Revista CIENCIA Y TECNOLOGÍA

ISSN 1810-6781 Rev. Cienc. Tecnol. 21(1): 45-50, (2025)

Red neuronal densa para la clasificación de llamadas de dos especies de ranas cristal

Fully Connected Neural Network for the Classification of Calls from Two Species of Glass Frogs

Byron Jiménez Oviedo*; Katalina Oviedo Rodríguez; Jorge Arroyo Hernández; Federico Mora Mora

Facultad de ciencias exactas y naturales, Escuela de Matemática Universidad Nacional de Costa Rica.

* Autor correspondiente: byron.jimenez.oviedo@una.cr (B. Jiménez)

DOI: 10.17268/rev.cyt.2025.01.04

RESUMEN

En este documento se presenta la aplicación de aprendizaje automático (AA) para la clasificación de llamadas de las especies de rana cristal, *Hyalinobatrachium fleischmanni* (Hf) y *Espadarana prosoblepon* (Ep) a partir de grabaciones de audio. Para esto se ha utilizado un dataset de datos acústicos obtenidos por medio de la manipulación de frecuencias (+4 semitonos para Hf y -4 semitonos para Ep) y la incorporación de ruidos ambientales (white noise/pink noise). El modelo de AA se ha entrenado con llamadas de ranas originales y modificadas con el objetivo de poder distinguir las llamadas ante variaciones en las señales acústicas. La evaluación del modelo se ha realizado mediante el uso de las métricas F1-score, precisión y recall. Los resultados obtenidos muestran la capacidad del modelo para clasificar con precisión (98 %) las llamadas de ranas.

Palabras clave: Aprendizaje Automático; Redes Neuronales Densas; Bioacústica; Llamadas de Ranas.

ABSTRACT

This document presents the application of machine learning (ML) for the classification of calls from glass frog species, *Hyalinobatrachium fleischmanni* (Hf) and *Espadarana prosoblepon* (Ep) based on audio recordings. For this, a dataset of acoustic data obtained through frequency manipulation (+4 semitones for Hf and -4 semitones for Ep) and the incorporation of environmental noise (white noise/pink noise) was used. The ML model was trained with original and modified frog calls in order to distinguish the calls under variations in the acoustic signals. Model evaluation was carried out using F1-score, precision, and recall metrics. The results show the model's ability to classify frog calls with high accuracy (98%).

Keywords: Machine Learning; Dense Neural Networks; Bioacoustics; Frog Calls.

1. INTRODUCCIÓN

El monitoreo acústico de los sonidos de los animales se ha convertido en una herramienta clave en la investigación ecológica (Blumstein et al., 2011; Mcloughlin et al., 2019). En este documento se presenta un estudio realizado respecto a las llamadas de las ranas cristal, las cuales, siendo bioindicadores, se utilizan como detectores de cambios en el ambiente (Willacy et al., 2015; Bosch & De la Riva, 2004). Sin embargo, estudios previos muestran que identificar las llamadas de diferentes especies puede ser una tarea compleja debido a su diversidad y variabilidad (Xie et al., 2022; Lapp et al., 2021). Estas limitaciones se han reducido gracias a los métodos de procesamiento de señales, los cuales son más robustos en la actualidad. Además, el aprendizaje automático y el procesamiento de grandes volúmenes de datos ha ayudado en la clasificación y detección de patrones en el monitoreo acústico de diversas especies (Ayoola et al., 2024; Rao et al., 2020). En particular, las redes neuronales densas (FCNN: Fully Connected Neural Network) y las convolucionales (CNN) han mostrado un rendimiento adecuado en la identificación de patrones acústicos complejos en espectrogramas (representaciones visuales de las frecuencias) y amplitudes de los sonidos en el tiempo (Xie et al., 2022; Chalmers et al., 2021; Jeantet & Dufourq, 2023; Chen et al., 2023). En este artículo se presenta la utilización de un aprendizaje automático, particularmente redes neuronales densas, para la clasificación de las llamadas de dos especies de ranas de cristal (o centrolénidos), *Hyalinobatrachium fleischmanni* (Hf) y *Espadarana prosoblepon* (Ep), a partir de grabaciones realizadas bajo condiciones controladas en laboratorio, específica-

Fecha de envío: 27-01-2025 Fecha de aceptación: 20-03-2025 Fecha de publicación: 28-03-2025

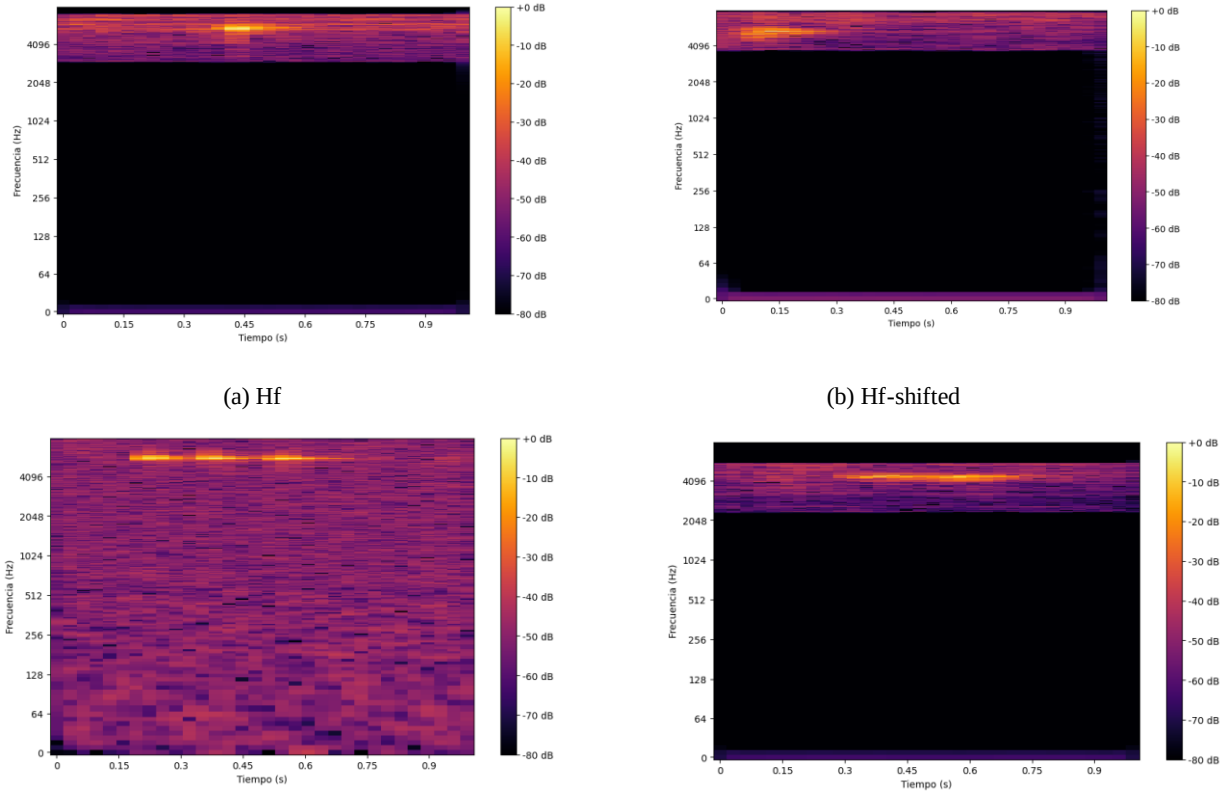
mente utilizando la base de datos “Glass Frog Calls”, que emplean técnicas de aumento de datos como el desplazamiento de frecuencia y la inyección de ruido (blanco y rosa) para incrementar la variabilidad del conjunto de datos (Vargas-Castro et al., 2024).

2. METODOLOGÍA

En esta sección se describen los métodos utilizados para realizar este estudio. Se explica cómo se recopilan los datos, y las herramientas que se utilizaron para analizar los resultados.

2.1 Base de datos

Para la fase de clasificación de las llamadas de las dos especies, Hf y Ep, se utilizó una base de datos “Glass Frog Calls” (Vargas-Castro et al., 2024), que contiene 5296 grabaciones acústicas (WAV) de ambas especies obtenidas bajo condiciones controladas en laboratorio. A las grabaciones se les añadió ruido blanco y ruido rosa. El ruido blanco es una señal que contiene todas las frecuencias audibles de manera uniforme, es decir, su energía se distribuye de forma igual a través del espectro de frecuencias. Esto hace que su sonido sea similar al de un "zumbido constante. El ruido rosa, por otro lado, es similar al blanco, pero su energía disminuye a medida que aumentan las frecuencias. Esto significa que las frecuencias bajas tienen más presencia que las altas, lo que produce un sonido más suave, similar al ruido de fondo en un entorno natural. Además, se aumentó la frecuencia de la Hf en 4 semitonos, elevando su tono, mientras que se disminuyó la frecuencia de las Ep en el mismo número de semitonos, bajando su tono. Esta modificación en las frecuencias permite un ajuste en las características tonales de las grabaciones, lo que podría tener efectos sobre cómo se perciben o se procesan los sonidos en diferentes contextos. De esta forma, la base de datos está constituida por 4 grupos etiquetados, a saber: Hf (1250 llamadas), Hf-shifted (1250 llamadas), Ep (1398 llamadas) y Ep-shifted (1398 llamadas). Un espectrograma de cada uno se puede ver en la Figura 1.



(a) Hf (b) Hf-shifted

(c) Ep **(d)** Ep-shifted

Figura 1. Espectrogramas de las cuatro llamadas de ranas.

2.2 Los coeficientes de frecuencia de Mel

La extracción de Coeficientes Cepstrales en las Frecuencias de Mel (MFCC por sus siglas en inglés) se utilizó para clasificar sonidos apelando a una percepción más humanizada de sonidos (Tiwari, 2010; Zheng et al.,

2001). Se ha utilizado la librería librosa de Python para cargar los datos de audio y calcular los MFCC correspondientes. Una vez extraídos los MFCC de los archivos, se almacenaron en una lista, junto con las etiquetas correspondientes a cada archivo de audio. Además, se realizó una normalización de las características (features) lo que permitió que todas las entradas estuvieran dentro de un rango similar.

2.3 Definición y Entrenamiento del Modelo

Para la clasificación de las señales de audio se define la FCNN usando TensorFlow/Keras, la cual es un tipo de red neuronal adecuada para la clasificación de los espectrogramas (Chen et al., 2023). La arquitectura de la FCNN fue diseñada con 64 neuronas iniciales con función de activación ReLU, luego se le aplicó un dropout de 0,5 para evitar el sobreajuste. En la siguiente capa se aplicó 32 neuronas, también con una función de activación ReLU. Por último, en la salida se utilizó una función de activación softmax que otorga probabilidades a cada clase etiquetada (Figura 2). El modelo se corre utilizando el optimizador Adam y la función de pérdida sparse categorical crossentropy, puesto que las etiquetas son categóricas y no necesitaron de un código disyuntivo completo. Finalmente, el modelo se entrena utilizando el conjunto de datos de entrenamiento (80%), y se prueba en un conjunto de prueba para evaluar su rendimiento (20%). Lo anterior permitió extraer características jerárquicas de las imágenes espectrales.

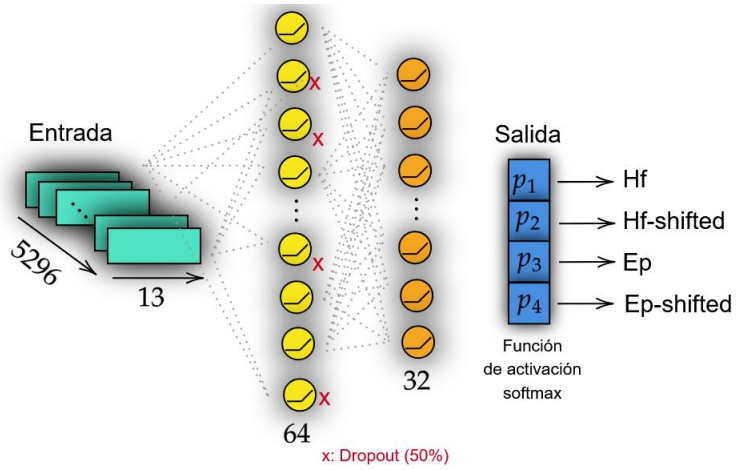


Figura 2. Diagrama de Red neuronal.

2.4 Principal código en Pyhon

En esta subsección se agrega el código que se ha generado en Python para su reproducibilidad (ver código completo en GitHub), como ya se mencionó, se debe importar la biblioteca librosa. Para extraer los coeficientes de Mel se utiliza el código dado en la figura 3.

```
#Función para extraer los MFCCs de un archivo de audio
def extract_mfcc(file_path, sample_rate=22050, n_mfcc=13, n_fft=2048, hop_length=512):
    # Carga el archivo de audio
    y, sr = librosa.load(file_path, sr=sample_rate)

    # Extrae MFCCs
    mfcc = librosa.feature.mfcc(y=y, sr=sr, n_mfcc=n_mfcc, n_fft=n_fft, hop_length=hop_length)

    # Promediar el MFCC
    mfcc = np.mean(mfcc, axis=1)

    return mfcc
```

Figura 3. Función que extrae los MFCC.

El código para FCNN según lo mencionado en la subsección 2.3 viene dado en la figura 4.

```
# FCNN
model = tf.keras.Sequential([
    tf.keras.layers.Dense(64, activation='relu', input_shape=(X.shape[1],)),
    tf.keras.layers.Dropout(0.5),
    tf.keras.layers.Dense(32, activation='relu'),
    tf.keras.layers.Dense(len(class_labels), activation='softmax')  # Salida para las 4 clases
])
```

Figura 4. Código del modelo FCNN.

Para la visualización de los espectrogramas basta usar la función specshow que posee la biblioteca librosa (Figura 5).

```
# Visualizar el espectrograma
plt.figure(figsize=(10, 6))
librosa.display.specshow(D, sr=sample_rate, x_axis='time', y_axis='log', cmap='inferno')
```

Figura 5. Código de visualización del espectograma.

3. RESULTADOS Y DISCUSIÓN

Una vez entrenado el modelo, se evalúa su desempeño sobre el conjunto de prueba. Para esto se utilizan la matriz de confusión que es una herramienta utilizada para evaluar el desempeño de un modelo de clasificación, comparando las predicciones realizadas por el modelo con las etiquetas reales del conjunto de datos de prueba. Esta matriz muestra los aciertos y errores del modelo, donde las filas representan las etiquetas reales y las columnas las predicciones del modelo, permitiendo identificar qué clases fueron correctamente clasificadas y cuáles fueron confundidas.

Además, a partir de esto se pueden obtener las siguientes métricas (Powers, 2020). La precisión

$$TP/(TP+FP), \tag{1}$$

que mide cuánto es la proporción de predicciones positivas verdaderas con respecto al total de predicciones positivas. El recall

$$TP/(TP+FN), \tag{2}$$

que mide cuántos de los casos positivos reales fueron identificados. Y el F1-score

$$2TP/(2TP+FP+FN), \tag{3}$$

la cual es una métrica más balanceada pues incorpora tanto a la precisión como al recall. En las fórmulas anteriores se utiliza la notación usual, donde TP se refiere a los verdaderos positivos, FP falsos positivos, FN falsos negativos. El reporte de las métricas y la matriz de confusión según figura 3, se realiza mediante scikit-learn, el cual realiza automáticamente la comparación entre las predicciones del modelo y las etiquetas reales del conjunto de datos.

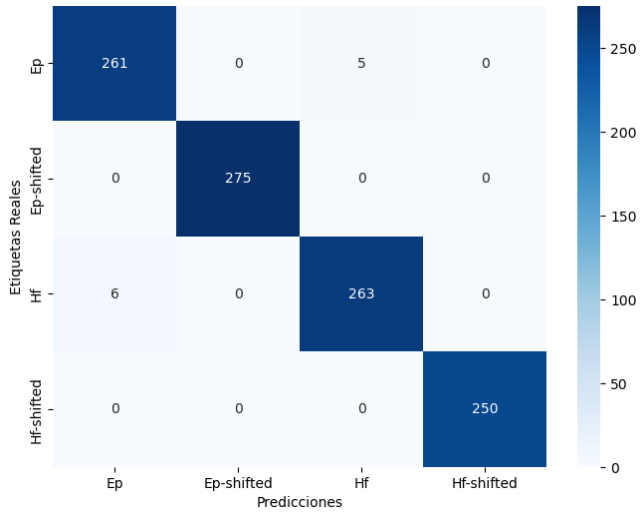


Figura 6. Matriz de confusión.

El reporte de las métricas y la matriz de confusión según figura 6, se realiza mediante scikit-learn, que compara las predicciones del modelo con las etiquetas reales. Las métricas se calcularon para cada clase individualmente, así como sus promedios globales y la exactitud total. Para la clase Ep, la precisión, el recall (a modo de ejemplo, el cálculo de la precisión fue 261/266, siguiendo (1)) y el F1-score dieron 0.98, lo que indica que el modelo clasificó correctamente el 98% de las instancias como positivas, identificó correctamente el 98% de las instancias positivas de esta clase y se tiene un buen balance entre precisión y recall en esta clase. Por otro lado, para la clase Ep-shifted, el modelo obtuvo un rendimiento perfecto, con una precisión de 1, un recall de 1 y un F1-score de 1. En el caso de la clase Hf y la clase Hf-shifted, los resultados son similares a los de Ep y la clase Hf-shifted. El modelo obtuvo una exactitud de 0,99, lo que indica que las instancias fueron clasificadas correctamente en casi todas las clases. Además, el promedio ponderado de las métricas es de 0,99, lo cual indica que las clases con más instancias no están dominando las métricas y que el rendimiento es consistente en todo el conjunto de datos. Estos resultados muestran que el modelo tiene un rendimiento adecuado lo que sugiere que es eficaz en la clasificación de las diferentes clases de llamada consideradas.

4. CONCLUSIONES

Las llamadas de las ranas pueden ser consideradas como indicadores sobre su comportamiento y el estado de su entorno. Los modelos de aprendizaje automático son aliados para identificar patrones en este tipo de vocalizaciones, aun cuando las condiciones de estos sonidos sean variadas y/o alteradas por ruidos ambientales, ya sean de fondo o por la cercanía o lejanía de este. Estos tipos de herramientas permiten un monitoreo continuo, evitando la intervención del investigador, lo cual puede colaborar con la protección de la biodiversidad. Este estudio resalta la importancia que tienen los datos originales o simulados en la investigación de este tipo. El acceso abierto y el aumento a los datos, tal como lo presenta la base “Glass Frog Calls”, es una técnica que ayuda a la capacidad de los modelos para reconocer variaciones de los llamados. El modelo de aprendizaje automático que se aplicó demostró eficiencia respecto a la clasificación de las llamadas realizadas por las dos especies de ranas. Las métricas de rendimiento utilizadas muestran que el modelo es capaz de clasificar correctamente las llamadas de las ranas, incluso cuando estas fueron alteradas por cambios en el tono o por la adición de ruido. Asimismo, el modelo también mostró la capacidad de manejar variaciones en las señales acústicas, algo crucial en la práctica del monitoreo ambiental, en las que las condiciones de grabación no siempre son las ideales. No obstante, todavía existen áreas de mejora. Consideramos que sería útil ampliar la base de datos de entrenamiento incluyendo más variabilidad, así como explorar otros modelos, como las redes neuronales convencionales o “support vector machine”. Todo esto podría nutrir al modelo con mejores capacidades para enfrentar diferentes tipos de ruido o distorsiones de las grabaciones acústicas.

5. AGRADECIMIENTOS

Agradecemos a la Universidad Nacional por brindar un entorno propicio para el desarrollo académico y profesional, así como por su constante motivación en la elaboración de artículos en el área de estudio.

REFERENCIAS BIBLIOGRÁFICAS

Ayoola, V. B., Idoko, I. P., Eromonsei, S. O., Afolabi, O., Apampa, A. R., & Oyebanji, O. S. (2024). The role of big data and AI in enhancing biodiversity conservation and resource management in the USA. World Journal of Advanced Research and Reviews, 23(2), 1851–1873. https://doi.org/10.30574/wjarr.2024.23.2.2350

Blumstein, D. T., Mennill, D. J., Clemins, P., Girod, L., Yao, K., Patricelli, G., Deppe, J. L., Krakauer, A. H., Clark, C., Cortopassi, K. A., Hanser, S. F., McCowan, B., Ali, A. M., & Kirschel, A. N. G. (2011). Acoustic monitoring in terrestrial environments using microphone arrays: applications, technological considerations and prospectus: Acoustic monitoring. The Journal of Applied Ecology, 48(3), 758–767. https://doi.org/10.1111/j.1365-2664.2011.01993.x

Bosch, J., & De la Riva, I. (2004). Are frog calls modulated by the environment? An analysis with anuran species from Bolivia. Canadian Journal of Zoology, 82(6), 880–888. https://doi.org/10.1139/z04-060

Chalmers, C., Fergus, P., Wich, S., & Longmore, S. N. (2021). Modelling animal biodiversity using acoustic monitoring and deep learning. In 2021 International Joint Conference on Neural Networks (IJCNN) (pp. 1–7). IEEE.

Chen, Z., Gao, D., Sun, K., Zhao, X., Yu, Y., & Wang, Z. (2023). Densely connected networks with multiple features for classifying sound signals with reverberation. Sensors (Basel, Switzerland), 23(16), 7225. https://doi.org/10.3390/s23167225

Jeantet, L., & Dufourq, E. (2023). Improving deep learning acoustic classifiers with contextual information for wildlife monitoring. Ecological Informatics, 77(102256), 102256. https://doi.org/10.1016/j.ecoinf.2023.102256

Rao, R., Montgomery, J., Garg, S., & Charleston, M. (2020). Bioacoustics data analysis – A taxonomy, survey and open challenges. IEEE access: practical innovations, open solutions, 8, 57684–57708. https://doi.org/10.1109/access.2020.2978547

Lapp, S., Wu, T., Richards-Zawacki, C., Voyles, J., Rodriguez, K. M., Shamon, H., & Kitzes, J. (2021). Automated detection of frog calls and choruses by pulse repetition rate. Conservation Biology: The Journal of the Society for Conservation Biology, 35(5), 1659–1668. https://doi.org/10.1111/cobi.13718

Mcloughlin, M. P., Stewart, R., & McElligott, A. G. (2019). Automated bioacoustics: methods in ecology and conservation and their potential for animal welfare monitoring. Journal of the Royal Society, Interface, 16(155), 20190225. https://doi.org/10.1098/rsif.2019.0225

Powers, D. M. W. (2020). Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. arXiv [cs.LG]. https://doi.org/10.48550/ARXIV.2010.16061

Tiwari, V. (2010). MFCC and its applications in speaker recognition. International Journal on Emerging Technologies, 1(1), 19–22.

Vargas-Castro, L. E., Hernández-Ledezma, J., & Pérez-Gómez, G. (2024). Glass Frog Calls. Dataset. https://doi.org/10.5281/zenodo.14251255

Willacy, R. J., Mahony, M., & Newell, D. A. (2015). If a frog calls in the forest: Bioacoustic monitoring reveals the breeding phenology of the endangered Richmond Range mountain frog (Philoria richmondensis). Austral Ecology, 40(6), 625–633. https://doi.org/10.1111/aec.12228

Xie, J., Zhu, M., Hu, K., Zhang, J., Hines, H., & Guo, Y. (2022). Frog calling activity detection using lightweight CNN with multi-view spectrogram: A case study on Kroombit tinker frog. Machine Learning with Applications, 7(100202), 100202. https://doi.org/10.1016/j.mlwa.2021.100202

Zheng, F., Zhang, G., & Song, Z. (2001). Comparison of different implementations of MFCC. Journal of Computer Science and Technology, 16(6), 582–589. https://doi.org/10.1007/bf02943243