



**Clasificación de uso del suelo y vegetación con redes neuronales convolucionales**  
**Classification of land use and vegetation with convolutional neural networks**

Rodolfo Montiel González<sup>1</sup>, Martín Alejandro Bolaños González<sup>1\*</sup>, Antonia Macedo Cruz<sup>1</sup>, Agustín Rodríguez González<sup>2</sup>, Adolfo López Pérez<sup>1</sup>

Fecha de recepción/Reception date: 11 de abril de 2022

Fecha de aceptación/Acceptance date: 28 de septiembre de 2022

<sup>1</sup>Colegio de Postgraduados. Campus Montecillo. México.

<sup>2</sup>Hidráulica y Agricultura Consultores S.A. México.

\*Autor para correspondencia; correo-e: bolanos@colpos.mx, martinb72@gmail.com

\*Corresponding author; e-mail: bolanos@colpos.mx, martinb72@gmail.com

**Abstract:**

The classification of land use and vegetation is a complex exercise difficult to perform with traditional methods, thus deep learning models constitute a viable alternative because they are highly capable of learning this complex semantics, a trait which allows their application in the automatic identification of land use and vegetation, based on spatiotemporal patterns derived from their appearance. The objective of this study was to propose and evaluate a deep learning convolutional neural network model for the classification of 22 different land covers and land use classes located in the *Atoyac-Salado* basin. The proposed model was trained using digital data captured in 2021 by the Sentinel-2 satellite; a different combination of hyperparameters was applied in which the accuracy of the model depends on the optimizer, the activation function, the filter size, the learning rate and the batch size. The results provided an accuracy of 84.57 % for the data set. A regularization method called Dropout was used to reduce overadjustment, with great effectiveness. It was proven with sufficient accuracy that deep learning with convolutional neural networks identifies patterns in the reflectance data captured by Sentinel-2 satellite images for land use and vegetation classification in intrinsically difficult areas of the *Atoyac-Salado* basin.

**Key Words:** Machine learning, automatic classification, *Atoyac-Salado* basin, Sentinel-2 images, artificial intelligence, remote sensing.

**Resumen**

La clasificación de uso del suelo y vegetación es un ejercicio complejo y difícil de realizar con métodos tradicionales, por lo que los modelos de aprendizaje profundo son una alternativa para su aplicación debido a que son altamente capaces de aprender esta semántica compleja, lo que hace plausible su aplicación en la identificación automática de usos del suelo y vegetación a partir de patrones espacio-temporales extraídos de su apariencia. El objetivo del presente estudio fue proponer y evaluar un modelo de red neuronal convolucional de aprendizaje profundo para la clasificación de 22 clases distintas de cobertura y uso del suelo ubicadas en la cuenca río Atoyac-Salado. El modelo propuesto se entrenó utilizando datos digitales capturados en 2021 por el satélite *Sentinel-2*; se aplicó una combinación diferente de hiperparámetros en la cual la precisión del modelo depende del optimizador, la función de activación, el tamaño del filtro, la tasa de aprendizaje y el tamaño del lote. Los resultados proporcionaron una precisión de 84.57 % para el conjunto de datos. Para reducir el

sobreajuste se empleó el método de regularización denominado *Dropout*, que resultó ser muy eficaz. Se comprobó con suficiente precisión que el aprendizaje profundo con redes neuronales convolucionales identifica patrones en los datos de la reflectancia captada por las imágenes del satélite *Sentinel-2* para la clasificación el uso de suelo y vegetación en áreas con una dificultad intrínseca en la cuenca del río Atoyac-Salado.

**Palabras clave:** Aprendizaje de máquina, clasificación automática, cuenca Atoyac-Salado, imágenes *Sentinel-2*, inteligencia artificial, sensores remotos.

## Introduction

Geographic information on Land Use and Vegetation (LUV) is an important input to support spatiotemporal studies of the behavior of plant communities present in the country. Thus, it contributes to the knowledge of the state of land cover (Inegi, 2017), which is essential for researchers and decision makers. The USV maps serve as a basis for the deduction of scenarios on the loss of natural capital or biodiversity, the generation of models of potential effects of global change and for the formulation of land use planning strategies (Mas *et al.*, 2009).

In Mexico, according to the LUV maps of the National Institute of Statistics and Geography (*Instituto Nacional de Estadística y Geografía*, Inegi), the average rate of land use change in forests and jungles during the 1992–2016 period was  $-133\ 000\ \text{ha/yr}^{-1}$ , with an evident decrease in the areas of primary vegetation and an increase in land uses associated with agricultural activities (irrigated agriculture, rainfed agriculture, induced pasture and cultivated pasture, mainly). However, the change rate stabilized in the last few years of analysis, from 2010 to 2016 (Paz-Pellat *et al.*, 2019).

Land cover change involves the modification of certain surface characteristics, such as the type of vegetation; whereas, land use change consists of an alteration in the way humans use or manage a certain area of the Earth (Patel *et al.*, 2019). Land

cover change has numerous ecological, physical and socioeconomic consequences (Pellikka *et al.*, 2013). Despite its importance, it is generally identified through expert classification, including visual interpretation of satellite images, which is costly, time-consuming and inaccurate. The implementation of computational methods allows for automatic, fast, accurate and cost-effective land cover classification with satellite imagery (Suárez *et al.*, 2017). Thus, remote sensing of land cover and land use change has the advantage of offering automated and repeatable large-scale methods for monitoring indicators of vegetation condition (Lawley *et al.*, 2015).

In recent years, there has been an increased interest in and need for reliable and updated land use and land cover information (Borràs *et al.*, 2017). Deriving land cover from remotely sensed data is essential for mapping, in addition to providing basic information to support scientific activities, since satellite images are freely and openly accessible and have greater storage and computational power (Hermosilla *et al.*, 2022). However, detailed classification is a strenuous task due to the unlimited amount of remotely sensed data, the complexity of species patterns and spatial compositions, and the lack of suitable approaches (Xie *et al.*, 2019).

This problem calls for the use of new techniques such as artificial intelligence, which centers on the research of multiple concepts that revolve around the imitation of the functions that humans perform (Ponce *et al.*, 2014). In this regard, machine learning stands out as a common tool for drawing information from large data sets (Shalev-Shwartz y Ben-David, 2014), suggesting the use of a machine or computer to learn in a manner analogous to the way in which the brain learns and predicts to automate operations in order to reduce human intervention in the automatic detection of meaningful pattern data (Theodoridis, 2015).

Deep learning is one of the most versatile modern techniques for feature extraction and classification (Bhosle and Musande, 2019), furthermore, it intelligently analyzes data on a large scale (Sarker, 2021). Deep learning algorithms extract complex

high-level abstractions (Najafabadi *et al.*, 2015), two types of algorithms—supervised and unsupervised— can be distinguished according to the data entry method utilized. Supervised learning is performed with known data (training data) of the class to be identified (Suárez *et al.*, 2017), whereas in unsupervised learning, no knowledge of the classes to be determined is required (Pérez y Arco, 2016). The input to a learning algorithm is training data and the output usually takes the form of another computer software that can perform a certain task (S Shalev-Shwartz y Ben-David, 2014).

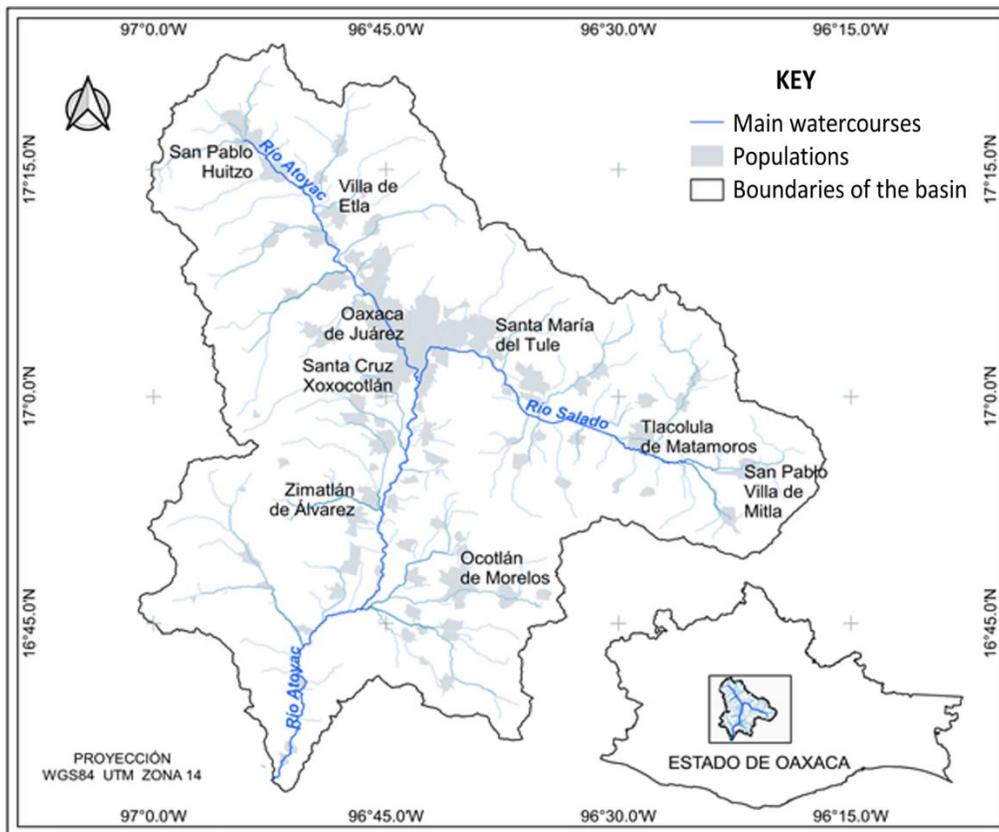
In the field of machine learning, convolutional neural networks (CNNs) have made considerable improvements and have aroused great interest in the academic and industrial communities (Krizhevsky *et al.*, 2017), because they use local connections to efficiently extract spatial information and shared weights (Chen *et al.*, 2016). CNN can extract more effective features with the help of class-specific information (Chen *et al.*, 2016). This requires large training data sets, and for multi-class issues, the data must be balanced (Suárez *et al.*, 2017).

There are several methods for classifying images, but not all are applicable to land cover classification (Macedo-Cruz *et al.*, 2010). Therefore, and in order to assess the accuracy with which deep learning can utilize convolutional neural networks to identify patterns in the classification of land use and vegetation based on reflectance data captured by remote sensors on board satellite platforms, we proposed carrying out the study in the conditions of the *Atoyac-Salado* basin. This is because its diversity of ecosystems and productive systems, urban development, orography and in particular the great diversity of LUV classes that converge in it make it a suitable and challenging area for applying classification methods supervised with artificial intelligence.

The objective was to propose and evaluate the performance of a computational method based on convolutional neural networks for the supervised classification of 22 different classes of LUV in the *Atoyac-Salado* basin in the state of *Oaxaca*.

## **Materials and Methods**

The *Atoyac-Salado* basin is located in the central part of the state of *Oaxaca* (Figure 1), between the parallels 16°49'25.86" and 17°11'34.09" N and the meridians 96°17'23.60" and 96°43 41.66" W. It extends from the source of the *Salado* river to the *Oaxaca* hydrometric station. This source is located in *San Francisco Telixtlahuaca*, where it bears the name of the *Nariz* river, at an altitude of approximately 2 418 masl. South of *San Pablo Huitzo*, it is called the *Atoyac*-river, and crosses the city of *Oaxaca de Juárez* up to the *Oaxaca* hydrometric station, at an altitude of approximately 1 500 m above sea level (Semarnat, 2017).



**Figure 1.** Location of the *Atoyac-Salado* basin and main watercourses.

The delimitation of the *Atoyac-Salado* basin was carried out in ArcSWAT™ (2012.10\_4.21) as an extension of the ArcGIS™ (14.4.1) software, from Inegi's high resolution digital elevation model LiDAR, 15 m resolution, Universal Transverse Mercator (UTM) Zone 14 projection. The outlet of the watershed is located at the *Paso Ancho* hydrometric station.

The units of analysis corresponded to the different land covers and land uses of the Series VI vector dataset at a scale of 1:250 000 (Inegi, 2017). Of the 22 classes of LUV (Table 1), two stand out for having the largest surface area: annual rainfed agriculture with 21.41 % of the total surface area, and secondary shrubby oak forest vegetation with 17.78 %. Three types of agricultural land were registered: rainfed, irrigated and humid, which were divided into annual, semi-permanent and

permanent, according to their duration. Based on this land use variability, the *Atoyac-Salado* basin was found to be suitable for the application of supervised classification methods with artificial intelligence.

**Table 1.** Assignment of class and key by type of LUV.

<b>Class</b>	<b>Code</b>	<b>Type of land use and vegetation</b>	<b>Surface area (ha)</b>
0	AH	Built urban	21 690.4
1	BP	Pine forest	1 384.1
2	BPQ	Pine-oak forest	10 157.7
3	BQP	Oak-pine forest	630.6
4	HS	Semi-permanent moisture agriculture	358.6
5	PI	Induced pastureland	41 935.9
6	RA	Annual irrigated agriculture	2 915.4
7	RAS	Annual and semi-permanent irrigated farming	3 8362
8	RS	Semi-permanent irrigated agriculture	1 032.2
9	TA	Annual rainfed agriculture	79 647.6
10	TAP	Annual rainfed and permanent agriculture	15 552.3
11	VSa/BP	Secondary shrub vegetation of pine forest	9 473.3
12	VSa/BPQ	Secondary shrub vegetation of pine-oak forest	4 378.7
13	VSa/BQ	Secondary shrub vegetation of oak forest	66 145.4
14	VSa/BQP	Secondary shrub vegetation of oak-pine forest	11 682.1
15	VSa/MK	Secondary shrub vegetation of mesquite forest	772.8
16	VSa/SBC	Secondary shrub vegetation of low deciduous forest	8 138.4
17	VSA/BP	Secondary arboreal vegetation of pine forest	10 569.2
18	VSA/BPQ	Secondary arboreal vegetation of pine-oak forest	20 728.7
19	VSA/BQ	Secondary arboreal vegetation of oak forest	19 473.6
20	VSA/BQP	Secondary arboreal of oak-pine forest vegetation	6 210.4
21	VSh/BQ	Secondary herbaceous vegetation of oak forest	726.5
Total			372 068.3

## Satellite imagery

The Copernicus Sentinel-2 mission consists of two identical satellites (2A and 2B) in the same orbit, developed by the European Space Agency (ESA, [https://www.esa.int/Applications/Observing\\_the\\_Earth/Copernicus/Sentinel-2](https://www.esa.int/Applications/Observing_the_Earth/Copernicus/Sentinel-2)). Equipped with an optical sensor, the multispectral instrument has a spatial resolution ranging from 10 to 60 m depending on the spectral band (Drusch *et al.*, 2012), with 13 bands in the visible, near-infrared and short-wave infrared ranges of the electromagnetic spectrum, and with a revisit time of 5 days at the equator (Gascon *et al.*, 2017).

The images used corresponded to Tiles T14QQD and T14QQE and the RGB and NIR spectral bands of multitemporal scenes with a 10 m spatial resolution, acquired on April 13, 2021, and May 3, 2021, respectively. Both images were captured by the Sentinel-2A satellite, with a 2-A processing level. Scenes with little or no cloud or haze were selected and downloaded from the Copernicus Open Access Hub website (<https://scihub.copernicus.eu/>). The sampling unit consisted of 20×20 pixel clipping. The sampling method was stratified random (Congalton and Green, 2009): prior knowledge of the study area from field trips made it possible to divide the area into groups or strata, which were randomly sampled.

## **Training samples**

QGIS (3.18.3) was used to extract training samples and delimit the study area. The size of the images, in *.tiff* format, was 20×20×4 pixels (height, width and number of bands), at least 80 % of the clipping were considered to belong to a single class.

A balanced training set was used to prevent classification based on unbalanced data (Gnip *et al.*, 2021). 6 000 training samples were extracted from each LUV class; however, in order to prevent oversampling, only 2 280 samples were drawn from the Semi-permanent cycle moisture agriculture class, and 4 356 samples from the Semi-permanent cycle irrigated agriculture, as these two classes cover a smaller surface area. In total, 126 636 training samples were generated.

## **CNN Model**

The CNN model algorithm was programmed with the Python language in a Jupyter notebook development environment, using the Tensorflow and Keras open-source libraries for machine learning. The model applied was of the Sequential type, the network layers were ordered and stacked linearly (Xie *et al.*, 2020). All the neurons in one layer connect to all the neurons in the next layer, based on sequences of three types of layers: convolutional, clustering and fully connected. The convolutional and fully connected layers are typically followed by a nonlinear activation function (Rousset *et al.*, 2021).

The model architecture consisted of three convolutional and three clustering layers, according to the suggestions of Chen *et al.* (2016), in order to balance the complexity and robustness of the network. 128 neurons, a Kernel size of 3 by 3, and the same padding were used in each layer. Zeros were added around the input images; the outputs of the layer had the same spatial dimensions as its inputs. Also utilized were an activation function called Rectified Linear Unit (ReLU), which returns 0 for each negative value in the input image and returns the same value for each positive value, and, subsequently, an Average-Pooling subsampling filter that considers the average

activation values of a window, plus a Dropout layer with 20 % potentiality of setting the inputs to zero.

Next, a (Flatten) layer that flattens the multidimensional outputs of the last convolutional layer in a one-dimensional format, and two dense layers: one of 512 hidden neurons with a ReLU activation and Dropout layer at 20 %, and the last one with 22 output neurons corresponding to the number of classes to be identified, with a softmax activation function to predict the potentiality of each class.

Training is the process of making inputs produce the desired outputs. It operates based on the establishment of previously known weights (Vinet and Zhedanov, 2011). For the adjustment of the connection weights, the data set was divided into two groups: training (80 %) and test (20 %). The former data, in turn, were divided into training (80 %) and evaluation (20 %), and were entered several times in the network, each repetition was called an epoch. The model was trained with 100 epochs.

During the learning phase, a transfer function was applied through a series of iterations in order to compare the predicted values with the observed values (Bocco *et al.*, 2007). The test set is not reviewed by the model in training but is used later, after adjusting the hyperparameters in order to provide an unbiased evaluation of the final model. Once the epochs have been created and the weights have been adjusted, the validation data are entered. The training ends when a low error is reached for all learning patterns (Bocco *et al.*, 2007). The following hyperparameters were used in the evaluation and testing: Kernel size, dropout rate, hidden layers, layer depth, and activation functions.

The model compilation included three parameters: optimizer, loss and metrics. Adam was used as the optimizing algorithm, as it is computationally efficient, has low memory requirements, is invariant to diagonal gradient scale change and is suitable for large problems in terms of data or parameters (Kingma and Ba, 2014). The model was compiled with the categorical cross-entropy function, and the

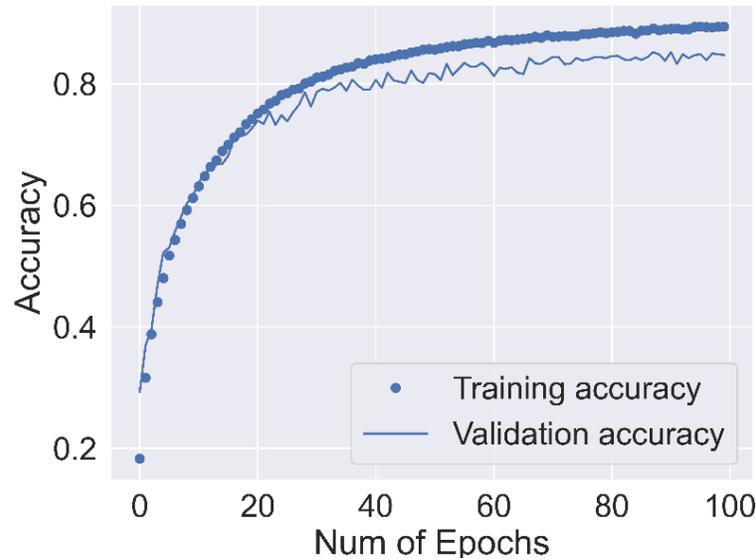
performance metric of interest was accuracy, which is related to the correctly predicted observation and to the total observations.

The classification was evaluated using the confusion matrix, as it summarizes the accuracy assessment and represents good practice (Olofsson *et al.*, 2014). The double-entry matrix confronts the actual values with the results of the classification, making it easy to detect where the two classes are being confused. Elements on the diagonal correspond to the correct prediction and those outside the diagonal correspond to incorrect predictions, both horizontally and vertically (Yeturu, 2020). The proportion of correctly assigned points expresses reliability (Mas *et al.*, 2003). In addition, other evaluation metrics were calculated: accuracy, sensitivity, and score.

In addition, the performance of the model was analyzed with the variations of its sensitivity and specificity using the receiver operating characteristic (*ROC*) curve, a parameter for assessing the goodness of the test. The accuracy of the test increases as the curve moves from the diagonal towards the upper left vertex. A higher value indicates that the model is capable of achieving a better performance (Liu *et al.*, 2022).

## **Results and Discussion**

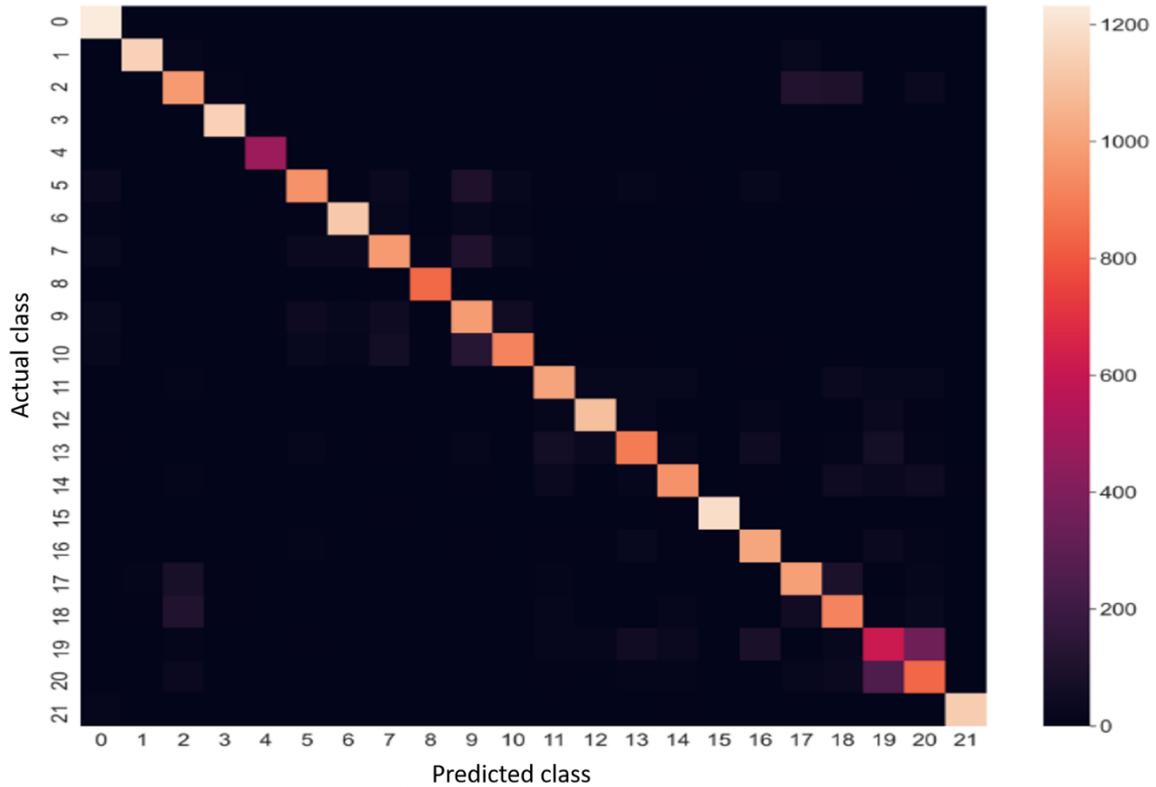
The training and validation data sets were used to provide an unbiased evaluation of the trained model, with hyperparameter tuning to obtain the best performance of the developed neural network model (Figure 2).



**Figure 2.** Model accuracy results.

The ratio between the total number of correctly identified entries and the total number of entries determined the overall classification accuracy, which reached a maximum of 89.44 % on training data and 84.57 % in validation over 100 epochs.

The results of the image classification were evaluated using the confusion matrix. Figure 3 shows in lighter color the classes with higher classification accuracy and the class for which the entered class was mistaken. In this case, the classes most often confused by the network were those labeled as 19 and 20, which corresponded to arboreal secondary vegetation of oak forest and arboreal secondary vegetation of oak-pine forest, because they are similar natural vegetation ecosystems, with a predominance of arboreal life forms. The floristic component differs partially between these systems, which may explain the confusion between the two classes.

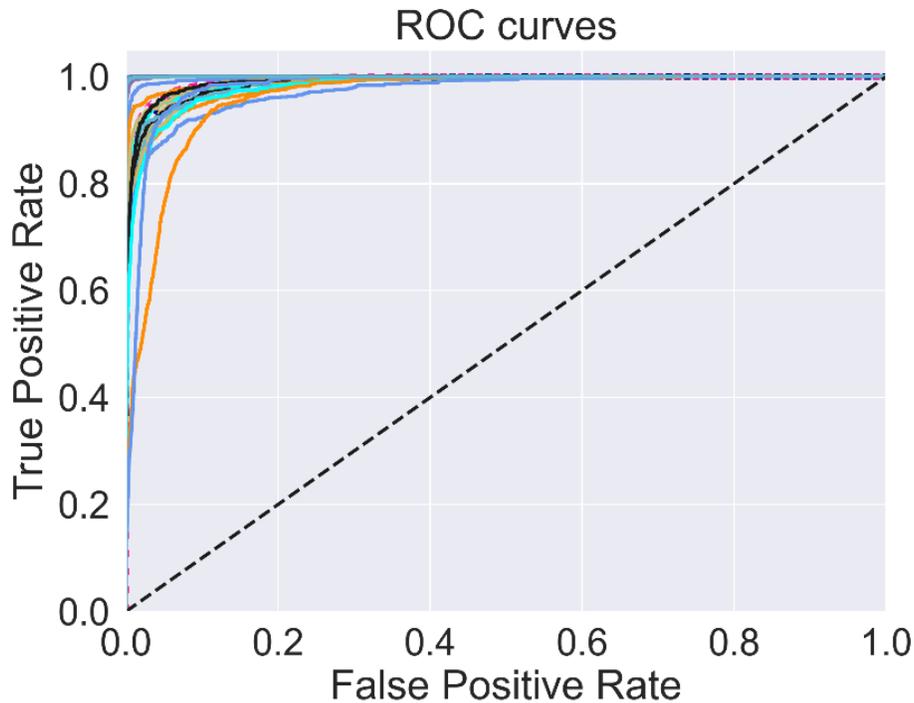


**Figure 3.** Confusion matrix.

According to the results for the evaluation metrics (Table 2 and Figure 4), a set of (macro and weighted) mean scores and accuracy with overall estimated performance of 85 % for all metrics is appreciated. Therefore, the land use and vegetation classification model was considered robust. These results indicate that the model has a low dispersion of the set of values obtained, with 85 % of positive cases that were correctly identified by the algorithm.

**Table 2.** Model assessment metrics.

<b>Class</b>	<b>Accuracy</b>	<b>Sensitivity</b>	<b>F1 Score</b>
0	0.89	1	0.94
1	0.97	0.96	0.97
2	0.78	0.79	0.79
3	0.97	0.99	0.98
4	0.98	0.98	0.98
5	0.86	0.77	0.82
6	0.93	0.93	0.93
7	0.83	0.79	0.81
8	0.98	0.99	0.99
9	0.72	0.82	0.77
10	0.88	0.77	0.82
11	0.84	0.85	0.85
12	0.91	0.90	0.91
13	0.82	0.74	0.78
14	0.86	0.81	0.83
15	0.99	0.98	0.99
16	0.84	0.89	0.86
17	0.81	0.82	0.82
18	0.72	0.78	0.75
19	0.56	0.51	0.53
20	0.62	0.69	0.65
21	0.99	0.98	0.98
Accuracy			0.85
Medium macro	0.85	0.85	0.85
Weighted mean	0.85	0.85	0.85



**Figure 4.** ROC curve by identified class.

During the training, the network experienced a positive change when a regulation layer (Dropout) was used with a 20 % chance of setting the inputs to zero, which allowed the model to fit the data while minimizing the error produced by the data at each epoch. Whereas, when not used, there was a point at which error increased and generated overtraining.

This research used 22 different classes, when usually approximately 10 are utilized. For example, Suárez *et al.* (2017) used four classes with 91.02 % of accuracy, Hu *et al.* (2018) employed seven classes with 82 % in accuracy, Bhosle and Musande (2019) classified 16 and four classes, respectively, with accuracies of 97.58 and 79.43 %.

The results obtained for performance were high, exhibited accurate answers and showed progress for the procedure performed with CNN in the automated classification of LUV with 22 classes, although the scale utilized for creating the series had the problem of generating large polygons of LUV classes not representative of

the local scale (Paz-Pellat *et al.*, 2019). The results of the present work were better than those of other previous CNN classification studies, in which an accuracy of 83.27 % in training and of 91.02 % in validation was registered for identifying four classes (Suárez *et al.*, 2017), and with accuracies of 90.18 % for vegetation cover classification and 87.92 % for land use in 12 classes (Zhang *et al.*, 2019).

The proposed model yielded satisfactory results on a very challenging dataset, even with the use of supervised learning alone. Once the data set was trained, the network experienced substantial overfitting when Dropout was omitted; however, no overfitting was reported when Dropout was added (Srivastava *et al.*, 2014).

It should be noted that the performance of the network becomes degraded when any of the intermediate layers is removed (Krizhevsky *et al.*, 2017), involving a loss of about 5 % when a single convolutional layer is removed. The depth setting of the CNN network is critical to the accuracy of the classification, as the quality of the learned features is influenced by the levels of representations and abstractions (Zhang *et al.*, 2019).

The results show the suitability of CNNs to classify LUVs in complex areas; however, their accuracy may vary as the number of classes increases, as in the case of Inegi's LUV maps, which consider 70 classes with 15 groupings (Paz-Pellat *et al.*, 2019). Thus, it would probably have to be grouped into spectrally similar classes in order to operationalize a deep learning classification scheme. Today, many other deep learning options with more complex architectures, could allow further advance in future research. In addition, the use of reflectance information exclusively from independent spectral bands can be limiting, therefore, we suggest adding layers of vegetation indices.

## **Conclusions**

The model correctly detects those classes that are furthest separated in spectral terms and that exhibit differential characteristics. Classes with less training data are not affected, although spectrally close classes register low recognition rates. Results improve as the network increases in number of layers and training time, but there are still orders of magnitude that must be overcome in order to increase classification accuracy.

It was proven with sufficient accuracy that deep learning with convolutional neural networks can identify patterns in the reflectance data captured by Sentinel-2 satellite images for land use and vegetation classification in intrinsically difficult areas in the *Atoyac-Salado* basin.

### **Acknowledgments**

The authors wish to express their gratitude to the College of Postgraduates (*Colegio de Postgraduados*) for providing the necessary resources to carry out this research, and to the National Council of Science and Technology (*Conacyt*) for the scholarship granted to Rodolfo Montiel González for his master's degree in science.

### **Conflict of interests**

The authors declare that they have no competing interests.

### **Contribution by author**

Rodolfo Montiel González: field work, code programming and writing of the manuscript; Martín Alejandro Bolaños González: conceptualization and elaboration of the manuscript; Antonia Macedo Cruz: code review, and revision and correction of the manuscript; Agustín Rodríguez González: general revision and correction of the manuscript; Adolfo López Pérez: general revision and correction of the manuscript.

## References

- Bhosle, K. and V. Musande. 2019. Evaluation of deep learning CNN Model for Land Use Land Cover classification and crop identification using hyperspectral remote sensing images. *Journal of the Indian Society of Remote Sensing* 47(11):1949–1958. Doi: 10.1007/s12524-019-01041-2.
- Bocco, M., G. Ovando, S. Sayago and E. Willington. 2007. Neural network model for land cover classification from satellite images. *Agricultura Técnica* 67(4):414–421. Doi: 10.4067/S0365-28072007000400009.
- Borràs, J., J. Delegido, A. Pezzola, M. Pereira, G. Morassi y G. Camps-Valls. 2017. Clasificación de usos del suelo a partir de imágenes Sentinel-2. *Revista de Teledetección* 48:55-66. Doi: 10.4995/raet.2017.7133.
- Chen, Y., H. Jiang, C. Li, X. Jia and P. Ghamisi. 2016. Deep feature extraction and classification of hyperspectral images based on Convolutional Neural Networks. *IEEE Transactions on Geoscience and Remote Sensing* 54(10):6232–6251. Doi: 10.1109/TGRS.2016.2584107.
- Congalton, R. G. and K. Green. 2009. *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*. CRC Press Taylor & Francis Group. Boca Raton, FL, USA. 192 p.

Drusch, M., U. Del Bello, S. Carlier, O. Colin, ... and P. Bargellini. 2012. Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. *Remote Sensing of Environment* 120:25–36. Doi: 10.1016/j.rse.2011.11.026.

Gascon, F., C. Bouzinac, O. Thépaut, M. Jung, ... and V. Fernandez. 2017. Copernicus Sentinel-2A calibration and products validation status. *Remote Sensing* 9(6):584-664. Doi: 10.3390/rs9060584.

Gnip, P., L. Vokorokos and P. Drotár. 2021. Selective oversampling approach for strongly imbalanced data. *PeerJ Computer Science* 7:e604. Doi: 10.7717/peerj-cs.604.

Hermosilla, T., M. A. Wulder, J. C. White and N. C. Coops. 2022. Land cover classification in an era of big and open data: Optimizing localized implementation and training data selection to improve mapping outcomes. *Remote Sensing of Environment* 268:1-17. Doi: 10.1016/j.rse.2021.112780.

Hu, Y., Q. Zhang, Y. Zhang and H. Yan. 2018. A deep convolution neural network method for land cover mapping: A case study of Qinhuangdao, China. *Remote Sensing* 10(12):1–17. Doi: 10.3390/rs10122053.

Instituto Nacional de Geografía y Estadística (Inegi). 2017. Guía para la interpretación de cartografía: uso del suelo y vegetación, Escala 1:250 000. Serie VI. Instituto Nacional de Geografía y Estadística, Inegi. Aguascalientes, AGS, México. 200 p. [https://books.google.com.mx/books?id=LCHZDwAAQBAJ&printsec=frontcover&hl=es&source=gbs\\_ge\\_summary\\_r&cad=0#v=onepage&q&f=false](https://books.google.com.mx/books?id=LCHZDwAAQBAJ&printsec=frontcover&hl=es&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false). (20 de abril de 2022).

Kingma, D. P. and J. B. 2014. Adam: A Method for Stochastic Optimization. Cornell University. Ithaca, NY, USA. 15 p.

Krizhevsky, A., I. Sutskever and G. E. Hinton. 2017. ImageNet classification with deep Convolutional Neural Networks. *Communications of the ACM* 60(6):84-90. Doi: 10.1145/3065386.

Lawley, V., M. Lewis, K. Clarke and B. Ostendorf. 2015. Site-based and remote sensing methods for monitoring indicators of vegetation condition: An Australian review. *Ecological Indicators* 60:1273–1283. Doi: 10.1016/j.ecolind.2015.03.021.

Liu, R., X. Yang, C. Xu, L. Wei and X. Zeng. 2022. Comparative study of Convolutional Neural Network and conventional machine learning methods for landslide susceptibility mapping. *Remote Sensing* 14(2):321-351. Doi: 10.3390/rs14020321.

Macedo-Cruz, A., G. Pajares-Martinsanz y M. Santos-Peñas. 2010. Clasificación no supervisada con imágenes a color de cobertura terrestre. *Agrociencia* 44(6):711–722. <https://agrociencia-colpos.org/index.php/agrociencia/article/view/833/833>. (25 de enero de 2022).

Mas, J. F., A. Velázquez y S. Couturier. 2009. La evaluación de los cambios de cobertura/uso del suelo en la República Mexicana. *Investigación Ambiental* 1(1):23–39. [https://www.ccmss.org.mx/wp-content/uploads/2014/10/La\\_evaluacion\\_de\\_los\\_cambios\\_de\\_cobertura-uso\\_de\\_suelo\\_en\\_la\\_Republica\\_Mexicana.pdf](https://www.ccmss.org.mx/wp-content/uploads/2014/10/La_evaluacion_de_los_cambios_de_cobertura-uso_de_suelo_en_la_Republica_Mexicana.pdf). (12 de mayo de 2022).

Najafabadi, M. M., F. Villanustre, T. M. Khoshgoftaar, N. Seliya, R. Wald and E. Muharemagic. 2015. Deep learning applications and challenges in big data analytics. *Journal of Big Data* 2(1):1–21. Doi: 10.1186/s40537-014-0007-7.

Olofsson, P., G. M. Foody, M. Herold, S. V. Stehman, C. E. Woodcock and M. A. Wulder. 2014. Good practices for estimating area and assessing accuracy of land change. *Remote Sensing of Environment* 148(25):42-57. Doi: 10.1016/j.rse.2014.02.015.

Patel, S. K., P. Verma and G. S. Singh. 2019. Agricultural growth and land use land cover change in peri-urban India. *Environmental Monitoring and Assessment* 191(9):600. Doi: 10.1007/s10661-019-7736-1.

Paz-Pellat, F., V. M. Romero-Benítez, J. A. Argumedo-Espinoza, M. Bolaños-González, B. de Jong, J. C. de la Cruz-Cabrera y A. Velázquez-Rodríguez. 2019. Dinámica del uso del suelo y vegetación. In: Paz-Pellat, F., J. M. Hernández-Ayón, R. Sosa-Ávalos y A. S. Velázquez-Rodríguez. (Edits.). Estado del Ciclo del Carbono en México, Agenda Azul y Verde, Primer Reporte. Programa Mexicano del Carbono. Texcoco, Edo. Méx., México. pp. 529–572.

Pellikka, P. K. E., B. J. F. Clark, A. G. Gosa, N. Himberg, ... and M. Siljander. 2013. Agricultural expansion and its consequences in the Taita Hills, Kenya. In: Paron, P., D. Ochieng O. and C. Thine O. (eds.). Kenya: a Natural Outlook: Geo-Environmental Resources and Hazards. Developments in Earth Surface Processes 16. Elsevier. Amsterdam, AMS, The Netherlands. pp. 65–179.

Pérez V., I. C. y L. Arco G. 2016. Una revisión sobre aprendizaje no supervisado de métricas de distancia. Revista Cubana de Ciencias Informáticas 10(4):43–67. <https://www.redalyc.org/articulo.oa?id=378349316004>. (14 de diciembre de 2021).

Ponce G., J. C., A. Torres S., F. S. Quezada A., A. Silva S., ... y O. Pedreño. 2014. Inteligencia Artificial. Iniciativa Latinoamericana de Libros de Texto Abiertos (LATIn). Montevideo, MO, Uruguay. 255 p.

Rousset, G., M. Despinoy, K. Schindler and M. Mangeas. 2021. Assessment of deep learning techniques for land use land cover classification in Southern New Caledonia. Remote Sensing 13(12):1–22. Doi: 10.3390/rs13122257.

Sarker, I. H. 2021. Machine Learning: Algorithms, Real-World applications and research directions. SN Computer Science 2(3):160. Doi: 10.1007/s42979-021-00592-x.

Secretaría de Medio Ambiente y Recursos Naturales (Semarnat). 2017. Acuerdo por el que se dan a conocer los resultados del estudio técnico de las aguas nacionales superficiales en las cuencas hidrológicas Río Papagayo 1, Río Petaquillas, Río Omitlán, Río Papagayo 2, Río Papagayo 3, Río Papagayo 4, Río Nexpa 1, Río Nexpa 2, Río Quetzala, Río Infiernillo, Río Santa Catarina, Río Ometepec 1, Río Ometepec 2, Río Ometepec 3, Río Cortijos 1, Río Cortijos 2, Río Cortijos 3, Río Cortijos 4, Río Ometepec

4, Río La Arena 1, Río La Arena 2, Río La Arena 3, Río Atoyac-Salado, Río Atoyac-Tlapacoyan, Río Sordo-Yolotepec, Río Atoyac-Paso de la Reina y Río Verde, pertenecientes a la Región Hidrológica número 20 Costa Chica de Guerrero. Diario Oficial de la Federación, Cuarta Sección. [https://dof.gob.mx/nota\\_detalle.php?codigo=5496053&fecha=04/09/2017#gsc.tab=0](https://dof.gob.mx/nota_detalle.php?codigo=5496053&fecha=04/09/2017#gsc.tab=0). (8 de noviembre de 2021).

Shalev-Shwartz, S. and S. Ben-David. 2014. Understanding machine learning: From theory to algorithms. Cambridge University Press. New York, NY, USA. 449 p.

Srivastava, N., G. Hinton, A. Krizhevsky, H. Sutskever and R. Salakhutdinov. 2014. Dropout: A simple way to prevent neural networks from overfitting. Journal of Machine Learning Research 15:1929–1958. [https://www.jmlr.org/papers/volume15/srivastava14a/srivastava14a.pdf?utm\\_content=buffer79b43&utm\\_medium=social&utm\\_source=twitter.com&utm\\_campaign=buffer](https://www.jmlr.org/papers/volume15/srivastava14a/srivastava14a.pdf?utm_content=buffer79b43&utm_medium=social&utm_source=twitter.com&utm_campaign=buffer). (28 de enero de 2022).

Suárez L., A. S., A. F. Jiménez L., M. Castro-Franco y A. Cruz-Roa. 2017. Clasificación y mapeo automático de coberturas del suelo en imágenes satelitales utilizando Redes Neuronales Convolucionales. ORINOQUIA 21(1):64–75. Doi: 10.22579/20112629.432.

Theodoridis, S. 2015. Introduction. In: Theodoridis, S. Machine Learning A bayesian and optimization perspective. Elsevier. Amsterdam, AMS, Netherlands. pp. 1–8.

Vinet, L. and A. Zhedanov. 2011. A "missing" family of classical orthogonal polynomials. Journal of Physics A Mathematical and Theoretical 44(8):1-16. Doi: 10.1088/1751-8113/44/8/085201.

Xie, G., A. Shanguan, R. Fei, W. Ji, W. Ma and X. Hei. 2020. Motion trajectory prediction based on a CNN-LSTM sequential model. Science China Information Sciences 63:1-21. Doi: 10.1007/s11432-019-2761-y.

Xie, Z., Y. Chen, D. Lu, G. Li and E. Chen. 2019. Classification of land cover, forest, and tree species classes with ZiYuan-3 Multispectral and Stereo Data. *Remote Sensing* 11(2):164-190. Doi: 10.3390/rs11020164.

Yeturu, K. 2020. Machine learning algorithms, applications, and practices in data science. In: Srinivasa R., A. S. R. and C. R. Rao (eds.). *Handbook of Statistics 43 Principles and Methods for Data Science*. Elsevier. Amsterdam, AMS, Netherlands. pp. 81-206.

Zhang, C., I. Sargent, X. Pan, H. Li, ... and P. M. Atkinson. 2019. Joint deep learning for land cover and land use classification. *Remote Sensing of Environment* 221:173–187. Doi: 10.1016/j.rse.2018.11.014.



Todos los textos publicados por la **Revista Mexicana de Ciencias Forestales** –sin excepción– se distribuyen amparados bajo la licencia *Creative Commons 4.0 Atribución-No Comercial (CC BY-NC 4.0 Internacional)*, que permite a terceros utilizar lo publicado siempre que mencionen la autoría del trabajo y a la primera publicación en esta revista.