

Lung noise recognition employing convolutional neural networks

Christian Salamea-Palacios, Ph.D.¹, Mario Torres-Intriago, Ing.², and Mateo Cordero-Izquierdo¹

^{1,2,3}Universidad Politécnica Salesiana, Grupo de Investigación en Interacción, Robótica y Automática, csalamea@ups.edu.ec, m_torres_i@hotmail.com, mateofcorderoi@gmail.com

Abstract– Currently, the use of artificial intelligence to aid medical diagnosis is increasingly accepted. This paper presents a new method for identifying the presence of asthma, pneumonia, or COPD in patients with respiratory diseases. The identification is based on auscultation-captured lung noise, using Mel cepstral coefficients and convolutional neural networks for classification. The proposed system achieved optimal performance with a minimum of 15 training epochs and a batch size of 32. It achieved up to 88% accuracy in detecting the aforementioned lung diseases, with a particularly high accuracy of 98.8% in identifying COPD and 54.5% in identifying normal cough.

Keywords - Auscultation, Asthma, Pneumonia, COPD, Networks, Neural, Convolutional, Spectrograms.

Reconocimiento de ruidos pulmonares utilizando redes neuronales convolucionales

Christian Salamea-Palacios, Ph.D.¹, Mario Torres-Intriago, Ing.², and Mateo Cordero-Izquierdo¹

^{1,2,3}Universidad Politécnica Salesiana, Grupo de Investigación en Interacción, Robótica y Automática, csalamea@ups.edu.ec, m_torres_i@hotmail.com, mateofcorderoi@gmail.com

Resumen– Hoy en día, el uso de inteligencia artificial para brindar apoyo al diagnóstico médico viene siendo cada vez más aceptado y con esa perspectiva, en este trabajo se presenta un novedoso método para identificar la presencia de asma, neumonía o EPOC en pacientes con enfermedades respiratorias. La identificación se realiza por medio del ruido pulmonar captado por auscultación, utilizando coeficientes cepstrales de Mel y redes neuronales convolucionales para la clasificación. El rendimiento óptimo del sistema propuesto se obtuvo con no menos de 15 épocas de entrenamiento y un batch de 32. En términos generales se ha obtenido hasta un 88% de precisión en la detección de las enfermedades pulmonares descritas anteriormente y específicamente se ha encontrado que el sistema funciona mejor identificando EPOC hasta con un 98,8% de los casos de acierto y la tos normal con hasta un 54,5%.

Palabras Clave - Auscultación, Asma, Neumonía, EPOC, Redes, Neuronales, Convolucionales, espectrogramas.

I. INTRODUCCIÓN

El apoyo al diagnóstico médico por medio de herramientas computacionales basadas en Inteligencia Artificial viene siendo bien visto por la comunidad científica y sus aplicaciones se han diversificado, ya sea con el procesamiento de imágenes o de señales audibles [1]. En este contexto, la pandemia del COVID-19 y el consecuente aislamiento social para evitar la propagación del virus, llevó a la humanidad a buscar soluciones desde diferentes ámbitos para combatir la enfermedad, las que iban desde vacunas hasta sistemas computacionales con Inteligencia Artificial para apoyo en el diagnóstico, y por ello, se crearon diferentes tipos de bases de datos en torno a patologías respiratorias que sirvieron para entrenar modelos de aprendizaje automático e identificar la presencia o no de COVID-19 [2]. Sin embargo, superada la pandemia, dichas bases de datos han empezado a ser útiles para otros fines ya que contienen información que puede ser de utilidad para apoyar en el diagnóstico de otras enfermedades respiratorias, como por ejemplo, el asma [3]. Este trabajo está enfocado en el uso de una base de datos de

ruidos pulmonares obtenida a partir de señales audibles conseguidas mediante auscultación pulmonar para entrenar sistemas de aprendizaje automático y reconocer ciertos tipos de patologías respiratorias.

La auscultación es un examen rutinario que utiliza un estetoscopio para percibir, por el oído humano, las vibraciones del sistema respiratorio provenientes de la caja torácica [4]. A día de hoy, gracias al avance tecnológico en la interpretación de ruidos y en la acústica respiratoria, es posible almacenar, procesar y analizar dichas señales [5].

Tomando como referencia que uno de los campos de la Inteligencia Artificial en donde se plasma el aprendizaje automático son las redes neuronales y que señales audibles han sido eficientemente procesadas mediante el uso de “espectrogramas” como atributos de entrada para redes neuronales convolucionales (CNNs), se han utilizado estas como clasificadoras de señales audibles de pacientes con enfermedades respiratorias.

Se ha podido distinguir el ruido auscultado en cuatro categorías: normal, asma, EPOC o neumonía, aprovechando que el sonido traqueal tiene un ancho de banda mayor que los sonidos traqueales con frecuencias entre los 70 a 2kHz aproximadamente [6]. Las CNNs son redes neuronales artificiales con un enfoque de aprendizaje profundo con el que pueden procesar imágenes y sonidos. A diferencia de las redes neuronales de enfoque de aprendizaje automático común, las CNNs utilizan la convolución para extraer información característica de las imágenes [7] además de que la repetición del proceso de convolución sobre imágenes cada vez más específicas posibilitan la extracción de características más concretas de la imagen, lo que las hace ideales para extraer información específica de los sonidos representados en los espectrogramas.

Las CNNs son un tipo de red neuronal profunda diseñada para el procesamiento de imágenes y señales, como sonidos y videos. A diferencia de las redes neuronales tradicionales, las CNNs utilizan una operación matemática llamada convolución en lugar de una multiplicación de matriz para extraer características importantes de las imágenes. Las CNNs se componen de capas convolucionales y de capas de pooling. El número y tamaño de las capas suficiente para una buena

Digital Object Identifier: (only for full papers, inserted by LACCEI).

ISSN, ISBN: (to be inserted by LACCEI).

DO NOT REMOVE

clasificación-predicción depende de la complejidad del problema y de la cantidad de datos disponibles. Para conseguir el mejor rendimiento y reducir la posibilidad de sobreajuste de la CNNs se aplican técnicas de normalización, y dropout, además de la modificación del número de capas y tamaño del kernel para mejorar la efectividad y la velocidad del entrenamiento [8].

La técnica descrita ha sido la utilizada para entrenar y evaluar un modelo de aprendizaje automático profundo que utiliza la información de ruidos pulmonares obtenidos por medio de auscultación como datos de entrada y es capaz de predecir que tipo de patología está asociada al ruido detectado, pudiendo ser un ruido normal, de neumonía, de EPOC o de asma [9], [10].

En lo que tiene que ver con los ruidos respiratorios, estos se producen en los pulmones durante la respiración y se dividen en dos tipos, ruidos respiratorios normales y anormales. Los primeros son sonidos producidos por el flujo de aire a través de las vías respiratorias y se pueden escuchar durante la inhalación y la exhalación, son generalmente suaves y simétricos en ambos lados del pecho [11]. Los segundos tienen varias clasificaciones según el ámbito de estudio.

Considerando su intensidad, pueden ser: ruidos disminuidos, aumentados y alterados. Los disminuidos son ruidos suaves o ausentes que se pueden escuchar cuando hay una obstrucción de las vías respiratorias, los aumentados son ruidos fuertes que se pueden escuchar cuando hay una obstrucción de las vías respiratorias o cuando se produce una constricción, y los alterados se producen cuando hay un problema en los tejidos pulmonares, como inflamación o edema [12].

Los ruidos adventicios son parte de una segunda clasificación de ruidos anormales y se dividen en ruidos continuos y discontinuos. Los ruidos adventicios continuos son sonidos que se escuchan de manera constante durante la respiración y se dividen en tres tipos: sibilancias, estridores y roncós. Los tipos sibilantes son sonidos agudos y silbantes que se producen cuando el aire pasa a través de vías respiratorias estrechas, como en el asma, el ruido estridor, por otra parte, es un sonido agudo y silbante que se produce cuando hay una obstrucción en las vías respiratorias superiores, como en la laringitis y el ruido ronco que es un sonido grave que se produce cuando hay una obstrucción en las vías respiratorias inferiores, como en la bronquitis [6].

Los ruidos adventicios discontinuos son sonidos que se producen de manera intermitente durante la respiración y se dividen en crepitantes y estertores, los crepitantes son sonidos suaves y discontinuos que se producen cuando hay líquido en los alvéolos pulmonares, como en la neumonía, los estertores son sonidos más fuertes y discontinuos que se producen cuando hay líquido en las vías respiratorias, como en el edema pulmonar [13].

Frente a toda esta variabilidad de ruidos pulmonares, se ha considerado la opción de utilizar herramientas de aprendizaje

automático aplicadas a las señales audibles de los ruidos pulmonares para reconocer las patologías asociadas en base de las características intrínsecas contenidas en las señales audibles. Para ello, se va a representar a cada señal audible mediante coeficientes cepstrales de Mel (MFCCs) que posibilitan el análisis en una escala de frecuencias que se ajustan a la percepción auditiva humana [14]. De esta manera, se pretende que las máquinas puedan tener una referencia similar de análisis a la que tienen los humanos expertos.

Los coeficientes cepstrales de Mel se obtienen mediante el siguiente procedimiento:

1. Se divide la señal en ventanas de 20ms y se aplica un filtrado Hamming para maximizar las zonas de más energía.
2. Se calcula la Transformada de Fourier de la señal ventaneada para extraer su contenido espectral.
3. Se calculan los MFCCs que llevan el espectro de cada ventana a la escala de Mel mediante el banco de filtros correspondiente y se eligen los 13 con más energía.
4. Se realiza una transformada inversa coseno para normalizar los MFCCs generados.

Los 13 coeficientes MEL con más energía van a aparecer cada ventana de análisis (20ms) a lo largo de toda la señal de audio analizada [15], y por ello, se van concatenando bidimensionalmente formando “espectrogramas” que son atributos sintéticos de mucha utilidad para un sistema de clasificación automática que se basa en imágenes.

Así, definidas las características de los ruidos pulmonares existentes y con la posibilidad de usar técnicas de aprendizaje profundo para la clasificación, se propone el reconocimiento de ruidos pulmonares de asma, epoc y neumonía obtenidos por medio de auscultación pulmonar usando redes neuronales convolucionales.

II. DESCRIPCIÓN DE LA TÉCNICA

A. Base de datos

La base de datos está formada por archivos de audio obtenidos a través de auscultación pulmonar en pacientes diagnosticados con asma, epoc o neumonía, todas validadas por expertos en el área de neumología. Se han definido estas cuatro categorías debido a que presentan ruidos característicos que podrían ser distinguibles por medio del sonido que emiten. Así, un ruido normal se identifica a partir de la evaluación de la inspiración y se confirma con la espiración, un ruido de asma se puede escuchar levemente en la inspiración pero lo primordial está en la espiración porque presenta un tono mayor, un ruido de neumonía se confirma en la inspiración y

en la espiración y un ruido EPOC se detecta en inspiración y se confirma en la espiración.

La base de datos ha sido dividida en dos partes, una para la etapa de entrenamiento del sistema y otra para la evaluación del mismo, con un total de 638 muestras tomadas de Kaggle [16] y 121 tomadas con un estetoscopio “Stemoscope”, de tipo digital aprobado por la FDA, con un rango de frecuencia de funcionamiento de 20Hz a 2KHz [17]. El tiempo de grabación promedio ha sido de 20 segundos y la información ha sido tomada desde diferentes puntos de la zona torácica. Su conexión Bluetooth posibilita que los sonidos sean tomados por medio de auriculares y conectados a un ordenador personal que almacena la información. Se han tomado 4 grabaciones por paciente y mediante el criterio de expertos se ha etiquetado cada fichero con la patología asociada. La distribución de la base de datos está descrita en la TABLA I.

TABLA I
DISTRIBUCIÓN DE LA BASE DE DATOS

	Kaggle	Stemoscope
Entrenamiento	509	82
Evaluación	129	39
TOTAL	638	121

En relación al número de observaciones para cada patología analizada, la base de datos se distribuye según como se muestra en la TABLA II.

TABLA II
BASE DE DATOS SEGÚN PATOLOGÍA

	Kaggle	Stemoscope	Total
Asma	49	14	63
Epoc	445	34	479
Neumonía	38	16	54
Normal	106	57	163

Los audios obtenidos de los pacientes han sido normalizados con el fin de asegurar de que tengan una amplitud y duración similar, factores que hacen posible el tratamiento de información por parte de la red neuronal. Los ruidos estudiados se ven como los de las figuras 1, 2, 3 y 4, donde se muestran las señales obtenidas del ruido normal, asma, neumonía y EPOC, respectivamente:



Fig. 1. Ruido normal

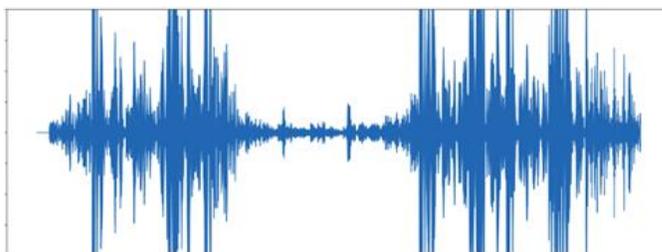


Fig. 2. Ruido de asma



Fig. 3. Ruido de neumonía



Fig. 4. Ruido de EPOC

B. Metodología

Como se mencionó anteriormente, la información auscultada ha sido obtenida por medio de un estetoscopio digital de marca “Stemoscope” que es capaz de enviar los sonidos a un teléfono inteligente o un cualquier dispositivo electrónico en forma de grabaciones de audio de 20 segundos como tiempo promedio por cada zona torácica (frontal izquierda y derecha y posterior izquierda y derecha), tal como se muestra en la figura 5.

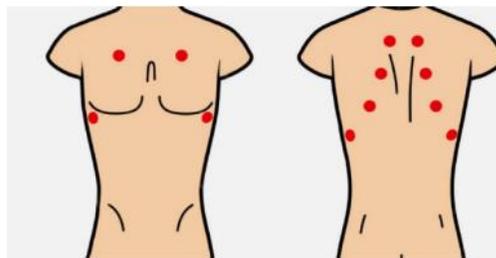


Fig. 5. Zonas de auscultación

Los sonidos ya extraídos pasan por un filtro de pre-énfasis y por un ventaneado Hamming que por sus características se consigue amplificar la señal de entrada en la mitad de cada trama. Luego se sigue todo el proceso necesario para extraer

los coeficientes cepstrales de Mel (MFCCs) que se utilizan debido a que funcionan de manera parecida a lo que funciona el oído humano. Posteriormente, los MFCCs se utilizan para construir los correspondientes espectrogramas mediante la concatenación de cada uno de los 13 MFCCs obtenidos desde los ficheros de audio en tramas de duración estándar de 20ms con traslape de 10ms, resultando en señales bidimensionales de tiempo versus frecuencia.

En este trabajo se ha utilizado una red neuronal convolucional por sobre otro tipo de redes neuronales, dado que es de mucho interés extraer la información producida por los sonidos pulmonares que se encuentra contenida espacialmente en los espectrogramas de Mel, siendo éstos los atributos que se procesan y que se utilizan en el proceso de clasificación.

Así, los espectrogramas se utilizan como entradas para la red neuronal convolucional que está configurada con tres bloques (Convolutivo-MaxPooling) [(42x11x128)-(21x6x128), (19x4x64)-(20x2x64) y (9x1x32)-(5x1x32)] y con tres capas MLP de [(160), (200) y (4)], donde el número 4 describe la capa de salida de la CNN, que define cada una de las clases de patologías pulmonares para las cuales ha sido entrenada la red neuronal. Así, el resultado puede ser: asma, neumonía, EPOC o una tos normal.

En la configuración global del sistema para la fase de experimentación se han utilizado hasta 100 épocas, un batch de 32 y una tasa de aprendizaje de 0.0001, valores con los que se ha evitado el sobre-entrenamiento del sistema y se han conseguido valores óptimos de clasificación. El diagrama de la figura 6 muestra en bloques el funcionamiento del sistema de reconocimiento partiendo de la señal audible de entrada en la parte baja del diagrama y con la clasificación de las enfermedades pulmonares en la parte superior.

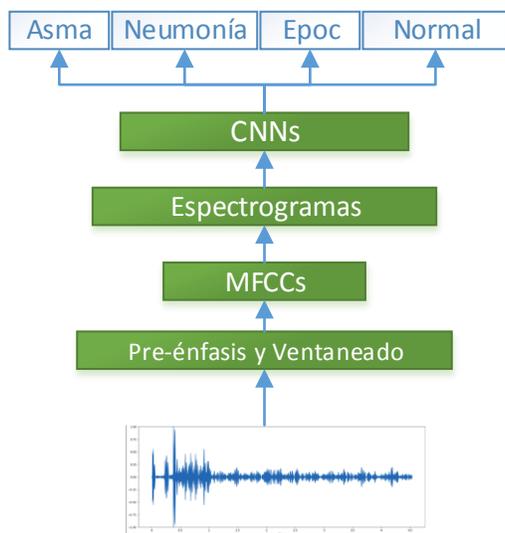


Fig.6. Diagrama de bloques del sistema de reconocimiento

III. RESULTADOS

Para cuantificar el rendimiento del sistema de reconocimiento se han utilizado funciones de medida de la pérdida producida en una predicción errónea y con ese fin se ha utilizado la entropía cruzada que permite establecer la capacidad de la red neuronal para predecir la patología correspondiente a un audio obtenido por auscultación de la caja torácica.

Para ello, se ha llevado a cabo el entrenamiento del sistema utilizando diferentes configuraciones de las CNNs y se han obtenido los diferentes modelos (h5), uno por cada patología asociada a los sonidos pulmonares analizados. Así se ha obtenido un modelo para asma, otro para EPOC, un tercero para neumonía y un último de tos normal. Estos modelos han sido utilizados posteriormente para evaluar sonidos pulmonares auscultados no utilizados en la fase de entrenamiento y predecir a que patología están asociados [18].

Estos resultados se muestran en términos de precisión, recall y F1, tal como muestra la Tabla III.

TABLA III
MÉTRICAS OBTENIDAS POR EL SISTEMA DE RECONOCIMIENTO

	Precision	Recall	F1-score
Asma	1	0.17	0.29
Epoc	1	0.93	0.96
Neumonia	0.75	0.5	0.6
Normal	0.59	1	0.74

A partir de la arquitectura y configuración inicial del sistema de reconocimiento, se han realizado los ajustes que se describen en la Tabla IV con el fin de determinar la configuración óptima con la que se pudieron obtener los mejores resultados con lo que se han obtenido los valores óptimos de configuración del sistema que permiten las mejores métricas de reconocimiento.

TABLA IV
CONFIGURACIONES DEL SISTEMA

N°	Filtro Conv1	Filtro Conv2	Filtro Conv3	Est1 MLP	Est2 MLP	Loss	Accuracy
1	64	32	32	160	200	0.41	88.8
2	128	32	32	160	200	0.65	84.13
3	128	64	32	160	200	0.54	85.71
4	128	64	32	160	250	0.57	83.31
5	128	64	32	160	200	0.52	86.51

Habiendo culminado la fase de experimentación se ha podido comprobar que el mejor resultado en cuanto a "accuracy" se ha obtenido con la configuración del experimento 2 y en base de éste, en la figura 7 se muestran las gráficas correspondientes al entrenamiento y evaluación del sistema. En ella se puede ver que no se requiere más de 20 épocas para obtener un valor óptimo de rendimiento, además que se puede ver que el

sistema es bastante estable frente a las diversas señales audibles de entrada.

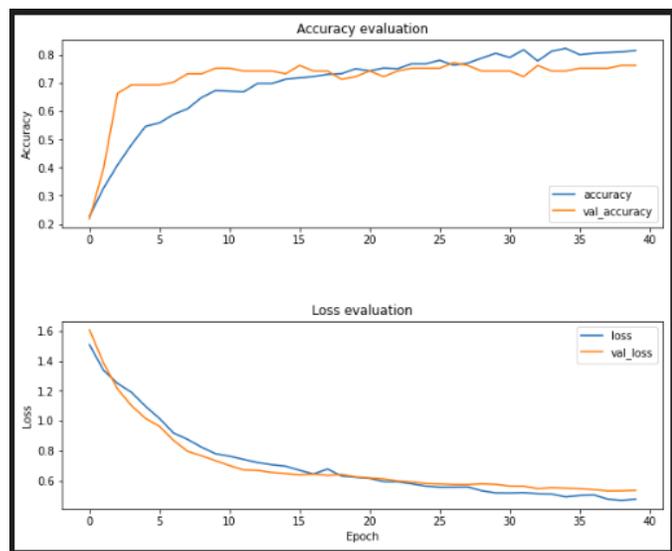


Fig. 7. Curvas de entrenamiento y evaluación del sistema

En lo que se refiere a fiabilidad del sistema más óptimo encontrado en la experimentación, se ha encontrado mediante una matriz de confusión que el sistema funciona bastante bien identificando EPOC con un 98,8% de acierto y menos bien identificando la tos normal donde acierta en un 54,5% y se confunde en un 21,2% con neumonía, un 18,2% con asma y un 6,1% con EPOC. El rendimiento del sistema en la detección de neumonía o asma no termina de ser determinante dado que las observaciones utilizadas para evaluar el sistema no son estadísticamente significativas. En la figura 8 se muestra la matriz de confusión donde se puede ver de manera explícita el mejor rendimiento del sistema de reconocimiento.

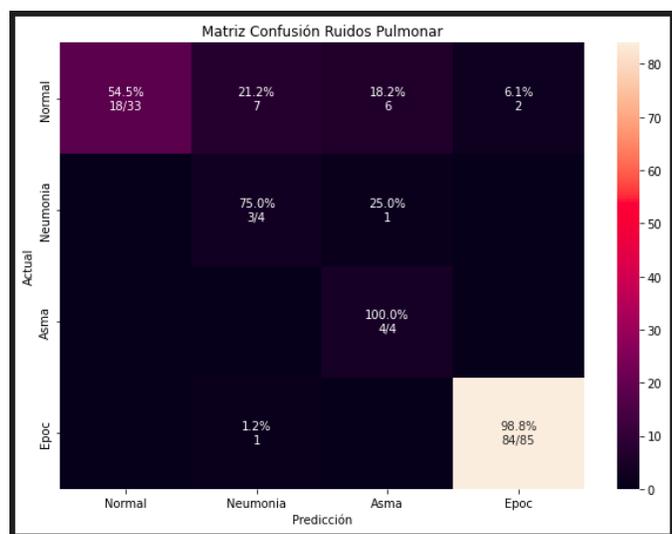


Fig. 8. Matriz de confusión del sistema de reconocimiento

IV. CONCLUSIONES

Se utiliza una red neuronal convolucional (CNN) para identificar patrones específicos de los ruidos pulmonares que indican la presencia de una patología o enfermedad en los pulmones.

La base de datos utilizada está formada por información extraída de la web y de auscultaciones reales, con estos datos se ha llegado a obtener hasta un 89% de precisión en la estimación de la patología por parte del sistema de reconocimiento, por ello, se le puede considerar como una herramienta de apoyo al diagnóstico médico que debe ser corroborado por un profesional de la salud.

Por otra parte, se ha demostrado la utilidad en el uso de bases de datos tomadas de más de una fuente de información y la factibilidad en el uso de imágenes (los espectrogramas) que representen a las señales audibles.

V. REFERENCIAS

- [1] (Gómez et al., 2018). Gómez, R. M., González-Díaz, S. N., Urrutia-Pereira, M., Valentín-Rostan, M., Yáñez, A., Jares, E., Cruz, Á., Calderón, M., & Bousquet, J. (2018). El Acuerdo Bruselas 2017 para Latinoamérica: Una iniciativa de la GARD y SLAai. *Revista Alergia México*, 65(3), 217-221.
- [2] Mazzei, J. A. (2020). COVID-19: Una nueva enfermedad respiratoria y una posible amenaza mundial. *Rev. am. med. respir*, 1-4.
- [3] Sánchez-Eraza, A. G., & Sánchez-Eraza, J. I. (2021). La Organización Mundial de la Salud, aciertos y desaciertos en relación con el COVID19 en Ecuador. *Polo del Conocimiento*, 6(1), 381-398.
- [4] Delgado, A. (1999). Aplicación de las redes neuronales en medicina. *Revista de la Facultad de Medicina*, 47(4), 221-223.
- [5] Gimeno-Santos, E., Herrero Cortina, B., Martí Romeu, J. D., & Vilaró Casamitjana, J. (2013). Auscultación pulmonar. Técnicas manuales e instrumentales para el drenaje de secreciones bronquiales en el paciente adulto. *Manual SEPAR de procedimientos*. Barcelona: Editorial Respira, 16-26.
- [6] Sánchez, I. (2003). Aplicaciones clínicas del estudio objetivo de los ruidos respiratorios en pediatría. *Revista chilena de pediatría*, 74(3), 259-268.
- [7] Matich, D. J. (2001). *Redes Neuronales: Conceptos básicos y aplicaciones*. Universidad Tecnológica Nacional, México, 41, 12-16.
- [8] Badillo, F. L., Hernández, C. A. R., Narváez, B. M., & Trillos, Y. E. A. (2021). Redes neuronales convolucionales: Un modelo de Deep Learning en imágenes diagnósticas. *Revisión de tema. Revista colombiana de radiología*, 32(3), 5591-5599.
- [9] Saca, F. L., Ramírez, A. F., & Cruz, C. A. (2018). Prototipo Funcional para Clasificación de Imágenes con Salida de Audio en un Sistema Embebido con Red Neuronal Convolutiva (Funcional Prototipo For Classification Of Images With Audio Output In An Embedded System Using Convolutional Neural Network). *Pistas Educativas*, 40(130).
- [10] Hernández Jaso, I. (2021). Desarrollo de un modelo de predicción de parámetros clínicos pulmonares mediante sensores de temperatura y humedad usando técnicas de inteligencia artificial.
- [11] Sarkar, M., Madabhavi, I., Niranjan, N., & Dogra, M. (2015). Auscultation of the respiratory system. *Annals of Thoracic Medicine*, 10(3), 158-168. <https://doi.org/10.4103/1817-1737.160831>
- [12] Bohadana, A., Izbicki, G., & Kraman, S. S. (2014). Fundamentals of Lung Auscultation. *New England Journal of Medicine*, 370(8), 744-751. <https://doi.org/10.1056/NEJMra1302901>

- [13] Sepúlveda, R. (2017). Las enfermedades respiratorias del adulto mayor en Chile: Un desafío a corto plazo. *Revista chilena de enfermedades respiratorias*, 33(4), 303-307.
- [14] Jung, S.-Y., Liao, C.-H., Wu, Y.-S., Yuan, S.-M., & Sun, C.-T. (2021). Efficiently classifying lung sounds through depthwise separable CNN models with fused STFT and MFCC features. *Diagnostics*, 11(4), 732.
- [15] Ramírez Sánchez, J. M., Montalvo Bereau, A. R., & Calvo de Lara, J. R. (2019). Evaluación de Rasgos Acústicos para el Reconocimiento Automático del Habla en Escenarios Ruidosos usando Kaldi. *Ingeniería Electrónica, Automática y Comunicaciones*, 40, 51-71.
- [16] Respiratory Sound Database. (s. f.). Recuperado 11 de marzo de 2023, de [HTTPS://WWW.KAGGLE.COM/DATASETS/VBOOKSHELF/RESPIRATORY-SOUND-DATABASE](https://www.kaggle.com/datasets/vbookshelf/respiratory-sound-database)
- [17] Stemoscope II is the best digital electronic bluetooth stethoscope for personal use. (s. f.). stemoscope. Recuperado 11 de julio de 2023, de <https://stemoscope.com/pages/stemoscope-ii-best-bluetooth-stethoscope-for-personal-use-fda-cleared>
- [18] Galvañ Sala, D. A. (2021). Comparativa de técnicas para la prevención del sobreajuste en redes neuronales.