

Identification and Solar Potential Estimation Using Machine Learning and Sentinel-2 Satellite Imagery in Rural Regions of Colombia and Chile

Ivan Carroll-Janer

Faculty of Economics, Los Andes University, Bogota, Colombia
Email: iecj@uniandes.edu.co

How to cite this paper: Carroll-Janer, I. (2026) Identification and Solar Potential Estimation Using Machine Learning and Sentinel-2 Satellite Imagery in Rural Regions of Colombia and Chile. *Energy and Power Engineering*, 18, 635-650. <https://doi.org/10.4236/epe.2026.1810029>

Received: September 3, 2026

Accepted: October 10, 2026

Published: October 13, 2026

Copyright © 2026 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0). <http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

This paper applies a MaxEnt classification workflow to Sentinel-2 spectral bands and derived indices to map rooftops, non-intervened land, and intervened land across selected rural areas in Colombia and Chile. The study compares detected rooftop counts with housing-census counts and links the classification output to a photovoltaic screening tool. It identifies cloud cover, snow-related spectral confusion, and roof geometry as important constraints on wider application.

Keywords

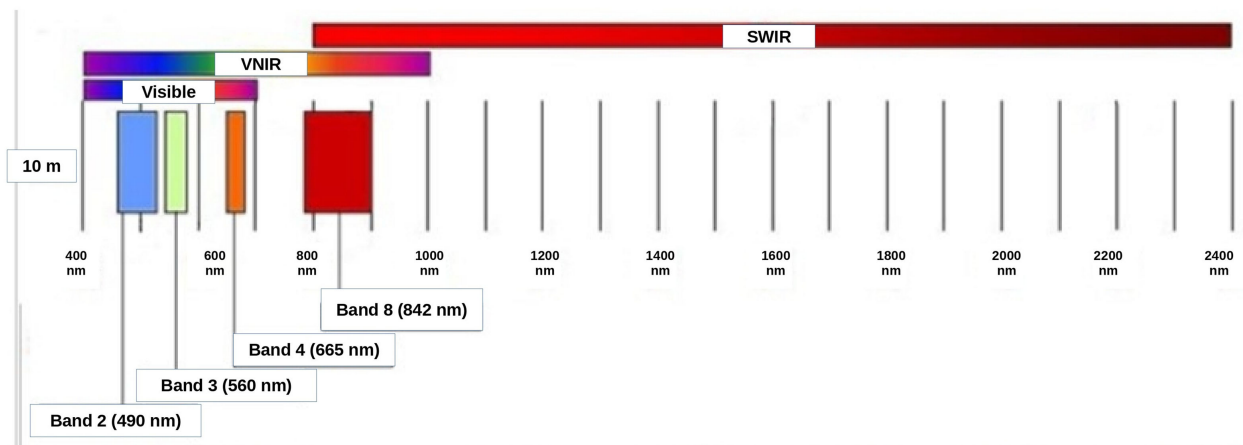
Sentinel-2, MaxEnt, Copernicus, Machine Learning, Photovoltaic Tool, Remote Sensing, Rooftop Solar PV, Distributed Generation, Energy Transition, Colombia, Chile

1. Tool Development for Estimating Rooftop Solar Potential

In Colombia, national land administration directives and remote sensing frameworks have increasingly prioritized satellite telemetry to democratize spatial technology and formalize rural infrastructure. Within this directive, multidisciplinary initiatives led by Los Andes University have advanced satellite telemetry applications using Sentinel-2 and data cube architectures across diverse domains, including deforestation tracking, agricultural monitoring [1] [2], biomass aerosol transport [3], ecosystem intervention [4], and land-use land-cover change from native forests to extensive agriculture [5]. Building directly upon these remote sensing workflows, the present study couples MaxEnt machine-learning roof extraction with solar radiation modeling across high-priority rural municipalities, providing a scalable spatial pipeline for decentralized photovoltaic potential

mapping.

Sentinel-2 imagery from the Copernicus program was ingested via Google Earth Engine for each study polygon. Median surface reflectance composites were computed for the period between January 1, 2023, and May 31, 2023, filtered for cloud coverage below 10%. As illustrated in **Figure 1**, Sentinel-2 spectral bands with 10-metre spatial resolution (B2 at 490 nm, B3 at 560 nm, B4 at 665 nm) as well as bands 8, 8A, 11, and 12 were extracted. Visible spectrum bands (B2 at 490 nm, B3 at 560 nm, B4 at 665 nm) and the Near-Infrared band (B8 at 842 nm) operate at a 10-meter spatial resolution.

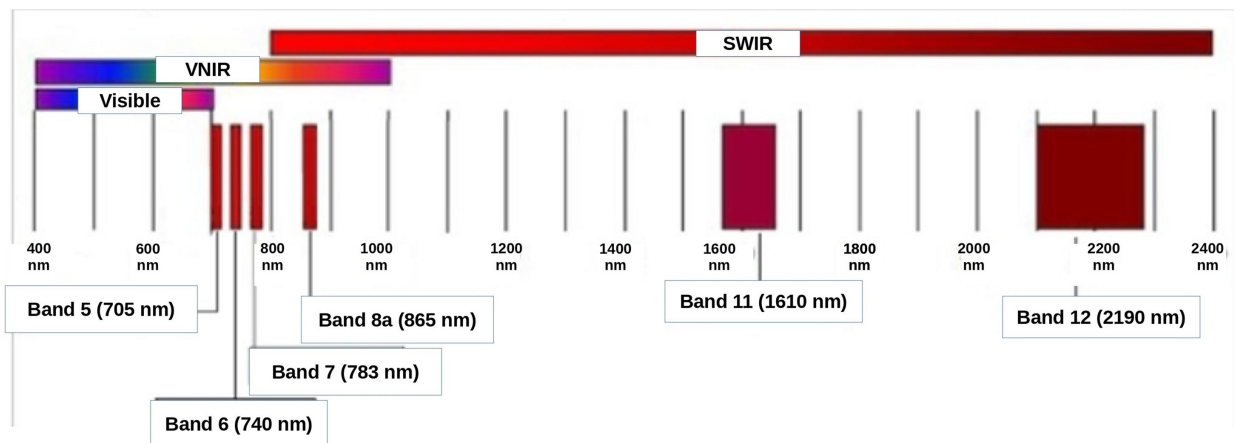


Adapted from European Space Agency online archive [6].

Figure 1. Sentinel-2 10 metre spatial resolution.

1.1. Bands and Indexes

As shown in **Figure 2**, bands 8A (865 nm), 11 (1610 nm), and 12 (2190 nm)—covering the narrow NIR and Short-Wave Infrared (SWIR) spectra—operate at a 20-metre spatial resolution.



Adapted from European Space Agency online archive [6].

Figure 2. Sentinel-2 20 metre spatial resolution.

Prior to model training and feature extraction, the 20 m spatial resolution bands (B8A, B11, and B12) were resampled to match the 10 m native grid of the visible and NIR bands (B2, B3, B4, B8). Resampling was executed in Google Earth Engine using nearest-neighbor interpolation to preserve raw spectral integrity without introducing synthetic values. All spectral layers were co-registered and snapped to the unified 10 m Copernicus Sentinel-2 UTM grid, establishing a consistent 10 m $\times$ 10 m pixel basis across all input features.

The inclusion of the SWIR channels (B11 and B12) is physical-chemically critical for structural classification: these wavelengths overlap directly with fundamental liquid water absorption features (H_2O molecular absorption bands). Consequently, they capture micro-scale moisture content in canopy foliage and upper soil layers, enabling robust spectral discrimination between moisture-rich natural surfaces (dense vegetation, moist soils) and spectrally dry, non-photosynthetic artificial materials (such as concrete, corrugated metal, and roofing tiles).

Using these spectral channels, four distinct spectral indices were generated through mathematical combinations of visible, NIR, and SWIR wavelengths: NDVI, NDWI, SAVI, and NDBI.

- NDVI: Normalized Difference Vegetation Index is calculated as the normalized ratio between the red band (Band 4) and the near-infrared band (Band 8).

$$NDVI = \frac{\text{Band 8} - \text{Band 4}}{\text{Band 8} + \text{Band 4}}$$

This index measures photosynthetic activity and plant vigor based on chlorophyll absorption characteristics [7]. NDVI values range between -1 and 1 ; negative values (near -1) indicate non-vegetated surfaces such as water, rock, or snow. Values between 0 and 0.5 correlate with sparse vegetation, shrubs, or grasslands, whereas values exceeding 0.5 correspond to dense forest canopies and active crops.

- NDWI: Normalized Difference Water Index is computed using the green band (Band 3) and NIR band (Band 8).

$$NDWI = \frac{\text{Band 3} - \text{Band 8}}{\text{Band 3} + \text{Band 8}}$$

Because open water surfaces strongly absorb NIR light, this ratio isolates aquatic features such as rivers and lakes [8]. Positive values (>0) represent open water bodies or high canopy water content, while negative values indicate dry or moisture-stressed surfaces.

- SAVI: Soil-Adjusted Vegetation Index integrates a soil-brightness correction factor ($L = 0.428$) into the standard vegetation index formula to minimize background soil reflectance influence [9].

$$SAVI = \frac{\text{Band 8} - \text{Band 4}}{\text{Band 8} + \text{Band 4} + 0.428} \times 1.428$$

SAVI is particularly effective in arid regions, areas with low vegetation density, or exposed soils [10]. Given Colombia's mountainous terrain and high suscepti-

bility to landslides, soil-adjusted metrics are critical for discriminating exposed, highly erodible terrain from built structures.

- NDBI: Normalized Difference Built-up Index—NDBI—isolates built-up infrastructure by leveraging the spectral contrast between SWIR (Band 11) and NIR (Band 8).

$$\text{NDBI} = \frac{\text{Band 11} - \text{Band 8}}{\text{Band 11} + \text{Band 8}}$$

Values close to +1 correlate with high built-up density and urban surfaces, whereas negative values correspond to natural land cover types [11].

In summary, the machine learning input feature space combined Sentinel-2 spectral bands (2, 3, 4, 8, 8A, 11, 12) with the derived indices (NDVI, NDWI, SAVI, NDBI).

1.2. MaxEnt Algorithm and Model Configuration

Supervised machine learning and reflectance classification were conducted using Maximum Entropy (MaxEnt) software. MaxEnt processing requires two main inputs: presence sample points and environmental layers. Ground-truth training samples were digitized using high-resolution imagery (<1 m) in Google Earth Pro (powered by Maxar Technologies, Airbus, and CNES). Environmental layers (spectral bands and derived indices) were prepared as ASCII grid files (.asc). Multi-class outputs were structured by evaluating parallel binary classifiers, capped at a maximum of three classifiers per sub-dataset to maximize structural accuracy [12] [13].

To ensure reproducibility and rigorous model calibration, MaxEnt was configured with the following parameters:

- Background Samples: 10,000 background points randomly sampled across non-built land-cover classes within each study polygon.
- Feature Classes: Linear (L), Quadratic (Q), and Product (P) features were selected to model non-linear responses across spectral bands and derived indices.
- Regularization & Replication: A regularization multiplier of $\beta = 1.0$ was applied to constrain overfitting, combined with a 5-fold cross-validation procedure.
- Threshold Selection: Continuous logistic outputs (Probability $\in [0, 1]$) were converted to binary predictions using the 10th percentile training presence threshold.
- Overlap Resolution: Overlapping predictions between parallel binary classifiers were resolved by assigning pixels the highest logistic probability output across competing models.

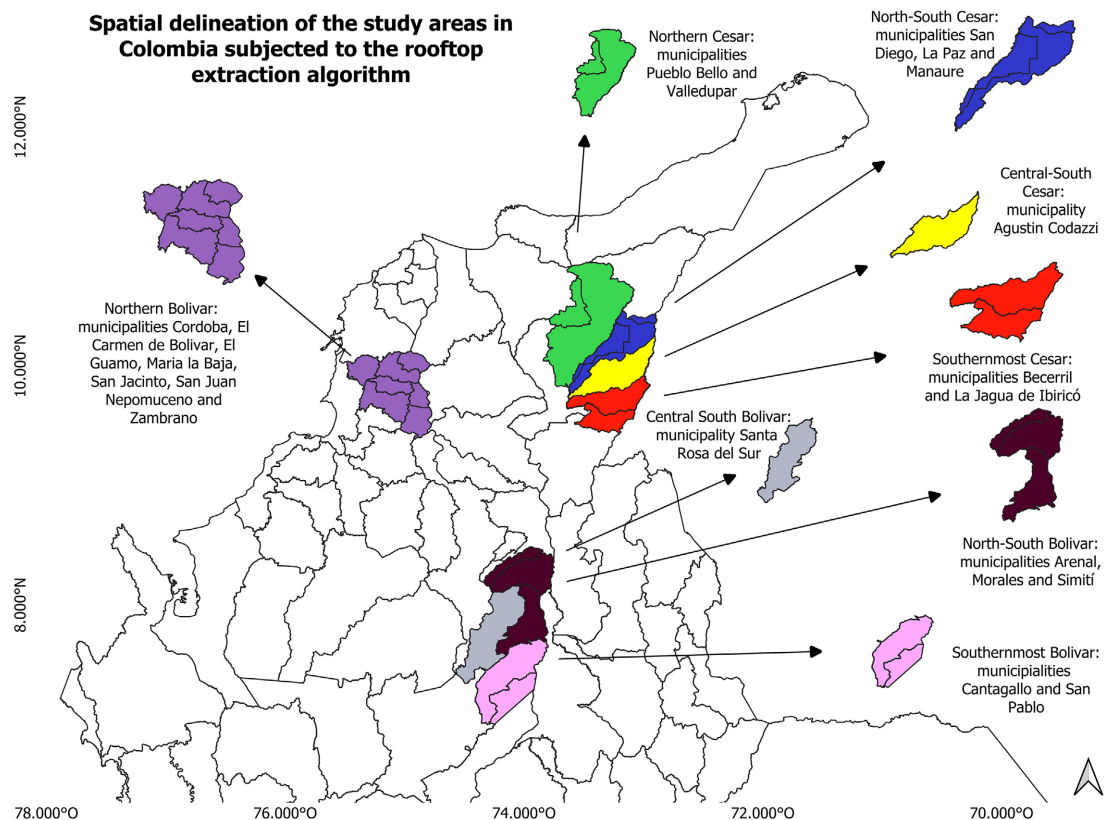
The mapped target of the classification algorithm is defined as a roof-containing pixel at a 10 m spatial resolution. Because a single 10 m $\times$ 10 m pixel (100 m²) may encompass multiple small structures, part of a single building, or mixed land cover (e.g., roof-vegetation-soil composites), continuous probability outputs from

MaxEnt were thresholded into a binary mask of roof-containing pixels. To derive discrete rooftop counts for administrative comparison, adjacent classified pixels were clustered into continuous spatial polygons using 8-neighbor vectorization. Small isolated clusters below a minimum structural threshold ($<20 \text{ m}^2$) were filtered out to eliminate speckle noise, while spatial centroid decomposition was applied to larger continuous roof clusters to approximate individual building units.

Data analysis, spatial manipulation, and visualization were executed using R Studio (v1.4.1717) and QGIS (v3.16.11). R processing utilized the “sp” and “raster” libraries, following spatial coding standards [14] [15].

2. Study Area Polygon Definition and Data Training

Figure 3 illustrates the spatial boundaries of the target study polygons across Colombia and Chile, where eleven distinct Sentinel-2 satellite images were processed.



Map prepared by the author using QGIS.

Figure 3. Division of study areas for the application of the rooftop identification algorithm.

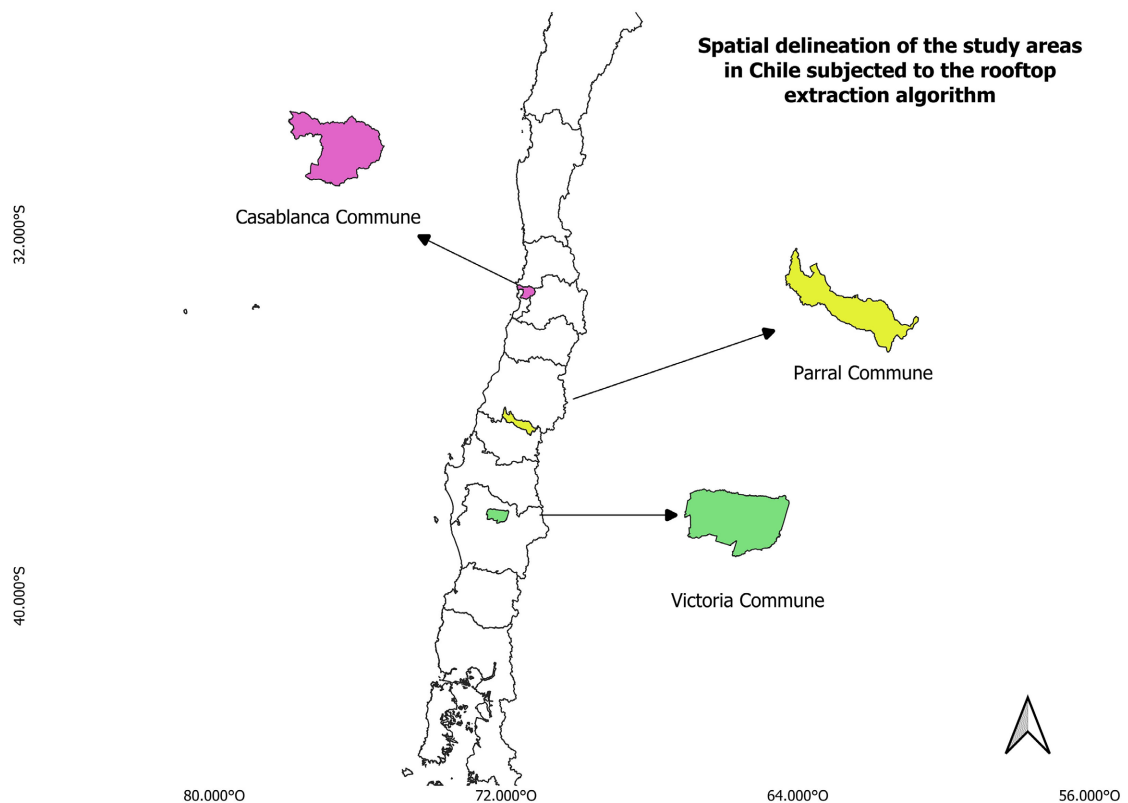
Model evaluations were re-run iteratively whenever unclassified pixel ratios exceeded 10% or rooftop spatial distribution deviated significantly from official housing census baselines [16] [17]. Official census data served as a structural baseline under the assumption that one residential dwelling corresponds to at least one primary rooftop. However, classified rooftop counts were expected to moderately ex-

ceed census figures, as rural homesteads frequently feature detached structures (e.g., separate kitchens, outdoor sanitation facilities, or livestock shelters).

Although iterative re-training increased processing time, it ensured high classification fidelity across heterogeneous rural landscapes and minimized unclassified pixels.

Individual Sentinel-2 scenes cover an area of $110 \text{ km} \times 110 \text{ km}$. Custom scripts executed via Google Earth Engine clipped imagery directly to target study polygons. Where study regions exceeded single scene footprints, target areas were subdivided. In Colombia, study zones were partitioned as follows:

- Bolivar Department (PDET Municipalities): Northern Bolivar (Cordoba, El Carmen de Bolivar, El Guamo, Maria la Baja, San Jacinto, San Juan Nepomuceno, Zambrano); North-South Bolivar (Arenal, Morales, Simití); Central South Bolivar (Santa Rosa del Sur); Southernmost Bolivar (Cantagallo, San Pablo).
- Cesar Department (PDET Municipalities): Northern Cesar (Pueblo Bello, Valledupar); North-South Cesar (San Diego, La Paz, Manaure); Central-South Cesar (Agustín Codazzi); Southernmost Cesar (Becerril, La Jagua de Ibirico).



Map prepared by the author using QGIS.

Figure 4. Map of Chile showing the division of study areas for the application of the rooftop identification algorithm.

The total area evaluated in Colombia comprised 125,340,904 pixels of 100 m^2 ($12,534 \text{ km}^2$) across Bolivar and 107,528,976 pixels of 100 m^2 ($10,752 \text{ km}^2$) across

Cesar, totaling 23,286 km² (*a spatial extent larger than the European nation of Slovenia equivalent to 20,273 km²*).

In Chile, the study targeted three agricultural and administrative communes: Casablanca, Parral, and Victoria, as illustrated in **Figure 4**.

A total of 47,580,000 pixels of 100 m² (4758 km²) were evaluated—breakdown by commune: Casablanca (1114 km²), Parral (2035 km²), and Victoria (1609 km²). Median composite scenes (Jan-May 2023, cloud cover < 10%) were processed per commune via Google Earth Engine. Boundary buffering was evaluated but discarded due to scene spatial limit constraints. To prevent spatial autocorrelation from artificially inflating model performance, accuracy metrics were evaluated using spatial block cross-validation (spatial holdout splits) rather than random point selection. Test points were geographically separated from training clusters using spatial buffering to guarantee independence.



High-spatial-resolution imagery <1 m obtained from Google Earth Pro.

Figure 5. Rooftop categories illustrating structural and color variations included in the training dataset.

As shown in **Figure 5**, high-spatial-resolution imagery (<1 m) obtained from Google Earth Pro covers all structural roofs regardless of color or material composition. Visible color variations (white, gray, red, blue, orange) and structural materials (corrugated zinc, clay tile, plastic, concrete) were incorporated into model feature spaces without requiring individual material-class pre-labeling.

For each study polygon, 300 supervised ground-truth points were established using high-resolution Google Earth Pro imagery: 100 points for “Rooftops”, 100 for “Non-Intervened land”, and 100 for “Intervened land”, as follows considerations:

- Non-Intervened: Natural, unbuilt land covers, including primary forests, open water bodies, paramo vegetation, and high-altitude snowpack. 10,000 background points randomly sampled across non-built land-cover classes within each study polygon.
- Intervened: Human-altered landscapes excluding buildings, such as agricultural crops, pastures, paved roads, and unpaved/dirt roads.
- Unclassified: Pixels remaining unassigned by the MaxEnt classification threshold.

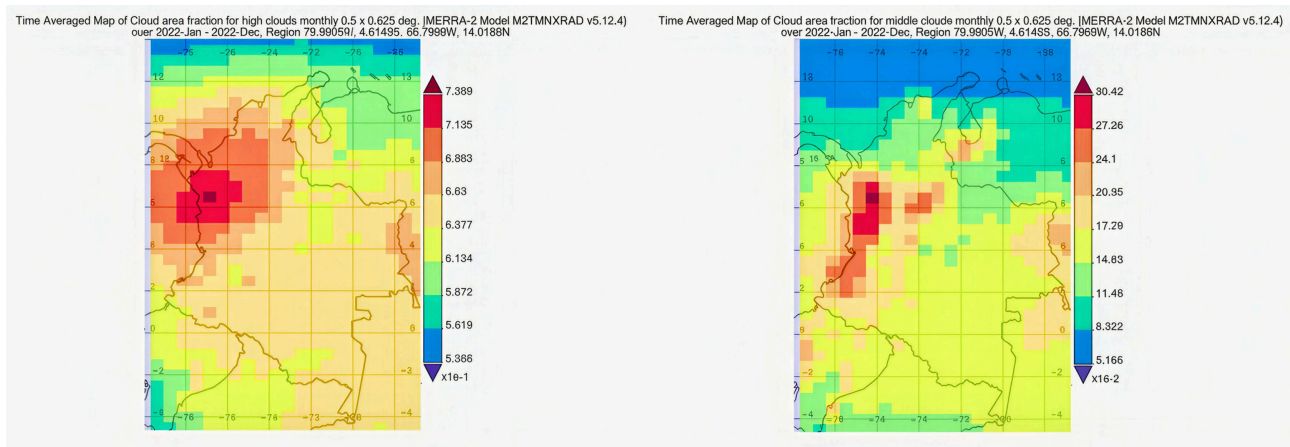
In Colombia, across 8 study polygons, a total of 2400 training points (800 rooftops, 800 non-intervened, 800 intervened) were carefully sampled, each maintaining a minimum concentric radius of 10 meters. In Chile, across 3 study polygons, 900 points were sampled following identical spatial buffering criteria.

3. Scenario Analysis of Distributed Energy Resources in the Study Regions

Prior field implementation experience in regions like the Choco department demonstrated that persistent tropical cloud cover poses significant constraints for optical satellite analysis. Atmospheric moisture accumulation along the mountain slopes of Colombia's three Andean ranges frequently obstructs optical sensors.

Consequently, Bolivar and Cesar departments were selected as robust proof-of-concept regions due to consistently low cloud cover during dry seasons. Scaling this framework nationwide will require cloud-masking algorithms or synthetic aperture radar (SAR) integration to mitigate persistent atmospheric interference.

According to long-term climatological observations from NASA's MERRA-2 (Modern-Era Retrospective Analysis for Research and Applications) satellite dataset, high- and mid-level cloud fractions across Colombia in 2022 were heavily concentrated over the Pacific coastal region, as shown in **Figure 6**.



Analysis and maps retrieved from the NASA Earth data Portal.

Figure 6. Annual average spatial distribution of high-altitude and mid-altitude cloud fractions over Colombia for 2022, derived from NASA MERRA-2 satellite data.

Rooftop raster outputs were integrated directly into the web-based photovoltaic visualization tool. Solar PV generation potential was calculated using the following formulation:

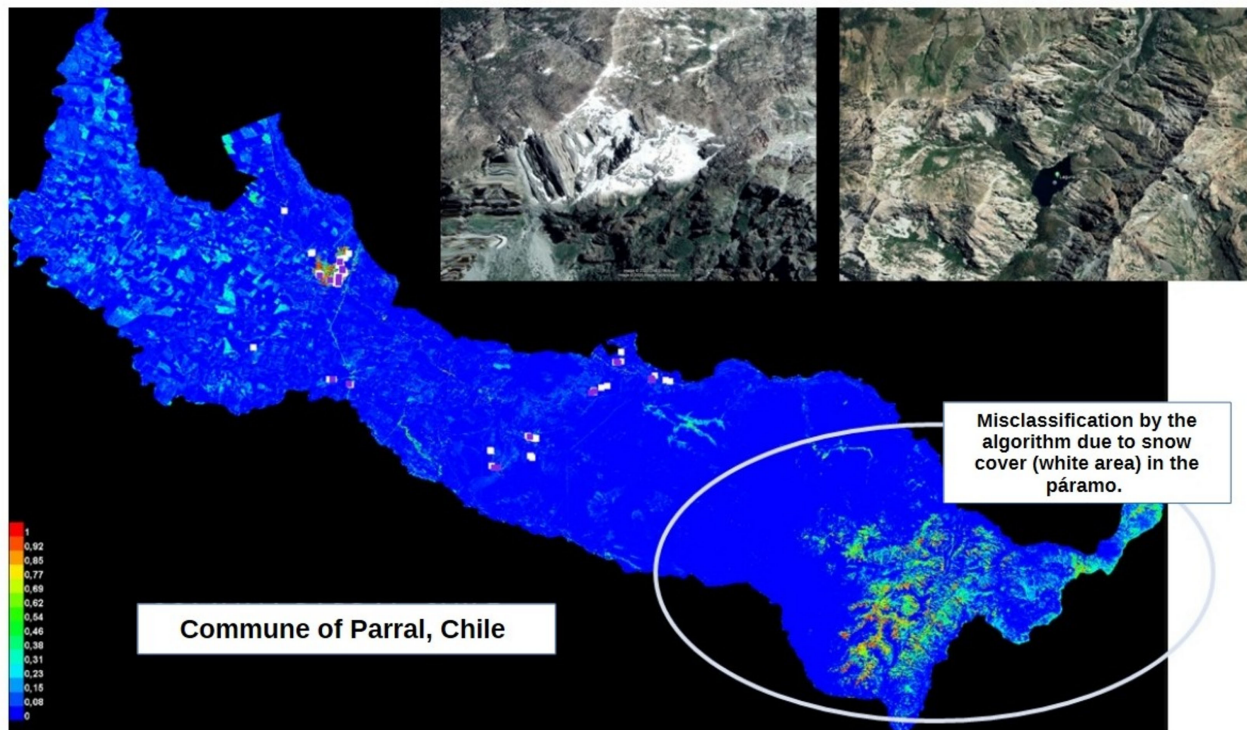
$$P_{PV} = G^* A^* \eta$$

where P_{PV} represents the solar photovoltaic power output (W), G is the global solar irradiance (W/m^2), A is the total rooftop panel area (m^2), and η is the system conversion efficiency ($\eta \approx 0.20$).

The satellite screen capture of the Maxent process shown in **Figure 7** highlights a prominent classification challenge: spectral confusion between high-altitude snowpack and light-colored rooftops due to near-identical visible and shortwave reflectance signatures.

A prominent classification challenge involved spectral confusion between high-altitude snowpack and light-colored rooftops due to near-identical visible and

shortwave reflectance signatures. This misclassification was observed in high-elevation areas of Parral (*Chile*) and Central-Southern Cesar (*Colombia*).



Satellite screen capture of Maxent process acquired by the author.

Figure 7. Classification error due to high-altitude snowpack interference during rooftop mapping. Model training was conducted in Google Earth Pro, and the Machine Learning model was executed using Copernicus Sentinel-2 satellite imagery (satellite image in MaxEnt captured by the author).

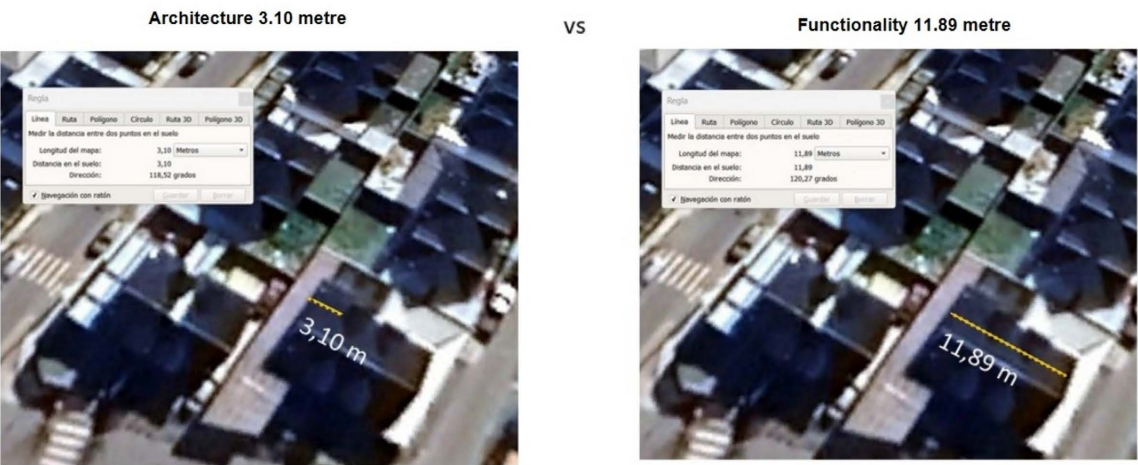
4. Opportunities for Improvement in Rooftop Solar Implementation

Iterative model tuning confirmed that combining multi-spectral bands with vegetation, water, and built-up indices provides a robust foundation for automated rooftop detection. However, as illustrated in **Figure 8**, spatial limitations in usable rooftop surface area often arise due to specific building design features, as observed in the Parral Commune, Chile.

Advanced spatial post-processing using commercial GIS platforms (e.g., ArcGIS Spatial Analyst) can enhance technical assessments by incorporating 3D building height models and rooftop pitch analysis. Tools such as “Area Solar Radiation” can calculate micro-level solar insolation by accounting for seasonal solar azimuth, shadows from adjacent structures, roof slope, and aspect orientation.

While rural rooftop structures in Colombia exhibit relatively homogeneous, flat geometries, residential architecture in Chile features complex roof slopes, gables, and dormer windows. These architectural features introduce spatial shading constraints and reduce effective usable panel area. High-resolution spatial analysis in Chilean communes showed that decorative architectural elements reduce contig-

uous roof plane lengths (e.g., from an overall roof dimension of 11.89 m down to 3.10 m usable space before encountering structural obstructions). and import your prepared text file. You are now ready to style your paper.



The architectural geometry of a structure with a 3.10-metre roof length limits the functionality of the algorithm.

A roof length of 11.89 metres is more suitable for the algorithm.

High-spatial-resolution imagery (<1 m) obtained from Google Earth Pro.

Figure 8. Spatial limitations in usable rooftop surface area due to building design features in the Parral Commune, Chile.



High-spatial-resolution imagery (<1 m) obtained from Google Earth Pro.

Figure 9. High-spatial-resolution image showing the roof surface of the Municipal Sports Complex in Cordoba (Bolívar, Colombia).

The methodology also facilitates identifying large-format public infrastructure suitable for community solar projects. For example, the roof of the Municipal Sports Complex in Cordoba (Bolívar, Colombia) measures 35 m × 38 m, as shown in **Figure 9**.

Deploying rooftop PV on large civic structures can accelerate rural electrification, lower public energy overhead, and support community development goals.

5. Results and Conclusions

When evaluating machine-learning outputs against national housing surveys (DANE in Colombia; INE in Chile), it is critical to distinguish between their underlying spatial constructs. Census baselines measure residential dwellings and households, whereas satellite-based classification identifies physical roof-containing structures (which include agricultural sheds, commercial units, industrial facilities, and multi-dwelling apartment blocks). Discrepancies between total pixel counts and census numbers stem from these structural divergences rather than classification error alone.

5.1. Colombian Results

Table 1 presents the rooftop detection results for the Bolívar Department using machine learning compared against official housing baseline data [16]. Specifically, the MaxEnt model identified 95,952 rooftops, whereas the 2018 DANE National Housing Census recorded 80,222 residential units.

Table 1. Rooftop detection results for the bolivar department using machine learning compared against official housing baseline data [16].

Colombia Analysis	Rooftop detection results								
	Northern Bolívar	% of Total Region Area	North-South Bolívar	% of Total Region Area	Central-South Bolívar	% of Total Region Area	Southernmost Bolívar	% of Total Region Area	Total Rooftops
Machine Learning Rooftop Count	56,403	0.14%	25,072	0.08%	8416	0.03%	6061	0.02%	95,952
2018 Housing Census (DANE) ^a	52,626	0.13%	11,210	0.03%	7342	0.03%	9044	0.03%	80,222

a. CNPV corresponds to the 2018 National Population and Housing Census of a Table footnote (*Censo Nacional de Población y Vivienda footnote is dispensable*). conducted by the National Administrative Department of Statistics.

- Northern Bolívar: 0.43% unclassified pixels; training accuracy: 98.68%; test accuracy: 98.58%.
- North-South Bolívar: 0.18% unclassified pixels; training accuracy: 98.74%; test accuracy: 98.58%.
- Central South Bolívar: 0.90% unclassified pixels; training accuracy: 98.09%; test accuracy: 97.51%.
- Southernmost Bolívar: 0.50% unclassified pixels; training accuracy: 99.30%;

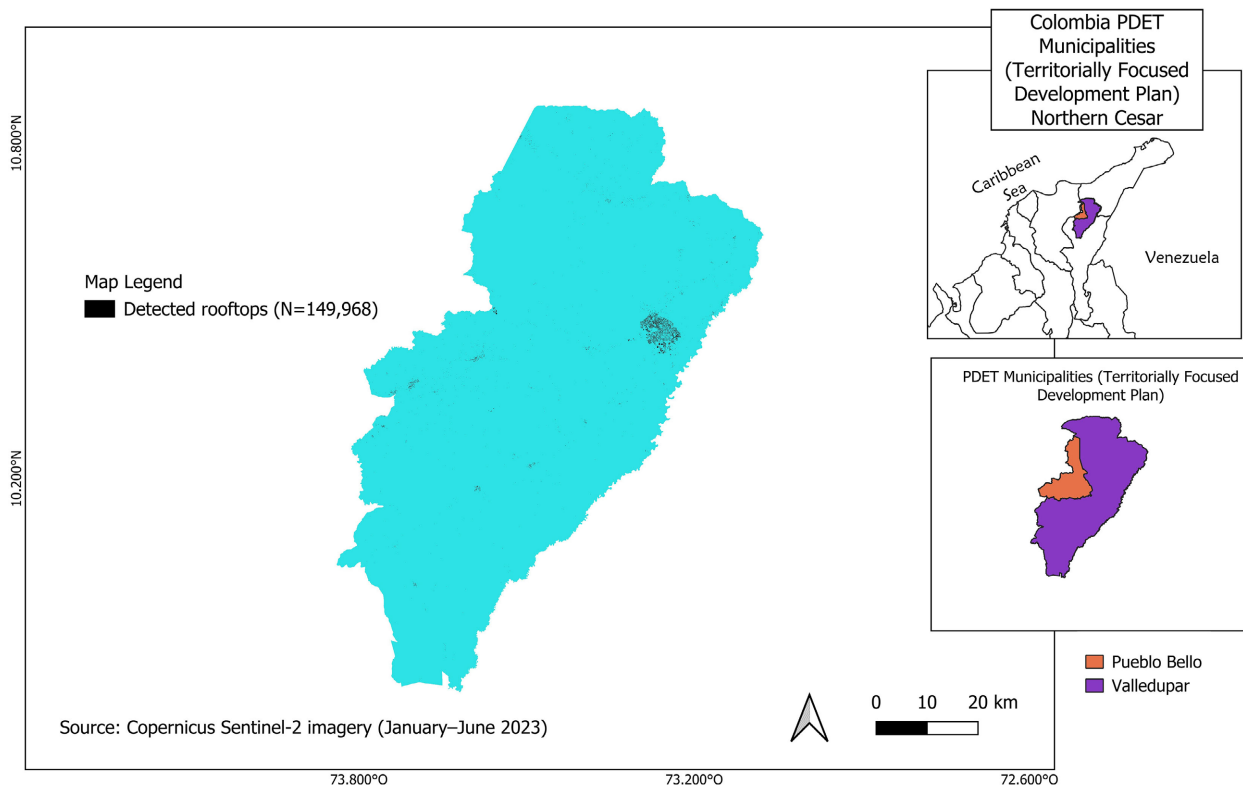
test accuracy: 99.17%.

In Cesar department, the model detected 243,815 rooftops, compared to 168,926 residential units in the 2018 DANE census.

- Northern Cesar: 1.99% unclassified pixels; training accuracy: 98.56%; test accuracy: 98.07%.
- North-South Cesar: 0.50% unclassified pixels; training accuracy: 99.19%; test accuracy: 98.85%.
- Central-South Cesar: 1.18% unclassified pixels; training accuracy: 99.09%; test accuracy: 98.90%.
- Southernmost Cesar: 3.04% unclassified pixels; training accuracy: 98.46%; test accuracy: 98.24%.

As shown in **Figure 10**, the spatial distribution map demonstrates the performance of the machine learning algorithm in northern Cesar. A total of 149,968 rooftops were identified across the region.

Algorithmic Detection of Rooftops Northern Cesar Polygon



Map prepared by the author using QGIS.

Figure 10. Spatial distribution of identified rooftops within the northern Cesar study polygon (*municipalities of Pueblo Bello and Valledupar*).

Following the same analysis for the Department of Cesar, **Table 2** shows the machine learning rooftop detection results compared against official housing baseline data [16].

Table 2. Rooftop detection results for the Cesar Department using machine learning compared against official housing baseline data [16].

Results in Colombia	Rooftop detection results								
	Northern Cesar	% of Total Region Area	North-South Cesar	% of Total Region Area	Central-South Bolivar	% of Total Region Area	Southernmost Bolivar	% of Total Region Area	Total Rooftops
Machine Learning Rooftop Count	149,968	0.3%	13,428	0.07%	70,842	0.4%	9577	0.48%	243,815
2018 Housing Census (DANE) ^b	126,871	0.26%	9112	0.05%	16,305	0.09%	16,638	0.08%	168,926

b. CNPV corresponds to the 2018 National Population and Housing Census of a Table footnote (*Censo Nacional de Población y Vivienda footnote is dispensable*). conducted by the National Administrative Department of Statistics.

5.2. Results for Study Regions in Chile

Across the evaluated Chilean communes (Victoria, Parral, Casablanca), the model detected 71,355 rooftops, compared to 41,882 residential units in the 2017 INE Housing Census.

- Victoria: 0.14% unclassified pixels; training accuracy: 98.91%; test accuracy: 98.75%.
- Parral: 9.69% unclassified pixels; training accuracy: 95.55%; test accuracy: 94.63% (impacted by high-altitude snowpack confusion).
- Casablanca: 0.43% unclassified pixels; training accuracy: 99.18%; test accuracy: 98.89%.

Extending this comparative analysis to Chile, **Table 3** shows the machine learning rooftop detection results across the evaluated communes compared against the official INE (Instituto Nacional de Estadísticas) housing census baseline [17].

Table 3. Rooftop detection results for evaluated communes in Chile using machine learning compared against official housing baseline data [17].

Results in Chile	Rooftop detection results						
	Victoria	% of Total Region Area	Parral	% of Total Region Area	Casablanca	% of Total Region Area	Total Rooftop
Machine Learning Rooftop Count	11,487	0.07%	3349	0.02%	56,519	0.49%	71,355
2017 Housing Census (INE) ^c	12,828	0.08%	17,448	0.03%	11,606	0.1%	41,882

c. National Statistics Institute [12].

Jackknife feature importance evaluations in MaxEnt revealed consistent variable performance across Colombia and Chile. Visible bands (B2, B3, B4) and environmental indices (NDWI, NDVI, SAVI) contributed most significantly to model gain. Within shortwave infrared channels, Band 12 (SWIR-2) provided the strongest relative contribution, whereas Band 8, Band 8A, and NDBI exhibited lower individual permutation importance. Nonetheless, maintaining this specific com-

bination of bands and indices proved necessary for stabilizing maximum-likelihood spectral separation across complex land covers. Under the Colombia Sostenible framework, rural development initiatives were established across vulnerable communities. Future phases should connect distributed rooftop PV generation directly to productive agricultural infrastructure (e.g., crop processing, refrigeration, and irrigation) rather than limiting energy usage strictly to basic household lighting.

6. Recommendations for Future Work

Community engagement and local stakeholder alignment remain critical for ensuring long-term technical adoption and social integration of distributed solar energy systems in rural areas.

Under the “Colombia Sostenible” framework, rural development initiatives were established across vulnerable communities. Future phases should connect distributed rooftop PV generation directly to productive agricultural infrastructure (e.g., crop processing, refrigeration, and irrigation) rather than limiting energy usage strictly to basic household lighting.

For example, in livestock management and silvopastoral systems, small-scale solar arrays can power electric fencing and water pumping systems, fostering sustainable agricultural transition.

Academia and public institutions must collaborate to present these spatial tools to local communities, demonstrating how decentralized clean energy reduces socioeconomic vulnerability and energy poverty.

Because the developed MaxEnt workflow effectively discriminates linear infrastructure, future research will focus on automated mapping and characterization of unpaved tertiary road networks. Mapping tertiary access routes will enable precise routing assessments for distribution lines, transformer placement, and grid extension planning in remote regions.

Finally, given the structural baseline utility and high spatial agreement within urban cores demonstrated by machine learning classification across Colombia (DANE) and Chile (INE), national statistical agencies are strongly advised to integrate remote sensing and automated land-cover modeling into census methodologies. While satellite mapping identifies physical structures rather than strictly residential households, deploying satellite-based structural identification offers significant strategic advantages:

- **Cost Reduction and Operational Efficiency:** It optimizes administrative budgets by significantly decreasing logistical overhead, paper-based surveys, and extensive field personnel deployment.
- **Risk Mitigation in Conflict and Remote Zones:** In regions characterized by active armed conflict, social instability, or ongoing post-peace-agreement transition, automated spatial mapping replaces physical site inspection, directly safeguarding field enumerators and researchers from physical security threats.
- **Transparency and Bureaucratic Streamlining:** By anchoring housing and in-

frastructure baselines to verifiable, open-access satellite telemetry, public agencies can streamline administrative workflows, minimize bureaucratic friction, and enhance data transparency, thereby mitigating risks of data manipulation or corruption during local census execution.

Consequently, integrating machine-learning workflows into national statistical protocols provides a scalable, secure, and cost-effective paradigm for modernizing population and housing updates in developing and conflict-affected nations.

Acknowledgements

I would like to express my sincere gratitude to Professors Fernando Jiménez, Guillermo Jiménez, and Rocío Sierra from the Faculty of Engineering at Los Andes University, as well as the Chilean team for their active engagement and valuable contributions, whose unwavering support, trust, and guidance, which made the execution of this research project possible.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

- [1] Lozano-Garzon, C., Bravo-Cordoba, G., Castro, H., Gonzalez-Rodriguez, G., Nino, D., Nunez, H., *et al.* (2022) Remote Sensing and Machine Learning Modeling to Support the Identification of Sugarcane Crops. *IEEE Access*, **10**, 17542-17555. <https://doi.org/10.1109/access.2022.3148691>
- [2] Pardo, C. (2020) Metodología para el monitoreo de cultivos a partir de imagenes satelitales con Machine Learning. Thesis Magister en Ingeniería de la Información, Universidad de los Andes.
- [3] Gouveia, N., Kephart, J.L., Dronova, I., McClure, L., Granados, J.T., Betancourt, R.M., *et al.* (2021) Ambient Fine Particulate Matter in Latin American Cities: Levels, Population Exposure, and Associated Urban Factors. *Science of the Total Environment*, **772**, Article ID: 145035. <https://doi.org/10.1016/j.scitotenv.2021.145035>
- [4] Peyre, G., Lenoir, J., Karger, D.N., Gomez, M., Gonzalez, A., Broennimann, O., *et al.* (2020) The Fate of Páramo Plant Assemblages in the Sky Islands of the Northern Andes. *Journal of Vegetation Science*, **31**, 967-980. <https://doi.org/10.1111/jvs.12898>
- [5] Carroll Janer, I. and Jimenez, F. (2022) Land-Use Changes by Old Colonies Mennonites in Mexico with Sentinel 2 and Trends Earth. *European Journal of Agriculture and Food Sciences*, **4**, 17-23. <https://doi.org/10.24018/ejfood.2022.4.3.498>
- [6] European Space Agency (2023) Sentinel-2 Data Collections. Copernicus Data Space Eco-System. <https://dataspace.copernicus.eu/data-collections/copernicus-sentinel-missions/sentinel-2>
- [7] Rouse, J.W., Haas, R.H., Schell, J.A. and Deering, D.W. (1973) Monitoring Vegetation Systems in the Great Plains with ERTS. *Proceedings of the 3rd ERTS Symposium*, Washington DC, 10-14 December 1973, 309-317.
- [8] McFeeters, S.K. (1996) The Use of the Normalized Difference Water Index (NDWI) in the Delineation of Open Water Features. *International Journal of Remote Sensing*, **17**, 1425-1432. <https://doi.org/10.1080/01431169608948714>

- [9] Huete, A.R. (1988) A Soil-Adjusted Vegetation Index (SAVI). *Remote Sensing of Environment*, **25**, 295-309. [https://doi.org/10.1016/0034-4257\(88\)90106-x](https://doi.org/10.1016/0034-4257(88)90106-x)
- [10] Bannari, A., Morin, D., Bonn, F. and Huete, A.R. (1995) A Review of Vegetation Indices. *Remote Sensing Reviews*, **13**, 95-120. <https://doi.org/10.1080/02757259509532298>
- [11] Zha, Y., Gao, J. and Ni, S. (2003) Use of Normalized Difference Built-Up Index in Automatically Mapping Urban Areas from TM Imagery. *International Journal of Remote Sensing*, **24**, 583-594. <https://doi.org/10.1080/01431160304987>
- [12] Fernández, I.C. and Morales, N.S. (2019) One-Class Land-Cover Classification Using MaxEnt: The Effect of Modelling Parameterization on Classification Accuracy. *PeerJ*, **7**, e7016. <https://doi.org/10.7717/peerj.7016>
- [13] Morales, N.S. and Fernández, I.C. (2020) Land-Cover Classification Using Maxent: Can We Trust in Model Quality Metrics for Estimating Classification Accuracy? *Entropy*, **22**, Article No. 342. <https://doi.org/10.3390/e22030342>
- [14] Kabacoff, R. (2015) *R in Action*. 2nd Edition, Manning.
- [15] Crawley, M. (2013) *The R Book*. 2nd Edition, Wiley.
- [16] Departamento Administrativo Nacional de Estadística [DANE] (2019) Censo Nacional de Población y Vivienda 2018: Resultados y caracterización de la vivienda y el hogar en Colombia. DANE, Bogotá.
- [17] Instituto Nacional de Estadísticas [INE] (2018) Censo de Población y Vivienda 2017: Síntesis de resultados. Instituto Nacional de Estadísticas de Chile, Santiago.