eISSN 2310-290X
Vol. 13, No. 2, 2024
https://revistas.unitec.edu/innovare
INNOVARE Revista de ciencia y tecnología

Original

Convolutional neural network for dactylological alphabet recognition of Honduran Sign Language (LESHO)

Red neuronal convolucional para el reconocimiento del alfabeto dactilológico de la Lengua de Señas Hondureña (LESHO)

Jorge Alejandro Ramos[1], Alicia María Reyes-Duke*[1], Alberto Max Carrasco-Bardales[2]

[1]Facultad de Ingeniería, Universidad Tecnológica Centroamericana, UNITEC, Honduras
[2]School of Advanced Technology, Algonquin College, Canada

Abstract / **Introduction** The critical problem of hearing and speech difficulties among thousands of Hondurans is a latent need, and in search of tools that allow the inclusion of the population with this condition is sought through the introduction of a convolutional neural network (CNN) designed for real-time detection and classification of the dactyl alphabet of the Honduran sign language (LESHO). This study represents an important step forward in promoting accessibility and inclusion of the Honduran deaf community, which faces few technological solutions adapted to their needs. **Methods** A proprietary dataset comprising more than 8,000 images with various angles and gestures was meticulously constructed, ensuring robust training and evaluation of the model. Spiral research methodology was employed to iteratively refine network performance, with an emphasis on accuracy and real-time deployment capabilities. **Results** The final model showed exceptional results during the testing, achieving a mean average precision (mAP) of 98.8%, a precision of 97.4%, and a recall of 97.7%. These metrics underscore the reliability of the CNN in recognizing both static and dynamic gestures with minimal errors. **Conclusion** The model's capacity to generalize indicates its potential for further applications, such as full sign language interpretation and expanded vocabulary training
Keywords Convolutional neural networks, Dactyl alphabet, Deep learning, Hand gestures recognition, Honduran sign language.

Resumen / **Introducción** El problema crítico de las dificultades auditivas y del habla entre miles de hondureños es una necesidad latente, y en busca de herramientas que permitan la inclusión de la población con esta condición se busca mediante la introducción de una red neuronal convolucional (CNN) diseñada para la detección y clasificación en tiempo real del alfabeto dáctilo de la lengua de señas hondureña (LESHO). Este estudio representa un importante paso adelante en la promoción de la accesibilidad y la inclusión de la comunidad sorda hondureña, que se enfrenta a escasas soluciones tecnológicas adaptadas a sus necesidades. **Métodos** Se construyó meticulosamente un conjunto de datos propio compuesto por más de 8.000 imágenes con diversos ángulos y gestos, lo que garantizó un entrenamiento y una evaluación robustos del modelo. Se empleó una metodología de investigación en espiral para perfeccionar iterativamente el rendimiento de la red, haciendo hincapié en la precisión y las capacidades de despliegue en tiempo real. **Resultados** El modelo final mostró resultados excepcionales durante la prueba, con una precisión media (mAP) del 98,8%, una precisión del 97,4% y una recuperación del 97,7%. Estas métricas subrayan la fiabilidad de la CNN a la hora de reconocer gestos estáticos y dinámicos con errores mínimos. **Conclusión** La capacidad de generalización del modelo indica su potencial para otras aplicaciones, como la interpretación completa de la lengua de signos y la formación de vocabulario ampliado.
Palabras Clave Alfabeto dáctilo, Aprendizaje profundo, Lenguaje de señas hondureño, Reconocimiento de gestos con las manos, Redes neuronales convolucionales.

Recepción: 14 diciembre 2024 / Aceptación: 30 diciembre 2024 / Publicación: 30 diciembre 2024
Cita: Ramos, J. A., Reyes-Duke, A. M., Carrasco-Bardales, A.M. (2024). Convolutional Neural Network (CNN) for Dactylological Alphabet Recognition of Honduran Sign Language (LESHO). *Innovare, Revista de ciencia y tecnología*, 13(2), 1-5. https://doi.org/10.69845/innovare.v13i2.426

INTRODUCTION

Hearing loss represents a significant public health concern. According to the World Health Organization, approximately one and a half billion individuals worldwide experience this condition (World Health Organization, 2021). In 2015, the non-profit organization SIL International cited a sociolinguistic survey conducted by the National Institute of Statistics (INE) in Honduras, which estimated that 2.6% of the Honduran population has some disability. Furthermore, SIL International posits that, of the percentage above, individuals with hearing and speech disabilities in Honduras are estimated to number between 18,000 and 70,000 (Williams, 2010). Therefore, a range of strategies have emerged to address this concern. Among the most advocated approaches, those involving

*Correspondencia: aliciareyes@unitec.edu

novel technologies stand out. Computer vision and deep learning have emerged as key enablers, distinguished by their high accuracy and effectiveness.

Computer vision has been profoundly impacted by the advent of deep learning models. Different artificial neural network (ANN) architectures, such as convolutional (CNN) and recurrent (RNN) neural networks, have enabled the development of machine learning models that are capable of automatically discovering the representations needed for image and video detection or classification (LeCun, 2015).

The origins of CNNs can be traced back to early artificial neural network models, including McCulloch and Pitts' MP model, Rosenblatt's perceptron, and Rumelhart's backpropagation networks (Li, 2021). Although the most significant breakthrough was achieved in 2012 (Krizhevsky, 2012), which demonstrated superior performance on the ImageNet LSVRC challenge through the use of deep CNN architecture also demonstrated by (Fernandez, 2021). Regarding their performance CNN employs convolution operations, rather than the fully connected connections of traditional neural networks, the operations that allow the detection of local features in the input data, such as edges, textures, and shapes in images. These architectures typically comprise a sequence of convolutional, pooling, and fully connected layers.

The convolution layer comprises neurons arranged to form filters. These filters are moved throughout the image and a dot product operation is performed between the input and the filter values (Daniels, 2021). The result of this process is a feature map, which is then passed to a pooling layer. The pooling layer is responsible for reducing the spatial dimensionality, keeping the most relevant features, so the fully connected layers can perform the final classification from the extracted features.

Most research endeavors aimed at advancing the development of novel technologies for the inclusion of individuals with hearing and speech disabilities have sought to align their solutions with one of three prevailing approaches within the broader landscape of scientific contributions. Figure 1 depicts these three prevailing approaches and the solutions that usually fall under them. However, the development of solutions within the computer vision and deep learning spectrum is much more (Amin *et al*., 2023).

The cost of implementing algorithms capable of processing images or video is relatively low in comparison to the high manufacturing cost that sensor-based prototypes would represent (Fuentes *et al*., 2020). Furthermore, deep learning models have also demonstrated remarkable efficiency in other computer vision applications, such as the detection of coral for reef protection in Cayo Blanco, Honduras (Jimenez-Nixon, 2022), the identification of bacteria, including E. coli, P. aeruginosa, and S. aureus present in the water of Lago de Yojoa, Honduras (Cortés, 2023), and to predict spinal abnormalities in neuroradiology images of Honduran patients (Interiano, 2023). These findings demonstrate that this technology is evolving and becoming increasingly capable of performing a growing number of tasks.

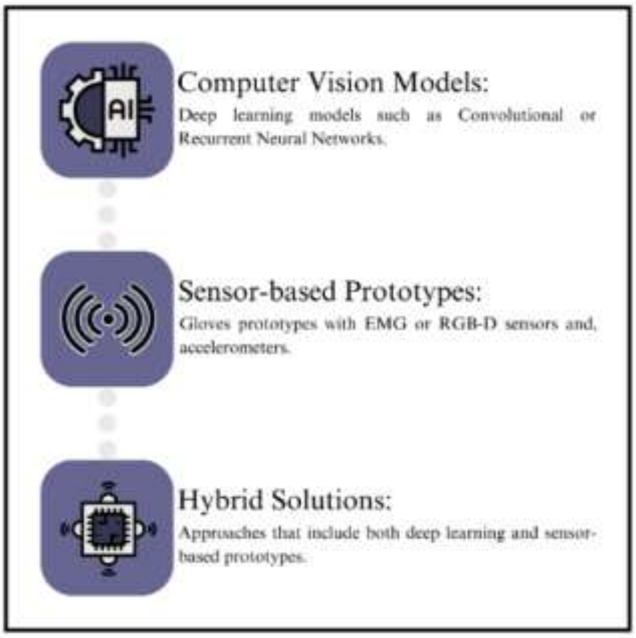


Figure 1. Prevailing approaches within Sing Language detection.

On that account, this research will develop a convolutional neural network to detect and classify the dactylological alphabet of the Honduran Sign Language (LESHO) through a deep learning model. A spiral methodology and a proprietary and specific dataset will be employed to achieve a mAP greater than 95% and an efficient real-time deployment.

METHODS

The spiral methodology: The spiral methodology was used. It was initially developed in 1986 and reported in paper by the mathematician and software engineering professor Barry Boehm (Boehm, 1995). The spiral methodology is designed to facilitate software development while also reducing the associated risks. The idea behind it is based on the hypothesis that an iterative cycle can be repeated until the desired objectives are achieved. Figure 2 illustrates the main stages involved in the implementation of Barry Boehm's spiral methodology.

Data Collection: A specific dataset was constructed for the purposes of the research. A total of 1,080 photographs were collected for each of the five healthy subjects who participated in the sample collection. Each batch of 1,080 photographs included representations of the 27 letters and the three digraphs that make up the Spanish alphabet, with both right and left hand separately. The photographs of each static gesture were captured from five distinct angular positions and two different photographic planes. Additionally, each dynamic gesture was represented four times as much as the corresponding static gestures, with the images taken at varying points in time.

Figure 2. Barry Boehm's spiral methodology diagram.

RESULTS

CNN trainings: The final model with prominent performance in its deployment, had six iterations in which the number of classes remained the same, but their pre-processing and augmentation parameters varied with the purpose of continuous improvement. Roboflow was always use d to complete the process, and it is identified that the main risk is to find low performance metrics.

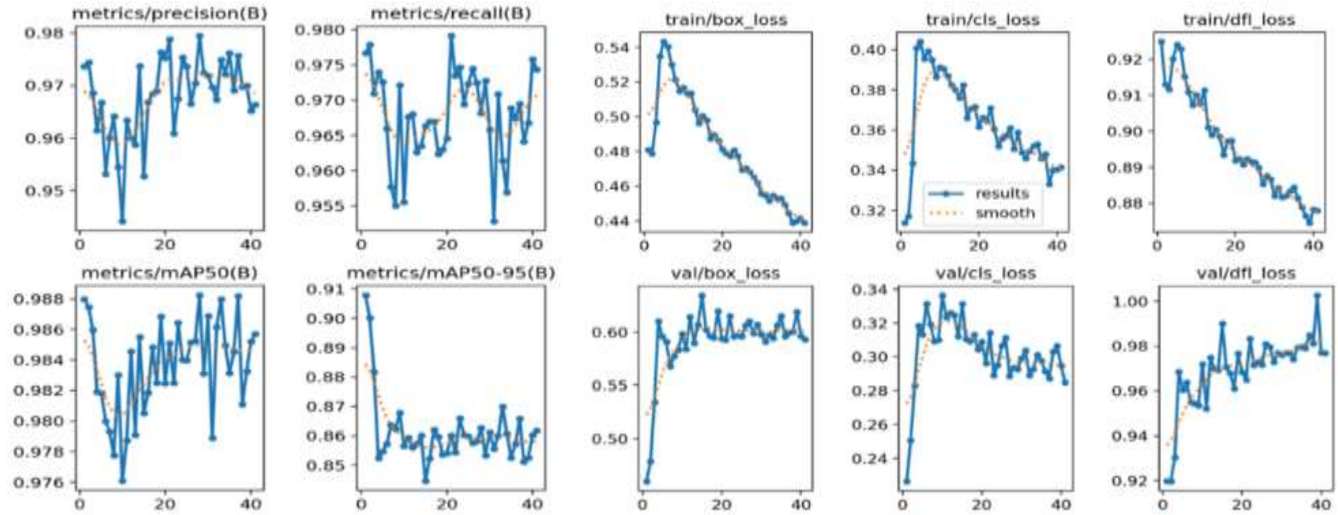
Analysis and comparison: The results of the training performance metrics across several iterations, with the final model demonstrating the highest values for all three metrics. The main metrics to evaluate the neural network in this project are precision, recall, mAP, box losses and class losses.

Model Deployment: A comparative analysis was done to deploy the model exhibiting the best results. Given the comparative analysis presented in Table 1, the decision to deploy the model exhibiting the most favorable performance metrics was taken, with the objective of achieving promising outcomes.

Table 1. Comparison Between Iterations Performance.

1	2	3	4	5	6
m: 92.1%	m: 91.7%	m: 97.7%	m: 98.0%	m: 98.5%	m: 98.8%
P: 74.5%	P: 83.2%	P: 91.0%	P: 95.7%	P: 97.1%	P: 97.4%
R: 83.9%	R: 80.1%	R: 92.3%	R: 96.4%	R: 96.9%	R: 97.7%

The following section will present the metrics corresponding to a model comprising six iterations, with 4,840 labeled images and a final set of 8,226 images after the application of various data enhancement parameters, including:



(a)

(b)

Figure 3. (a) Box, class and object loss for the validation set (b) Detection examples.

The images were rotated between -15° and +15°, the saturation was adjusted between -25% and +25%, the brightness between -20% and +20%, the exposure between -10% and +10%, and the noise was limited to 1% of the pixels. The results of the most important evaluation metrics of the Convolutional Neural Network for the Honduran sign language dactylological alphabet detection are graphed as shown in Figure 3. Where the values for mAP, precision, and recall were found to be 98.8%, 97.4%, and 97.7%.

ANALYSIS AND DISCUSSION

A convolutional neural network capable of detecting and classifying hand gestures presents a model with an accuracy of 94.04% and 83.82% in terms of depth on the EgoGesture and NVIDIA datasets, respectively (Köpüklü, 2019). Although the study did not focus on hand gestures belonging to any specific sign language, it nonetheless sets a precedent for all future works within the framework of computer vision and convolutional neural networks involving hand gesture recognition.

Addressing deep learning models for sign language detection, some authors suggest the use of RNNs of the long short-term memory (LSTM) type to overcome the limitations ANNs have within the framework of sign language interpretation which is the detection of dynamic gestures (i.e., those that involve movement) (Grif, 2021). The RNN model achieved an f-score of 91% on the test data pertaining to the Russian sign language.

However, in regard of CNNs employed for the purpose of this investigation, the authors employed a pre-trained GoogleNet architecture, trained on the ILSVRC2012 dataset and the ASL datasets from the universities of Surrey and Massey for real-time American Sign Language detection (Garcia & Viesca, 2016). This study demonstrated that a computer vision model based on a convolutional neural network offers the advantage of generalization and a seamless integration in the context of the development of tools for the inclusion of individuals with hearing and speech disabilities.

The class loss parameter corresponds to the classification loss of objects. The fluctuating curves indicate that the model is experiencing difficulties in correctly classifying objects at certain stages, yet overall demonstrates a consistent and progressive improvement as training progresses. Consequently, object loss is defined as the loss related to the prediction of displacement or additional dimensions of the detected objects.

The observed behavior of the graphs indicates that the model is undergoing a learning process, exhibiting enhanced capabilities in the detection, classification, and accurate prediction of object features.

A confusion matrix is a widely utilized tool in machine learning for the evaluation of classifier quality. It is employed to tabulate the predicted and actual decision classes. The elements on the diagonal of the matrix contain the number of correctly classified elements, while the off-diagonal elements represent the incorrect classifications (Düntsch & Gediga, 2019). It is crucial to highlight that the model has a couple of limitations regarding its performance and class prediction. The classes with a higher risk of underperformance are letters "h", "m", "ñ", and "s" due to a high percentage of confusion between other classes when the confidence threshold is increased.

In 2014, the legislature power of Honduras approved LESHO as the official language system for all people with hearing and speech disabilities through Decree Law No. 321-2013, cited in reference (Poder Legislativo de Honduras, 2014). Consequently, for the preceding ten years, the deaf-mute community in Honduras lacked a symbol of identity that would ensure their ability to communicate and represent themselves with dignity. Furthermore, there were no inclusive tools that could facilitate communication and interpretation.

The alphabet is the fundamental building block of any language. The Honduran dactylological alphabet, which comprises the 27 letters of the Spanish alphabet and three digraphs (i.e., the combination of consonants "ch," "ll," and "rr"), exemplifies this principle. The Honduran dactylological alphabet serves as the foundation of the Honduran sign language, LESHO.

Conclusion

The development of a convolutional neural network (CNN) model for recognizing the dactylological alphabet of the Honduran Sign Language (LESHO) represents a promising approach to promote accessibility and inclusion for the Honduran deaf community. Computer vision and deep learning techniques, such as CNNs, have demonstrated significant potential for advancing sign language recognition technologies. A proprietary and specific data set, coupled with a spiral research methodology, can be instrumental in developing robust and specialized CNN models for sign language recognition.

Author contributions

All authors participated in the research, preparing the manuscript and approved its final version.

Conflicts of interest

None.

Ethics approval

The five healthy subjects participating consented to be part of photographic data collection and testing.

IA use

None.

REFERENCES

Amin, M., Rizvi, S., Mazzei, A., & Anselma, L. (2023). Assistive Data Glove for Isolated Static Postures Recognition in American Sign Language Using Neural Network. *Electronics* 2023, 12, 1904. doi: https://doi.org/10.3390/electronics12081904

Boehm, B. (1995). A Spiral Model of Software Development and Enhancement. *Interactive Technologies*, 281-292.

Cortés, C. A. B., & Duke, A. M. R. (2023, November). A Comparative Study of CNN Trainings for the Detection of E. Coli, P. Aeruginosa and S. Aureus: A Raspberry Pi-Based Prototype. *In 2023 IEEE 41st Central America and Panama Convention (CONCAPAN XLI)* (pp. 1-6). https://doi.org/10.1109/CONCAPANXLI59599.2023.10517539

Daniels, S., Suciati, N., & Fathichah, C. (2021, February). Indonesian sign language recognition using yolo method. In IOP Conference Series: Materials Science and Engineering (Vol. 1077, No. 1, p. 012029). https://iopscience.iop.org/article/10.1088/1757-899X/1077/1/012029

Düntsch, I., & Gediga, G. (2019, May). Confusion matrices and rough set data analysis. In Journal of Physics: Conference Series (Vol. 1229, No. 1, p. 012055). IOP Publishing. http://dx.doi.org/10.1088/1742-6596/1229/1/012055

Fernandez, O. A., Ordóñez-Ávila, J. L., & Magomedov, I. A. (2021, December). Evaluation of parameters in a neural network for detection of red ring pest in oil palm. In AIP Conference Proceedings (Vol. 2442, No. 1). https://doi.org/10.1063/5.0076481

Fuentes, M. S., Zelaya, N. A. L., & Avila, J. L. O. (2020, April). Coffee fruit recognition using artificial vision and neural networks. In 2020 5th International Conference on Control and Robotics Engineering (ICCRE) (pp. 224-228). https://doi.org/10.1109/ICCRE49379.2020.9096441

Garcia, B., & Viesca, S. A. (2016). Real-time American sign language recognition with convolutional neural networks. Convolutional Neural Networks for Visual Recognition, 2(225-232), 8.

Grif, M. G., & Kondratenko, Y. K. (2021, October). Development of a software module for recognizing the fingerspelling of the Russian Sign Language based on LSTM. In Journal of Physics: Conference Series (Vol. 2032, No. 1, p. 012024). IOP Publishing. https://iopscience.iop.org/article/10.1088/1742-6596/2032/1/012024

Interiano, A. A. A., Palma, M. A. M., & Leiva, K. M. R. (2023, December). Prediction of Spinal Abnormalities in Neuroradiology Images Applying Deep Transfer Learning. In 2023 IEEE International Conference on Machine Learning and Applied Network Technologies (ICMLANT) (pp. 1-7). https://doi.org/10.1109/ICMLANT59547.2023.10372991

Jimenez-Nixon, D. A., Corrales, J. F. M., & Reyes-Duke, A. M. (2022, December). Coral detection using artificial neural networks based on blurry images for reef protection in cayo blanco, honduras. In 2022 IEEE International Conference on Machine Learning and Applied Network Technologies (ICMLANT) (pp. 1-6). https://doi.org/10.1109/ICMLANT56191.2022.9996481

Köpüklü, O., Gunduz, A., Kose, N., & Rigoll, G. (2019, May). Real-time hand gesture detection and classification using convolutional neural networks. In 2019 14th IEEE international conference on automatic face & gesture recognition (FG 2019) (pp. 1-8). https://doi.org/10.1109/FG.2019.8756576

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. Communications of the ACM, 60(6), 84-90. https://doi.org/10.1145/3065386

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. nature, 521(7553), 436-444. https://doi.org/10.1038/nature14539

Li, Z., Liu, F., Yang, W., Peng, S., & Zhou, J. (2021). A survey of convolutional neural networks: analysis, applications, and prospects. IEEE transactions on neural networks and learning systems, 33(12), 6999-7019. https://doi.org/10.1109/TNNLS.2021.3084827

Poder Legislativo de Honduras. (2014, May 22). Decreto 321/2013. Ley de la Lengua de Señas Hondureña (LESHO). *Diario Oficial de la República de Honduras "La Gaceta".*

Terven, J., Cordova-Esparza, D. M., Ramirez-Pedraza, A., & Chavez-Urbiola, E. A. (2023). Loss functions and metrics in deep learning. A review. arXiv preprint arXiv:2307.02694.

Williams, H. (2010). A Sociolinguistic survey of the Honduran Deaf community. p. 24. Retrieved from https://www.sil.org/resources/publications/entry/9181

World Health Organization. (2021). World report on hearing. Retrieved april 27, 2024, from https://www.who.int/publications/i/item/9789240020481