How to cite: Huacani, W., Meza, N.P., Sanchez, D.D., Huanca, F., Calizaya, E.E., Calizaya, F.G., Huanca, R. (2022) Land Use Mapping Using Machine Learning, Apurímac-Peru Region. *2022 "Air and Water – Components of the Environment" Conference Proceedings*, Cluj-Napoca, Romania, p. 176-187, DOI: 10.24193/AWC2022_17.

LAND USE MAPPING USING MACHINE LEARNING, APURÍMAC-PERU REGION

Walquer HUACANI[1], Nelson P. MEZA[1], Darío D. SANCHEZ[1], Fernando HUANCA[2], Elmer E. CALIZAYA[3], Fredy G. CALIZAYA[4], Richar HUANCA[5]

DOI: 10.24193/AWC2022_17

ABSTRACT. The objective of the research is to develop a global land use / land cover map (LULC) of the Apurímac Region, from ESA Sentinel-2 images with a resolution of 10 m. to predict 10 soil type classes throughout the year in order to generate a representative snapshot of 2020. The methodology used in the analysis is the machine learning model, for the classification it was based on Artificial Intelligence (AI). For the processing, 6 bands of Sentinel-2 surface reflectance data were used: visible blue, green, red, near-infrared and two short-wave infrared bands, to create the final map, the model is run on multiple dates of images throughout the year on the Google Earth Engine (GEE) platform. The results of the study determine the total area is 2 111 415.29 ha, where the water represents 9 392.84 ha. (0.44%), on the other hand, snow/ice occupies 227.89 ha, representing 0.01%, while cultivated land occupies an area of 34 408.09 ha, (1.63%), bushes/shrubs occupy most of 1 740 486.69 ha, which represents 82.435% of the total area.

Keywords (4-6): Machine learning, Land cover, Google Earth Engine (GEE), Artificial Intelligence (AI)

1. INTRODUCTION

Information on land cover at a global level, as used by the scientific community, governments and international organizations among others, is essential to know environmental changes, food security, conservation, and the coordination of actions necessary to mitigate and adapt to global change. [1]–[3]. These data also play an important role in improving the performance of ecosystem, hydrology and atmosphere models [4]. This accurate and reliable information on land cover at a global level is necessary, therefore, it is urgently needed [5].

[1] Professional Academic School of Mining Engineering, Micaela Bastidas National University of Apurímac Perú; whuacani@unamba.edu.pe, nmeza@unamba.edu.pe, dsanchez@unamba.edu.pe
[2] Academic Department of Mathematics and Statistics, San Antonio Abad del Cusco National University, Perú; fernando.huanca@unsaac.edu.pe
[3] Professional Academic School of Topographic Surveying Engineering, Universidad Nacional del Altiplano Puno, Perú; ecalizaya@unap.edu.pe
[4] Faculty of Agricultural Sciences, Agronomy Engineering Universidad Nacional del Altiplano Puno, Perú; fcalizaya@unap.edu.pe
[5] E.I. Our Lady of Perpetual Help Huanta Ayacucho, Peru; richar.cardena@ayacucho.edu.pe

Soil is the basis of human livelihood and well-being, it is the main provider of food, water and many services among others, including biodiversity and the ecosystem, anthropogenic activities have caused extensive changes in land cover [6]. Changes in land cover are a determining factor and a consequence of global environmental change, and they have generated a great impact on our society and the ecosystem at the national and global level [7], [8]. The "Earth" is defined as a place where human activities take place, composed of water, soil, forests and farmland, described as Ground cover (LC) [9], while the use of the soil (LU) indicates how humans use the land [7]. Changes in LU and LC over a period of time are collectively known as land use land cover change (LULCC). Land, water and the environment have apparently been affected due to rapid growth in human population and technological advances [10]–[12]. Accurate information about LU, LC, and their changes are of great importance for decision makers and scientists to effectively plan and manage in sustainable development [6], [13].

In this study, we analyze and quantify land cover use with algorithms powered by artificial intelligence. The analysis is from global land use / land cover map (LULC) samples. Produced by a deep learning model trained with more than 5 billion Sentinel-2 pixels, they compose in a final representative map of 2020, with a detail of 10 land cover classes as follows:

1.- Water. Areas where water was predominantly present throughout the year; cannot cover areas with sporadic or ephemeral water; contains little or no sparse vegetation, no rocky outcrop or elements built like piers; Examples: rivers, ponds, lakes, oceans, flooded salt plains.

2.- Trees. Any significant grouping of tall dense vegetation (15 m or more), typically with a closed or dense canopy; examples: wooded vegetation, groups of tall and dense vegetation within savannas, plantations, swamps or mangroves.

3.- Grass. Open areas covered with homogeneous grasses with little or no taller vegetation; cereals and wild grasses without an obvious human layout (ie, without a mapped field); Examples: meadows and natural fields with little or no tree cover, open savanna with few or no trees, parks / golf courses / lawns, pastures.

4.- Flooded vegetation. Areas of any type of vegetation with evident intermixing of water during most of the year; seasonally flooded area that is a mixture of grass / shrubs / trees.

5.- Crops Cereals, grasses and crops planted in plots by humans that are not at the height of the trees; Examples: corn, wheat, soybeans, fallow patches of structured land.

2. MATERIALS AND METHODS

2.1. Study area

The study area corresponds to the Apurímac region (Fig. 1), its surface area is approximately 2 111 415.20 ha. Natural vegetation comprises a diversity of categories, ranging from high Andean grasslands, natural and planted forests, to a series of scrub associations. However, the dominance of grasslands and bushes (45% of the extension) is notorious, they are mainly found in the high areas to the south of the department, in the provinces of Aymaraes, Antabamba, Grau and Cotabambas.

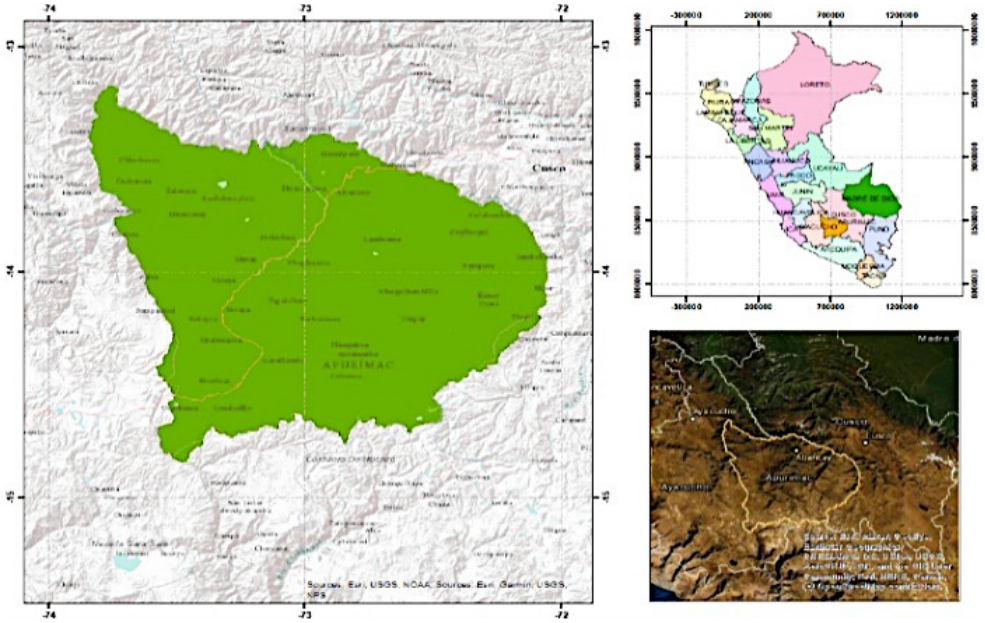

Fig. 1. Geographical location of the Apurímac Region.

2.2. Theoretical Framework

2.2.1. Impact Observatory

It is a mission-driven technology company that provides artificial intelligence (AI) - based algorithms and data-on-demand for sustainability and environmental risk analysis for governments, nonprofits, businesses and markets. Founded in 2020 in Washington, DC, the goal of Impact Observatory is to train and provide information to users for global decision making to be planetary analysts. Available at: impactobservatory.com.

2.2.2. ESRI

ESRI Company is a world market leader in Geographic Information System (GIS) software, mapping and location intelligence, founded in 1969 in Redlands, California, USA, ESRI software is implemented in more than 350,000 organizations globally and in more than 200,000 institutions in the Americas, Asia and the Pacific, Europe, Africa, and the Middle East, including Fortune 500 companies, government agencies, non-profit organizations, and universities, has regional offices, international distributors and partners providing local support in more than 100 countries on six continents. With its pioneering commitment to geospatial information technology, ESRI designs the most innovative solutions for digital transformation, the Internet of Things.

2.2.3. Google Earth Engine (GEE)

It is a cloud-based platform for planetary-scale work that features geospatial analysis capabilities and provides the computational power of Google. It provides capacities to influence a variety of high-impact social issues including deforestation,

floods, drought, natural disasters, disease and food security, water management, climate monitoring and environmental protection [14].

2.2.4. Sentinel Images

It is a multi-satellite project developed by the European Space Agency (ESA) within the framework of the Copernicus Program [15]. It is made up of five satellites with different objectives that go from terrestrial surveillance to marine observation: 1) Sentinel-1, its objective is the observation of the Earth and the oceans; 2) Sentinel-2, its main objective is Land observation, and it consists of two satellites that provide high resolution images; 3) Sentinel-3, its main objective is marine observation; 4) Sentinel-4, dedicated to air quality monitoring; 5) Sentinel-5, like its predecessor, is dedicated to air quality monitoring.

2.2.5. Land Use and Land Cover

Information based on Artificial Intelligence (AI), has a resolution of 10 m, based on Sentinel-2 satellite images, with 85% accuracy and 10 categories of land cover (land use). Leaders in governments, NGOs, finance, and industry need reliable, actionable information about the changing world to understand opportunities, identify threats, and measure the impacts of actions. (IO) Land Use & Land Cover (LULC) meets this need with up-to-date maps, uses artificial intelligence-based algorithms to create user-designated maps on demand, automatically, anywhere on Earth, from local to global scale [16].

2.2.6. Related jobs

The use of these resources is considered a very important tool in the field of research since it contributes to obtaining homogeneous and precise data for its application in various areas. Such is the case of [12] who carried out an automatic classification of land cover from satellite images using a machine learning method based on Neural Networks (CNN); obtained an accuracy of 83.52% in training and 91.02% in validation. Similarly [6], presents the update of the land cover cartography, performed the calculation of the Normalized Difference Vegetation Index (NDVI) on a mosaic of Sentinel 2B-1C satellite images in the Google Earth Engine (GEE) platform; carried out an unsupervised classification by clusters obtaining an accuracy of 83%. [17] performed an analysis of satellite images of different spatial resolutions to interpret the behavior of lagoons in a region using the Standardized Precipitation Index (IEP). [18], cited by [16], in the evaluation of the precision of the 2020 world map by the Impact Observatory, the surface estimates were adjusted for each class, arriving A 95% confidence interval for each area estimate, giving users a clearer picture of the precision and total area of each class.

2.3. Materials and Methodology

2.3.1. Materials

For the present study, we have considered data from the technology company Impact Observatory, which developed its land classification model with artificial

intelligence using a training data set of five billion image pixels tagged by humans. This model received the Sentinel-2 scene collection for classification, processing more than 400,000 Earth observations to produce the final map, available at: https://www.arcgis.com/home/item.html?id=fc92d38533d440078f17678ebc20e8e2

2.3.2. Data used

For data processing, they have direct access through the web services in ArcGIS Living Atlas, and ESRI, available on the platform of https://www.arcgis.com/apps/mapviewer/index.html?layers=d6642f8a4f6d4685a24ae2dc0c73d4ac

Fig. 2. Land Use / Global Land Cover Map (LULC).

2.3.3. Data availability

Global maps (LULC), provide information on land use, this dataset can be used to visualize land use / land cover anywhere on Earth. This application provides access to 700 individual 10-meter resolution GeoTIFF files of the 2020 Esri Land Cover map produced by Impact Observatory. The product generated in this work is available for download, it corresponds to zone 18L at: https://www.arcgis.com/apps/instant/media/index.html?appid=fc92d38533d440078f17678ebc20e8e2 , with IMAGE ID: projects / sat- io / open-datasets / landcover / ESRI_Global-LULC_10m / 18L_20200101-20210101

2.3.4. Description and characteristics

The format that contains it is GeoTIFF, with a resolution of 10 individual meters of the Esri 2020 Land Cover map produced by Impact Observatory, the data belongs to zone 18L, with a period range that varies from 2020-01-01 to 2021- 01-01, is available for viewing on the Code: https://code.earthengine.google.com/514a294747ee5e7a136372b7e947d7bc

2.3.5. Mapping with Artificial Intelligence (AI)

Impact Observatory's deep learning artificial intelligence land classification model, used a massive training data set of billions of image pixels tagged by humans,

developed by the National Geographic Society, the global map was developed at Starting Sentinel-2 image for 2020 on Microsoft Planetary Computer.

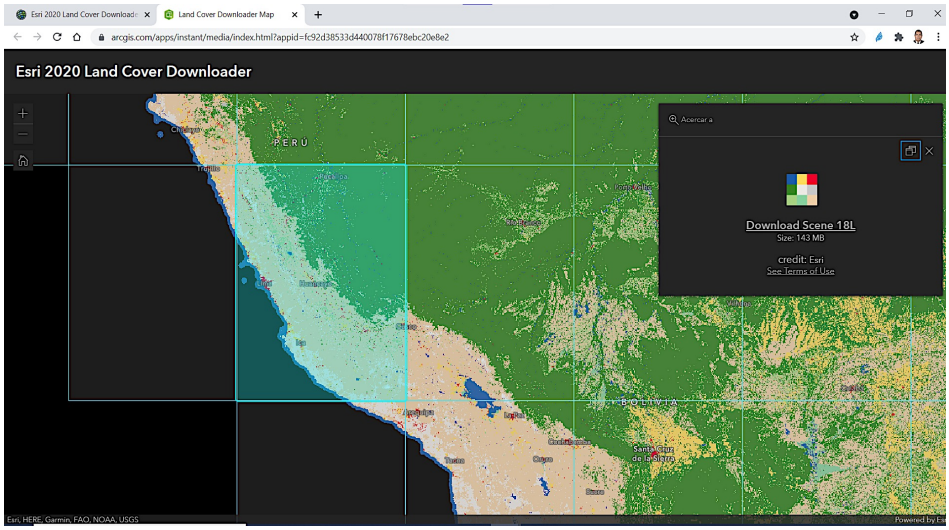

Fig. 3. ESRI 2020 Map Land Cover Downloader.

2.3.6. Learning Model

The deep learning model uses 6 bands of Sentinel-2 surface reflectance data: visible blue, green, red, near-infrared, and two short-wave infrared bands. To create the final map, the model is run on multiple dates throughout the year, and the results are compiled into a final representative map of 2020. For the process, data from the Sentinel-2 processing platform was accessed, at through Microsoft's planetary computer, produced by a deep learning model trained with more than 5 billion hand-tagged Sentinel-2 pixels, sampled at more than 20,000 sites [16].

2.3.7. Classification Process

The methodology used for the classification process is based on expert scientists to manually digitize the land use entities from satellite images, due to the size, this process took a long time, later decision tree models were used to infer land use characteristics from ancillary data; However, this method did not provide an accurate representation, spectral information from satellite images was also applied to perform a supervised classification, but this procedure could not successfully distinguish entities. However, in recent years, machine learning, a sub-discipline of artificial intelligence (AI), has progressed to the point that the use of computer vision and deep learning in analysis and classification imaging is now viable (Fig. 4).

2.3.8. Analysis of Space Variation

For the analysis of space variation, algorithms based (AI) were used, where the computer assigns a label to an image to categorize geotagged images, how water, trees, dense vegetation, bare soils, among others, have been classified, where the

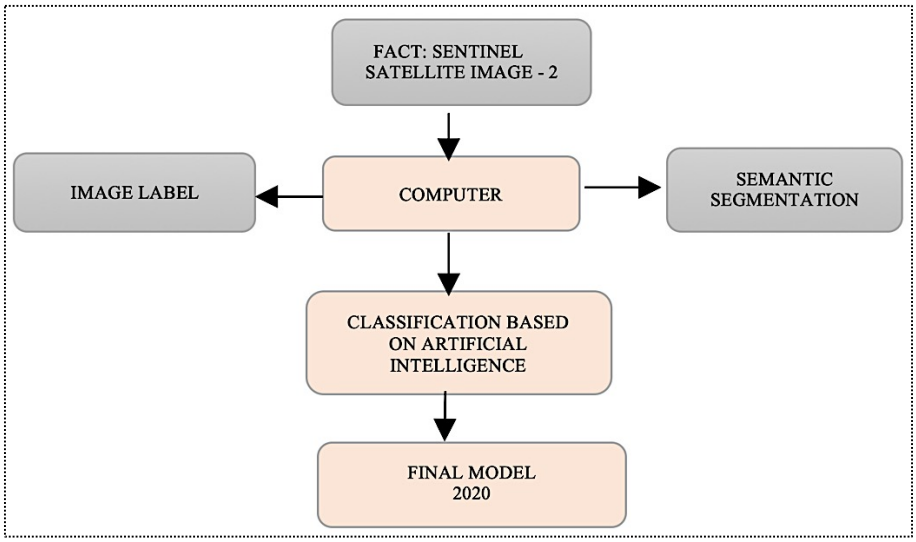

Fig. 4. Land use classification process workflow.

computer needs to find different types of segmentation within an image, as well as their location. Another important task in computer vision is semantic segmentation, in which each pixel in an image is classified as belonging to a particular class. In GIS, semantic segmentation can be used for land cover classification or to extract road networks from satellite images (Fig. 5).

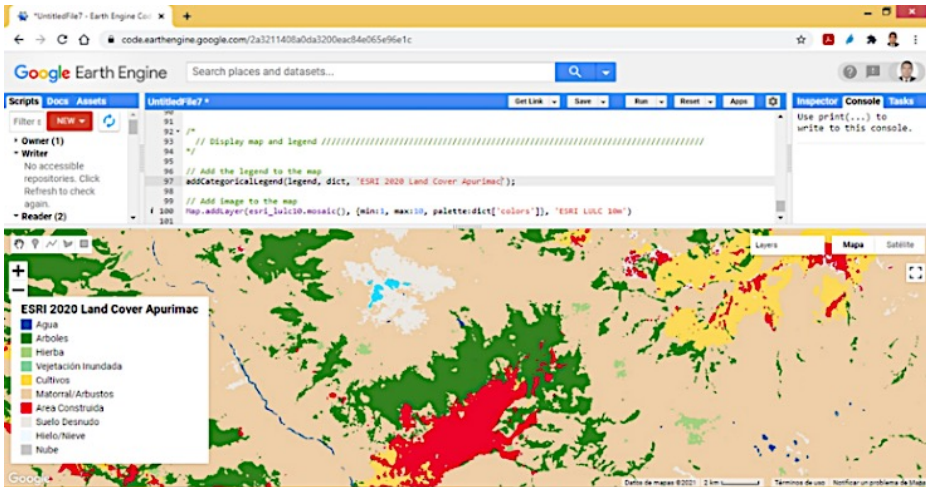

Fig. 5. Data Processing Land Cover ESRI 2020, GEE platform.

3. RESULTS AND DISCUSSION

3.1. Data validation

The validation of the methodology is a fundamental stage that should not be evaded for any reason, because it confirms and gives confidence to the information

presented by the satellite images and the certainty of the methodology [19]. The map has been generated using Sentinel-2 images, working with more than 400,000 observations to reclassify the territory into 10 classes. The tools used are highly reliable for the evaluation of land use, in order to make decisions for management and conservation (Fig. 6).

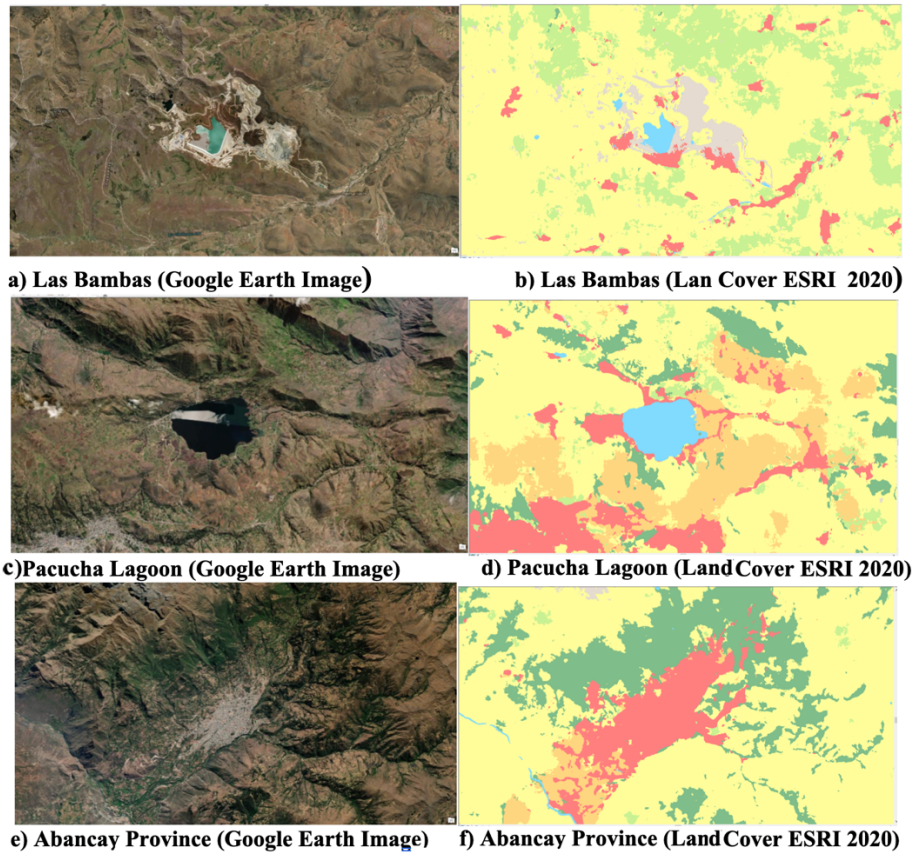

Fig. 6. Visual results of "validation" experiments.

3.2. Results

The results of mapping of soil types, applying deep learning algorithms, are summarized in Table 1, we can see that the total area is 2 111 415.29 ha, the water represents an area of 9 392.84 ha which represents 0.44%, while the cultivated land occupies an area of 34 408.09 ha, representing 1.63%, it is noted that the bushes/shrubs that are a mixture of small clusters of plants scattered in a landscape, show soil or exposed rock, cover of scarce shrubs, they occupy the largest area of 1 740 486.69 ha, which represents 82.435% of the total area. To view the code training process: https://code.earthengine.google.com/1996db9717f38659e12b7ff2ebbd13a3 (Fig. 7).

The use of natural resources must be supported by adequate territorial planning, in this sense, land cover is a key input for understanding the territorial-environmental processes and dynamics of a certain region [20].

Table 1. ***Types of land use period 2020, Apurímac Region***

Type of Land	Pixel	Resol. (10*10)	Area (ha)	%
Water	939284	100	9392.84	0.44
Trees	8072151	100	80721.51	3.82
Herb	11536132	100	115361.32	5.46
Flooded Vegatation	174399	100	1743.99	0.08
Crops	3440809	100	34408.09	1.63
Scrub/Shrubs	174048669	100	1740486.69	82.43
Build Area	2727120	100	27271.2	1.29
Bare ground	10178540	100	101785.4	4.82
Snow/Ice	22789	100	227.89	0.01
Clouds	1636	100	16.36	0.00
Total	211141529	100	2111415.29	100

Source: Satellite Images Landsat. Processing on the Google Earth Engine platform.

A remarkable growth in the number of analytical data sets (Fig. 8), with the desire and need to study large areas, has brought the concept of Big Data into the field of Earth observation in recent years. [21]; [22]. This accompanied by Google Earth Engine (GEE), which provides planetary scale calculation functions of raster and vector data, can effectively handle data in mapping LULC [23].

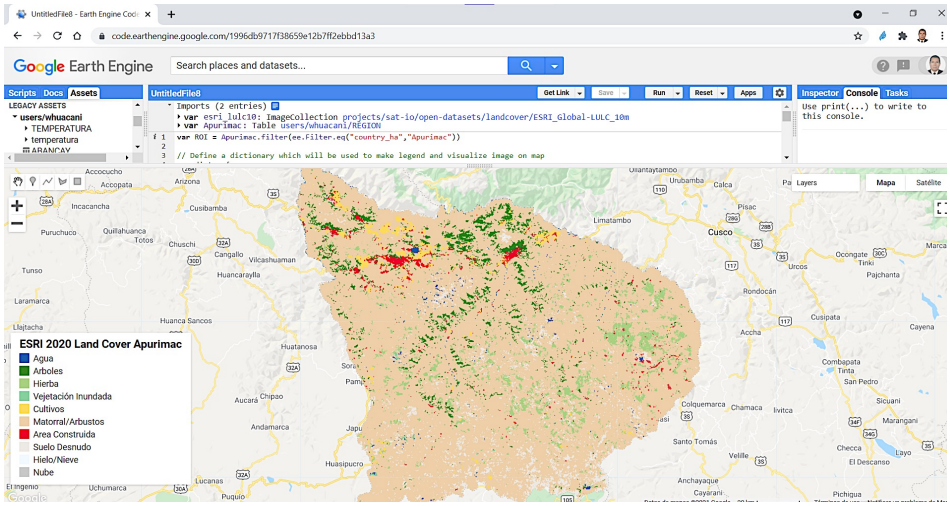

Fig. 7. Data processing results in (GEE) in high resolution coverage of land use in Apurimac Region.

As a result of this analysis of the type of land use, the majority of the occupied part of Apurímac Region is scrub/shrubs according to the land cover data.

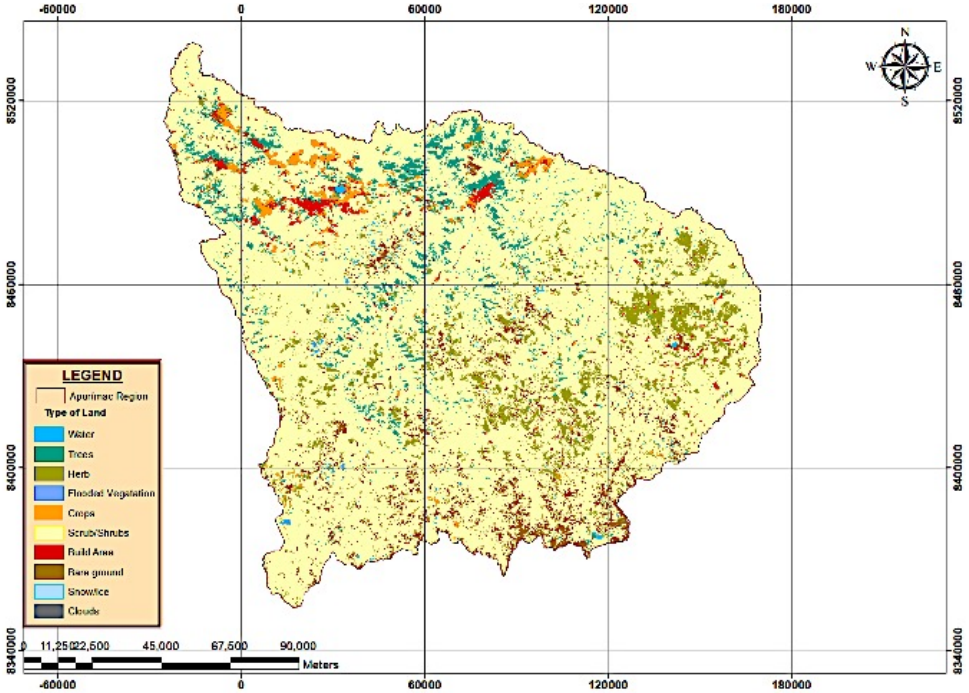

Fig. 8. Land use map of Apurímac region 2020.

5. CONCLUSIONS

In this research, the type of use of coverage of 10 classes was determined, from land cover data (LULC), applying the machine learning algorithms of Artificial Intelligence (AI) in the GEE platform, of the Apurímac region to the 2020 period.

According to the analysis, the total area is 2 111 415.29 ha, each class occupies a space, the water represents an area of 9 392.84 ha of 0.44%, on the other hand the snow/ice occupies 227.89 ha, representing 0.01% of the total, while the cultivated land occupies an area of 34 408.09 ha, which represents 1.63%, it is clearly noted that the bushes/shrubs that are a mixture of small clusters of plants or individual plants scattered in a landscape, showing exposed soil or rock, scarce shrub cover, occupying the largest area of 1 740 486.69 ha, which represents 82.435% of the total area.

Furthermore, the findings also produce high precision efficiency when applying algorithms driven by artificial intelligence (AI).

REFERENCES

1. Y. Ban, P. Gong, and C. Giri, “Global land cover mapping using Earth observation satellite data: Recent progresses and challenges,” ISPRS J. Photogramm. Remote Sens., vol. 103, no. February 2020, pp. 1–6, 2015, [CrossRef]
2. J. Chen et al., “Global land cover mapping at 30 m resolution: A POK-based operational approach,” ISPRS J. Photogramm. Remote Sens., vol. 103, pp. 7–27, 2015, [CrossRef]

3. M.-R. Rujoiu-Mare and B.-A. Mihai, “Mapping Land Cover Using Remote Sensing Data and GIS Techniques: A Case Study of Prahova Subcarpathians,” Procedia Environ. Sci., vol. 32, pp. 244–255, 2016, {CrossRef]
4. P. Gong et al., “Annual maps of global artificial impervious area (GAIA) between 1985 and 2018,” Remote Sens. Environ., vol. 236, p. 111510, Jan. 2020, [CrossRef]
5. D. Yang, C. S. Fu, A. C. Smith, and Q. Yu, “Open land-use map: a regional land-use mapping strategy for incorporating OpenStreetMap with earth observations,” Geo-Spatial Inf. Sci., vol. 20, no. 3, pp. 269–281, 2017, [CrossRef]
6. C. Morales, E. Helena, and J. Mosciaro, “Determinación De Cobertura De Suelo Con Imágenes Sentinel 2B-1C En Entorno Google Earth Engine .,” Periurbanos hacia el concenso, no. 1, pp. 2–5, 2016, [Online]. Available: https://inta.gob.ar/sites/default/files/inta_resumen_ampliado_sentinel.pdf.
7. J. A. Foley et al., “Global consequences of land use,” Science (80-.)., vol. 309, no. 5734, pp. 570–574, 2005, [CrossRef]
8. G. B. Bonan, “Forests and climate change: Forcings, feedbacks, and the climate benefits of forests,” Science (80-.)., vol. 320, no. 5882, pp. 1444–1449, 2008, [CrossRef]
9. E. F. Lambin and P. Meyfroidt, “Global land use change, economic globalization, and the looming land scarcity,” Proc. Natl. Acad. Sci. U. S. A., vol. 108, no. 9, pp. 3465–3472, 2011, [CrossRef]
10. P. K. Langat, L. Kumar, and R. Koech, “Understanding water and land use within Tana and Athi River Basins in Kenya: opportunities for improvement,” Sustain. Water Resour. Manag., vol. 5, no. 3, pp. 977–987, 2019, [CrossRef]
11. M. Li, S. Zang, B. Zhang, S. Li, and C. Wu, “A review of remote sensing image classification techniques: The role of Spatio-contextual information,” Eur. J. Remote Sens., vol. 47, no. 1, pp. 389–411, 2014, [CrossRef]
A. S. Suárez, A. F. Jiménez, M. Castro, and A. Cruz, “Clasificación y mapeo automático de coberturas del suelo en imágenes satelitales utilizando Redes Neuronales Convolucionales,” Orinoquia, vol. 21, no. 1, pp. 64–75, 2017.
12. M. Feng and X. Li, “Land cover mapping toward finer scales,” Sci. Bull., vol. 65, no. 19, pp. 1604–1606, 2020, [CrossRef]
13. N. Gorelick, M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, and R. Moore, “Google Earth Engine: Planetary-scale geospatial analysis for everyone,” Remote Sens. Environ., vol. 202, pp. 18–27, 2017, [CrossRef]
14. F. Pech May, J. V. Sánchez Hernández, and H. Sánchez Jacinto, “Análisis de zonas de cultivo y cuerpos de agua mediante el cálculo de índices radiométricos con imágenes Sentinel-2,” Lámpsakos, no. 24, p. 48, 2021, [CrossRef]
15. et al Karra, Kontgis, “ESRI 2020 Global Land Use and Land Cover layer,” IGARSS 2021-2021 IEEE Int. Geosci. Remote Sens. Symp. IEEE, 2021, p. 2021, 2020, [Online]. Available: https://www.arcgis.com/home/item.html?id=d6642f8a4f6d4685a24ae2dc0c73d4ac.
16. V. S. Aliaga, F. Ferrelli, V. Y. Bohn, and M. C. Piccolo, “Utilización de imágenes satelitales para comprender la dinámica lagunar en la Región Pampeana,” Rev. Teledetección, no. 46, p. 133, 2016, [CrossRef]
17. P. Olofsson, G. M. Foody, S. V. Stehman, and C. E. Woodcock, “Making better use of accuracy data in land change studies: Estimating accuracy and area and quantifying uncertainty using stratified estimation,” Remote Sens. Environ., vol. 129, pp. 122–131, Feb. 2013, [CrossRef]
18. G. E. C.-R. y F. E. F.-U. Karen V. Suárez-Parra, “Validación de la metodología Corine Land Cover (CLC) para determinación espacio-temporal de coberturas: caso

microcuenca de la quebrada Mecha (Cómbita, Boyacá), Colombia," http://repository.humboldt.org.co/bitstream/handle/20.500.11761/9467/01%20Validaci%c3%b3n-Corine%20Land%20Cover_Su%c3%a1rez%20et%20al%202016.pdf?sequence=3&isAllowed=y, 2016, DOI: https://doi.org/http://dx.[CrossRef]
19. D. Vargas, P. A. Bernal, J. Leal Villamil, and M. A. Quimbayo C, "Cobertura del suelo bajo metodología Corine Land Cover para el Bosque de Galilea y su área de influencia, Tolima (Colombia)," IX Jornadas Educ. en Percepción Remota y SIG para Centroamérica y el Caribe, no. 15, p. 34, 2019, [CrossRef]
20. N. X. Bian J., Li A., Lei G., Zhang Z., "Global high-resolution mountain green cover index mapping based on Landsat images and Google Earth Engine," ISPRS J. Photogramm. Remote Sensing, Vol. 162, April 2020, 2020, doi: https://[CrossRef]
21. H. Guo, "Big Earth data: A new frontier in Earth and information sciences," Big Earth Data, vol. 1, no. 1–2, pp. 4–20, 2017, [CrossRef]
22. J. Delegido, J. Verrelst, L. Alonso, and J. Moreno, "Evaluation of sentinel-2 red-edge bands for empirical estimation of green LAI and chlorophyll content," Sensors, vol. 11, no. 7, pp. 7063–7081, 2011, [CrossRef]