Journal of Research High School

Volume 2023 (1) 228 - 234
ISSN 2688-3651 [online]

Scaling an Ensemble ML Algorithm for the Classification of Tree Species Through Satellite Imagery

Neiv Gupta[1] *

[1]Monta Vista High School, Cupertino, CA USA

Received March 12, 2023; Revised April 17, 2023; Accepted, June 9, 2023

Abstract

Dry conditions in the Western United States have increased the frequency and severity of forest fires in the Sierra Nevada Mountain Range. Organizations and cities are actively working towards developing a better understanding of forest structure and dynamics. While tree species classification models in the past have dealt with smaller regions and fewer trees, we hypothesized that it is possible to scale the area and number of trees analyzed by our model without sacrificing model accuracy by adding additional variables to satellite imagery, such as Normalized Difference Vegetation Index (NDVI), Normalized Difference Moisture Index (NDMI), Soil-Adjusted Vegetation Index (SAVI), crown ratio, tree height, and tree diameter. We compared the results of applying the Random Forest (RF) Machine Learning (ML) algorithm to a dataset containing satellite imagery alone and with a dataset containing satellite imagery augmented with object-specific attributes (OSA) such as crown shape, tree height, and tree diameter. We then trained and tested the algorithm across two large and different regions with similar tree species prevalence. After the addition of OSA to training data, the results from the experiment demonstrated a mean classification accuracy increase from 66.4% to 90.2%, thus allowing the ML model to scale over larger regions.

Keywords: Environmental Engineering, Remote Sensing, Forest Management, Machine Learning, Classification

1. Introduction

With the recent increase in frequency, intensity, and duration of forest fire events in the Western United States, cities and researchers are looking to better understand forest structures to prevent and mitigate large fire events. Classifying and mapping tree species provides an efficient and effective way to manage forest inventories and protect forest resources. Accurate maps are also necessary for effectively monitoring drought and fire conditions, which could severely threaten a forest ecosystem (Talukdar, et al., 2020; Ballanti, et al., 2016). These maps could help firefighters better understand 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, providing synoptic views and information over large areas at very high resolutions. 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 (Immitzer, et al., 2012). LiDAR allows for the capture of spatial patterns of on-the-ground features through multiple spectral bands, which makes it a very useful tool in the field of remote

* Corresponding Author
neiv06@gmail.com

Advisor: Jeff Li Wen
jlwen@stanford.edu

sensing, specifically for the classification of remotely sensed objects. 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 is a possible alternative to hyperspectral LiDAR imagery, despite its inability to reach the detail and spectral band variety of hyperspectral lidar imagery (Immitzer, et al., 2012; Wang, et al., 2021). The terms hyperspectral and multispectral refer to the electromagnetic spectral band variety of the image. Hyperspectral imagery encompasses more spectral bands, making it more sophisticated than multispectral imagery (Wang, et al., 2021).

The application of machine learning in classification algorithms used in the general field of remote sensing has been increasing in popularity. These algorithms have become increasingly important for general object classification through hyperspectral imagery and multispectral satellite imagery. For example, past research and applications have used the RF machine learning algorithm to classify land cover, map ecological zones and landslides, create forest canopy fuel maps for fire forecasting, and analyze urban tree species inventories (Ballanti, et al. 2016; Immitzer, et al., 2012). 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.

While RF has performed successfully with LiDAR and spectral data in past research (Ballanti, et al., 2016; Ghimire, 2010; Clark and Roberts, 2012), our experimentation demonstrates RF providing 66.4% mean classification accuracy when using satellite imagery alone. We also trained and validated across different regions with similar prevalence of tree species. In this study, we experimented with including OSA, such as crown ratio, tree height, and tree diameter, with satellite imagery to improve the classification accuracy across larger and geologically diverse regions. Our goal was to scale the model by improving model performance over a more extensive area with variations in topographical features and vegetation. We hypothesized that RF would demonstrate higher classification accuracy with the addition of OSA.

2. Materials and Methods

2.1 Study Area and Data

Our region of study was the Greater Lake Tahoe region/El Dorado National Forest, California (39°58'N, -121°24' W). Our satellite image, downloaded from the United States Geological Survey (USGS) website, was captured from the Landsat 8 Operational Land Imager (OLI). 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 significant variance among tree species in the area. Of this larger region, we broke up the dataset into two small subregions, one in the Northern Greater Lake Tahoe Region (R1) and the other in the El Dorado National Forest (R2), south of the Greater Lake Tahoe Region. We optimized our data this way 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.

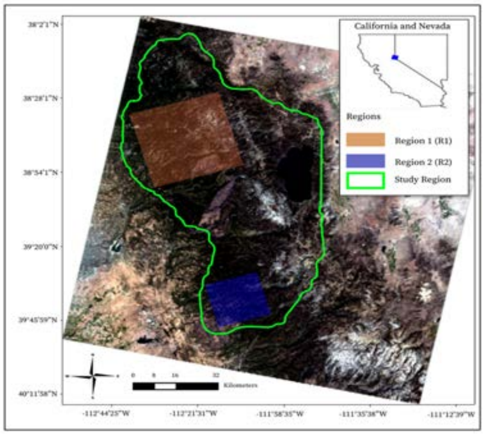


Figure 1. Study site and location of testing regions. The image was acquired from the United States Geological Survey database and captured by the Landsat 8 OLI satellite.

Our ground truth data came from the US Department of Agriculture (USDA) Forest Service TreeMap2016, a tree-level model of forests in the conterminous United States. Both regions have a similar distribution of tree species, with the tree species discussed in our study being the most prevalent in the region. 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.

Table 1. The distribution of tree species in the regions of study. The table also includes the total number of trees of the selected species in R1 and R2. 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 study regions and computed the distributions for the most prevalent tree species in the regions.

Scientific Name	R1 Count		R2 Count	
	(‘000)	%	(‘000)	%
Abies concolor (AC)	377.1	24.3	84.1	23.9
Arbutus menziesii (AM)	30.5	2.0	17.7	22.7
Calocedrus decurrens (CD)	179.3	11.6	148.6	13.6
Cornus nuttallii (CN)	68.7	4.4	11.1	13.5
Pinus lambertiana (PL)	148.5	9.6	63.3	10.2
Pinus ponderosa (PP)	88.5	5.7	24.5	4.4
Pseudotsuga menziesii (PM)	467.6	30.2	140.9	4.0
Quercus chrysolepis (QC)	125.3	8.1	84.7	3.0
Quercus douglasii (QD)	25.4	1.6	27.5	2.9
Quercus kelloggii (QK)	38.3	2.5	18.7	1.8
Total	1,549.4	100.0	621.0	100.0

2.2 Pre-processing and Data Formatting

The satellite imagery we used for our analysis were the second band (blue, 0.450 - 0.51 µm), third band, (green, 0.53 - 0.59 µm), fourth band (red, 0.64 - 0.67 µm), fifth band (near infra-red (NIR), 0.85-0.88 µm) and sixth band (Short-wave infrared (SWIR1), 1.57-1.65 µm). NDVI, NDMI, and SAVI were computed using NIR and SWIR1. Crown Ratio, Tree Height, and Tree Diameter were obtained using the USDA TreeMap2016 data set.

We also re-projected the satellite imagery on the WGS84 coordinate reference system (CRS) to match the TreeMap2016 raster image's CRS, NAD83 Conus Albers. 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.

We balanced our datasets using random undersampling to prevent data imbalances and an uneven dataset. 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. The data points removed from the majority classes are chosen randomly (Hasanin and Khoshgoftaar, 2018).

2.3 Classification

For model training, we used a cross-validation approach. We trained our models on Region 1 and validated on Region 2 (train-test pair of R1, R2), and trained on Region 2 and validated on Region 1. We performed this experiment using satellite imagery alone as well as satellite imagery with OSA data. The inputs for our models were NDMI, NDVI, SAVI, the strength values of the red, green, and blue bands of satellite imagery represented as a 16-bit digital notation, as well as OSA data using the USGS TreeMap2016 dataset.

For our classification, we applied the RF machine learning algorithm. RF is a non-parametric ensemble learning algorithm consisting of a large number of decision trees, which enhances traditional decision trees. An individual bootstrapping sample (sampling with replacement) is utilized to construct each decision tree. At each node of the tree, the split determination is based on the Gini criterion. With standard decision trees, nodes are split by the variable that provides the best split or the highest decrease in Gini. However, RF randomly selects a subset of variables at each node and chooses the best splitting variable. New data are classified from a majority vote among the classification outcomes of all constructed decision trees. For determining a rough estimate of the classification error, the out-of-bag data (OOB), the samples not in the bootstrapping sample, are used. Each decision tree is used to classify the samples with the OOB dataset. 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 (Immitzer, et al., 2012; Breiman, 2001). For our model, we set our parameters such that warm_start=False, n_estimators=100, and max_depth=100. These parameters ensure the algorithm uses adequate decision trees with significant depth to increase the robustness of the model and improve model performance.

3. Results

For each train-test pair for both satellite imagery and satellite imagery + OSA, we calculated the classification accuracy, precision, recall, F1-score, and Cohen's Kappa coefficient. We then constructed a confusion matrix for each train-test pair for satellite imagery + OSA to determine the tree species with the highest mean classification accuracies among all algorithms. Model precision indicates the accuracy of the model in terms of how many instances that the model classified as a certain label were actually correct. Model recall indicates the accuracy of the model in terms of how many instances were correctly classified over the total number of instances for that specific label. The F1-score is the harmonic mean of precision and recall. Kappa reflects the model's true accuracy without the addition of correct classifications due to random chance (Yacouby and Axman, 2020).

For training and validation on satellite imagery + OSA, our model exhibited a mean classification accuracy of 90.20%, mean precision of 90.90%, mean recall of 90.14%, mean F1-Score of 90.14% and mean Kappa of 89.07%.

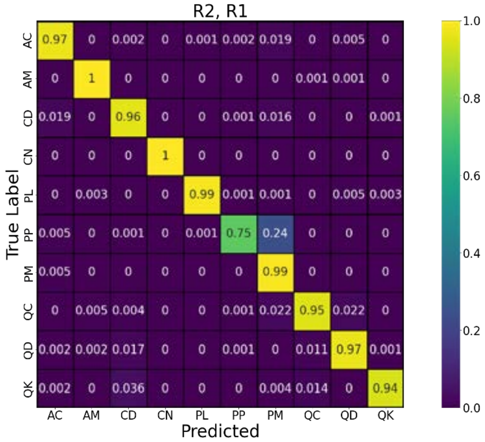
Table 2. Percentage Accuracy Metrics Table, Satellite Imagery + OSA

Train-Test	*Accuracy*	*Precision*	*Recall*	*F1-Score*	*Kappa*
(R1, R2)	95.17	95.75	95.07	95.09	94.53
(R2, R1)	85.23	86.04	85.2	85.18	83.6

For training and validation on satellite imagery alone, our model exhibited a mean classification accuracy of 66.40%, mean precision of 69.04%, mean recall of 66.25%, mean F1-Score of 65.72% and mean Kappa of 62.50%.

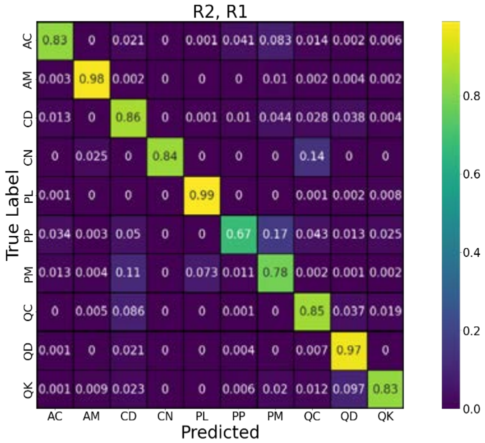
Table 3 (all figures in %): Accuracy Metrics Table, Satellite Imagery Only

Train-Test	*Accuracy*	*Precision*	*Recall*	*F1-Score*	*Kappa*
(R1, R2)	73.52	75.5	72.46	71.72	69.38
(R2, R1)	59.28	62.58	60.04	59.72	55.61



	AC	AM	CD	CN	PL	PP	PM	QC	QD	QK
AC	0.97	0	0.002	0	0.001	0.002	0.019	0	0.005	0
AM	0	1	0	0	0	0	0	0.001	0.001	0
CD	0.019	0	0.96	0	0	0.001	0.016	0	0	0.001
CN	0	0	0	1	0	0	0	0	0	0
PL	0	0.003	0	0	0.99	0.001	0.001	0	0.005	0.003
PP	0.005	0	0.001	0	0.001	0.75	0.24	0	0	0
PM	0.005	0	0	0	0	0	0.99	0	0	0
QC	0	0.005	0.004	0	0	0.001	0.022	0.95	0.022	0
QD	0.002	0.002	0.017	0	0	0.001	0	0.011	0.97	0.001
QK	0.002	0	0.036	0	0	0	0.004	0.014	0	0.94

Figure 2. Confusion Matrix: Classification using satellite imagery + OSA. Training-Testing R1/R2



	AC	AM	CD	CN	PL	PP	PM	QC	QD	QK
AC	0.83	0	0.021	0	0.001	0.041	0.083	0.014	0.002	0.006
AM	0.003	0.98	0.002	0	0	0	0.01	0.002	0.004	0.002
CD	0.013	0	0.86	0	0.001	0.01	0.044	0.028	0.038	0.004
CN	0	0.025	0	0.84	0	0	0	0.14	0	0
PL	0.001	0	0	0	0.99	0	0	0.001	0.002	0.008
PP	0.034	0.003	0.05	0	0	0.67	0.17	0.043	0.013	0.025
PM	0.013	0.004	0.11	0	0.073	0.011	0.78	0.002	0.001	0.002
QC	0	0.005	0.086	0	0	0.001	0	0.85	0.037	0.019
QD	0.001	0	0.021	0	0	0.004	0	0.007	0.97	0
QK	0.001	0.009	0.023	0	0	0.006	0.02	0.012	0.097	0.83

Figure 3. Confusion Matrix: Classification using satellite imagery + OSA. Training-Testing R2/R1

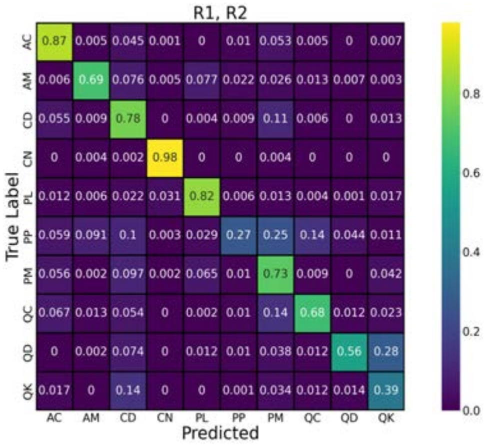


Figure 4. Confusion Matrix: Classification using satellite imagery only. Training-Testing R1/R2

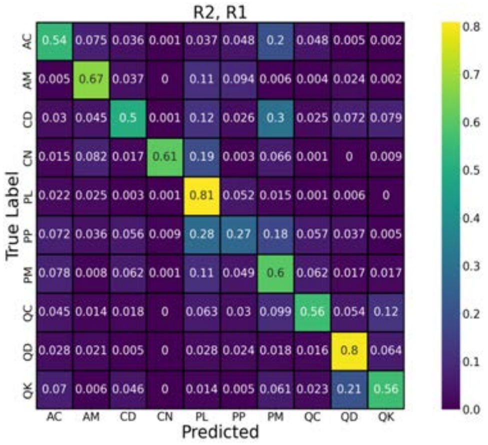


Figure 5. Confusion Matrix: Classification using satellite imagery only. Training-Testing R2/R1

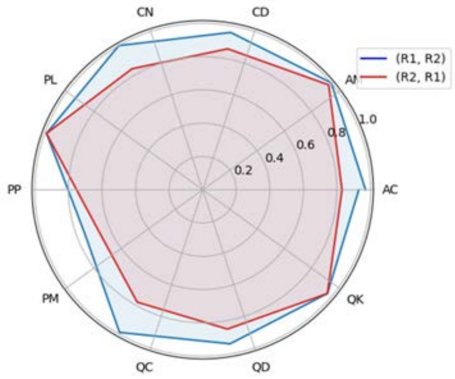


Figure 6. Spider chart comparison of satellite imagery + OSA tree species-specific classification accuracy for both Train-Test Pairs. The spider chart compares RF's overall performance for each Train-Test pair. The distance of an algorithm's polygon's edge to the end of the spoke reflects the accuracy the algorithm demonstrated for that specific Train-Test pair.

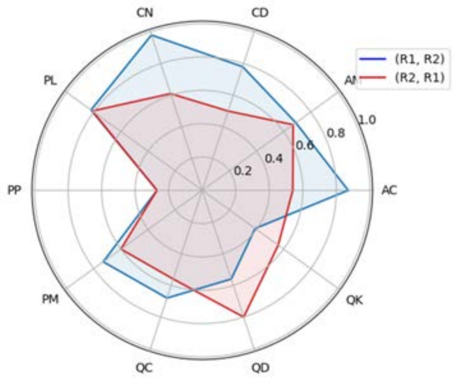


Figure 7. Spider chart comparison of satellite imagery only tree species-specific classification accuracy for both Train-Test Pairs.

4. Discussion

The addition of object-specific attributes to satellite imagery significantly improved classification accuracy across regions. Tables 2 & 3 show a summary of the performance metrics of the RF classification for cross-region Training and Testing split (Train-R1:Test-R2, Train-R2:Test-R1), under both scenarios, using satellite imagery alone and satellite imagery augmented with OSA. When adding OSA, the average classification accuracy increased from 66.4% to 90.2%. In addition, the Kappa values went from indicating moderate to strong agreement between the classification results and reference data by increasing from 62.50% to 89.07%. When comparing the confusion matrices in Figures 2 & 3 to the confusion matrices in Figures 4 & 5, the percentage of correct classifications in Figures 2 & 3 is higher across all tree species than those in Figures 4 & 5. The comparison between Figures 6 & 7 reflects the same observation, as the distance between the labels and edges of the R1, R2 & R2, R1 polygons in Figure 6's spider chart is significantly less than that of Figures 7's spider chart, indicating higher tree species-specific classification accuracy for satellite imagery + OSA. Overall, introducing OSA in our models improved scalability by increasing model accuracy when training and validating on separate regions. The difference in accuracy metrics between train-test pairs is possibly due to variations in the geographical features or environmental dynamics between regions. In addition, the disparity in total instances, displayed in

Table 1, could potentially explain the difference in accuracy metrics, as Region 1 having more instances than Region 2 allows the model to possess more knowledge when training on Region 1 and validating on Region 2 than it does when training on Region 2 and validating on Region 1, thus yielding a higher classification accuracy for the R1, R2 train-test pair.

There could be many underlying reasons for reduced classification accuracy when using satellite imagery values only. This includes a minimal distinction between RGB strength values pixels between different tree species, a complex forest structure of the study region, and a top-viewed pixel-based classification approach for tree species based on a large plot of land. However, adding OSA provides better distinction for RF algorithms to iterate, resulting in better classification accuracy.

When processing the data for analysis, we ran into numerous memory issues because of the size and scale of our datasets. We initially tried to encompass a significantly larger cutout of the Greater Lake Tahoe/El Dorado National Forest region to ensure ample geographic diversity but continuously ran into errors because of our system's limitations on memory. To bypass this issue, we experimented with incremental learning and k-fold cross-validation as possible solutions, but our system continued to run into memory issues. Especially for studies concerning large datasets, incremental learning allows 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. A stronger system could potentially process a larger study area with higher OSA specificity, allowing models to encompass larger regions with more geographical diversity without sacrificing performance. Further research that implements incremental learning, k-fold cross-validations, and a stronger processing system could potentially help construct more sophisticated models that are more accurate and encompass larger regions.

5. Conclusion

Mapping tree species provide an effective way to manage forest inventories and resources. While high classification accuracy for tree species is possible for small regions using satellite imagery, this research concludes that scaling of RF ML algorithm across a wider region is possible with high classification accuracy by including OSA such as crown shape, tree diameter, and tree height to satellite imagery. Since this research focused on the Greater Lake Tahoe region, additional investigations should explore the applicability of these findings in other regions and introduce incremental or k-fold cross-validation approaches to further improve model performance and scalability.

Acknowledgments

I would like to thank my mentor, Mr. Jeff Wen (Stanford School of Earth, Energy & Environmental Sciences), for his guidance throughout this project. I would also like to thank Dr. Jeff Liu (MIT Lincoln Laboratory) for helping me build foundational knowledge for this research and Mr. Ryan Malhoski (GIS Manager, City of South Lake Tahoe) for helping identify potential datasets.

References

Ballanti, L., et al. (2016). Tree Species Classification Using Hyperspectral Imagery: A Comparison of Two Classifiers. *Remote Sensing* 8(6) 445
https://doi.org/10.3390/rs8060445

Breiman, L. (2001). *Machine Learning* 45(1), 5–32.
https://doi.org/10.1023/a:1010933404324

Clark, M. L., & Roberts, D. A. (2012). Species-Level Differences in Hyperspectral Metrics among Tropical Rainforest Trees as Determined by a Tree-Based Classifier. *Remote Sensing* 4(6), 1820–1855.
https://doi.org/10.3390/rs4061820

Ghimire, B., Rogan, J., & Miller, J. (2010). Contextual land-cover classification: incorporating spatial dependence in land-cover classification models using random forests and the Getis statistic. *Remote Sensing Letters* 1(1), 45–54. https://doi.org/10.1080/01431160903252327

Hasanin, T., & Khoshgoftaar, T. (2018). The Effects of Random Undersampling with Simulated Class Imbalance for Big Data. *IEEE International Conference on Information Reuse and Integration (IRI).* https://doi.org/10.1109/iri.2018.00018

Immitzer, M., Atzberger, C., & Koukal, T. (2012). Tree Species Classification with Random Forest Using Very High Spatial Resolution 8-Band WorldView-2 Satellite Data. *Remote Sensing* 4(9), 2661–2693. https://doi.org/10.3390/rs4092661

Talukdar, S., et al. (2020). Land-Use Land-Cover Classification by Machine Learning Classifiers for Satellite Observations—A Review. *Remote Sensing*, 12(7), 1135. https://doi.org/10.3390/rs12071135

Wang, Y., et al. (2021). Classification of Street Tree Species Using UAV Tilt Photogrammetry. *Remote Sensing* 13(2), 216. https://doi.org/10.3390/rs13020216

Yacouby, R., & Axman, D. (2020). Probabilistic Extension of Precision, Recall, and F1 Score for More Thorough Evaluation of Classification Models. *Proceedings of the First Workshop on Evaluation and Comparison of NLP Systems, Association for Computational Linguistics*, 79-91 doi:10.18653/v1/2020.eval4nlp-1.