Application

In an additional test of the effectiveness of CHM data in vegetation class separation, we created CHM-based height masks to apply post-classification. Our goal was to reduce commission of agricultural and mature conifer regions. The CHM rasters were reclassified, with values between 5 feet and 90 feet assigned a value of one, and all other heights assigned a value of zero. The minimum value was selected to exclude areas of low, non-forest vegetation while preserving young WVPP stands, and the maximum value was selected to exclude mature, non-pine conifer stands. This range exceeds the expected height of most planted WVPP stands, because most stands targeted by this study are under 30 years old as a result of the conservation efforts of the 1990s [32]. However, the 90-foot ceiling was established because mature conifers tend to be more spectrally similar to WVPP than young conifers, and conifers often reach heights within this range upon maturity. We assumed that conifers under 90 feet would be spectrally distinct from WVPP. These masks were applied to the 4-band and 5-band classifications of both parametric and non-parametric methods. Lidar data were not available for the entire extent of the classified focus area. For no-data lidar regions, the 4-band classification data were used. In areas with available lidar data, we multiplied the 4- or 5-band parametric or non-parametric classified rasters by the CHM mask rasters. This raster algebra assigned a value of zero to areas that fell outside the acceptable height range, while areas that fell within the positive mask range maintained their original classification value.

#### 2.4.3. Vegetation Index Band Incorporation

In a preliminary attempt to increase class separability, we tested the impact of incorporating two vegetation indices as fifth bands: The normalized difference vegetation index (*NDVI*) and the visible-band difference vegetation index (*VDVI*), similar to Wang Xiaoqin et al. [27].

$$NDVI = \frac{(\rho_{red} - \rho_{nir})}{(\rho_{red} + \rho_{nir})}$$
(1)

$$VDVI = \frac{\left(2 * \rho_{green} - \rho_{red} - \rho_{blue}\right)}{\left(2 * \rho_{green} + \rho_{red} + \rho_{blue}\right)}$$
(2)

The *NDVI* and *VDVI* are calculated with pixel values and do not rely on sensor-specific coefficients, such as the enhanced vegetation index. The *NDVI* and *VDVI* are extensively used in unmanned aircraft systems (UAS), which do not experience significant atmospheric interference, and often only collect visible band imagery, due to the wide availability of high spatial resolution cameras [33]. We predicted that *NDVI* would help differentiate drought-tolerant WVPP from moisture-adapted conifers in the Valley, and that *VDVI* would help accentuate the visible color variations between WVPP plantations and other tree species. This trial was carried out on a smaller subset of rasters than the CHM band addition and mask application.

#### 2.5. Parametric Classification

#### 2.5.1. Algorithm Selection for Parametric Method

To identify the WVPP using the parametric method, we chose three algorithms based on their popularity in image analysis: ML [34], minimum distance (MD) [35], and spectral angle mapper (SAM) [36]. The three algorithms were initially assessed based on their ability to delineate the WVPP from the spectral RGBI data only. We selected, for further analysis, the most accurate algorithm, the accuracy being measured in this situation by the correctly identified pines and the size of the false positives. Each trial used identical training datasets to classify the same raster, and the required parameters were tuned through trial and error. The large size of the aerial imagery dataset recommended a classification method that uses a spectral library to classify the dataset, rather than developing training data specific to each image. SAM has this capability, so an additional training dataset of pine, non-pine mixed conifer, and non-forest vegetation was compiled with samples across the southern Willamette Valley. We performed a second trial of SAM classification on the sample region using this spectral library training dataset.

## 2.5.2. ML Classification Process

The ML algorithm distributes pixels among the classes based on the probability of class membership. The version of ML used in the study uses non-normalized discriminant function  $g_i(x)$  rather than true probabilities [37]:

$$g_i(x) = \ln p(\omega_i) - \frac{1}{2} \ln |\Sigma_i| - \frac{1}{2} (x - m_i)^T \Sigma_i^{-1} (x - m_i)$$
(3)

## i = class

x = n-dimensional data (where n is the number of bands)  $p(\omega_i) =$  probability that class  $\omega_i$  occurs in the image and is assumed the same for all classes  $|\Sigma_i| =$  determinant of the covariance matrix of the data in class  $\omega_i$   $\Sigma_i^{-1} =$  its inverse matrix  $m_i =$  mean vector To allocate a pixel to a class, the discriminant functions are compared with the threshold, which leads to the inequality:

$$\ln p(\omega_i) - \frac{1}{2} \ln |\Sigma_i| - \frac{1}{2} (x - m_i)^T {\Sigma_i}^{-1} (x - m_i) > T_i$$
(4)

where  $T_i$  is the threshold for class  $\omega_i$ .

The rearrangements of terms from (4) ends with a separation that can be interpreted in probabilistically:

$$(x-m_i)^T \Sigma_i^{-1} (x-m_i) < -2T_i + 2ln \ p(\omega_i) - ln|\Sigma_i|$$
(5)

In eventuality that *x* is normally distributed,  $(x - m_i)^T \Sigma_i^{-1} (x - m_i)$  follows a  $\chi^2$  distribution with the degrees of freedom equal the number of bands [37]. The  $\chi^2$  distribution helps in deciding whether a pixel belong to a class or not, as larger values of  $(x - m_i)^T \Sigma_i^{-1} (x - m_i)$  place the pixel in the tail of the distribution. In our study, we have selected the thresholds by choosing the probability of rejecting a pixel from a class thru trial and error. We found that a value of 5% for the probability of rejecting a pixel as belonging to a class supplies the expected results, irrespective of the class. Therefore, the thresholds were computed as:

$$T_{i} = \ln p(\omega_{i}) - \frac{1}{2}\ln|\Sigma_{i}| - 4.744$$
(6)

The ML classifications were performed using the ENVI implementation of the ML algorithm, which was applied on the enhanced NAIP imagery with and without the CHM mask on the four-band spectral NAIP data and the five-band NAIP + CHM data. Therefore, a set of four ML classifications were executed (Figure 2).



Figure 2. Methods workflow for final classification techniques.

## 2.6. Non-Parametric Classification

The ML classifier is not the only algorithm that provides accurate results for image analysis [14,38]. Decision trees, or ensemble methods, are another tool often used for image classification [39–41]. Among decision tree methods, RF is likely the most popular [42–44]. RF is a non-parametric stochastic supervised learning algorithm suitable for image classification [45–47]. Mirroring most supervised classification algorithms, RF contains a training phase during which multiple decision trees are built. The results produced by each decision tree are aggregated, and the output is the class with the largest

number of occurrences. The RF algorithm has the ability to correct for overfitting [46,48], a property not common to other machine learning algorithms, such as neural networks [49,50]. Toloşi and Lengauer [51] pointed out that even though the RF algorithm is robust to multi-collinearity and overfitting, its accuracy likely depends on the relationship between the number of features and the number of samples. Among the available implementations of the RF algorithm, we selected the randomForest package, which is an R enactment of the original code written by Breiman and Cutler in Fortran [52]. The RF algorithm depends on a multitude of parameters [53], chief among them being the number of trees and the number of variables used at each split of a tree. To identify the most appropriate values, we have run a factorial experiment with four levels for the number of trees (i.e., 100, 200, 300, and 400) and three levels for the number of variables used at each split of a tree (i.e., 2, 3, and 4).

Mirroring the ML method, we performed four RF variations within the non-parametric classification method. The 4-band NAIP and 5-band NAIP + CHM rasters were classified in R, and the CHM binary mask was applied to each of these (Figure 2).

Irrespective of the classification approach (i.e., parametric or non-parametric), the NAIP images were proven to provide superior results than classifications applied on images with lower spatial resolution [54], such as Landsat. The reduction of the pixel size can increase the classification accuracy by more than 10% compared with larger pixels, as it will be able to delineate classes such as shade or ground [55–57]. Smaller pixels are able to capture landscape details invisible to larger pixels; therefore, they are more adequate to delineate spectral signatures within the ponderosa pine plantations, particularly the ground. Considering that the crown of the trees will cover most of the ground, the sparse non-pine pixels within the plantation are easily separated by using contiguity and minimum surface constraints.

#### 2.7. Training, Validation, and Accuracy Assessment

The NAIP images used for assessment contain less than 0.02% of identified ponderosa pine. Because of the relatively low abundance of WVPP in naturally or artificially mixed stands, and the difficulty in validating the accuracy of a classified mixed stand, only plantations that would provide an uninterrupted signature of WVPP were considered for training and validation. Across the 69,722 ha of the area, all of the 12.8 ha of pine pixels were used to train and validate the classifiers, with a randomized 7:3 ratio of training to validation pine stands. For the ML parametric classifier, we randomly selected non-pine pixels covering 86.3 ha. Because of the reduced, unbalanced presence of the pine in the study area, there were no assessment data outside the 12.8 ha, meaning that the accuracy was carried out on the validation data. For each tile and classification method, a set of three randomizations were executed, which produced 243 combinations, i.e., 3<sup>5</sup>.

We assess the performances of the nonparametric classifier RF with a 5-fold cross-validation approach, similar to Goetz et al. [58] and Heung et al. [59]. Each run was executed by splitting the pixels identified as ponderosa pine 70% for training and 30% for validation, as recommended by Chicco [60]. Considering the unbalanced data, we have considered four sample sizes for the non-pine class, all reported to the size of pine class: the same, twice as large, 10 times larger, and 100 times larger. Because the images were acquired in slightly different illumination conditions, we conducted the non-parametric classifications via two strategies: One in which the RF classification was carried out on each image, independently (i.e., training and validation from one image at a time), and one in which the RF classification was executed by combining all the images (i.e., training and validation data extracted from all images simultaneously).

All classifications were assessed using overall accuracy, kappa coefficient, and pine commission and omission error (Equations (4)–(7)) [61,62]. We placed a significant emphasis on the overall accuracy, because of the unbalanced ratio of non-pine area to pine area (at least 100 times according to Mueller-Warrant et al. [56] and 1000 times according to the selected study area). For the unbalanced data, the kappa coefficient is a more suitable metric for accuracy assessment, because it accounts for the

uneven distribution of the pixel among classes. Although pine commission and omission error provide a one-sided perspective on the classification abilities, they do supply a precise metric for our objective.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$
(7)

$$Omission = \frac{FN}{FN + TN}$$
(8)

$$Commission = \frac{FP}{FP + TP}$$
(9)

$$Kappa = \frac{(TP + FN) \times All - ((TP + FN)(TP + FP) + (FP + TN))(FN + TN))}{All^2 - ((TP + FN)(TP + FP) + (FP + TN))(FN + TN))}$$
(10)

where:

TP = number of pixels with pine predicted with pine (true positive) TN = number of pixels without pine predicted without pine (true negative) FP = number of pixels without pine predicted with pine (false positive) FN = number of pixels with pine predicted without pine (false negative) All = total number of pixels: All = TP + TN + FP + FN

## 3. Results

## 3.1. Aerial Imagery Pre-Processing Results

To achieve the best smoothing results in Orfeo Toolbox, we used a spatial radius of four pixels, a range (spectral) radius of 256 DN values, and a 0.1 mode convergence threshold. Iterations were capped at 100. These values were settled upon through trial and error. The spatial radius was selected for its ability to smooth stand canopies while preserving the edges of larger features. The range radius value impacted trial classifications to a lesser degree, so a relatively large value was selected to ensure spectral consistency within features. The mode convergence and max iteration values were kept at default.

## 3.2. Spectral Data Enhancement Results

#### 3.2.1. Preliminary CHM Band Incorporation

We found that incorporating the CHM data as a fifth band generated greater accuracies for both parametric and nonparametric methods in this preliminary test. However, the benefit was not consistent between both classification methods. Adding the CHM band increased overall accuracy from 62.8% to 65.2% and kappa coefficient from 0.48 to 0.52 for ML, but for the RF algorithm, the CHM band increased overall accuracy from 86.0% to 96.3% and kappa coefficient from 0.46 to 0.75 (Table 1).

**Table 1.** Accuracy statistics for the addition of a canopy height models (CHM) band to the maximum likelihood (ML) and random forest (RF) classifications.

	4-Band ML	5-Band ML	4-Band RF	5-Band RF
Overall Accuracy	62.8%	65.2%	86.0%	96.2%
Kappa Coefficient	0.48	0.52	0.46	0.75
Pine Omission	40.2%	7.4%	17.5%	16.1%
Pine Commission	25.9%	43.8%	14.8%	3.6%

#### 3.2.2. Preliminary CHM Mask Application

While the addition of a CHM band was most pronounced in the non-parametric method, the application of a CHM mask seemed most effective in the parametric method. The preliminary

mask application increased the parametric overall accuracy from 62.8% to 92.4% and kappa coefficient from 0.48 to 0.60 (Table 2). The effect of the mask on the non-parametric method was not as impressive as was on the parametric method; overall accuracy rose from 86.0% to 90.2%, and the kappa coefficient rose from 0.46 to 0.49 (Table 2).

	4-Band ML	Masked 4-Band ML	4-Band RF	Masked 4-Band RF
Overall Accuracy	62.9%	92.4%	86.0%	90.2%
Kappa Coefficient	0.48	0.60	0.46	0.49
Pine Omission	40.2%	38.3%	17.5%	34.7%
Pine Commission	25.9%	2.8%	14.8%	6.5%

Table 2. Accuracy statistics for the application of a CHM mask to the ML and RF classifications.

The large increase in pine omission to RF following the application of the CHM mask was initially worrisome, but the likely cause of the increase is canopy gap lidar returns. Any pixel falling within the pine class validation polygon was flagged as "true pine" in the accuracy assessment, but the high resolution of the CHM resulted in canopy gap ground returns being classified as "non-pine" after masking, because the pixel had a true height of less than five feet. This was flagged as an error in the accuracy assessment, but it should not be considered an innate shortcoming of the RF classification, because of the spectral classification was correct, and was later changed to non-pine because of the ground return. This could be remedied by adjusting the mask parameters.

## 3.2.3. Vegetation Index Incorporation

The incorporation of *NDVI* and *VDVI* as bands did not improve the classification accuracy by an appreciable margin during preliminary testing. The addition of an *NDVI* band did improve overall accuracy by 2.4% and the kappa statistic by about 0.02 (Table 3), and was considered insufficient to justify processing and adding *NDVI* bands for all rasters. *VDVI* decreased overall accuracy by 1.9% and kappa by 0.02. In all instances of index incorporation, pine omission was decreased by several percentages, but this improvement came at the cost of pine commission.

	<b>Overall Accuracy</b>	Kappa	Pine Omission	Pine Commission
No Indices	61.1%	0.478	23.8%	12.9%
NDVI	63.5%	0.507	19.4%	15.6%
VDVI	59.2%	0.451	19.2%	18.8%
NDVI + VDVI	60.5%	0.468	17.9%	18.3%

Table 3. Accuracy statistics for the addition of indices to the ML classification.

After considering the results of the CHM and vegetation index band additions, we concluded that the parametric and nonparametric methods would benefit from the incorporation of CHM data through pixel-level fusion and post-classification masking, and that the inclusion of vegetation index bands would not appreciably improve the results of either method. Therefore, the subsequent analysis was executed without including *NDVI* and *VDVI*.

## 3.3. Parametric Classification Results

## 3.3.1. Selection of the Parametric Algorithm

Of the three parametric algorithms, ML most accurately classified the known WVPP stands present in the sample region (Table 4). Although it did not attain the highest overall accuracy, it classified the pine class with the highest producer accuracy by 13% and achieved a user accuracy consistent with the other two algorithms. SAM correctly identified most pine areas but had a very high rate of commission error in both non-pine classes; its relatively high overall accuracy was inflated by high accuracies in the non-pine classes. MD was unable to distinguish the WVPP stands from certain agricultural crops and mixed conifer stands. Some of the most prominent WVPP stands were a challenge for MD to classify.

	<b>Overall Accuracy</b>	Kappa Coefficient	Pine Omission	Pine Commission
ML	53.9%	0.41	19.8%	15.5%
SAM	56.6%	0.44	33.0%	17.0%
MD	35.9%	0.23	40.0%	19.3%

Table 4. Comparative statistics for parametric classifiers.

In addition to the SAM results listed above, which used individual training datasets, the SAM trial with the spectral library training dataset performed extremely poorly at 24.8% overall accuracy with a 0.095 kappa coefficient. Pine was massively overrepresented on the landscape, yet was poorly predicted in the true stand locations, with a commission rate of 36.24% and an omission rate of 42.14%. It was this finding that affirmed our decision to use ML for the parametric classification, in which each image had its own finely tuned training dataset.

## 3.3.2. Parametric Classification with ML Algorithm

The parametric classification involved multiple iterations, as new training areas were identified or unreliable training areas removed. This process produced higher accuracies, as additional adjustments were implemented on the training areas. The most effective ENVI class threshold was 0.05 for all classes; a threshold of 0.1 or higher resulted in very high commission error, while thresholds lower than 0.05 resulted in a high level of unclassified areas, likely causing omission error.

The ML classifications produced a broad range of accuracies across the various combinations of spectral and CHM data. The simple four-band aerial imagery classification produced a fairly low-accuracy image at 67.73% overall accuracy and 0.29 kappa (Table 5). Pine commission was very high at 47.48%, and pine omission was at its highest at 13.59%. The addition of the CHM band increased the overall accuracy and kappa coefficient of the parametric method substantially. Overall accuracy increased by about 15%, and the kappa coefficient increased by 0.22. Pine omission was nearly unchanged, but the application of the CHM mask removed nearly 30% of pine class commission (Table 5).

	4-Band	4-Band Masked	5-Band	5-Band Masked
Overall Accuracy	67.73%	83.14% (+15.41%)	78.08% (+10.35%)	97.22% (+29.49%)
Kappa Coefficient	0.29	0.51 (+0.22)	0.47 (+0.18)	0.89 (+0.60)
Pine Omission	13.59%	13.48% (-0.11%)	10.98% (-2.61%)	10.83% (-2.76%)
Pine Commission	47.48%	18.84% (-28.64%)	6.42% (-41.06%)	0.83% (-46.65)

Table 5. Confusion matrix for the ML classifications (in parenthesis, change from four-band statistics).

The classification of the five-band images, which incorporated the CHM band, reached higher accuracies on average. The un-masked five-band images were classified with a 78.08% overall accuracy and a kappa of 0.47 (Table 5). Pine omission was a few percentage points lower than the four-band classifications, but pine commission saw an even greater decrease to an average of 6.42%. The application of the CHM mask increased overall accuracy by nearly 20% and nearly doubled the kappa coefficient. Pine omission was nearly unchanged, but pine commission nearly disappeared, achieving an average of just 0.83% (Table 5). A visual representation of the five-band masked classification can be seen in Figure 3.



**Figure 3.** ML classification of five-band raster with CHM mask applied. Red areas are contiguous regions of pine class greater than 1000 square meters. White areas are smaller regions of pine class. Black areas are classified as non-pine.

## 3.4. Non-Parametric Classification Results

The non-parametric classifications were conducted in the same four combinations as the parametric classifications, with the addition of one extra variation. Each tile was classified once using only its spatially related training data (individual classification—Table 6) and once using training information from all five images (collective classification—Table 7).

Table 6.	Mean statistic	s of all	RF	individual	images	classifications	(in	parenthesis,	change	from
four-band	d statistics).									

	4-Band	4-Band Masked	5-Band	5-Band Masked
Overall Accuracy	95.63%	95.62% (-0.01%)	95.96% (+0.33%)	97.28% (+1.65%)
Kappa Coefficient	0.90	0.90 (0.00)	0.91 (+0.01)	0.94 (+0.04)
Pine Omission	4.27%	4.23% (-0.04%)	3.63% (-0.64%)	2.35% (-1.92%)
Pine Commission	4.42%	4.43% (+0.01%)	4.24% (-0.18%)	2.91% (-1.51%)

**Table 7.** Mean statistics of all RF collective classifications (in parenthesis, change from four-band statistics.).

	4-Band	4-Band Masked	5-Band	5-Band Masked
Overall Accuracy	93.43%	93.36% (-0.07%)	95.93% (+0.50%)	97.30% (+1.87%)
Kappa Coefficient	0.85	0.85 (0.00)	0.91 (+0.06)	0.94 (+0.09)
Pine Omission	7.17%	6.94% (-0.23%)	3.74% (-3.43%)	2.25% (-4.92%)
Pine Commission	6.27%	6.42% (+0.15%)	4.23% (-2.04%)	2.94% (-3.33%)

Overall accuracy for all RF classification variations exceeded 90%, with kappa values ranging between 0.85 and 0.94. In most cases, exclusively using training data that were spatially related to the

raster gave better results, but only by at the most 0.05 kappa (Table 6). The collective classifications only showed superior results in the five-band masked classification (Table 7).

The application of a CHM mask did not benefit the four-band classification, and in fact reduced accuracy by a very small margin. When applying the mask to the five-band classification, accuracy increased by 0.03 kappa. Both pine omission and pine commission were reduced in the five-band masked classification (Figure 4).



**Figure 4.** RF individual classification of five-band raster with CHM mask applied. Red areas are contiguous regions of pine class greater than one hectare. White areas are smaller regions of pine class. Black areas are classified as non-pine.

## 4. Discussion

## 4.1. Spectral Data Enhancement

The addition of height data to the NAIP aerial imagery at both pixel level (included in classification) and decision level (included post-classification) improved classification accuracies across all variations of parametric and non-parametric methods. The incorporation of vegetation indices had minimal effects upon the accuracy of the preliminary ML classification, and in fact reduced overall accuracy and kappa for two of the three index combinations. The number of indices that were available for use were extremely limited, as EVF and other indices requiring band coefficients have not yet been established for NAIP imagery. *NDVI* and *VDVI* use simple band algebra to emphasize differences in vegetation reflectance. Such indices can be valuable in reducing excess spectral information and exaggerating differences between classes, but when classified in combination with all original input bands, there is little to no new information added to the pixel stack. The minor improvement in overall accuracy offered by the *NDVI* band did not provide sufficient evidence for its usefulness in class separation, and the decrease in accuracy among the other index combinations supported our decision not to include these indices further.

The four-band ML classification on average had the lowest accuracy of all classification methods (67.73% overall accuracy and 0.29 kappa coefficient). This low average can be attributed to several influences. The biggest challenge for the four-band classification is the difficulty to represent the pine class fully, with its spectral variations introduced by age and tree heath, while also excluding all other types of vegetation. Additionally, each raster contains varying quantities of pine, exacerbating the difficulty in adequately representing the pine class within each raster when few stands are available to be sampled. Images with fewer pine stands available, consequently training data, generally produced worse results.

The accuracy of the four-band classification was greatly improved by applying the CHM mask, primarily by its removal of pine commission in the non-forest vegetation class. By masking the four-band ML classified images, mean pine commission was reduced by about 29%, resulting in an overall accuracy of 83.14% and kappa coefficient of 0.51. Visually, the similarities between WVPP stands and certain agricultural crops are stronger than the similarities between WVPP stands and other deciduous or coniferous forest species, making the lower height mask of five feet particularly effective.

Incorporation of the CHM band to create the five-band had a similar, but slightly weaker effect on the ML classification in terms of average overall accuracy and kappa coefficient. However, pine class commission was reduced by a substantial 41%, while pine class omission was reduced by about 3%. Masking cannot truly affect pine omission, as the mask does not add any pixels excluded from the pine class. Any decrease in pine omission from masking is the result of masking a different iteration of the four-band or five-band classified image. This large decrease in pine commission suggests that the ML algorithm was successfully able to remove non-pine areas that matched the spectral signature but did not match the average height of the pine training areas. Some of these non-pine areas that were ultimately included may have been included at the fringe of the class parameters, lacking a hard limit in height parameters. There may have been points within the pine training data that fell outside the 5 to 90-foot limit, creating leeway in height discrimination.

While the generalized accuracy metrics suggest that a mask application is more effective in isolating the pine class, the substantial decrease in pine commission in the five-band classification compared to the mask's effect on commission reinforces the impact of the unbalanced data. Because the non-pine class is an amalgam of every single land cover other than pine, its less impressive accuracy lowered the overall accuracy and kappa of the five-band classification, while the mask automatically transferred any pixels outside of height range to the non-pine class, boosting its accuracy automatically.

The steady improvement in accuracy throughout the ML methods suggests that neither incorporation of CHM data should be ruled out in a parametric classification. Certainly, a combination of spectral and height bands is superior to an entirely spectral classification, but the application of a decision-level height mask helps correct excess commission error within a binary classification, even when the same data are incorporated at the pixel level. This is particularly visible in the jump from 0.47 to 0.89 kappa coefficient between the unmasked and masked five-band ML classification.

### 4.3. Non-Parametric Classification and the Effects of CHM Incorporation

The four-band RF classification applied to individual images surpassed the accuracy of the four-band ML classification by nearly 30%, achieving an average overall accuracy of 95.63% (Table 6). This incredible increase in accuracy speaks to the effectiveness of non-parametric algorithms in classifying pure spectral data. While the four-band RF classification was quite impressive, increases in accuracy were practically inexistent with the addition of height data, as opposed to the dramatic increases seen between ML methods.

The application of the height mask to the four-band classification was effectively null, while the incorporation of the CHM band as the five-band raster improved accuracy by a scant 0.33% (0.01 kappa). Pine omission and commission were likewise weakly improved. A slightly larger, but otherwise negligible improvement was made by masking the five-band classification.

We found that better accuracy was achieved when classifying images with their own spatially related training data (individual classification) as opposed to a collective classification (Table 7). This is likely due to the spectral variability between images, which has a pronounced effect upon the attempt to separate the precise spectral signatures of varying vegetation species. The minimum accuracy for collective classification was more than 5% smaller than the smallest accuracy obtained for individual images (Table 7), even that the standard deviation was almost the same (i.e., 0.021). The decrease in accuracy with more than 2% from individual images to all the images is probably the results of the RF accounting for the increase variability associated with the larger landscape considered simultaneously rather than separately. The reduction in the ability of the RF to delineate ponderosa pine over large areas, which inheritably have not only larger variability but also was recorded from different position of the sun and atmospheric conditions, was present in the omission and commission errors, which also increased with more than 2% compared with the classification of individual images (Tables 6 and 7). The addition of CHM did not improve the accuracy (i.e., 93.43% vs. 93.36%) but reduced the variability (i.e., standard deviation decreased from 0.0205 to 0.0203, meaning 1%), even insignificantly.

#### 4.4. Parametric v. Nonparametric Accuracy

While the RF algorithm generally outperformed the ML method in most variations in terms of overall accuracy, most notably when classifying four-band spectral data, the accuracies of ML and RF began to converge with the application of CHM data through pixel-level fusion and mask application in the ML classifications. Overall accuracies were nearly equal between the five-band masked classifications by each algorithm, with a slightly lower kappa for ML (Table 5). The use of CHM data in the ML classifications was particularly effective in reducing commission in the pine class. This came at the price of increasing pine omission, which is visible in Figures 3 and 4; the ML image is much less dense than the RF image. The contiguous pine stands were not classified as completely by ML as they were by RF, and therefore were highlighted by a larger sieve for the ML image (areas greater than 1000 m highlighted for ML, areas greater than one hectare highlighted for RF).

When classifying large areas, the acquisition of training and validation regions can be tedious. While the collective RF classifications were technically lower in accuracy than the individual classifications, the differences were small enough that a collective RF classification could require a lower density of training data across the landscape, enabling the user to accurately classify regions where training data are scant or non-existent. This is a major drawback of ML, which requires that training data be provided from the image to be classified.

The unbalanced proportion of WVPP in respect to the rest of the land use classes was successfully identified by the selected parametric and non-parametric algorithm. However, the non-parametric algorithm provided results that were consistently superior to the parametric algorithm, even if the difference in overall accuracy was operationally negligible. The non-parametric algorithm had a kappa coefficient consistently larger than 0.90—values unattained by the parametric algorithm. Nevertheless, the only situation when the two types of algorithms were comparable were when the elevation data were included, which in many areas is not available. In the case when only spectral data are available, the non-parametric algorithm is superior in binary classify landscapes with highly unbalanced vegetation classes.

### 4.5. Sources of Error

The limited spectral resolution, large size of the imagery dataset, and the unbalanced targeted classes contributed to many of classification errors. Considering that the training data could not be fully ground-verified, it is likely that some of the pine training regions that could not be verified with street-view imagery were not truly pine but were instead Douglas-fir or another spectrally or structurally similar species. Furthermore, the training data could only be obtained by visual assessment of the imagery, rather than in-person sampling. In as many cases as possible, we only used training data that could be validated through street-view imagery. However, these data are not available in

rural parts of the study area, so for some regions, training data were created solely through assessment of high-resolution satellite imagery in various color composites. Some training data were composed of plantations that may have been too thin to provide a solid spectral signature, thereby including understory pixels that contributing to pine commission. Age differences between pine stands also contributed to spectral variability within the pine class.

The most common confusion was between pine and mixed-conifer plantations that shared similar planting and growth patterns, and certain agricultural crops. Widely spaced Douglas-fir plantations, especially mature Douglas-fir, is visually similar to many WVPP plantations. As WVPP and Douglas-fir stands age, they become more similar spectrally. WVPP stands tend to have a duller, browner reflectance than other conifer stands. This is partially because WVPP is planted at a wider spacing and has low-density crown characteristics, resulting in a higher proportion of ground visibility within the stand and through the canopy. We aimed to minimize this confusion through the integration of CHM data with the spectral data.

## 5. Conclusions

We found that the highest accuracies, considering overall average, kappa coefficient, omission, and commission, are achieved by fusing CHM data with spectral data and performing a non-parametric classification, followed by a height mask. However, if height data are not available for the area of interest, a non-parametric classification of spectral data alone provides near-equivalent results.

If a parametric algorithm is to be used, the classification is dramatically improved through the addition of height data and masking, as the addition of non-spectral data offers greater class separability in spectrally similar areas. Because our classification targets were forest stands rather than individual trees, height data of moderate resolution was useful and more readily available for such a large study size. The unbalanced nature of the dataset was sufficiently handled through under-sampling of the majority non-pine class.

RF was the superior classification method for our data, because the non-parametric algorithm was ultimately the best fit for the unbalanced, non-parametric nature of our study area. Isolating such a small segment of the spectral landscape within a massive study area required the on-the-fly adaptations of RF, while ML was limited by our defined classes and training areas.

Although RF classification of five-band imagery achieved the highest accuracies of the six methods, the ML method still reached a reasonably high kappa coefficient and extremely low commission rates with the addition of the CHM mask. While RF is the preferable method when classifying small-scale, high-resolution aerial imagery, the ML method can still achieve high accuracies if height data is available to enhance it.

Our findings suggest that the parametric and non-parametric method used in the study are appropriate for unbalanced binary classifications. However, the non-parametric algorithm is better suited for heterogeneous land cover, and if necessary, can be modified to use a standardized training dataset on an iteration of images, whereas the parametric method requires training data that are spatially related to each image in the iteration. The non-parametric method seems to be suitable for both standalone spectral and height-fused data, although the addition of height data may increase the classification accuracy across nearly all metrics.

**Author Contributions:** Conceptualization, S.A.F.; methodology, A.P.R. and B.M.S.; software, A.P.R., B.M.S., and C.Q.; validation, A.P.R.; formal analysis, A.P.R. and B.M.S.; investigation, A.P.R., S.A.F., and B.M.S.; resources, S.A.F. and B.M.S.; data curation, A.P.R.; writing—original draft preparation, A.P.R.; writing—review and editing, B.M.S. and S.A.F.; visualization, A.P.R.; supervision, S.A.F. and B.M.S.; project administration, A.P.R.; funding acquisition, S.A.F. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received funding from the Forestry and Natural Resources Extension Program at Oregon State University, the U.S. Department of Agriculture, grant number 2019-67019-29462, and the McIntire Stennis project OREZ-FERM-875.

**Acknowledgments:** We would like to thank Glenn Howe for his significant contributions in recommending the usage of RF. We would also like to thank Matt Gregory for his support in the beginning stages of the study and for his help in troubleshooting code issues.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

## References

- 1. Martin, M.E.; Newman, S.D.; Aber, J.D.; Congalton, R.G. Determining forest species composition using high spectral resolution remote sensing data. *Remote Sens. Environ.* **1998**, 65, 249–254. [CrossRef]
- 2. Brandtberg, T. Individual tree-based species classification in high spatial resolution aerial images of forests using fuzzy sets. *Fuzzy Sets Syst.* **2002**, *132*, 371–387. [CrossRef]
- 3. Carleer, A.; Wolff, E. Exploitation of very high resolution satellite data for tree species identification. *Photogramm. Eng. Remote Sens.* **2004**, *70*, 135–140. [CrossRef]
- 4. Dalponte, M.; Bruzzone, L.; Gianelle, D. Fusion of Hyperspectral and LIDAR Remote Sensing Data for Classification of Complex Forest Areas. *IEEE Trans. Geosci. Remote Sens.* **2008**, *46*, 1416–1427. [CrossRef]
- 5. Ballanti, L.; Blesius, L.; Hines, E.; Kruse, B. Tree species classification using hyperspectral imagery: A comparison of two classifiers. *Remote Sens.* **2016**, *8*, 445. [CrossRef]
- Jones, T.G.; Coops, N.C.; Sharma, T. Assessing the utility of airborne hyperspectral and LiDAR data for species distribution mapping in the coastal Pacific Northwest, Canada. *Remote Sens. Environ.* 2010, 114, 2841–2852. [CrossRef]
- 7. Ghosh, A.; Fassnacht, F.E.; Joshi, P.K.; Kochb, B. A framework for mapping tree species combining hyperspectral and LiDAR data: Role of selected classifiers and sensor across three spatial scales. *Int. J. Appl. Earth Obs. Geoinf.* **2014**, *26*, 49–63. [CrossRef]
- Fassnacht, F.E.; Latifi, H.; Stereńczak, K.; Modzelewska, A.; Lefsky, M.; Waser, L.T.; Straub, C.; Ghosh, A. Review of studies on tree species classification from remotely sensed data. *Remote Sens. Environ.* 2016, 186, 64–87. [CrossRef]
- 9. Dechesne, C.; Mallet, C.; Le Bris, A.; Gouet-Brunet, V. Semantic segmentation of forest stands of pure species combining airborne lidar data and very high resolution multispectral imagery. *ISPRS J. Photogramm. Remote Sens.* **2017**, *126*, 129–145. [CrossRef]
- Anderson, J.E.; Plourde, L.C.; Martin, M.E.; Braswell, B.H.; Smith, M.L.; Dubayah, R.O.; Hofton, M.A.; Blair, J.B. Integrating waveform lidar with hyperspectral imagery for inventory of a northern temperate forest. *Remote Sens. Environ.* 2008, *112*, 1856–1870. [CrossRef]
- 11. Shen, S.S. Summary of Tyles of Data Fusion Methods Utilized in Workshop Papers. In *Proceedings of the NASA Conference Publication 3099: Multisource Data Integration in Remote Sensing*; Tilton, J.C., Ed.; NASA: Greenbelt, MD, USA, 1990; pp. 145–149.
- 12. Pohl, C.; Van Genderen, J.L. Review article Multisensor image fusion in remote sensing: Concepts, methods and applications. *Int. J. Remote Sens.* **1998**, *19*, 823–854. [CrossRef]
- 13. Erbek, F.S.; Özkan, C.; Taberner, M. Comparison of maximum likelihood classification method with supervised artificial neural network algorithms for land use activities. *Int. J. Remote Sens.* **2004**, *25*, 1733–1748. [CrossRef]
- 14. Shafri, H.Z.M.; Suhaili, A.; Mansor, S. The Performance of Maximum Likelihood, Spectral Angle Mapper, Neural Network and Decision Tree Classifiers in Hyperspectral Image Analysis. *J. Comput. Sci.* 2007, *3*, 419–423. [CrossRef]
- 15. Lu, D.; Weng, Q. A survey of image classification methods and techniques for improving classification performance. *Int. J. Remote Sens.* **2007**, *28*, 823–870. [CrossRef]
- 16. Zhu, X.X.; Tuia, D.; Mou, L.; Xia, G.-S.; Zhang, L.; Xu, F.; Fraundorfer, F. Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geosci. Remote Sens. Mag.* **2017**, *5*, 8–36. [CrossRef]
- 17. Lary, D.J.; Alavi, A.H.; Gandomi, A.H.; Walker, A.L. Machine learning in geosciences and remote sensing. *Geosci. Front.* **2016**, *7*, 3–10. [CrossRef]
- Dalponte, M.; Bruzzone, L.; Gianelle, D. Tree species classification in the Southern Alps based on the fusion of very high geometrical resolution multispectral/hyperspectral images and LiDAR data. *Remote Sens. Environ.* 2012, 123, 258–270. [CrossRef]

- 19. Bhowan, U.; Johnston, M.; Zhang, M.; Yao, X. Evolving Diverse Ensembles Using Genetic Programming for Classification with Unbalanced Data. *IEEE Trans. Evol. Comput.* **2013**, *17*, 368–386. [CrossRef]
- 20. Ali, A.; Shamsuddin, S.M.; Ralescu, A.L. Classification with class imbalance problem: A review. *Int. J. Adv. Soft Comput. Appl.* **2015**, *7*, 176–204.
- 21. Leevy, J.L.; Khoshgoftaar, T.M.; Bauder, R.A.; Seliya, N. A survey on addressing high-class imbalance in big data. *J. Big Data* **2018**, *5*. [CrossRef]
- 22. Hansen, H.P. Postglacial Forest Succession, Climate, and Chronology in the Pacific Northwest. *Am. Philos. Soc.* **1947**, *37*, 1–130. [CrossRef]
- 23. Hibbs, D.E.; Wilson, M.V.; Bower, A.L. Ponderosa pine of the Willamette Valley, western Oregon. *Northwest Sci.* **2002**, *76*, 80–84.
- 24. Boyd, R. Strategies of Indian Burning in the Willamette Valley. In *Indians, Fire and the Land in the Pacific Northwest;* Boyd, R., Ed.; Oregon State University Press: Corvallis, OR, USA, 1999; pp. 96–138.
- 25. Fletcher, R. Willamette Valley Ponderosa Pine—A Primer. 2007. Available online: https://www.westernfores try.org/wvppca/2008/WillametteValleyPonderosaPinePrimer2007.pdf (accessed on 12 October 2020).
- 26. OTB Team. OTB CookBook Documentation, Release 6.2.0. 2017. Available online: https://www.orfeo-toolbo x.org/CookBook-6.2/ (accessed on 12 October 2020).
- 27. Xiaoqin, W.; Miaomiao, W.; Shaoqiang, W.; Yundong, W. Extraction of vegetation information from visible unmanned aerial vehicle images. *Trans. Chin. Soc. Agric. Eng.* **2015**, *31*, 152–159. [CrossRef]
- 28. Omernik, J.M. Perspectives on the Nature and Definition of Ecological Regions. *Environ. Manag.* **2004**, *34*, S27–S38. [CrossRef]
- 29. Omernik, J. Ecoregions of the Conterminous United States. Ann. Assoc. Am. Geogr. 1987, 77, 118–125. [CrossRef]
- 30. Oregon Lidar Consortium Oregon Department of Geology and Mineral Industries (DOGAMI). Available online: https://www.oregongeology.org/lidar/ (accessed on 8 June 2020).
- Ohmann, J.L.; Gregory, M.J.; Henderson, E.B.; Roberts, H.M. Mapping gradients of community composition with nearest-neighbour imputation: Extending plot data for landscape analysis. *J. Veg. Sci.* 2011, 1–17. [CrossRef]
- 32. Fletcher, R. *Establishing and Managing Ponderosa Pine in the Willamette Valley;* Oregon State University Extension Service: Corvallis, OR, USA, 2003.
- 33. Xue, J.; Su, B. Significant remote sensing vegetation indices: A review of developments and applications. *J. Sens.* **2017**, 2017, 1353691. [CrossRef]
- Pratt, J.W. FY Edgeworth and RA Fisher on the Efficiency of Maximum Likelihood Estimation. *Ann. Stat.* 1976, 4, 501–514. [CrossRef]
- 35. Wolfowitz, J. The Minimum Distance Method. Ann. Math. Stat. 1957, 28, 75-88. [CrossRef]
- Kruse, F.A.; Lefkoff, A.B.; Boardman, J.W.; Heidebrecht, K.B.; Shapiro, A.T.; Barloon, P.J.; Goetz, A.F.H. The Spectral Image Processing System (SIPS)—Interactive Visualization and Analysis of Imaging Spectrometer Data. *Remote Sens. Environ.* 1993, 44, 145–163. [CrossRef]
- 37. Richards, J.A. Remote Sensing Digital Image Analysis, 5th ed.; Springer: Berlin/Heidelberg, Germany, 2013.
- 38. Li, M.; Zang, S.; Zhang, B.; Li, S.; Wu, C. A review of remote sensing image classification techniques: The role of Spatio-contextual information. *Eur. J. Remote Sens.* **2014**, 47, 389–411. [CrossRef]
- Belgiu, M.; Drăguţ, L. Random forest in remote sensing: A review of applications and future directions. ISPRS J. Photogramm. Remote Sens. 2016, 114, 24–31. [CrossRef]
- 40. Gómez, C.; White, J.C.; Wulder, M.A. Optical remotely sensed time series data for land cover classification: A review. *ISPRS J. Photogramm. Remote Sens.* **2016**, *116*, 55–72. [CrossRef]
- 41. Yang, G.; Fang, S. Improving remote sensing image classification by exploiting adaptive features and hierarchical hybrid decision trees. *Remote Sens. Lett.* **2017**, *8*, 156–164. [CrossRef]
- 42. Xia, J.; Falco, N.; Benediktsson, J.A.; Du, P.; Chanussot, J. Hyperspectral Image Classification With Rotation Random Forest Via KPCA. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 1601–1609. [CrossRef]
- 43. Hakkenberg, C.R.; Peet, R.K.; Urban, D.L.; Song, C. Modeling plant composition as community continua in a forest landscape with lidar and hyperspectral remote sensing. *Ecol. Appl.* **2018**, *28*, 177–190. [CrossRef]
- 44. Zhang, X.M.; He, G.J.; Zhang, Z.M.; Peng, Y.; Long, T.F. Spectral-spatial multi-feature classification of remote sensing big data based on a random forest classifier for land cover mapping. *Cluster Comput.* **2017**, *20*, 2311–2321. [CrossRef]

- 45. Ho, T.K. Random Decision Forests. In Proceedings of the 3rd International Conference on Document Analysis and Recognition, Montreal, QC, Canada, 14–16 August 1995; IEEE Computer Society Press: Montreal, QC, Canada, 1995; pp. 278–282.
- 46. Kleinberg, E. An overtraining-resistant stochastic modeling method for pattern recognition. *Ann. Stat.* **1996**, 24, 2319–2349. [CrossRef]
- 47. Breiman, L. Random Forests. Mach. Learn. 2001, 45, 5–32. [CrossRef]
- 48. Hastie, T.; Tibshirani, R.; Friedman, J. The Elements of Statistical Learning; Springer: New York, NY, USA, 2009.
- Castelluccio, M.; Poggi, G.; Sansone, C.; Verdoliva, L. Training convolutional neural networks for semantic classification of remote sensing imagery. In Proceedings of the 2017 Joint Urban Remote Sensing Event (JURSE), Dubai, UAE, 6–8 March 2017; pp. 1–4.
- 50. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958. [CrossRef]
- 51. Toloşi, L.; Lengauer, T. Classification with correlated features: Unreliability of feature ranking and solutions. *Bioinformatics* **2011**, *27*, 1986–1994. [CrossRef] [PubMed]
- 52. Liaw, A.; Wiener, M. Classification and regression by randomforest. R News 2002, 2, 18–22. [CrossRef]
- 53. Huang, B.F.F.; Boutros, P.C. The parameter sensitivity of random forests. *BMC Bioinf.* **2016**, *17*, 331. [CrossRef] [PubMed]
- 54. Fisher, J.R.B.; Acosta, E.A.; Dennedy-Frank, P.J.; Kroeger, T.; Boucher, T.M. Impact of satellite imagery spatial resolution on land use classification accuracy and modeled water quality. *Remote Sens. Ecol. Conserv.* **2018**, *4*, 137–149. [CrossRef]
- Roth, K.L.; Roberts, D.A.; Dennison, P.E.; Peterson, S.H.; Alonzo, M. The impact of spatial resolution on the classification of plant species and functional types within imaging spectrometer data. *Remote Sens. Environ.* 2015, 171, 45–57. [CrossRef]
- Mueller-Warrant, G.W.; Whittaker, G.W.; Griffith, S.M.; Banowetz, G.M.; Dugger, B.D.; Garcia, T.S.; Giannico, G.; Boyer, K.L.; McComb, B.C. Remote sensing classification of grass seed cropping practices in western Oregon. *Int. J. Remote Sens.* 2011, 32, 2451–2480. [CrossRef]
- 57. Comber, A.; Umezaki, M.; Zhou, R.; Ding, Y.; Li, Y.; Fu, H.; Jiang, H.; Tewkesbury, A. Using shadows in high-resolution imagery to determine building height. *Remote Sens. Lett.* **2012**, *3*, 551–556. [CrossRef]
- 58. Goetz, J.N.; Brenning, A.; Petschko, H.; Leopold, P. Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling. *Comput. Geosci.* **2015**, *81*, 1–11. [CrossRef]
- Heung, B.; Ho, H.C.; Zhang, J.; Knudby, A.; Bulmer, C.E.; Schmidt, M.G. An overview and comparison of machine-learning techniques for classification purposes in digital soil mapping. *Geoderma* 2016, 265, 62–77. [CrossRef]
- 60. Chicco, D. Ten quick tips for machine learning in computational biology. BioData Min. 2017, 10, 35. [CrossRef]
- 61. Kaartinen, H.; Hyyppä, J.; Yu, X.; Vastaranta, M.; Hyyppä, H.; Kukko, A.; Holopainen, M.; Heipke, C.; Hirschmugl, M.; Morsdorf, F.; et al. An International Comparison of Individual Tree Detection and Extraction Using Airborne Laser Scanning. *Remote Sens.* **2012**, *4*, 950. [CrossRef]
- 62. Strimbu, V.F.; Strimbu, B.M. A graph-based segmentation algorithm for tree crown extraction using airborne LiDAR data. *ISPRS J. Photogramm. Remote Sens.* **2015**, *104*, 30–43. [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/).