

# Evaluating machine learning algorithms to classify forest tree species through satellite imagery

Neiv Gupta<sup>1</sup>, Jeff Li Wen<sup>2</sup>

<sup>1</sup> Monta Vista High School, Cupertino, California

<sup>2</sup> School of Earth, Energy & Environmental Sciences, Stanford University, Stanford, California

## SUMMARY

Recent events indicate an uptick in forest fires in the western United States, prompting cities and organizations to develop a better understanding of forests and how to manage them. Tree species classification is important for forest management and carbon sequestration analysis. Currently, remote sensing stands as the prevalent method for the classification of tree species, land cover, etc. Researchers often use machine learning techniques for classification and general remote sensing. We hypothesized that it is possible to classify forest tree species with high classification accuracy using solely RGB values as the inputs for the machine learning models. We experimented with different machine learning algorithms such as Random Forest (RF), k-Nearest Neighbors (kNN), Gradient Boosting (GB), and Linear Discriminant Analysis (LDA) to classify forest tree species, specifically through multispectral Landsat 8 satellite imagery. Each algorithm was trained and validated using the same dataset and satellite imagery of the same region. Our findings indicated RF had the highest classification accuracy of 95.4% for validation on the same general region it trained on. kNN, GB, and LDA had classification accuracies of 81.6%, 76.4%, and 64.6%, respectively. Based on these results, we concluded that RF is the more accurate algorithm for classifying tree species through RGB satellite imagery. Our findings also indicate that model training and inference on the same general region result in higher classification accuracy. However, as the inference region changes, the classification accuracy reduces. In such cases, additional predictor variables, including trunk diameter, crown shape, and vegetation indices, could be introduced to improve classification accuracy.

## INTRODUCTION

Recent increases in the prevalence and frequency of forest fires in the western United States cause reductions in forest cover and the release of sequestered carbon (1). Cities across the United States are implementing climate action plans that call to establish a baseline and periodically update their greenhouse gas inventories (2). City officials and researchers are also looking to better understand forest structure and forest changes over time for various mitigations, such as reducing high-intensity burns during forest fire events (3).

Classifying and mapping tree species provides an efficient and effective way to construct carbon budget models, manage forest inventories, and protect forest resources (2). Accurate maps and classification are also necessary for effectively monitoring drought and fire conditions, which could pose a severe threat to a forest ecosystem (2, 3). These maps could potentially help firefighters grasp a better understanding of a forest's vegetation and characteristics, which are essential variables to consider when attempting to predict and assess the behavior of an active fire.

Remote sensing is a perfect technique for such tasks, as it provides synoptic views and information over large areas at very high resolutions (4). Specifically for tree species classification, remote sensing through high spectral bands of imagery provides the highest resolution and detail for tree species classification. As a result, airborne hyperspectral light detection and ranging (LiDAR) imagery satisfies the optimal conditions for sensors best suited for tree species classification (5). However, airborne LiDAR is not a practical source of imagery due to its high costs and limited availability. As a result, alternative sources of remotely sensed imagery must be considered. Multispectral satellite imagery, which is widely available, stands as a possible alternative to hyperspectral LiDAR imagery, despite its inability to reach the detail and spectral band variety of hyperspectral LiDAR imagery (4, 6). The terms hyperspectral and multispectral refer to the electromagnetic spectral band variety of the image (6). Hyperspectral imagery encompasses more spectral bands, making it more sophisticated than multispectral imagery (6).

Both LiDAR and multispectral satellite imagery are among the most widely used data sources in remote sensing, which occasionally involves the use of machine learning. The application of machine learning in classification algorithms used in remote sensing has been increasing in popularity because of its processing power and ability to automate the classification processes (2). The algorithms used have been divided into two subcategories: supervised and unsupervised techniques (2). Unsupervised learning algorithms train by making predictions based on the data and actively adjusting for the correct answer, while supervised learning algorithms train from labeled inputs and outputs (2).

These algorithms have become increasingly important for general object classification through hyperspectral imagery and multispectral satellite imagery. For example, the random forest (RF) supervised machine learning algorithm has been used to classify land cover, map ecological zones and landslides, create forest canopy fuel maps for fire forecasting, and analyze urban tree species inventories (2, 3). In these applications, RF has been used with both hyperspectral data and multispectral satellite imagery because of the large number of input variables provided for the algorithm and RF being

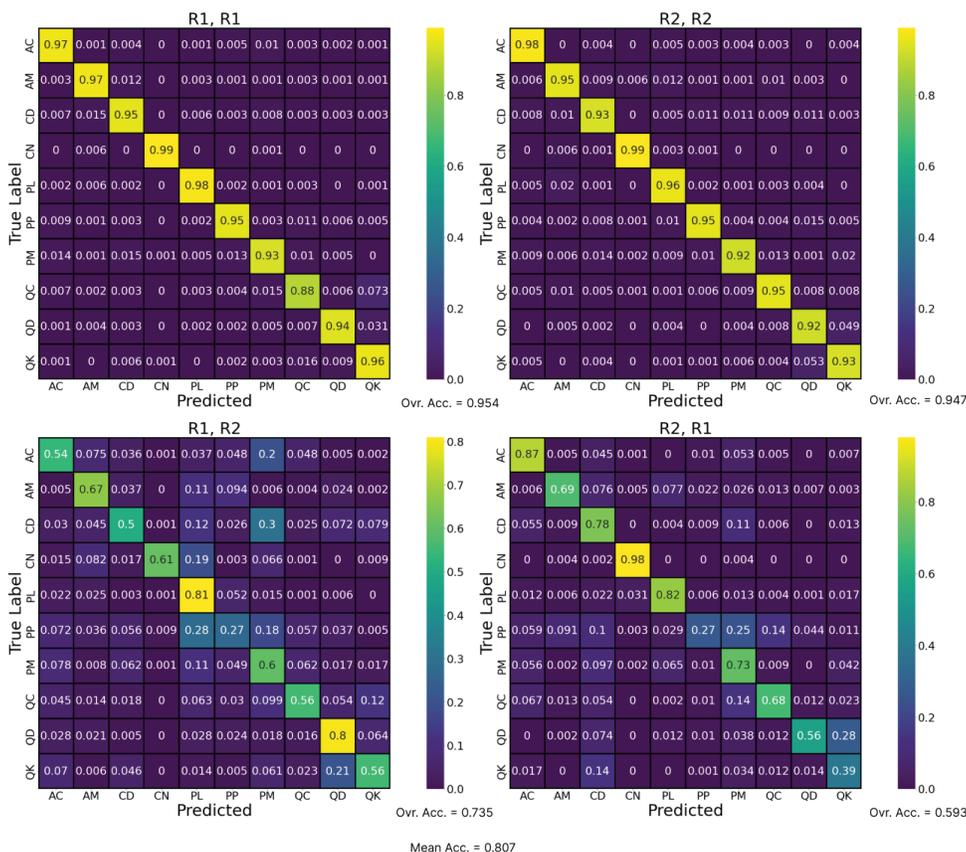
relatively insensitive when faced with small sample sizes (i.e., the amount of training samples for each class is dramatically smaller than the dimension of the feature space) (4, 7, 8). Nevertheless, there have been numerous studies where RF has performed successfully with LIDAR and spectral data (3, 7–9). As for the other algorithms, k-nearest neighbors (kNN), gradient boosting (GB), and linear discriminant analysis (LDA) have all been used for classification through hyperspectral imagery and multispectral satellite imagery to perform similar tasks (10–12).

While researchers in the past have used machine learning algorithms to classify forest tree species through satellite imagery, there lacks a simpler approach, using only red, green, and blue (RGB) strength values and a comparison of performance between these algorithms. In this study, we compared the RF, kNN, GB, and LDA machine learning algorithms for the classification of forest tree species through multispectral satellite imagery and determined if this is possible with high classification accuracy. The algorithms categorized as supervised are RF, kNN, and GB, while LDA is the only unsupervised machine learning algorithm. Generally, supervised learning is more efficient with labeled data, which is why RF, kNN, and GB could potentially prevail as better methods for our type of classification. We also believed that our machine learning models would yield a high classification accuracy because of their performance in

similar classifications in past research (2–9). While object-based classification approaches have proven to be more successful than pixel-based classification approaches, our work examined a more practical pixel-based classification approach with various algorithms to determine the best classification algorithm for this scenario. We trained each model to classify the forest tree species of a certain plot based on the strength of the red, green, and blue bands of the plot’s respective satellite imagery pixel. The results of our study show RF exhibited the highest classification accuracy when training and validating on the same general region and the highest mean classification accuracy when training and validating across regions. kNN, GB, and LDA exhibited mean classification accuracies definitively lower than RF when training and validating across regions. Potential reasons why our models did not perform better include the large scale of our data, complex forest structure, or lack of specific tree data.

### RESULTS

We conducted our study on the Greater Lake Tahoe region/ El Dorado National Forest, California. Our machine learning analysis was conducted on this region using tree species data from the U.S. Department of Agriculture (USDA) Forest Service and satellite imagery from the Landsat 8 Operational Land Imager (OLI). The bounds of the satellite image were



**Figure 1: Confusion matrix results for Random Forest (Training Region, Testing Region).** The y-axis represents the true instances of the tree species designated in Table 2, and the x-axis represents the instances of the same tree species predicted by the algorithm. The overall classification accuracies for train-test pairs (R1, R1), (R2, R2), (R1, R2), and (R2, R1) were 95.4%, 94.7%, 73.5%, and 59.3% respectively. The mean classification accuracy was 80.7%.

very large, encompassing cities, dry grasslands, and desert environments, which could confound our machine-learning models into a wrongful or needless classification. To avoid this, we cropped the dataset to the two further subregions within the larger image, with each region containing a very similar forest structure and dynamics. Not only was the region split necessary for reducing the processing load on our models and increasing performance, but it also allowed us to see if the model was scalable on a different, yet similar, region, even though the model was trained on a separate region that was not included in the inference. Region 1 (R1) was in the Northern Greater Lake Tahoe region, while Region 2 (R2) was in the El Dorado National Forest.

We defined the mean-classification accuracy as the average of the classification accuracy between training on Region 1 and validating on Region 1 (train-test pair of R1, R1), training on Region 2 and validating on Region 2 (R2, R2), training on Region 1 and validating on Region 2 (R1, R2), and training on Region 2 and validating on Region 1 (R2, R1). For training and testing on the same region, a random sample of data points was subset for solely testing and excluded from training to avoid overfitting. We trained our models to classify the forest tree species of a certain plot based on the strength of the red, green, and blue bands of the plot's respective pixel from the satellite image. For each model, we calculated the mean-classification accuracies. We then constructed a

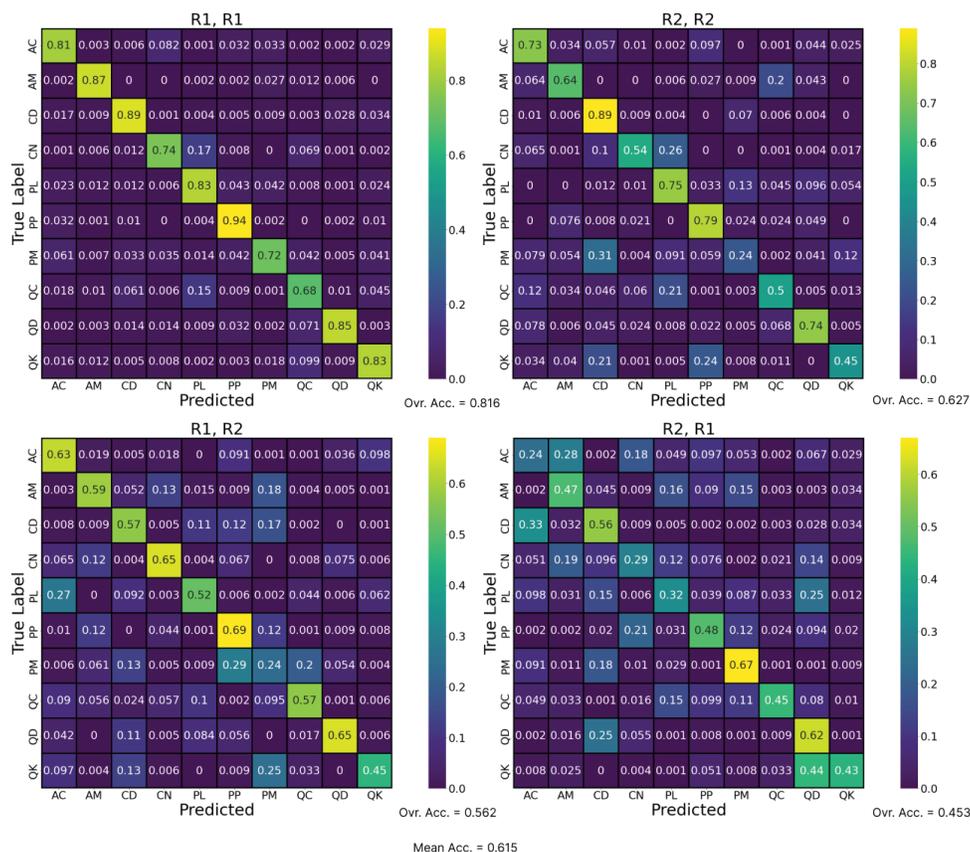
confusion matrix for each train-test pair validation of each model to determine the tree species with the highest mean classification accuracies among all algorithms.

Our findings show RF had the highest mean classification accuracy of 80.7%. For training and validation on the (R1, R1) and (R2, R2) train-test pairs, RF had classification accuracies of 95.4% and 94.7%, respectively. For training and validation on the R1, R2 and R2, R1 train-test pairs, RF had classification accuracies of 73.5% and 59.3%, respectively. The classification accuracies for all four train-test pairs were the highest out of the three other algorithms (Figure 1).

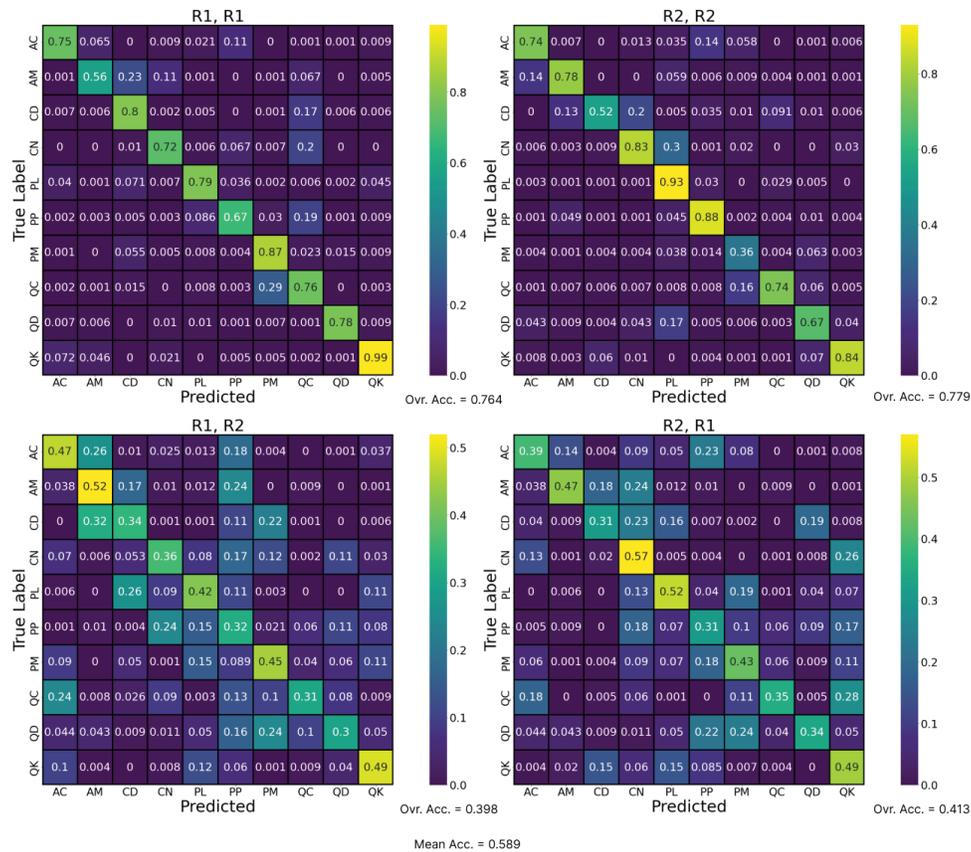
kNN had a mean classification accuracy of 61.5%, lower than RF. kNN had classification accuracies of 81.6% and 62.7% for training and validation on the (R1, R1) and (R2, R2) train-test pairs, respectively. For training and validation on the (R1, R2) and (R2, R1) train-test pairs, kNN had classification accuracies of 56.2% and 45.3%, respectively (Figure 2).

GB performed slightly worse than kNN, having a mean classification accuracy of 58.9%. For training and validation on the (R1, R1) and (R2, R2) train-test pairs, GB had classification accuracies of 76.4% and 77.9%, respectively. GB classification accuracies were 39.8% and 41.3% for training and validation on the (R1, R2) and (R2, R1) train-test pairs, respectively (Figure 3).

LDA performed the worst out of the three other algorithms, with the lowest mean classification accuracy of 49.4%. LDA



**Figure 2: Confusion matrix results for k-Nearest Neighbor (Training Region, Testing Region).** The y-axis represents the true instances of the tree species designated in Table 2, and the x-axis represents the instances of the same tree species predicted by the algorithm. The overall classification accuracies for train-test pairs (R1, R1), (R2, R2), (R1, R2), and (R2, R1) were 81.6%, 62.7%, 56.2%, and 45.3% respectively. The mean classification accuracy was 61.5%.



**Figure 3: Confusion matrix results for Gradient Boosting (Training Region, Testing Region).** The y-axis represents the true instances of the tree species designated in Table 2, and the x-axis represents the instances of the same tree species predicted by the algorithm. The overall classification accuracies for train-test pairs (R1, R1), (R2, R2), (R1, R2), and (R2, R1) were 76.4%, 77.9%, 39.8%, and 41.3% respectively. The mean classification accuracy was 58.9%.

had classification accuracies of 64.6% and 62.6% for training and validation on the (R1, R1) and (R2, R2) train-test pairs, respectively. For training and validation on the (R1, R2) and (R2, R1) train-test pairs, LDA had classification accuracies of 43.0% and 27.4%, respectively (Figure 4).

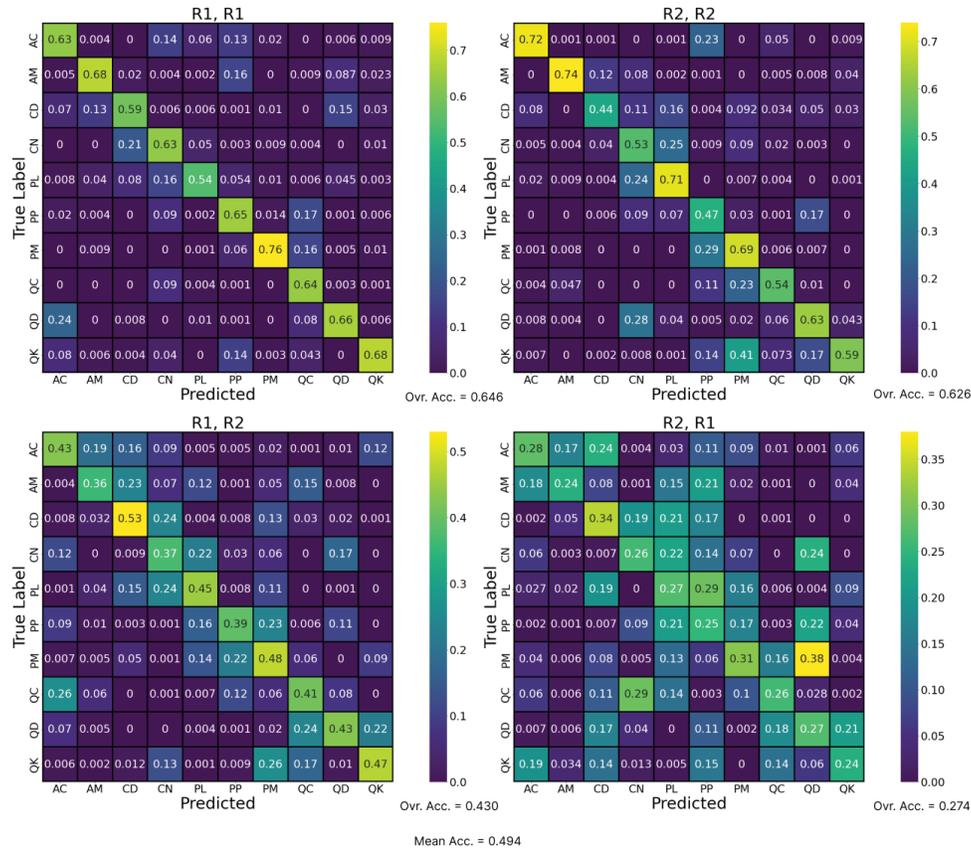
Across all algorithms, the classification accuracies for training and testing on different regions (train test pairs of R1, R2, and R2, R1) were lower than the classification accuracies for training and testing on the same region (train test pairs of R1, R1, and R2, R2) (Figure 5). In addition, the classification accuracies for individual tree species were lower when training and testing on different regions (Figure 5). Altogether, our

results support the conclusion that tree species classification with high classification accuracy using solely RGB values is possible when training and validating machine learning algorithms on the same general region and that RF performs with higher classification accuracy compared to GB, kNN, and LDA for this type of classification.

A potential explanation for the low classification accuracies of our models could be the minimal distinction between RGB strength values pixels between different tree species, which is best observed when looking at the final collapsed and concatenated dataset between the tree species reference data and satellite imagery pixel data (Table 1).

Row	x-coordinate*	y-coordinate*	Blue**	Green**	Red**	Tree species
1	-2.110728e+06	2.057456e+06	8227	9044	8706	<i>Pinus lambertiana</i>
2	-2.110728e+06	2.057456e+06	8678	9501	9711	<i>Quercus chrysolepis</i>
3	-2.062677e+06	2.102086e+06	7902	8732	8322	<i>Abies concolor</i>
4	-2.062677e+06	2.102086e+06	8390	9266	9415	<i>Pinus ponderosa</i>
...	...	...	...	...	...	...

**Table 1: Example rows of the final concatenated and collapsed dataset used for machine learning analysis.** The algorithms use the values of the Blue, Green, and Red columns to predict the label in the Tree species column. The designated tree species and RGB strength values come through a concatenation of the satellite imagery and the USDA Forest Service TreeMap2016. \*Geo coordinates in NAD83 Conus Albers. \*\* Strength of color bands in Landsat 8 satellite imagery using 16-bit digital notation.



**Figure 4: Confusion matrix results for Linear Discriminant Analysis (Training Region, Testing Region).** The y-axis represents the true instances of the tree species designated in Table 2, and the x-axis represents the instances of the same tree species predicted by the algorithm. The overall classification accuracies for train-test pairs (R1, R1), (R2, R2), (R1, R2), and (R2, R1) were 64.6%, 62.6%, 43.0%, and 27.4% respectively. The mean classification accuracy was 49.4%.

## DISCUSSION

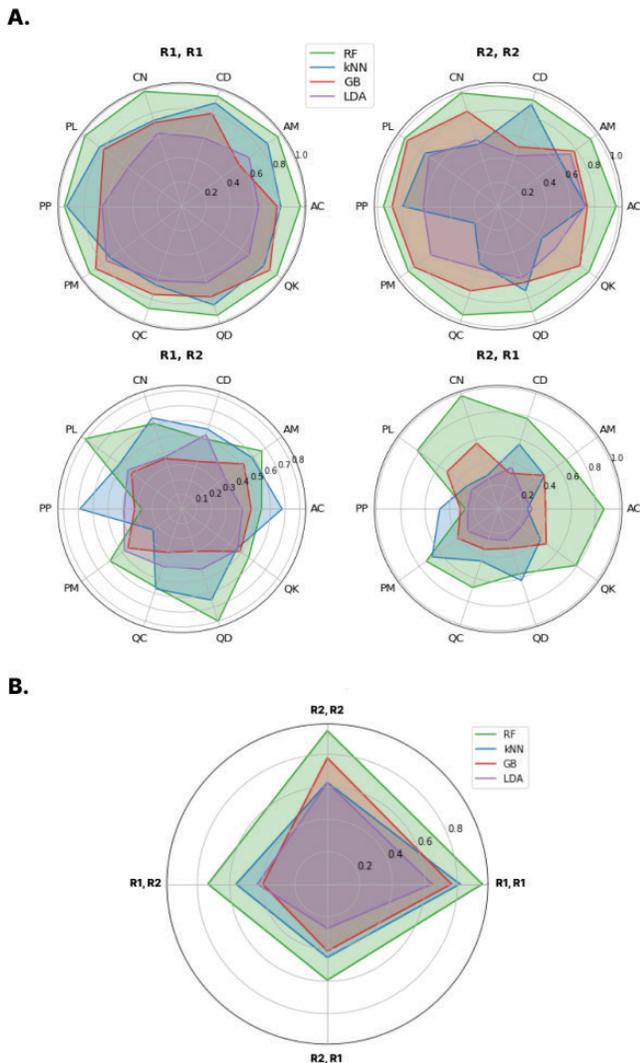
In order to study the applicability of machine learning in tree species classification through satellite imagery, we tested four widely-utilized machine learning algorithms: RF, kNN, GB, and LDA. Among them, we found the RF classification algorithm to have the highest precision compared to the other algorithms. Conversely, the algorithm with the lowest performance was LDA. We also found that training and testing on the same general region result in higher classification accuracies than training and testing on different regions across all algorithms. For instance, when validating the R2-trained RF model, the classification accuracy for validation on R2 was 36.1% higher than validation on R1. Across all algorithms, the percentage drop-off in classification accuracy between validation on the same and different regions ranged from 17.4-36.6%. In addition, the classification accuracies for models validated on different regions were remarkably lower than models training on the same region, with no model resulting in a classification accuracy above 74%. This demonstrates that the algorithms perform with higher precision when training and validating on the same general region.

There could be many underlying reasons for such a substantial drop-off in the classification accuracy when validating on a different region. A few possible reasons include the size of our cropped datasets, the complex forest structure of the study region, a top-viewed pixel-based

classification approach for specific tree species based on a large plot of land, and, as mentioned, very minimal distinction and delineation between RGB strength values pixels between different tree species.

The strength values of the red, green, and blue bands of satellite imagery for PS and QC tree species have little distinction, with the 16-bit digital notation having at most a difference of 700 among all three RGB strength values (Table 1). Such minute differences between tree species were widely spread across the dataset, which increases the probability of misclassification from the machine learning algorithms. Adding more data, such as vegetation and land cover indexes, could draw a stronger distinction between separate tree species, regardless of the environment the tree is surrounded by, and potentially increase the classification accuracy of the models.

While pixel-based classification is a viable method for classifying three-dimensional objects, numerous studies have shown object-based classification prevails as a better method, especially for tree species and land cover classifications. In addition, pixel-based classification fails to account for spatial patterns in a three-dimensional object, limiting its power and accuracy for classification compared to object-based classification (4, 19, 20). For instance, object-based classification allows for the utilization of crown shape, height, and trunk dimensions, while pixel-based classification



**Figure 5: Classification accuracies for each model for individual tree species and train-test pairs (Training Region, Testing Region).** The spider charts compare overall performance and train/test pair-specific performance for each algorithm. The distance of an algorithm's polygon's edge to the end of the spoke reflects the accuracy the algorithm demonstrated for that specific label. (A) Tree species-specific classification accuracies for each model on all train/test pairs. (B) Train/test pair-specific classification accuracies for each model (Training Region, Testing Region).

using solely RGB strength values only encompasses color variation and possibly crown delineation between trees. An object-based classification approach would not only allow for more predictor variables to be used and draw a stronger distinction between tree species but also isolate the tree to solely spatial patterns/features. In addition, this approach will substantially subsidize the influence of the surrounding environment on the model's prediction, potentially increasing the classification accuracy for training and testing on different general regions. Past research has demonstrated success with applying satellite imagery to object-based approaches, but it required predefining object-specific attributes in complement with satellite imagery (4). Nevertheless, when possessing the necessary spectral band variety, satellite

imagery does present the scope for the inclusion of vegetation indices and chlorophyll content as predictor variables, which could potentially improve classification accuracy by increasing specie to specie distinction (4, 21).

Using RGB satellite imagery alone to classify forest tree species did not yield favorable results as the region changes and expands, but combining RGB satellite imagery and other data points, such as vegetation indexes, foliage height rasters, and more bands of satellite imagery, could potentially yield a more accurate model. By bringing in more variables, we can potentially draw a stronger distinction among tree species and isolate them from their surrounding environment, which will ultimately strengthen our models to classify distinct tree species.

A possible alternative to using satellite imagery would be LiDAR imagery. Past research has demonstrated that classification using LiDAR imagery results in higher classification accuracies compared to classification using satellite imagery (21, 22). LiDAR imagery allows for models to leverage metrics such as tree height, trunk diameter, and crown shape, which will help draw a more robust distinction between tree species rather than solely RGB values without being too heavily influenced by the surrounding climate and environment. However, LiDAR imagery is not a largely available data source and requires expensive infrastructure to be put in place for data collection. In addition, data on object-specific attributes of tree species is not largely available and requires field data collection, which also requires expensive infrastructure and resources (4). At the same time, LiDAR data and object-based classification methods would be a far more robust approach to classifying tree species. With adequate infrastructure and resources, this approach is feasible.

When processing the data for analysis, we ran into numerous memory issues because of the size and scale of our datasets. To bypass this issue, we initially experimented with incremental learning and k-fold cross-validation as a possible solution, but our system continued to run into memory issues. Especially for studies concerning large datasets, incremental learning allows for the model to be trained from a series of batches, compared to the entire dataset at once, which could pose issues depending on the strength of the system used for data analysis and processing. Specifically, incremental learning is learning through streaming data, which arrives over time without sacrificing the model's accuracy. As a result, the models' overall accuracy when training and validating different general regions could potentially have improved with a stronger system designed for handling larger datasets and a successful implementation of incremental learning (23).

A potential reason LDA performed the worst out of the algorithms is that LDA is an unsupervised learning algorithm. Unsupervised learning algorithms train by making predictions based on the data and actively adjusting for the correct answer (24). On the other hand, supervised learning algorithms train from labeled inputs and outputs (24). For applications like tree species classification using labeled data, supervised algorithms like RF, kNN, and GB can learn through pattern recognition and actively measure their accuracy through loss functions, which the algorithms use to minimize error (24). This ultimately makes them more efficient than LDA, as LDA is not tailored for labeled data and does not actively minimize error (24). Therefore, the performance of LDA was not as strong as the supervised learning algorithms.

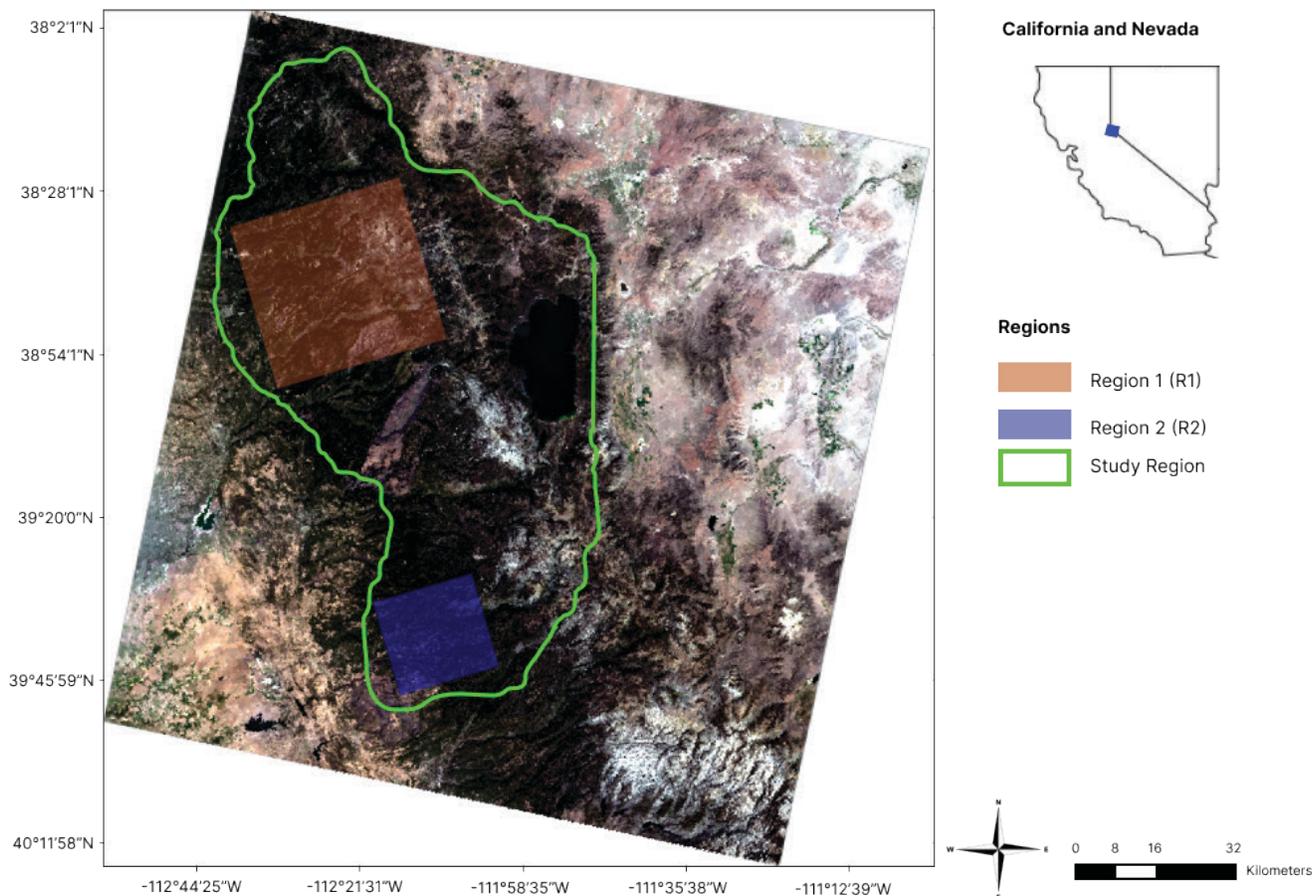
This study focused on experimenting with various machine learning algorithms to classify forest tree species through satellite imagery. The machine learning algorithms used included RF, kNN, GB, and LDA classifiers. Our models were trained to classify the forest tree species of a certain plot based on the strength of the red, green, and blue bands of a pixel from the satellite image. Our findings indicated that classification based solely on RGB values is possible with high accuracy when training on the same general region. RF performed the best, having the highest classification accuracy of 95.4% when training on the same general region and the highest mean classification accuracy of 80.7%. For all algorithms, however, there was a substantial decrease in classification accuracy when validating on a different general region. When trained and validated on the same general region, our models provide accurate classifications of forest tree species, which are important for carbon sequestration analysis, forest management, and fuel treatment. Further research that uses vegetation indexes, object-based classification approaches, or incremental learning approaches could potentially yield higher classification accuracies and help construct more robust models.

## MATERIALS AND METHODS

### Study Area and Data

Our region of study was located in the western United States, specifically the Greater Lake Tahoe region/El Dorado National Forest, California (39°58'N, -121°24' W). The area is a mix of mountainous terrain and dense temperate forest with elevations ranging from 0 m to 1898 m above sea level, which adds to the marked variance among tree species in the area. Our satellite image, downloaded from the United States Geological Survey database, was captured from the Landsat 8 OLI (Figure 6) (26). The tree species discussed in our study, which were the most prevalent in our region of study, are white fir (*Abies concolor*, AC), Pacific madrone (*Arbutus menziesii*, AM), California incense-cedar (*Calocedrus decurrens*, CD), Pacific dogwood (*Cornus nuttallii*, CN), sugar pine (*Pinus lambertiana*, PL), ponderosa pine (*Pinus ponderosa*, PP), Douglas fir (*Pseudotsuga menziesii*, PM), canyon live oak (*Quercus chrysolepis*, QC), blue oak (*Quercus douglasii*, QD), and California black oak (*Quercus kelloggii*, QK) (Table 2).

We broke up the dataset into two small subregions, one in the northern region and one in the southern region of the Greater Lake Tahoe area. We optimized our data this way



**Figure 6: Study site and location of testing regions.** The image was acquired from the United States Geological Survey database and captured by the Landsat 8 OLI satellite.

Scientific Name	Common Name	Abbreviation	R1 Count (# of trees)	R1 Proportion (%)	R2 Count (# of trees)	R2 Proportion (%)
<i>Abies concolor</i>	White fir	AC	377,133	24.3	84,118	23.9
<i>Arbutus menziesii</i>	Pacific madrone	AM	30,488	2.0	17,695	22.7
<i>Calocedrus decurrens</i>	California incense-ceda	CD	179,279	11.6	148,561	13.6
<i>Cornus nuttallii</i>	Pacific dogwood	CN	68,678	4.4	11,065	13.5
<i>Pinus lambertiana</i>	Sugar pine	PL	148,517	9.6	63,330	10.2
<i>Pinus ponderosa</i>	Ponderosa pine	PP	88,539	5.7	24,528	4.4
<i>Pseudotsuga menziesii</i>	Douglas fir	PM	467,736	30.2	140,873	4.0
<i>Quercus chrysolepis</i>	Canyon live oak	QC	125,265	8.1	84,662	3.0
<i>Quercus douglasii</i>	Blue oak	QD	25,377	1.6	27,541	2.9
<i>Quercus kelloggii</i>	California black oak	QK	38,338	2.5	18,666	1.8
<b>Total</b>			<b>1,549,350</b>		<b>621,039</b>	

**Table 2: The distribution of tree species in the regions of study.** The total number of trees of the selected species in R1 and R2 is also included. The data were imputed from the Forest Inventory and Analysis database, which the USDA Forest Service matched onto a raster grid. We processed the raster data of our regions of study and computed the distributions for the most prevalent tree species in the regions.

because the scale of the study site was too large to train and cross-validate our machine learning models. The bounds of the full satellite image also contained non-forested land, such as shrubland, agricultural land, urban areas, etc., which could confound our models and lead to misclassification. In addition, we specifically selected the two subregions in the northern and southern regions of the Greater Lake Tahoe region because it allowed us to see if the models were scalable on very similar, but not identical, regions. Both regions had a similar distribution of tree species, with the discussed tree species being the most prevalent in the regions. The reason for the distribution of tree species not being identical across both regions is that the distribution of tree species varies due to environmental factors such as altitude, mean temperature, forest density, weather patterns, etc. (Figure 6) (13).

Our ground truth data came from the USDA Forest Service TreeMap2016, a tree-level model of the forests in the conterminous United States (14). The authors of the dataset matched Forest Inventory and Analysis plot data to a 30x30m

raster grid. The dataset comprises a raster map GeoTIFF file, which contains plot identifiers for each pixel, and a TreeMap table CSV file that, when referenced to the plot identifier in the raster data, can report the most prevalent tree species for that plot (14).

### Preprocessing and Data Formatting

The bands of satellite imagery we used for our analysis were the second band (blue wavelength, 450–510 nm), third band (green wavelength, 530–590 nm), and fourth band (red wavelength, 640–670 nm). The image had a resolution of 30 meters, which is the default resolution for Landsat 8 OLI satellite imagery. To create an RGB satellite image that we could input into our machine learning models to classify the tree species in the area, we concatenated and overlaid the bands of imagery into a single GeoTIFF image.

To prevent challenges while data processing, we also re-projected the satellite imagery, formerly onto the WGS84 coordinate reference system (CRS), to match the CRS of the

TreeMap2016 raster image, which is on the NAD83 Conus Albers CRS. In addition, we cropped the dataset to match the bounds of the satellite image, considering the TreeMap2016 dataset contained data on the entire conterminous United States. Due to the bounds of the full satellite image encompassing non-forested land and our system running into image processing constraints, we cropped the dataset to the two further subregions within the larger image.

To prevent data imbalances and an uneven dataset, which could confound our machine learning models, we balanced our datasets using random undersampling. Random undersampling balances an uneven dataset by keeping all data points in a minority class and decreasing the size of the majority class to match the size of the minority class (25). The data points that are removed from the majority classes are chosen at random (25).

### Classification

For each machine learning algorithm, we used a cross-validation approach. We trained our models on Region 1 and validated on Region 1 (train-test pair of R1, R1), trained on Region 2 and validated on Region 2 (train-test pair of R2, R2), trained on Region 1 and validated on Region 2 (train-test pair of R1, R2), and trained on Region 2 and validated on Region 1 (train-test pair of R2, R1). For training and testing on the same region (train-test pairs of R1, R1, and R2, R2), we subset a random sample of data points solely testing and excluded this sample from training to remove the possibility of overfitting. All classification and data processing was conducted in Google Colab using Python 3.6.

For our machine learning analysis, we used pixel-based classification methods to classify forest tree species through satellite imagery. Our inputs for our models were the strengths values of the red, green, and blue bands of satellite imagery represented as a 16-bit digital notation. All classification models were trained and validated using the Python library scikit-learn version 1.2.1.

We started by using the RF classifier. RF is a non-parametric ensemble learning algorithm consisting of a large number of decision trees, which enhances traditional decision trees (4, 15). An individual bootstrapping sample (sampling with replacement) is utilized to construct each decision tree (4, 15). At each node of the tree, the split determination is based on the Gini criterion (4, 15). With standard decision trees, nodes are split by the variable that provides the best split or the highest decrease in Gini (4, 15). However, RF randomly selects a subset of variables at each node and chooses the best splitting variable (4, 15). New data are classified from a majority vote among the classification outcomes of all constructed decision trees (4, 15). For determining a rough estimate of the classification error, the out-of-bag data (OOB), the samples not in the bootstrapping sample, are used (4, 15). With the OOB dataset, each decision tree is used to classify the samples (4, 15). Finally, for each sample in the original data set, the majority vote of the corresponding decision trees is compared with the truth labels, resulting in an estimate of the misclassification rate (4,15). For our model, we set our parameters such that `warm_start=False`, `n_estimators=100`, and the `max_depth=100`.

We also applied the kNN classifier. kNN was developed from the need to perform discriminant analysis when reliable parametric estimates of probability densities are

difficult to determine or unknown (16). It is a nonparametric learning algorithm that makes no initial assumptions about a primary dataset (16). The classification involves classifying an unknown sample based on the classifications of the neighboring samples (16). The optimal choice of the chosen number of neighbors (k) depends on the metrics used for classification and regression purposes (16). For our analysis, we used 1000 neighbors to classify tree species because of our data's large scale and wide range of possible class labels (16).

Another tree-based ensemble algorithm we applied was the GB algorithm. GB is a powerful ensemble algorithm that employs decision trees to construct the classifiers (10, 17). In addition, the algorithm applies iteration by adding new models to correct weaknesses found in former models, improving the overall performance accuracy of the model (10, 17). Essentially, a decision tree or linear regression that improves the model most is added to an ensemble at each iteration until the set number of estimators (`n_estimators`) has been reached (10, 17). The main difference between GB and RF is that GB utilizes shallow trees with fewer splits than RF, which uses trees to their maximum extent (10, 17). For our model, we set our parameters such that `n_estimators=100`, `learning_rate=0.1`, and the `max_depth=100`.

The unsupervised machine learning algorithm we experimented with was LDA. LDA is a dimensionality reduction algorithm that maximizes the ratio of between-class variance to the within-class variance, guaranteeing maximum separability (18). Essentially, LDA is an algorithm that tries to maximize the distinction between multiple classes, which allows it to classify input data into a specific class (18). For our analysis, we used a class-dependent transformation, which involves maximizing the separability between classes and utilizing two optimizing criteria for the independent transformation of the data sets (18). For our model, we used a singular value decomposition solver, no shrinkage, and set `n_components=None`.

**Received:** June 18, 2022

**Accepted:** November 2, 2022

**Published:** March 18, 2023

### REFERENCES

1. Brown, Timothy J., *et al.* "The Impact of Twenty-First Century Climate Change on Wildland Fire Danger in the Western United States: An Applications Perspective." *Climatic Change*, vol. 62, no.1, Jan. 2004, pp. 365–388. doi:10.1023/B:CLIM.0000013680.07783.de.
2. Talukdar, Swapan, *et al.* "Land-Use Land-Cover Classification by Machine Learning Classifiers for Satellite Observations—A Review." *Remote Sensing*, vol.12, no. 7, 2 Apr. 2020., pp. 1135. doi:10.3390/rs12071135.
3. Ballanti, Laurel, *et al.* "Tree Species Classification Using Hyperspectral Imagery: A Comparison of Two Classifiers." *Remote Sensing*, vol. 8, no. 6, 24 May 2016, pp. 445. doi:10.3390/rs8060445.
4. Immitzer, Markus, *et al.* "Tree Species Classification with Random Forest Using Very High Spatial Resolution 8-Band WorldView-2 Satellite Data." *Remote Sensing*, vol. 4, no. 9, 14 Sep. 2012, pp. 2661–2693. doi:10.3390/rs4092661.
5. Koukal, Tatjana and Atzberger, Clement. "Potential

- of Multi-Angular Data Derived From a Digital Aerial Frame Camera for Forest Classification.” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 5, no. 1, Feb. 2012, pp. 30–43. doi:10.1109/JSTARS.2012.2184527.
6. Wang, Yutang, *et al.* “Classification of Street Tree Species Using UAV Tilt Photogrammetry.” *Remote Sensing*, vol. 13, no. 2, 10 Jan. 2021, pp. 216. doi:10.3390/rs13020216.
  7. Ghimire, B. “Contextual land-cover classification: incorporating spatial dependence in land-cover classification models using random forests and the Getis statistic.” *Remote Sensing Letters*, vol. 1, no. 1, 22 Jan. 2010, pp. 45–54. doi:10.1080/01431160903252327.
  8. Clark, Matthew L. and Roberts, Dar A. “Species-Level Differences in Hyperspectral Metrics among Tropical Rainforest Trees as Determined by a Tree-Based Classifier.” *Remote Sensing*, vol. 4, no. 6, 18 June 2012, pp. 1820–1855. doi:10.3390/rs4061820.
  9. Raczko, Edwin and Zagajewski, Bogdan. “Comparison of support vector machine, random forest and neural network classifiers for tree species classification on airborne hyperspectral APEX images.” *European Journal of Remote Sensing*, vol. 50, no.1, 9 Mar. 2017, pp. 144–154. doi:10.1080/22797254.2017.1299557.
  10. Sun Fei, *et al.* “Efficiency of Extreme Gradient Boosting for Imbalanced Land Cover Classification Using an Extended Margin and Disagreement Performance.” *ISPRS International Journal of Geo-Information*, vol. 8, no. 7, 23 July 2019, pp. 315. doi:10.3390/ijgi8070315.
  11. Zhang, Yanchao, *et al.* “Fusion of Multispectral Aerial Imagery and Vegetation Indices for Machine Learning-Based Ground Classification.” *Remote Sensing*, vol. 13, no. 8, 7 Apr. 2021, pp. 1411. doi:10.3390/rs13081411.
  12. Davidson, Scott J., *et al.* “Mapping Arctic Tundra Vegetation Communities Using Field Spectroscopy and Multispectral Satellite Data in North Alaska, USA.” *Remote Sensing*, vol. 8, no. 12, 26 Nov. 2016, pp. 978. doi:10.3390/rs8120978.
  13. Ricklefs, Robert E. and He, Fangliang. “Region effects influence local tree species diversity.” *Proceedings of the National Academy of Sciences*, vol. 113, no. 3, 5 Jan. 2016. pp. 674–679. doi:10.1073/pnas.1523683113.
  14. Riley, Karin L., *et al.* “TreeMap 2016: A tree-level model of the forests of the conterminous United States circa 2016.” *Forest Service Research Data Archive*, 2021. doi:10.2737/RDS-2021-0074.
  15. Breiman, Leo. “Random Forests.” *Machine Learning*, vol. 45, no.1, 1 Oct. 2001, pp. 5–32. doi:10.1023/A:1010933404324.
  16. Peterson, Leif. “K-nearest neighbor.” *Scholarpedia*, vol. 4, no. 2, 21 Feb. 2009, 1883. doi:10.4249/scholarpedia1883.
  17. Natekin, Alexey and Knoll, Alois. “Gradient boosting machines, a tutorial.” *Frontiers in Neuro Robotics*, vol. 7, no. 21, 04 Dec. 2013, doi:10.3389/fnbot.2013.00021.
  18. Fisher, Adrian, and Danaher, Tim. “A Water Index for SPOT5 HRG Satellite Imagery, New South Wales, Australia, Determined by Linear Discriminant Analysis.” *Remote Sensing*, vol. 5, no. 11, 13 Nov. 2013, pp. 5907–5925. doi:10.3390/rs5115907.
  19. Cleve, Casey, *et al.* “Classification of the wildland-urban interface: A comparison of pixel- and object-based classifications using high-resolution aerial photography.” *Computers, Environment and Urban Systems*, vol. 32, no. 4, July 2008, pp. 317–326. doi:10.1016/j.compenvurbsys.2007.10.001.
  20. Sheeren, David, *et al.* “Tree Species Classification in Temperate Forests Using Formosat-2 Satellite Image Time Series”. *Remote Sensing*, vol. 8, no. 9, 7 Sep. 2016, pp. 734. doi:10.3390/rs8090734.
  21. Ma, Quin, *et al.* “Comparison of Canopy Cover Estimations From Airborne LiDAR, Aerial Imagery, and Satellite Imagery.” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 9, 13 June 2017, pp. 4225–4236. doi:10.1109/JSTARS.2017.2711482.
  22. Maxa, Melissa and Bolstad, Paul. “Mapping northern wetlands with high-resolution satellite images and LiDAR.” *Wetlands*, vol. 29, no.1, 1 Mar. 2009, pp. 248. doi:10.1672/08-91.1.
  23. He, Chen, *et al.* “Incremental Learning From Stream Data.” *IEEE Transactions on Neural Networks*, vol. 22, no. 12, 31 Oct. 2011, pp. 1901-1914. doi:10.1109/TNN.2011.2171713.
  24. Alloghani, Mohamed, *et al.* “A Systematic Review on Supervised and Unsupervised Machine Learning Algorithms for Data Science.” *Unsupervised and Semi-Supervised Learning, Springer International Publishing*, 5 Sep. 2019, pp. 3–21. doi:10.1007/978-3-030-22475-2\_1.
  25. Hasanin, Tawfiq, and Taghi Khoshgoftaar. “The Effects of Random Undersampling with Simulated Class Imbalance for Big Data.” *2018 IEEE International Conference on Information Reuse and Integration (IRI)*, 9 July 2018, doi:10.1109/IRI.2018.00018.
  26. U.S. Geological Survey, *EarthExplorer*, 8 Sep. 2016, earthexplorer.usgs.gov

**Copyright:** © 2023 Gupta and Wen. All JEI articles are distributed under the attribution non-commercial, no derivative license (<http://creativecommons.org/licenses/by-nc-nd/3.0/>). This means that anyone is free to share, copy and distribute an unaltered article for non-commercial purposes provided the original author and source is credited.