

Does adding an elevation layer make a meaningful difference in land cover classification accuracy?

USGIF Posters to Papers
Samantha Velasco
Bethune-Cookman University

BLUF

Elevation should be treated as a standard input into classification models. Adding a 10-meter Digital Elevation Model (DEM) to radar and optical-based land cover classification produced a statistically significant improvement in overall accuracy across every sensor combination tested (average gain of 7.7 percentage points, $p < 0.001$). Within a 17-band stack, elevation contributed more to the classification than any other individual band tested. Elevation's benefit was largest when only one sensor was available, and, though smaller once multiple sensors were already combined, remained statistically significant. This study is not methodologically novel, but it carries broader significance for GEOINT applications: rather than incorporating more bands, a GEOINT analyst may be better off incorporating elevation into their classification stack to improve accuracy.

Top-Level Analysis

The analysis compared seven combinations of three remotely-sensed datasets used for land use and land cover (LULC) studies, each tested with and without elevation. The datasets include SAR (Sentinel-1), optical data (Sentinel-2), and higher resolution optical data (PlanetScope). SAR is an active sensor that transmits microwave signals and measures their return, data collection through cloud cover and at night^{1,2}. Optical sensors record reflected sunlight across multiple wavelengths, providing finer detail but requiring clear skies². The elevation data comes from a 2015 Light Range and Detection (LiDAR)-derived DEM from the United States Geological Survey (USGS). A Digital Elevation Model (DEM) is a raster grid of ground-surface elevations and is typically the least computationally expensive, most static layer because it does not require reacquisition for each date³.

LC was classified using supervised classification, in which an algorithm learns from operator-labeled training points and applies the resulting model to the remaining pixels^{4,5}. The Random Forest (RF) classifier builds multiple decision trees from random training subsets and combines their votes into a final class label^{6,7}.

RF classification was performed using five LC classes (open water, urban, bare land, upland vegetation, and wetland) for all seven sensor combinations (Table 1), each with and without a DEM, across nine dates representing five hurricanes and four baseline conditions in the Mosquito Control District (MCD) of Volusia County, Florida (2016–2025). Table 1 summarizes mean accuracy and Cohen's Kappa. A Wilcoxon signed-rank test also compared paired accuracies for each sensor combination (Table 1).

Table 1. Mean overall accuracy and Kappa coefficient, with and without the DEM, for each of the seven sensor combinations tested (n = 9 dates each). Wilcoxon signed-rank p-values are for the paired accuracy comparison. SAR refers to Sentinel-1 data; S2 refers to Sentinel-2.

Combination	Number of Sensors	Accuracy without DEM	Accuracy with DEM	Change in Accuracy	Kappa without DEM	Kappa with DEM	Change in Kappa	Wilcoxon p
SAR	1	0.478	0.680	+0.202	0.348	0.598	+0.250	0.004
Sentinel-2	1	0.808	0.861	+0.053	0.759	0.825	+0.066	0.012
PlanetScope	1	0.635	0.778	+0.144	0.543	0.720	+0.177	0.008
SAR + S2	2	0.832	0.858	+0.026	0.788	0.822	+0.034	0.250
SAR + PlanetScope	2	0.762	0.809	+0.047	0.698	0.757	+0.059	0.055
S2 + PlanetScope	2	0.814	0.848	+0.034	0.762	0.807	+0.045	0.250
SAR + S2 + PlanetScope	3	0.828	0.859	+0.031	0.783	0.822	+0.038	0.020

The DEM's improvement was statistically significant for every single-sensor baseline (Wilcoxon $p \leq 0.012$) and for the full three-sensor stack ($p = 0.02$). The three two-sensor combinations (SAR + S2, SAR + PlanetScope, and S2 + PlanetScope) showed consistent accuracy gains that did not reach significance.

A mixed effects model pooling all observations tested the DEM's effect on accuracy and Kappa, controlling for date and sensor combination as random effects. This confirmed a significant DEM effect on both accuracy ($\beta = 0.077$, $p < 0.001$) and Kappa ($\beta = 0.095$, $p < 0.001$). A DEM $\times$ sensor-count interaction showed that the DEM's benefit decreased as more spectral sensors were included ($\beta = -0.063$, $p < 0.001$), but the DEM remained significant even with all three sensors. These results suggest that the DEM is worth including regardless of the other data being incorporated.

Secondary Analysis

The MCD lies on Florida's central Atlantic coast and includes a mix of dense, urban development, coastal wetlands, and agricultural land.

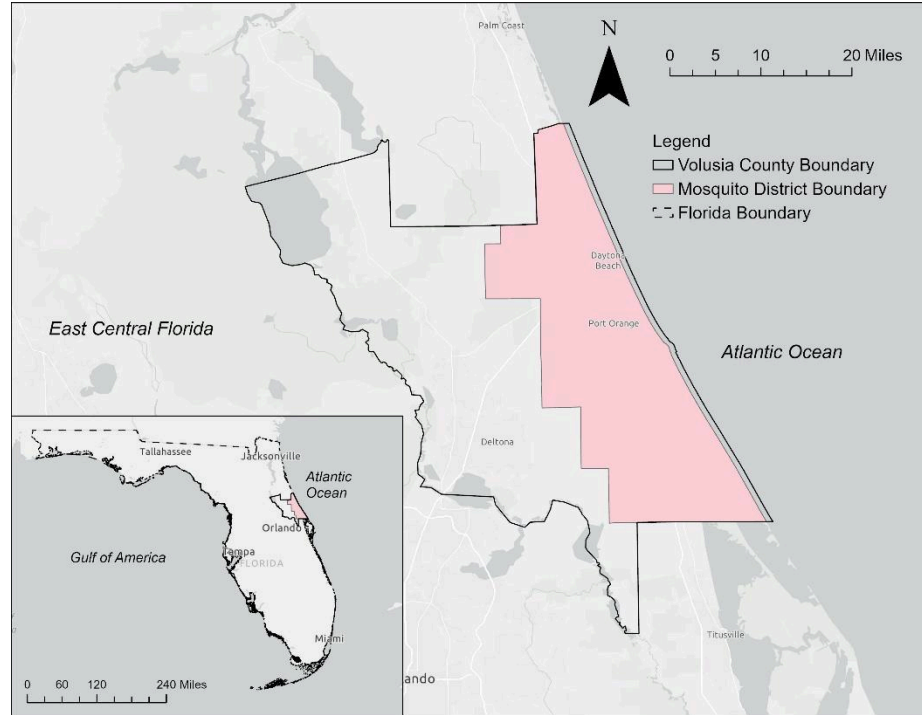


Figure 1. The MCD is shown in pink in reference to Volusia County and the state of Florida.

Table 2 summarizes each dataset used as an input to the RF classification and their access.

Table 2. Summary of remotely sensed datasets used as inputs to the RF LC classification model. All datasets were accessed and processed within Google Earth Engine (GEE) except PlanetScope, ingested via the Planet-GEE Pipeline CLI. The DEM is a static layer applied consistently across all dates.

Sensor	Mission	Provider	Spatial Resolution	Temporal Resolution	Bands Used	Preprocessing	Data Access
SAR	Sentinel-1 GRD	ESA / Copernicus	10 m	~ 6 days	VV, VH	Radiometric calibration, terrain correction, speckle filtering (applied in GEE)	GEE: COPERNICUS/S1_GRD
Optical	Sentinel-2	ESA / Copernicus	10-60 m (band dependent)	~ 5 days	All 13 bands	Surface reflectance, cloud masking (applied in GEE)	GEE: COPERNICUS/S2_SR_HARMONIZED
High-resolution optical	PlanetScope	Planet Labs	3 m	Daily (variable)	Red, Green, Blue, NIR	Radiometric calibration, surface reflectance	Planet Labs
Elevation	USGS 3DEP DEM	USGS	10 m	Static (single acquisition)	N/A	None required	GEE: USGS/3DEP/10 m

For ground-truth date (see Table 3), 100 ground-truth points per class (500 total) were collected across the LC classes and split 70/30 for training and validation.

Table 3. Ground-truth collection dates, sensor acquisition dates, and associated hurricane events for the study period (2016–2025).

Ground Truth Date	Sentinel-2 Collection	PlanetScope Collection	Sentinel-1 Collection	Event
2025-11-12	2025-11-13	2025-11-13	2025-11-15	N/A
2024-10-15	2024-10-11	2024-10-14	2024-10-14	Hurricane Milton
2024-04-19	2024-04-17	2024-04-17	2024-04-18	N/A
2023-02-04	2023-02-17	2023-02-06	2023-01-30	N/A
2022-11-15	2022-11-29	2022-11-11	2022-11-10	Hurricane Nicole
2022-10-08	2022-10-05	2022-10-05	2022-10-02	Hurricane Ian
2022-03-27	2022-03-29	2022-03-28	2022-03-24	N/A
2017-09-12	2017-09-01	2017-09-13	2017-09-12	Hurricane Irma
2016-10-09	2016-10-21	2016-10-14	2016-10-09	Hurricane Matthew

Following classification, per-band Gini importance, which is the standard RF variable-importance measure, computed as the cumulative decrease in node impurity attributable to each band across all trees in the forest⁸, was exported from a 17-band full-stack classification, using GEE's `smileRandomForest` classifier and `explain()` output (Table 4).

Table 4. Normalized per-band Gini importance (mean across 9 dates), full 17-band stack, ranked highest to lowest.

Band	Normalized Gini importance (%)
DEM	10.21
B11 (SWIR-1, S2)	7.19
VV (SAR)	6.34
B12 (SWIR-2, S2)	5.96
VH (SAR)	5.86
PS-NIR	5.74
B8A (S2)	5.59
B4 (Red, S2)	5.55
B2 (Blue, S2)	5.52
B8 (NIR, S2)	5.46
PS-Green	5.44
PS-Blue	5.39
B5 (S2)	5.39
B3 (Green, S2)	5.25
PS-Red	5.21
B6 (S2)	5.16
B7 (S2)	4.76

Figure 2 displays the per-band Gini importance spread across all classification dates.

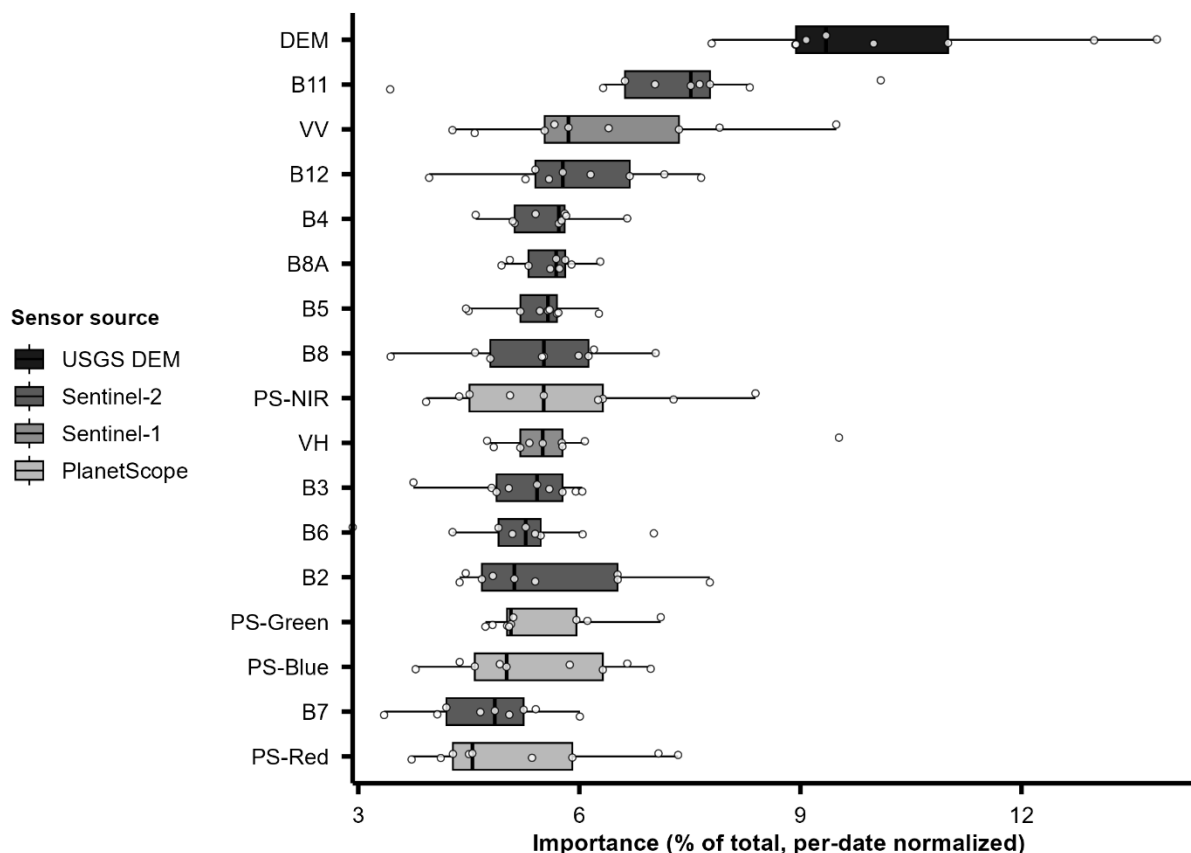


Figure 2. Per-band Gini importance across all classification dates. The DEM exceeds every individual spectral band, although with wider spread across dates than most spectral bands, indicating its contribution varied somewhat by date but was consistently high.

Table 4 and figure 4 show that the DEM has the highest Gini importance of any individual band.

Impact and Relation to GEOINT's EBK

Accurately classifying LULC is foundational to many GEOINT applications, including natural disaster monitoring^{5, 8, 9} and environmental vulnerability assessment^{5, 9-12}. Knowing which data source drives a classification decision and which sensors are worth prioritizing ultimately improves the analyst's confidence. Further, when classification accuracy improves, client-side decisions can be made faster and more reliably.

This study contributes to the GEOINT Essential Body of Knowledge (EBK) by illustrating how remote sensing datasets should be weighed in future multi-sensor classification models. It aligns with EBK 3.0 Competency 2, which addresses converging multiple sensor types for maximum benefit, and Emerging Competency ML 3.3, which addresses best practices for geospatial machine learning.

Confidence

Confidence in this assessment is moderate. The statistical tests and Gini importance rankings support the same conclusion about the DEM's contribution. This finding is consistent

with several other classification studies^{3,12}, however, the site-specific classification across these dates has not been corroborated by an independent dataset.

In terms of this study's methods, RF is a well-established classifier⁵. Gini importance can favor variables with many distinct values¹³, but bias is unlikely because all bands are continuous measurements with similar value ranges.

Bias

Several limitations should be considered when interpreting these results. The temporal sample is limited to 9 dates across a 9-year period at a single site in east-central Florida. Generalization to other geographies with different hydrologies requires further validation.

Aggregating importance by sensor rather than individual band produces different results (Table 5).

Table 5. Total Gini importance aggregated by sensor source (sum of that source's per-band importance, mean across 9 dates).

Sensor Source	Number of Bands	Total importance (%)
Sentinel-2	10	55.8
PlanetScope	4	21.8
Sentinel-1 (SAR)	2	12.2
DEM	1	10.2

Sentinel-2's ten bands account for over half of total importance, simply because it contributes the most bands, not because any single Sentinel-2 band outperforms the DEM.

Finally, while RF is well supported in the literature for LULC classification, a different classifier could produce different results⁶.

References

1. Tarpanelli, Angelica, Alessandro C. Mondini, and Stefania Camici. 2022. "Effectiveness of Sentinel-1 and Sentinel-2 for Flood Detection Assessment in Europe." *Natural Hazards and Earth System Sciences* 22 (8): 2473–89. <https://doi.org/10.5194/nhess-22-2473-2022>.
2. Guliyeva, Sona, and Piero Boccardo. 2026. "Geospatial Technologies for Flood and Drought Management: A Review of Earth Observation Data, Procedures, and Their Operational Effectiveness." *Aerotecnica Missili & Spazio* 105 (5): 707–28. <https://doi.org/10.1007/s42496-026-00309-4>.
3. Muhadi, Nur Atirah, Ahmad Fikri Abdullah, Siti Khairunniza Bejo, Muhammad Razif Mahadi, and Ana Mijic. 2020. "The Use of LiDAR-Derived DEM in Flood Applications: A Review." *Remote Sensing* 12 (14): 2308. <https://doi.org/10.3390/rs12142308>.
4. Talukdar, Swapan, and Swades Pal. 2018. "Wetland Habitat Vulnerability of Lower Punarbhaba River Basin of the Uplifted Barind Region of Indo-Bangladesh." *Geocarto International* 35 (8): 857–86. <https://doi.org/10.1080/10106049.2018.1533594>.
5. Breiman, Leo. 2001. "Random Forests." *Machine Learning* 45 (1): 5–32. <https://doi.org/10.1023/a:1010933404324>.
6. Gislason, Pall Oskar, Jon Atli Benediktsson, and Johannes R. Sveinsson. 2005. "Random Forests for Land Cover Classification." *Pattern Recognition Letters* 27 (4): 294–300. <https://doi.org/10.1016/j.patrec.2005.08.011>.
7. Che, Tao, Lin Xiao, and Yuei-An Liou. 2014. "Changes in Glaciers and Glacial Lakes and the Identification of Dangerous Glacial Lakes in the Pumqu River Basin, Xizang (Tibet)." *Advances in Meteorology* 2014 (January): 1–8. <https://doi.org/10.1155/2014/903709>.
8. Dao, Phuong, and Yuei-An Liou. 2015. "Object-Based Flood Mapping and Affected Rice Field Estimation With Landsat 8 OLI and MODIS Data." *Remote Sensing* 7 (5): 5077–97. <https://doi.org/10.3390/rs70505077>.
9. Zhang, Yonghong, Taotao Ge, Wei Tian, and Yuei-An Liou. 2019. "Debris Flow Susceptibility Mapping Using Machine-Learning Techniques in Shigatse Area, China." *Remote Sensing* 11 (23): 2801. <https://doi.org/10.3390/rs11232801>.
10. Nguyen, Kim-Anh, and Yuei-An Liou. 2019. "Mapping Global Eco-environment Vulnerability Due to Human and Nature Disturbances." *MethodsX* 6 (January): 862–75. <https://doi.org/10.1016/j.mex.2019.03.023>.
11. Talukdar. 2019. "Effects of Damming on the Hydrological Regime of Punarbhaba River Basin Wetlands." *Ecological Engineering* 135 (May): 61–74. <https://doi.org/10.1016/j.ecoleng.2019.05.014>.
12. Konapala, Goutam, Sujay V. Kumar, and Shahryar Khalique Ahmad. 2021. "Exploring Sentinel-1 and Sentinel-2 Diversity for Flood Inundation Mapping Using Deep Learning." *ISPRS Journal of Photogrammetry and Remote Sensing* 180 (August): 163–73. <https://doi.org/10.1016/j.isprsjprs.2021.08.016>.
13. Nembrini, Stefano, Inke R König, and Marvin N Wright. 2018. "The Revival of the Gini Importance?" *Bioinformatics* 34 (21): 3711–18. <https://doi.org/10.1093/bioinformatics/bty373>.
14. Deng, J. S., K. Wang, Y. H. Deng, and G. J. Qi. 2008. "PCA-based Land-use Change Detection and Analysis Using Multitemporal and Multisensor Satellite Data." *International Journal of Remote Sensing* 29 (16): 4823–38. <https://doi.org/10.1080/01431160801950162>.