A peer-reviewed version of this preprint was published in PeerJ on 28 February 2019.

View the peer-reviewed version (peerj.com/articles/6101), which is the preferred citable publication unless you specifically need to cite this preprint.

Sumsion GR, Bradshaw MS, Hill KT, Pinto LDG, Piccolo SR. 2019. Remote sensing tree classification with a multilayer perceptron. PeerJ 7:e6101 https://doi.org/10.7717/peerj.6101

Remote sensing tree classification with a multilayer perceptron

G Rex Sumsion [1] , **Michael S Bradshaw** [1] , **Kimball T Hill** [1] , **Lucas D G Pinto** [1] , **Stephen R Piccolo** [Corresp. 1]

[1] Department of Biology, Brigham Young University, Provo, UT, United States

Corresponding Author: Stephen R Piccolo
Email address: stephen_piccolo@byu.edu

To accelerate scientific progress on remote tree classification—as well as biodiversity and ecology sampling—The National Institute of Science and Technology created a community-based competition where scientists were invited to contribute informatics methods for classifying tree species and genus using crown-level images of trees. We predicted tree species and genus at the pixel level using hyperspectral and LIDAR observations. We compared three algorithms that have been implemented extensively across a broad range of research applications: support vector machines, random forests, and multilayer perceptron. At the pixel level, the multilayer perceptron algorithm predicted species or genus with high accuracy (92.7 and 95.9%, respectively) on the training data and performed better than the other algorithms (85.8-93.5%). This indicates promise for the use of the MLP algorithm for tree-species classification and coincides with a growing body of research in which neural network-based algorithms outperform other types of classification algorithms for machine vision. To aggregate patterns across the images, we used an ensemble approach that averages the pixel-level outputs of the MLP algorithm to predict species at the crown level. The accuracy of these predictions on the test set was 68.8% for species.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26971v1 | CC BY 4.0 Open Access | rec: 1 Jun 2018, publ: 1 Jun 2018

Remote sensing tree classification with a multilayer perceptron

G. Rex Sumsion[1], Michael S. Bradshaw[1], Kimball T. Hill[1], Lucas D. G. Pinto[1], and Stephen R. Piccolo[1,*]

Brigham Young University, Department of Biology, Provo, UT 84602 (USA)

* Please address correspondence to Stephen R. Piccolo, stephen_piccolo@byu.edu.

Abstract

To accelerate scientific progress on remote tree classification—as well as biodiversity and ecology sampling—The National Institute of Science and Technology created a community-based competition where scientists were invited to contribute informatics methods for classifying tree species and genus using crown-level images of trees. We predicted tree species and genus at the pixel level using hyperspectral and LIDAR observations. We compared three algorithms that have been implemented extensively across a broad range of research applications: support vector machines, random forests, and multilayer perceptron. At the pixel level, the multilayer perceptron algorithm predicted species or genus with high accuracy (92.7 and 95.9%, respectively) on the training data and performed better than the other algorithms (85.8-93.5%). This indicates promise for the use of the MLP algorithm for tree-species classification and coincides with a growing body of research in which neural network-based algorithms outperform other types of classification algorithms for machine vision. To aggregate patterns across the images, we used an ensemble approach that averages the pixel-level outputs of the MLP algorithm to predict species at the crown level. The accuracy of these predictions on the test set was 68.8% for species.

Introduction

As the earth's population grows, understanding global ecosystems is becoming more critical to preserving biodiversity and answering ecological questions. Methods for answering such questions have been advanced with improvements in technology. For example, many researchers have begun to explore the effectiveness of using remote-sensing techniques to improve sampling for biodiversity and ecology research.

Many disciplines have accelerated scientific progress through community-based competitions (Marbach et al., 2010; Prill et al., 2011; Wan & Pal, 2014; Seyednasrollah et al., 2017). Such competitions help to overcome limitations of individual scientific studies that evaluate only a single approach. Many such papers fail to adequately compare their methods against alternative approaches, compare against only a single alternative approach, fail to use common datasets for comparison, or use inconsistent evaluation metrics (Marconi et al., 2018). Therefore, to advance the development of quantitative methods for sampling biodiversity, the National Institute of Standards and Technology (NIST) developed a competition series that allows researchers to evaluate a common high-quality, remote-sensing dataset (provided by the National Ecological Observatory Network) using different quantitative methods. By aggregating and evaluating evidence across these contributions, they hope that more accurate methods will be developed to improve future analyses.

We focused on task III in this competition: classifying the species or genus of tree crowns from provided hyperspectral and LIDAR (light radar) pixel data. In addition to classifying at the tree crown level, our study compares several machine-learning algorithms' abilities to classify at the pixel level as done in other studies (Clark, Roberts & Clark, 2005; Castro-Esau et al., 2006; Carlson et al., 2007; Dalponte et al., 2009).

Methods

Data Preprocessing

The data from NEON included the following data products: 1) Woody plant vegetation structure (NEON.DP1.10098); 2) Spectrometer orthorectified surface directional reflectance - flightline (NEON.DP1.30008); 3) Ecosystem structure (NEON.DP3.30015); and 4) High-resolution orthorectified camera imagery (NEON.DP1.30010). We used both the LIDAR and hyperspectral data to train the implemented algorithms. Since the hyperspectral data column band_426 contained values that were missing for the majority of the data rows, we determined that it would be more simple and consistent to exclude the column entirely from the training data than using an imputation method to deal with missing values. We used all other features in training the implemented algorithms. These features include data associated with the crowns labeled as "other". For additional details about the image data that were provided as part of the competition, we refer the reader to Marconi, et al.

Algorithm Training

For diagnostic purposes during training, we used the cross_val_score function in scikit-learn to evaluate classification accuracy via k-fold cross validation (Pedregosa et al., 2011). In performing these evaluations, we used the default of 3 folds.

Classification Algorithms

Several prior tree species classification studies used random forests (RF) and support vector machines (SVM) classifiers (Ghosh et al., 2014; Baldeck et al., 2015; Ferreira et al., 2016). Other studies have shown that the multi-layer perceptron (MLP)—a less used algorithm in this field—may be beneficial for tree classification problems that use both LIDAR and hyperspectral imaging. In this study, we implemented the MLP algorithm and compared its performance against that of the SVM and RF algorithms.

The SVM algorithm creates a hyperplane (a barrier) between observations from two classes, attempting to separate the classes by a maximal margin. This margin is adjusted throughout training until classification error is minimized (Scholkopf et al.). This algorithm is used frequently when there are complicated patterns in data and has been applied frequently to RNA expression data, for example (Statnikov et al., 2005; Statnikov & Aliferis, 2010; Feig et al., 2012; Attur et al., 2015).

RF classifiers operate by creating a number of decision trees. Each decision tree is constructed by selecting k features randomly; among the *k* features, a node is created to divide samples based on those features, thus forming a branch in the tree. Node creation is repeated iteratively based on the remaining features. Training occurs via bagging with replacement—that is, random samples are selected from a training set, and trees are fit to those samples. The algorithm then classifies test samples on each tree in the forest and outputs a mode classification. This approach theoretically prevents overfitting because the forest is based on a distribution of decision trees consisting of diverse sets of features (Breiman, 2001).

Neural networks use several layers of nodes that fall into one of three categories: input, hidden, or output (Kuncheva; Haindl, Kittler & Roli, 2007; Du et al., 2012; Woźniak & Graña, 2014). A node on a given layer will receive a weighted average of the outputs of the previous layer and given its specific weight will contribute to a new weighted average which is propagated down the network until a final output is reached. During training, each node's weights in the network are optimized with the goal of minimizing error. The assumption is that a network of nodes can gain insight in supervised learning that a single node cannot.

Algorithm Implementations

Each of the classification algorithms provide many hyperparameters that must be selected. Depending on how these hyperparameters are set, the algorithm's performance can change dramatically. These settings are often optimized by trial and error—not by structured rules. Some studies have searched to find better ways to optimize certain hyperparameters more effectively, such as the SVM (Chapelle et al., 2002), but it is difficult to know which values will be most effective. For this study, to simplify implementation—and to equally compare the previously mentioned algorithms—we employed these algorithms with default parameters (except as noted below) using the scikit-learn Python library (Pedregosa et al., 2011). For the RF classifier, we used bootstrapping with replacement and no maximum depth for the decision trees. For the SVM algorithm, the default settings included a C value of 1.0 and the radial basis function kernel. The MLP algorithm was implemented using three inner layers. Each inner layer was structured to have exactly 40 nodes. Important default settings to note are the maximum number of training epochs capped at 200 to avoid overfitting the data and a learning rate of 0.001.

From the Pixel Level to the Crown Level

First, we used the hyperspectral and LIDAR data as inputs to the classification algorithms to predict species and genus for each pixel; next we applied a custom ensemble-averaging method to the pixel-level predictions and used these averaged predictions to make tree crown-level predictions (Figure 1). This ensemble method (Dietterich, 2000) averages the probabilistic predictions across all pixels in a given tree crown. This allows for each input pixel to have a modest impact on the final predictions for the respective tree crown. An assumption of this approach is that although some individual pixels may be predicted incorrectly, a considerable proportion of individual pixels would be predicted correctly, and aggregating evidence across all pixels would lead to correct crown-level predictions.

The code that we used in this study is available in a GitHub repository at https://github.com/byubrg/NIST-Competition-Fall-2017. For details of other methods that were used in the competition, please see the overview paper by Marconi et al..

Results

First we evaluated the three classification algorithms at the pixel level on the training data. Because many studies in the past only evaluated their algorithms at the pixel level, this study aimed to assess the MLP's ability to make accurate predictions in this context in relation to the SVM and RF algorithms, which have been used more commonly. Overall, the algorithms predicted genus and species with high accuracy (85.8-95.9%) and attained higher accuracy for genus than for species (Figure 2). One reason for these differing levels of accuracy may be that species within a genus are often quite similar biologically and therefore can also be expected to have similar traits and appear similar in remote sensing imagery. Also, for genus-based classification, the algorithms only needed to differentiate among 5 class labels whereas they needed to distinguish among 9 class labels for species classification. In this comparison, the MLP algorithm's performance dropped least from genus to species (95.9% vs. 92.7%). The SVM and RF algorithms attained classification accuracies of 91.1% and 93.5%, respectively, for genus prediction and 85.8% and 86.8% for species predictions (Figure 2). The differences in performance between MLP and the other algorithms are substantial enough to suggest that the multilayer perceptron should be explored further for tree classification through remote sensing—perhaps especially when using a relatively large number of labels.

Our final model used an ensemble-based approach to average pixel-level predictions for the MLP algorithm only. When applied to the competition's test data, our solution obtained an accuracy of 68.8% for crown-level classification (pixel-level predictions were not assessed as part of the final evaluation). Although our solution exceeded the baseline expectation of 66.7% accuracy, our approach failed to generalize well. To better understand these results and how our results compare to other participants' in the competition, please see the description by Marconi, et al..

Discussion

As early as 1998, computer vision techniques have been used to answer biological questions. In some of these studies, hyperspectral imagery has been used to differentiate between similarly colored items, such as chlorophyll a and chlorophyll b (Blackburn, 1998). However, it wasn’t until 2005 that computer vision was explored specifically for tree species classification. One of the first such studies explored tree-species classification and evaluated accuracies of pixel-level predictions on the leaf and crown scales (Clark, Roberts & Clark, 2005). In 2006 and 2007, researchers then analyzed the effectiveness of using various wavelengths in an image. These studies found a correlation between higher prediction accuracy and the use of more wavelengths (Castro-Esau et al., 2006; Carlson et al., 2007). It wasn’t until 2009 that a study’s results specifically supported the hypothesis that hyperspectral images provide the highest accuracy (Dalponte et al., 2009). Since that time, many studies have analyzed hyperspectral images to find even more effective algorithms and forms of data representation for remote tree-species classification. One of the most interesting studies combined hyperspectral and LIDAR data to obtain higher accuracies with their algorithms (Alonzo, Bookhagen & Roberts, 2014). This approach is used in our study.

Conclusion

We selected the MLP algorithm for our final predictive model as a result of its growing popularity in computer vision and its relatively superior performance on our training data. The relatively high accuracy of neural-network based algorithms, in general, has led to their use in many recent computer vision studies (Simonyan & Zisserman, 2014; Rawat & Wang, 2017). Its accuracy frequently outperforms other methods (Ciresan, Meier & Schmidhuber, 2012). We found that the MLP algorithm is an effective method for tree classification using hyperspectral and LIDAR imagery. However, when attempting to aggregate those predictions to crown-level observations, the accuracy dropped considerably, even though the prediction accuracy still exceeded random-chance expectations. This drop could be due to oversimplification of our ensemble method in that it did not account for spatial relationships among the pixels and did not correct for outlier effects. Alternative approaches that may have led to better results include 1) using a convolutional neural network to aggregate the pixel-level predictions and account for spatial relationships (Krizhevsky, Sutskever & Hinton, 2012), 2) use an ensemble method that is more robust to outliers (Kuncheva; Haindl, Kittler & Roli, 2007; Du et al., 2012; Woźniak &

Graña, 2014), and/or 3) include all three classification algorithms in our ensemble, thus potentially reducing the effect of outliers and incorrect pixel-level predictions. Alternatively, it may be more effective to make crown-level predictions directly using hyperspectral and LIDAR values rather than using a hierarchical approach.

Another potential limitation of our approach is that we used default hyperparameter values for the classification algorithms. Due to the high potential shown by the MLP algorithms, in future studies it would be valuable to optimize hyperparameters and potentially to use a deep learning architecture to fine tune the algorithm's performance as much as possible.

Bibliography

Alonzo M., Bookhagen B., Roberts DA. 2014. Urban tree species mapping using hyperspectral and lidar data fusion. Remote Sensing of Environment148:70–83. DOI: 10.1016/j.rse.2014.03.018.

Attur M., Krasnokutsky S., Statnikov A., Samuels J., Li Z., Friese O., Hellio Le Graverand-Gastineau M-P., Rybak L., Kraus VB., Jordan JM., Aliferis CF., Abramson SB. 2015. Low-Grade Inflammation in Symptomatic Knee Osteoarthritis: Prognostic Value of Inflammatory Plasma Lipids and Peripheral Blood Leukocyte Biomarkers. Arthritis & Rheumatology67:2905–2915. DOI: 10.1002/art.39279.

Baldeck CA., Asner GP., Martin RE., Anderson CB., Knapp DE., Kellner JR., Wright SJ. 2015. Operational Tree Species Mapping in a Diverse Tropical Forest with Airborne Imaging Spectroscopy. PLOS ONE 10:e0118403. DOI: 10.1371/journal.pone.0118403.

Blackburn GA. 1998. Quantifying Chlorophylls and Caroteniods at Leaf and Canopy Scales. Remote Sensing of Environment66:273–285. DOI: 10.1016/S0034-4257(98)00059-5.

Breiman L. 2001. Random Forests. Machine Learning 45:5-32. DOI: 10.1023/A:1010933404324.

Carlson KM., Asner GP., Hughes RF., Ostertag R., Martin RE. 2007. Hyperspectral Remote Sensing of Canopy Biodiversity in Hawaiian Lowland Rainforests. Ecosystems10:536–549. DOI: 10.1007/s10021-007-9041-z.

Castro-Esau KL., Sanchez-Azofeifa GA., Rivard B., Wright SJ., Quesada M. 2006. Variability in leaf optical properties of Mesoamerican trees and the potential for species classification. American Journal of Botany93:517–530. DOI: 10.3732/ajb.93.4.517.

Chapelle O., Vapnik V., Bousquet O., Mukherjee S. 2002. Choosing Multiple Parameters for Support Vector Machines. Machine Learning 46:131-159. DOI: 10.1023/A:1012450327387.

Ciresan D., Meier U., Schmidhuber J. 2012. Multi-column deep neural networks for image classification. In: *2012 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 3642–3649. DOI: 10.1109/CVPR.2012.6248110.

Clark ML., Roberts DA., Clark DB. 2005. Hyperspectral discrimination of tropical rain forest tree species at leaf to crown scales. Remote Sensing of Environment96:375–398. DOI: 10.1016/j.rse.2005.03.009.

Dalponte M., Bruzzone L., Vescovo L., Gianelle D. 2009. The role of spectral resolution and classifier complexity in the analysis of hyperspectral images of forest areas. Remote Sensing of Environment113:2345–2355. DOI: 10.1016/j.rse.2009.06.013.

Dietterich TG. 2000. Ensemble Methods in Machine Learning. In: Springer, Berlin, Heidelberg, 1–15. DOI: 10.1007/3-540-45014-9_1.

Ferreira MP., Zortea M., Zanotta DC., Shimabukuro YE., de Souza Filho CR. 2016. Mapping tree species in tropical seasonal semi-deciduous forests with hyperspectral and multispectral data. Remote Sensing of Environment 179:66-78. DOI: 10.1016/j.rse.2016.03.021.

Du P., Xia J., Zhang W., Tan K., Liu Y., Liu S. 2012. Multiple classifier system for remote sensing image classification: a review. Sensors (Basel, Switzerland)12:4764–92. DOI: 10.3390/s120404764.

Feig JE., Vengrenyuk Y., Reiser V., Wu C., Statnikov A., Aliferis CF., Garabedian MJ., Fisher EA., Puig O. 2012. Regression of Atherosclerosis Is Characterized by Broad Changes in the Plaque Macrophage Transcriptome. PLoS ONE7:e39790. DOI: 10.1371/journal.pone.0039790.

Ghosh A., Fassnacht FE., Joshi PK., Kochb B. 2014. A framework for mapping tree species combining hyperspectral and LiDAR data: Role of selected classifiers and sensor across three spatial scales. International Journal of Applied Earth Observation and Geoinformation 26:49-63. DOI: 10.1016/j.jag.2013.05.017.

Grossberg S. 1988. Nonlinear neural networks: Principles, mechanisms, and architectures. Neural Networks1:17–61. DOI: 10.1016/0893-6080(88)90021-4.

Haindl M., Kittler J., Roli F. 2007. Multiple classifier systems : 7th international workshop, MCS 2007, Prague, Czech Republic, May 23-25, 2007 : proceedings. Springer.

KLEENE., C. S. 1956. Representations of events in nerve nets and finite automata. Automata Studies [Annals of Math. Studies 34].

Krizhevsky A., Sutskever I., Hinton GE. 2012. ImageNet Classification with Deep Convolutional Neural Networks. :1097–1105.

Kuncheva LI. Multiple Classifier Systems. In: Combining Pattern Classifiers. Hoboken, NJ, USA: John Wiley & Sons, Inc., 101–110. DOI: 10.1002/0471660264.ch3.

Marconi S., Graves SJ., Gong D., Nia MS., Bras M Le., Dorr BJ., Fontana P., Gearhart J., Greenberg C., Harris DJ., Kumar SA., Nishant A., Prarabdh J., Rege SU., Bohlman SA., White EP., Wang DZ. 2018. A data science challenge for converting airborne remote sensing data into ecological information. DOI: 10.7287/peerj.preprints.26966v1.

Marbach D., Prill RJ., Schaffter T., Mattiussi C., Floreano D., Stolovitzky G. 2010. Revealing strengths and weaknesses of methods for gene network inference. Proceedings of the National Academy of Sciences of the United States of America107:6286–91. DOI: 10.1073/pnas.0913357107.

McCulloch WS., Pitts W. 1943. A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*5:115–133. DOI: 10.1007/BF02478259.

Pedregosa F., Varoquaux G., Gramfort A., Michel V., Thirion B., Grisel O., Blondel M., Prettenhofer P., Weiss R., Dubourg V., Vanderplas J., Passos A., Cournapeau D., Brucher M., Perrot M., Duchesnay É. 2011. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research 12:2825–2830.

Prill RJ., Saez-Rodriguez J., Alexopoulos LG., Sorger PK., Stolovitzky G. 2011. Crowdsourcing network inference: the DREAM predictive signaling network challenge. Science signaling4:mr7. DOI: 10.1126/scisignal.2002212.

Rawat W., Wang Z. 2017. Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review. *Neural Computation* 29:2352–2449. DOI: 10.1162/neco_a_00990.

Scholkopf B., Simard P., Smola A., Vapnikt V. Prior Knowledge in Support Vector Kernels.

Seyednasrollah F., Koestler DC., Wang T., Piccolo SR., Vega R., Greiner R., Fuchs C., Gofer E., Kumar L., Wolfinger RD., Kanigel Winner K., Bare C., Neto EC., Yu T., Shen L., Abdallah K., Norman T., Stolovitzky G., Soule HR., Sweeney CJ., Ryan CJ., Scher HI., Sartor O., Elo LL., Zhou FL., Guinney J., Costello JC., Community PCDC. 2017. A DREAM Challenge to Build Prediction Models for Short-Term Discontinuation of Docetaxel in Metastatic Castration-Resistant Prostate Cancer. JCO Clinical Cancer Informatics:1–15. DOI: 10.1200/CCI.17.00018.

Simonyan K., Zisserman A. 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014).

Statnikov A., Aliferis CF. 2010. Analysis and Computational Dissection of Molecular Signature Multiplicity. PLoS Computational Biology6:e1000790. DOI: 10.1371/journal.pcbi.1000790.

Statnikov A., Aliferis CF., Tsamardinos I., Hardin D., Levy S. 2005. A comprehensive evaluation of multicategory classification methods for microarray gene expression cancer diagnosis. Bioinformatics21:631–643. DOI: 10.1093/bioinformatics/bti033.

Wan Q., Pal R. 2014. An Ensemble Based Top Performing Approach for NCI-DREAM Drug Sensitivity Prediction Challenge. PLoS ONE9:e101183. DOI: 10.1371/journal.pone.0101183.

Woźniak M., Graña M. 2014. A survey of multiple classifier systems as hybrid systems. Information Fusion16:3–17. DOI: 10.1016/J.INFFUS.2013.04.006.

Acknowledgements

The ECODSE competition was supported, in part, by a research grant from NIST IAD Data Science Research Program to D.Z. Wang, E.P. White, and S. Bohlman, by the Gordon and Betty Moore Foundation's Data-Driven Discovery Initiative through grant GBMF4563 to E.P. White, and by an NSF Dimension of Biodiversity program grant (DEB-1442280) to S. Bohlman. The National Ecological Observatory Network is a program sponsored by the National Science Foundation and operated under cooperative agreement by Battelle Memorial Institute. This material is based in part upon work supported by the National Science Foundation through the NEON Program. We also acknowledge the Brigham Young University Fulton Supercomputing Laboratory for computational resources that were used for this analysis.

Author Contributions

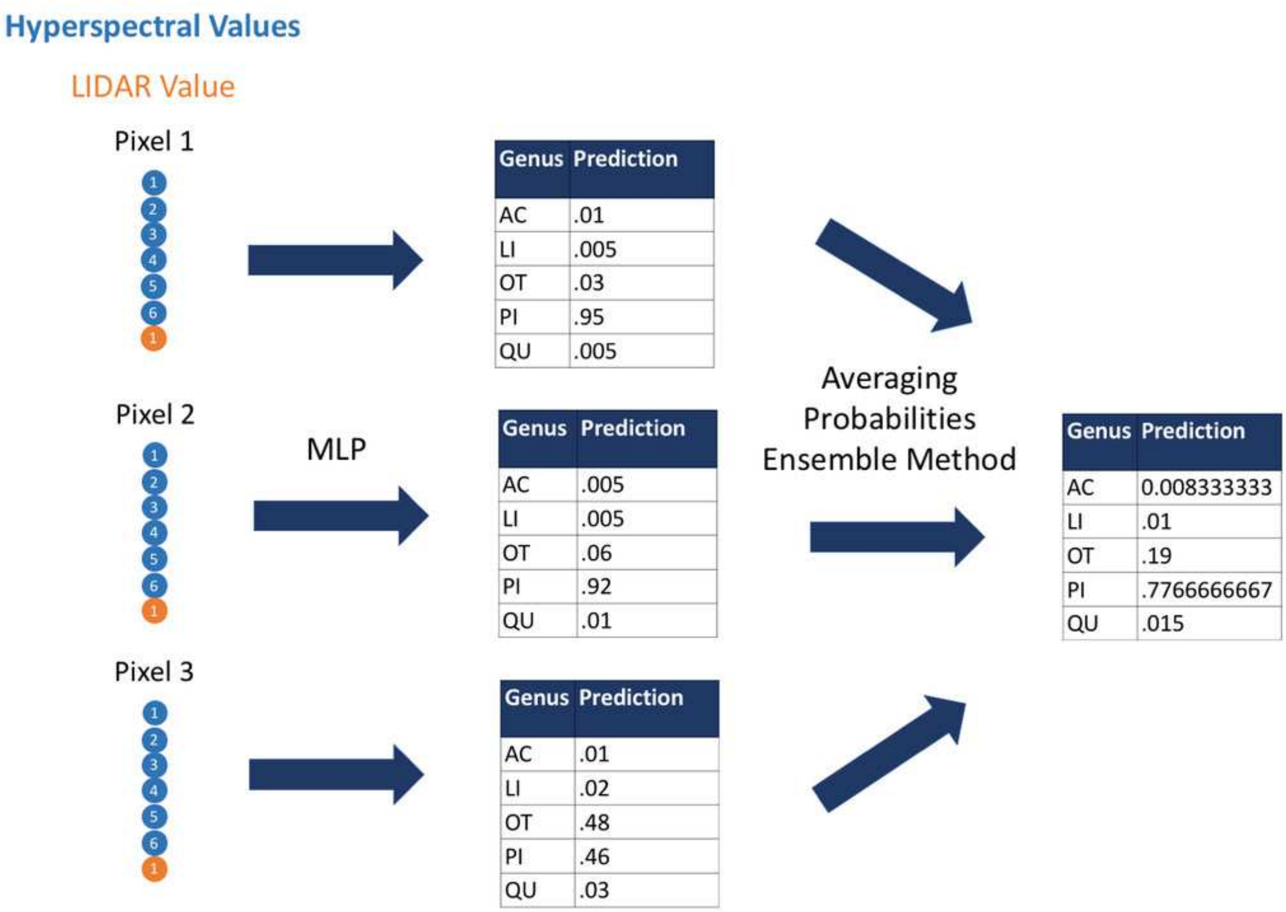
GRS, MB, KTH, and LP conceived the methodological approach. MB implemented the MLP algorithm and preprocessed the data. GRS implemented the SVM and RF algorithms and wrote the custom ensemble method. GRS created the figures and drafted the manuscript. LP, MB, KTH, and SRP helped in drafting and revising the manuscript. SRP helped in interpreting results.

All authors read and approved the final manuscript.

Figure 1

Visual representation of the ensemble method

We used the multilayer perceptron algorithm to derive predictions of species and genus based on hyperspectral and LIDAR values at the pixel level. We then aggregated these predictions to crown-level predictions using an ensemble approach that averaged the probabilistic, pixel-level predictions.

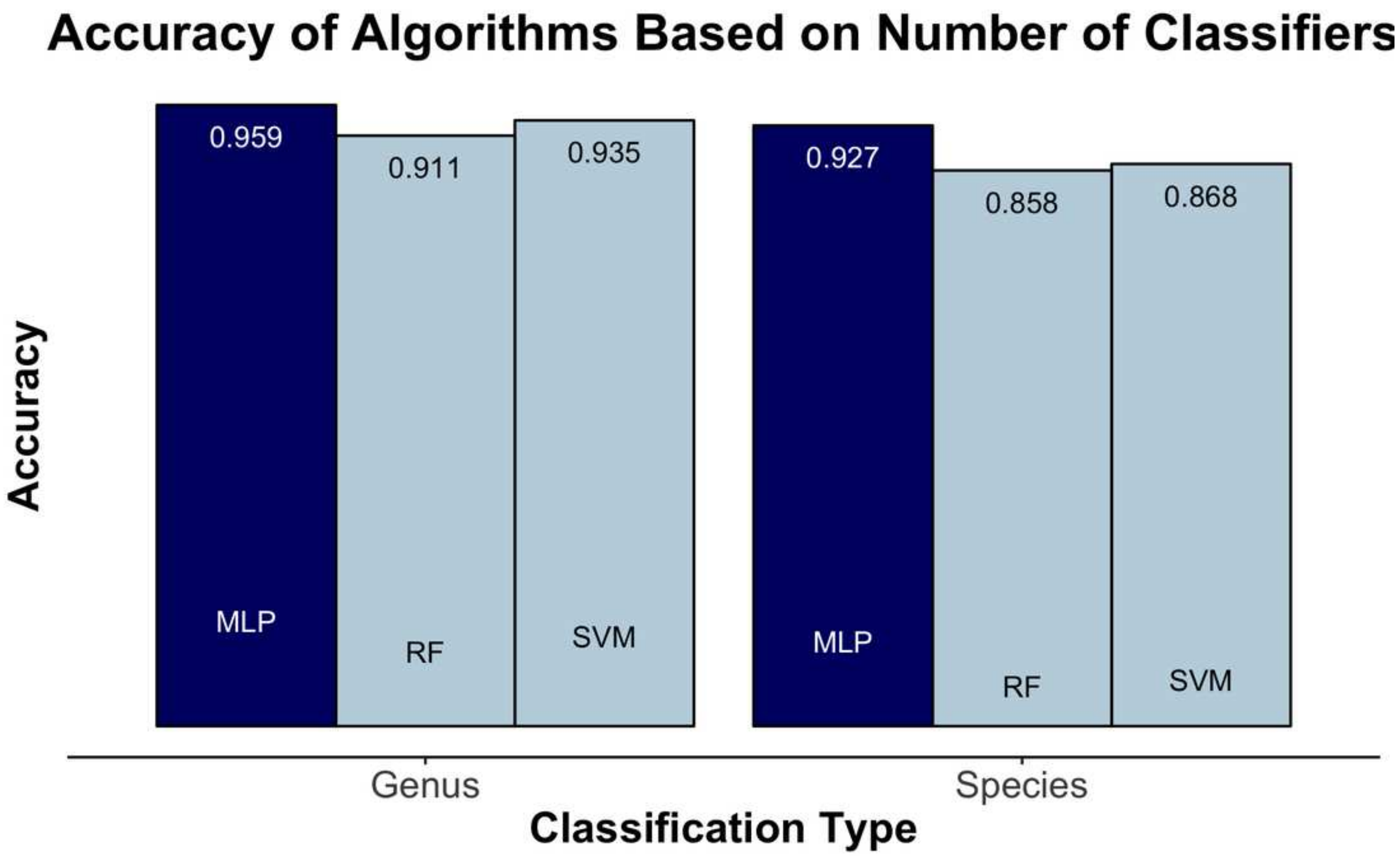


PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26971v1 | CC BY 4.0 Open Access | rec: 1 Jun 2018, publ: 1 Jun 2018

Figure 2

Bar plots illustrating classification accuracy for the classifiers on species (9 labels) and genus (5 labels)

MLP = multilayer perceptron. RF = random forests. SVM = support vector machines.



PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26971v1 | CC BY 4.0 Open Access | rec: 1 Jun 2018, publ: 1 Jun 2018