A Multimodal Attention-Based Model for Tree Species Classification Using LiDAR and Satellite Imagery

[1]Sablon Hadrien & [2]Rajen Bajgain

[1]Vegetation Asset Strategy & Analytics (VASA), [2]Electric Risk Management & Analytics (ERMA)
Pacific Gas and Electric (PG&E)

INTRODUCTION

- ❖Accurate mapping of tree species is crucial for wildfire mitigation, biodiversity conservation, and sustainable forest management under climate change. While advances in remote sensing and deep learning have improved species classification, scarcity of high-quality ground truth data, low-resolution sensors, and small study areas with limited species diversity hinder scalability and generalization.
- ❖While advances in remote sensing and deep learning have improved species classification, scarcity of high-quality ground truth data, low-resolution sensors, and small study areas with limited species diversity hinder scalability and generalization.
- ❖To address these limitations, we assembled a dataset of half a million data points from five distinct level-III ecoregions in California. Ground truth labels across more than 20 species were obtained from arborist-supported tree inventories.
- ❖We developed three deep learning models: a LiDAR-derived depth-view model (DVM) that exploits structural characteristics, a satellite-based Surface Reflectance model (SRM) that leverages spectral information, and a novel multimodal framework (MXAT) built with attention mechanisms to learn interdependencies between these complementary data modalities.

OBJECTIVES

This study presents a deep learning multimodal framework that integrates airborne LiDAR data and satellite Surface Reflectance (SR) data across five Level III ecoregions in California.

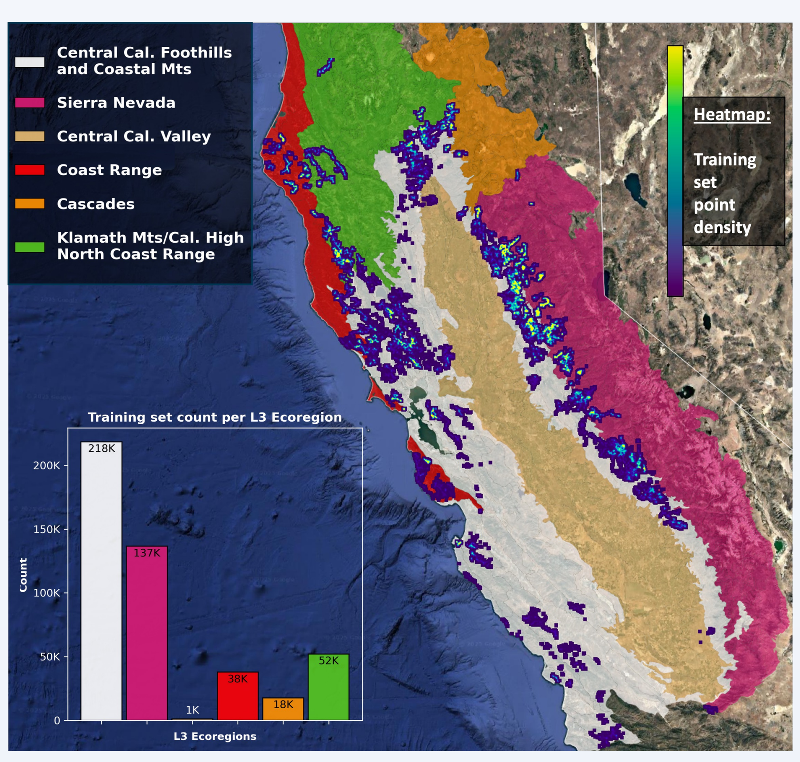

Fig. 1: Training set distribution across California visualized as a heatmap, with embedded histogram detailing the distribution per level III ecoregion

MATERIALS AND METHODS

DATASETS

- ❖PlanetScope SuperDove SR basemaps (4.7 m resolution, 8 spectral bands)
- ❖Airbone LiDAR data acquired by Sharper Shape (26,000 miles across five level III ecoregions)
- ❖Tree inventory data (450k training samples)

MODEL ARCHITECTURE

We developed three deep learning models for tree species classification:

(1) DVM, a Convolutional Neural Network (CNN)-based architecture for LiDAR-derived depth-views

DVM is a deep learning framework based on the architecture introduced by SimpleView an approach leveraging six orthogonal camera projections that has shown promising performance on limited tree point cloud samples.

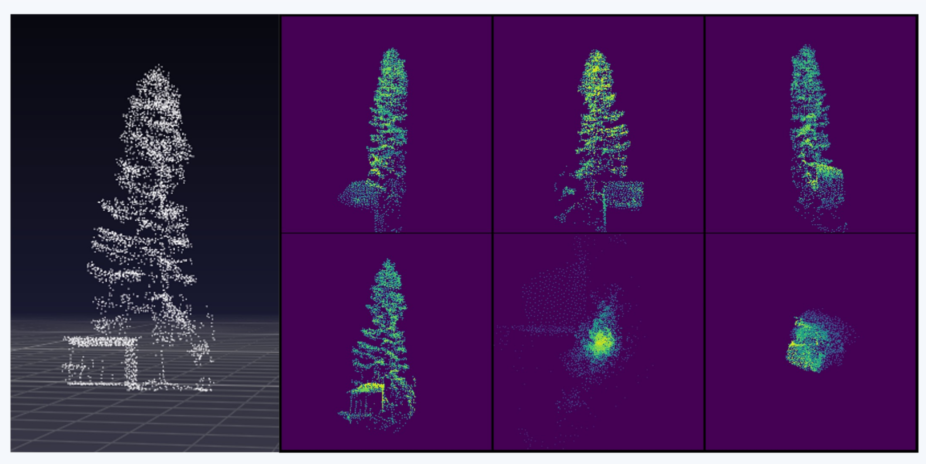

Fig. 2: LiDAR point cloud and projected depth-views (4r1t1b). From top-left to bottom-right: 1st radial view to 4th radial view (4r), followed by top-down view (t) and bottom-up view (b)

(2) SRM, a lightweight CNN for PlanetScope Surface Reflectance imagery

SRM is a shallow CNN comprising three convolutional layers, followed by batch normalization, a dense layer, and a classification head. The network processes 7×7-pixel patches extracted from Planet Lab's SR imagery, centered on LiDAR-derived tree-tops, with the eight spectral bands serving as input channels.

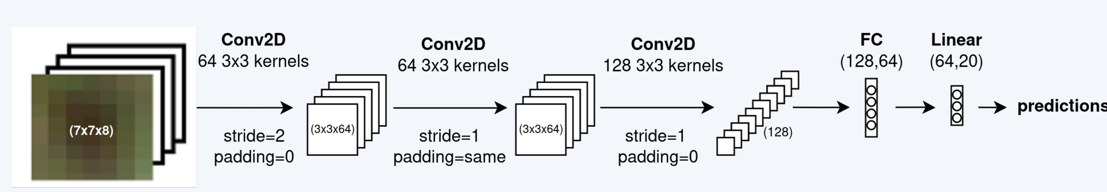

Fig. 3: Overview of the proposed Surface Reflectance Model (SRM) architecture. Each Conv2D and fully connected (FC) layer is followed by batch normalization and a ReLU activation function.

(3) MXAT, a multimodal model that integrates both modalities

To effectively integrate complementary information from LiDAR and satellite data, we developed a transformer-based multimodal fusion model. we introduced an attention-driven fusion strategy, and evaluated it against multiple fusion approaches, including simple concatenation, element-wise operations, and co-attention-based mechanisms

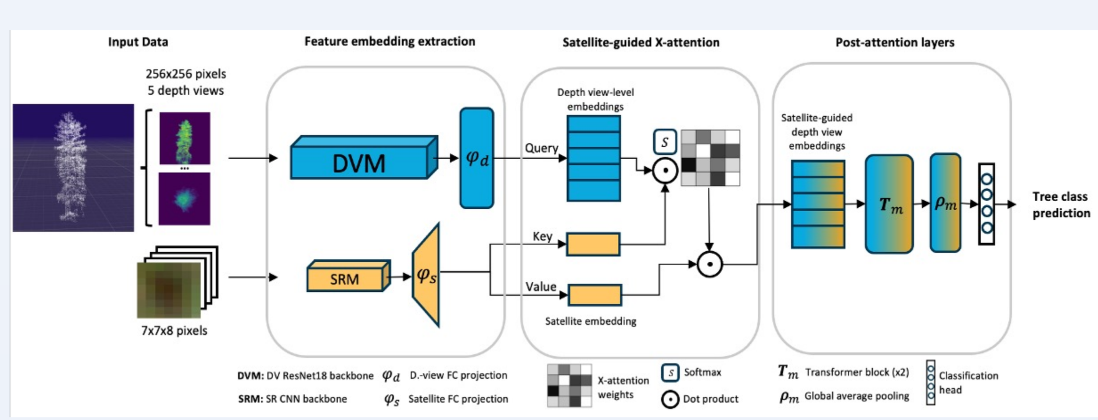

Fig. 4: Overview of the proposed Multimodal Cross-Attention model (MXAT) architecture

RESULTS

- ❖All models were evaluated on a consistent testing set, focusing on classification accuracy for over16,000 data points spanning 20 custom tree taxa
- ❖The impact of sample size on classification performance deserves attention, as underrepresented classes tend to exhibit lower recall
- ❖The three models exhibit varying levels of sensitivity across species(confusion matrix)
- ❖DVM demonstrates robust overall performance, with five species surpassing 80\% sensitivity and ten exceeding 60%
- ❖In contrast, SRM, trained exclusively on moderate-resolution satellite imagery, achieves limited overall accuracy but excels in detecting specific species such as Monterey pines (93%), true redwoods (78%), eucalyptus (73\%), and, to a lesser extent, tan oaks (60%).

Table 1: Sample size per custom taxon and sensitivity % on testing set for DVM, SRM and MXAT. Best performance across models is highlighted in gray .

Tree custom taxa	Sample size (% total)	DVM (lidar)	SRM (satellite)	MXAT (multimodal)
Monterey pines	1.0	88	93	**97**
Ponderosa pines	16.3	**80**	10	**80**
Gray pines	6.0	87	52	**89**
Eucalyptus trees	1.1	81	72	**89**
True redwoods	3.8	88	78	**93**
Other oaks	14.7	61	24	**65**
Live oaks	13.0	53	20	**57**
Douglas firs	16.4	**45**	13	34
Incense cedars	5.8	**72**	13	68
Liquidambar trees	0.2	**63**	37	**63**
Sugar pines	0.8	62	42	**67**
Black oaks	6.4	58	39	**60**
Umbellularia trees	2.1	46	35	**59**
Walnuts	0.4	43	18	**50**
Tan oaks	2.7	46	60	**61**
Fir trees	1.0	37	22	**44**
Valley oaks	4.0	60	52	**63**
Poplars	0.2	30	9	**38**
Arbutus trees	2.0	22	12	**30**
Other	2.2	28	7	**39**
Mean sensitivity		**57.5**	**35.4**	**62.2**

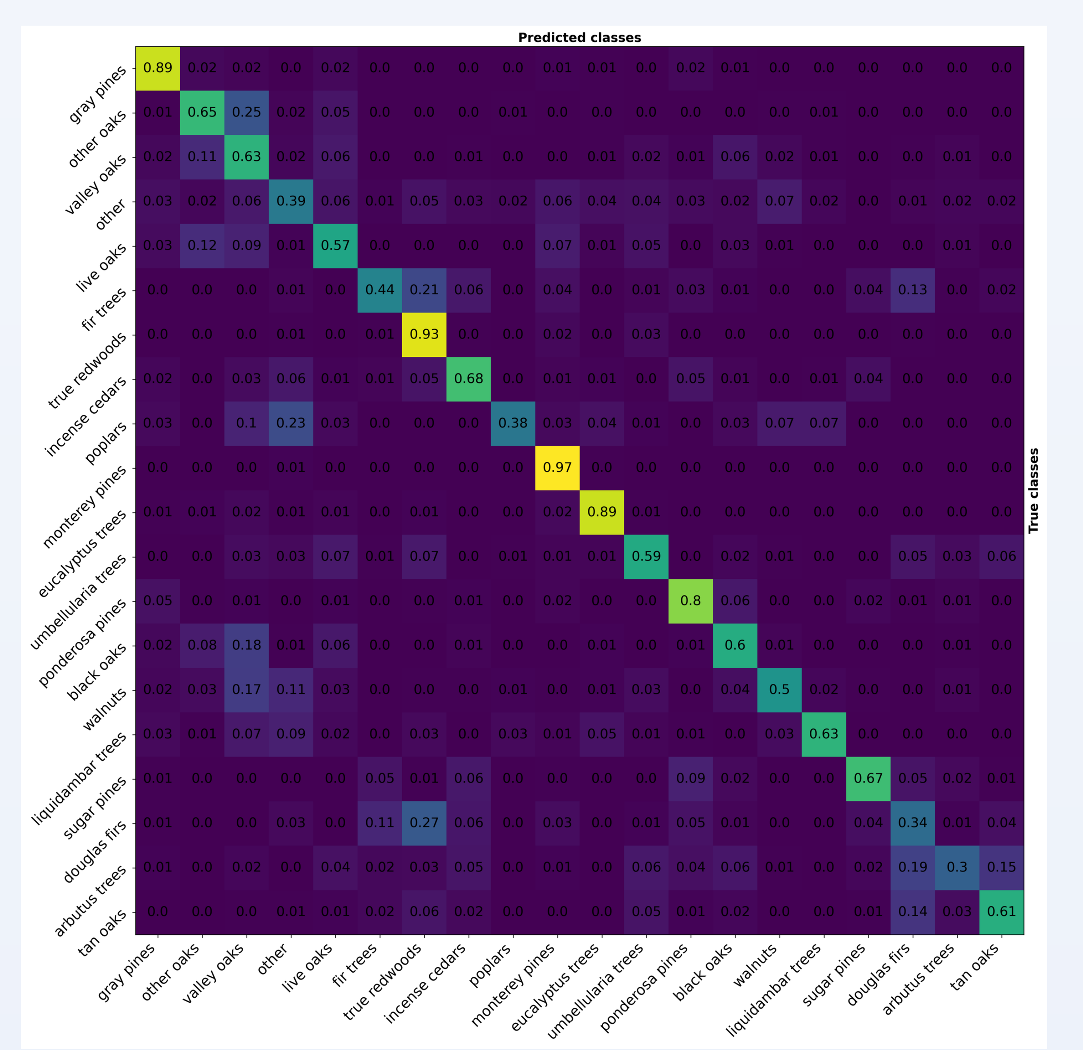

Fig. 5: MXAT Confusion Matrix on test set (normalized)

CONCLUSIONS

- ❖Overall LiDAR-based representations significantly benefits from multispectral fusion.
- ❖MXAT surpassed the DVM baseline by nearly 5\% in mean sensitivity across all predicted taxa.
- ❖These results demonstrate that even with moderate-resolution imagery, synergy between LiDAR and satellite data can be actively exploited to considerably improve classification performance

REFERENCES

- ❖[Allen et al.] Allen, M. J., Grieve, S. W. D., Owen, H. J. F., and Lines, E. R. Tree species classification from complex laser scanning data in mediterranean forests using deep learning. 14(7):1657–1667.
- ❖[Dixon et al.] Dixon, D. J., Zhu, Y., Brown, C. F., and Jin, Y. Satellite detection of canopy-scale tree mortality and survival from California wildfires with spatio-temporal deep learning. 298:113842.
- ❖[Goyal et al.] Goyal, A., Law, H., Liu, B., Newell, A., and Deng, J. Revisiting point cloud shapeclassification with a simple and effective baseline. (arXiv:2106.05304).
- ❖[Neyns et al]. Neyns, R., Efthymiadis, K., Libin, P., and Canters, F. Fusion of multi-temporal PlanetScope data and very high-resolution aerial imagery for urban tree species mapping. 99:128410.

CONTACT

HOS1@pge.com
RB19@pge.com
Pacific Gas and Electric. Oakland, California