Diagnóstico Automatizado de Auscultación Pulmonar Pediátrica Usando Redes Neuronales Profundas
Jorge Lopez Perez, Damián Taire and Claudio Delrieux
SADIO Electronic Journal of Informatics and Operations Research (EJS) Vol. 24 No. 1 (2025) e-ISSN 1514-6774
https://doi.org/10.24215/15146774e072 | https://revistas.unlp.edu.ar/ejs
Sociedad Argentina de Informática e Investigación Operativa | Universidad Nacional de La Plata | Buenos Aires | Argentina

Diagnóstico Automatizado de Auscultación Pulmonar Pediátrica Usando Redes Neuronales Profundas

Automated Diagnosis of Pediatric Pulmonary Auscultation Using Deep Neural Networks

Jorge I. Lopez Perez[1][0000-0001-7132-676X], Damián L. Taire[2 3][0000-0001-6505-1560] y Claudio Delrieux [1][0000-0002-2727-8374]

[1] Universidad Nacional Del Sur, Departamento de Ing. Eléctrica y Computadoras (DIEC), Instituto de Ciencias e Ingeniería de la Computación - ICIC (UNS-CONICET), Av. Alem 1253 (B8000CPB), Bahía Blanca, Buenos Aires, Argentina

[2] Instituto Patagónico de Ciencias Sociales y Humanas "Dra. María Florencia del Castillo Bernal" (IPCSH), Consejo Nacional de Investigaciones Científicas y Técnicas, Puerto Madryn, Argentina

[3] Departamento de Neumonología Pediátrica, Hospital Zonal “Dr. Andrés R. Isola”, Puerto Madryn, Argentina

jorgeivanl620@gmail.com
dtaire@cenpat-conicet.gob.ar
cad@uns.edu.ar

Resumen. En este trabajo se investiga la implementación de redes neuronales profundas en la clasificación de sonidos respiratorios, una tarea determinante para el diagnóstico de enfermedades pulmonares. Para esta labor, se emplea la arquitectura VGG-16, reconocida por su eficacia en la clasificación de imágenes, la cual ha sido adaptada para procesar datos de audio. Se realizaron la recopilación y preprocesamiento del conjunto de datos de sonidos respiratorios, utilizando coeficientes cepstrales de frecuencia de Mel (MFCC´s) como entrada de la red. Los resultados obtenidos revelan un rendimiento significativo, con una precisión del 79% en la clasificación de sonidos respiratorios. Este resultado resalta el potencial de las redes neuronales convolucionales pre entrenadas en el campo médico. Sin embargo, persisten desafíos por superar, como la necesidad de conjuntos de datos más amplios y una comprensión más profunda de los resultados para su implementación clínica efectiva.

Palabras clave: Redes neuronales profundas, Sonidos respiratorios, Arquitectura VGG-16, Coeficientes cepstrales en frecuencia de Mel (MFCC´s), Diagnóstico de patologías respiratorias.

Abstract. This study investigates the implementation of deep neural networks in the classification of respiratory sounds, a crucial task for diagnosing pulmonary

Received September 2024; Accepted December 2024; Published March 2025

diseases. For this purpose, the VGG-16 architecture, renowned for its effectiveness in image classification, was adapted to process audio data. The respiratory sound dataset was collected and preprocessed using Mel-frequency cepstral coefficients (MFCCs) as input to the network. The results reveal significant performance, achieving 79% accuracy in classifying respiratory sounds. This outcome highlights the potential of pre-trained convolutional neural networks in the medical field. However, challenges remain, such as the need for larger datasets and a deeper understanding of the results for effective clinical implementation.

Keywords: Deep neural networks, Respiratory sounds, VGG-16 architecture, Mel-frequency cepstral coefficients (MFCCs), Diagnosis of respiratory diseases

1 Introducción

Las enfermedades respiratorias a lo largo de la vida constituyen una carga global significativa para los individuos, las familias, los sistemas de atención médica y las sociedades. Estas condiciones generan impactos que incluyen costos directos asociados a servicios de salud, como hospitalizaciones y tratamientos farmacológicos, así como costos indirectos relacionados con la pérdida de productividad laboral y escolar, años de vida ajustados por discapacidad y el impacto emocional en los cuidadores y el núcleo familiar [1, 2, 3]. Además, estas enfermedades no solo representan un desafío médico, sino también económico y social, afectando de manera desproporcionada a las comunidades menos favorecidas y exacerbando las desigualdades existentes en los sistemas de salud. Este impacto se magnifica en poblaciones vulnerables como niños, ancianos y personas con bajos ingresos, quienes enfrentan barreras significativas para acceder a diagnósticos tempranos y tratamientos efectivos.

La auscultación pulmonar, una herramienta diagnóstica no invasiva, ha demostrado ser crucial en la identificación y seguimiento de diversas patologías respiratorias [4, 5]. Los sonidos respiratorios adventicios (SRA), como roncus, sibilancias y crepitantes, son indicadores comunes en enfermedades como el asma, la enfermedad pulmonar obstructiva crónica (EPOC), la enfermedad pulmonar intersticial, la fibrosis quística (FQ), las bronquiectasias no FQ, la tuberculosis pulmonar, la bronquiolitis y la neumonía. Estos hallazgos clínicos no solo guían la toma de decisiones médicas, como la prescripción de antibióticos o la derivación a especialistas, sino que también son esenciales para evaluar la evolución de los pacientes y monitorear la respuesta al tratamiento [6]. En el caso del asma, una de las patologías respiratorias más prevalentes, la mayoría de los casos se manifiestan clínicamente durante los primeros años de vida. En esta etapa, la obstrucción reversible de las vías aéreas es difícil de medir objetivamente, lo que complica el diagnóstico temprano y oportuno [7, 8, 9]. En consecuencia, las observaciones de los padres y cuidadores sobre síntomas como las sibilancias adquieren un rol crucial. Sin embargo, existe una brecha significativa entre las sibilancias percibidas con un estetoscopio por parte de los profesionales de la salud y aquellas identificadas por los cuidadores [10, 11]. Los sonidos respiratorios (SR), obtenidos mediante auscultación,

poseen características únicas que los convierten en una herramienta diagnóstica accesible, económica y efectiva. Además de ser no invasivos y cómodos para los pacientes, su adquisición no requiere equipos costosos, puede repetirse tantas veces como sea necesario y demanda una cooperación mínima del paciente. Estas cualidades los posicionan como un recurso valioso en entornos clínicos de recursos limitados. Sin embargo, la subjetividad inherente a la auscultación ha sido ampliamente reconocida como una limitación crítica, lo que ha impulsado el desarrollo de métodos computarizados para el análisis automatizado de sonidos respiratorios [12]. En las últimas décadas, la automatización de la clasificación de sonidos respiratorios anormales ha avanzado considerablemente, con una amplia investigación enfocada en métodos innovadores y tecnologías emergentes [13]. La mayoría de estos métodos sigue en fase de desarrollo y típicamente comprende dos etapas fundamentales: el preprocesamiento de los audios y la extracción de características relevantes. El preprocesamiento incluye la eliminación de ruidos no deseados para mejorar la calidad de los datos, utilizando técnicas como filtros pasa bandas. La extracción de características, por otro lado, se centra en identificar parámetros clave como coeficientes cepstrales de frecuencia Mel (MFCC´s) [14, 15], características espectrales, medidas de energía, entropía y coeficientes de wavelet. La precisión de estas técnicas se ve limitada no solo por la calidad de los datos, sino también por la subjetividad en la interpretación de los resultados, lo que refuerza la necesidad de una estandarización en las metodologías empleadas.

Entre los algoritmos de aprendizaje automático aplicados en esta área se encuentran las máquinas de soporte vectorial (SVM), las redes neuronales artificiales (ANN), los modelos de mezcla gaussiana (GMM) y las técnicas de regresión logística, entre otros [16]. Aunque estos enfoques han mostrado resultados satisfactorios en estudios controlados, su desempeño depende en gran medida del tamaño y la calidad de los conjuntos de datos utilizados para su entrenamiento. Una limitación común es la baja disponibilidad de datos en el ámbito médico, donde los estudios suelen incluir un número reducido de pacientes. Esto restringe la generalización de los algoritmos y plantea desafíos adicionales en términos de aplicabilidad clínica [17]. En este contexto, el uso de redes neuronales convolucionales (CNN) preentrenadas representa una solución estratégica. Estas arquitecturas, como la VGG-16 [18, 19], permiten capitalizar el conocimiento adquirido previamente a partir de grandes conjuntos de datos, reduciendo la carga computacional y los tiempos de entrenamiento. Al adaptar estas redes a problemas específicos mediante el ajuste de sus capas finales, es posible lograr un alto rendimiento con conjuntos de datos relativamente pequeños. Este enfoque minimiza el riesgo de sobreajuste y acelera el desarrollo de modelos, haciendo más eficiente su implementación en aplicaciones prácticas.

En este estudio, utilizamos la red preentrenada VGG-16 con el conjunto de datos de ImageNet, adaptada para procesar representaciones de MFCC´s extraídas de sonidos respiratorios. Nuestro objetivo es clasificar pacientes con o sin patologías respiratorias utilizando una base de datos proporcionada por un especialista en neumonología infantil del Departamento de Neumonología Pediátrica del Hospital Zonal “Dr. Andrés R. Isola” en Puerto Madryn, Argentina. Este enfoque tiene el potencial de mejorar significativamente la precisión y rapidez del diagnóstico clínico en enfermedades respiratorias pediátricas, particularmente en entornos de bajos recursos.

2 Metodología

2.1 Base de datos: Origen y Características

La creación de la base de datos de sonidos respiratorios fue llevada a cabo por un equipo de investigadores argentinos, quienes diseñaron un estudio exhaustivo para recolectar un amplio rango de sonidos pulmonares. Esta base de datos contiene un total de 342 grabaciones de audio, cada una con una duración de 20 segundos, realizadas bajo condiciones controladas y en entornos clínicos. Las grabaciones se obtuvieron en siete sitios específicos de auscultación, como se detalla en la **Fig. 1**: cinco ubicaciones en la parte frontal del tórax y dos en la parte dorsal. El estudio incluyó a 138 pacientes, entre los que se encontraban lactantes, niños y adolescentes, tanto con síntomas respiratorios como sin ellos. Se seleccionaron individuos con diversas patologías pulmonares para asegurar una muestra representativa de sonidos respiratorios. Los sitios de auscultación fueron definidos de la siguiente manera: los sitios 1 y 2 correspondieron a los lóbulos inferiores derecho e izquierdo; los sitios 3 y 4 se ubicaron en los lóbulos superiores derecho e izquierdo; los sitios 5 y 6 en el lóbulo medio derecho y la língula, respectivamente; mientras que el sitio 7 se localizó en la tráquea. Cada archivo de audio contiene entre 5 y 12 ciclos respiratorios completos, lo que resultó en un total acumulado de 1.54 horas de grabaciones.

El conjunto de audios abarca una amplia gama de sonidos respiratorios, tales como roncus, estridor, crepitantes y sibilancias. Algunas grabaciones incluyen simultáneamente varios tipos de sonidos, lo que enriquece la diversidad de la base de datos. Además, se incluyeron tanto grabaciones sin interferencias como grabaciones con ruido ambiental, con el fin de reproducir de manera realista las condiciones a las que se enfrentan los profesionales de la salud durante la auscultación en entornos clínicos reales.

La recolección de las grabaciones se llevó a cabo durante tres años consecutivos: en 2018 se registraron 103 grabaciones, en 2019 se obtuvieron 234, y en 2020 se sumaron otras 5, todas realizadas en el consultorio externo de neumonología infantil del Hospital Zonal "Dr. Andrés R. Isola" de Puerto Madryn, Chubut, Argentina. Este enfoque temporal permitió incorporar registros a lo largo de diferentes estaciones del año, lo que contribuye a la validez y diversidad del conjunto de datos. El Comité de Ética de la Investigación (CEI) perteneciente al Comité de Bioética del Área Programática Norte de la provincia de Chubut aprobó el estudio de investigación (N.º de ingreso 1510, con fecha 08/11/2018).

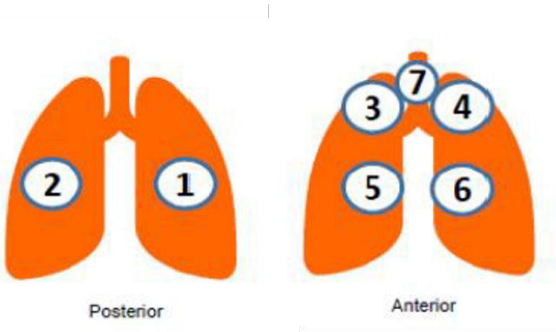

Fig. 1. Sitios de auscultación, elaboración propia.

Cada archivo de audio fue cuidadosamente etiquetado por un equipo de expertos con el fin de asegurar la precisión en la clasificación de los diferentes sonidos respiratorios presentes en las grabaciones. Para una identificación clara y uniforme, se estructuraron los nombres de los archivos utilizando cinco elementos separados por guiones bajos ("_"). Estos elementos incluyen el número de paciente, el índice de grabación, la ubicación específica de auscultación en el tórax, el modo de adquisición empleado —Secuencial o Canal Único (SC)—, y el tipo de equipo utilizado durante la recolección de datos. El equipo consistió en un micrófono Shure MVL-3M acoplado a un estetoscopio Littmann Classic II Pediatric Stethoscope (ShuLittC2 PS), garantizando así una alta calidad de sonido durante las auscultaciones. Posteriormente, las grabaciones fueron clasificadas en dos grupos según la presencia o ausencia de sibilancias detectadas: el grupo con presencia de sibilancias se etiquetó con un valor de 1, mientras que el grupo sin sibilancias se marcó con un valor de 0. Esta categorización binaria permite un análisis más preciso de los sonidos respiratorios patológicos y facilita su uso en futuros estudios de machine learning orientados a la detección automática de anomalías respiratorias.

2.2 Preprocesamiento de auscultaciones

Dado que los audios grabados en entornos clínicos suelen presentar diversos tipos de ruido, fue necesario implementar un proceso de preprocesamiento detallado para mejorar la calidad de los datos y asegurar la precisión en el análisis de las señales respiratorias. Los ruidos presentes incluyen tanto los generados por el entorno, como conversaciones y movimientos, así como sonidos internos del cuerpo, tales como latidos cardíacos, ruidos intestinales y aquellos producidos por la expansión y contracción de la caja torácica. Para garantizar la precisión en el análisis de estas grabaciones, se llevó a cabo un proceso de preprocesamiento (ver **Fig. 2**). En primer lugar, los archivos fueron resampleados a una frecuencia de muestreo de 4000 Hz (4 kHz). Esta frecuencia fue seleccionada estratégicamente para optimizar el análisis al reducir la complejidad computacional y enfocarse en el rango de frecuencias clave donde se encuentran los sonidos respiratorios. Este ajuste permite mantener un equilibrio adecuado entre la calidad del audio y la eficiencia del procesamiento. A continuación, se diseñó y aplicó un

filtro Butterworth pasa bandas de quinto orden, con frecuencias de corte entre 100 Hz y 400 Hz, para eliminar eficazmente los ruidos de fondo y las interferencias de baja y alta frecuencia, preservando únicamente las señales respiratorias de interés. Este proceso fue llevado a cabo mediante el uso de una librería de Python especializada en la manipulación y procesamiento de señales de audio. La aplicación de un filtro de este tipo asegura la uniformidad y calidad de los datos procesados, facilitando su análisis en etapas posteriores.

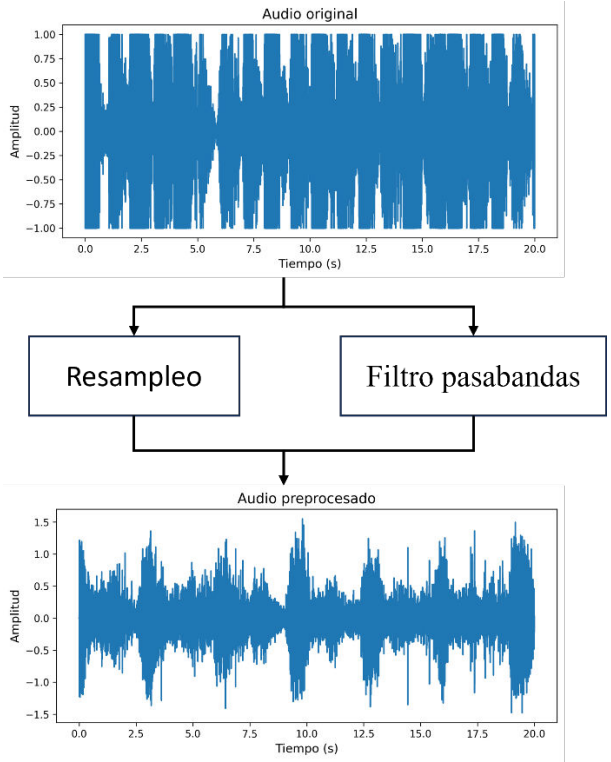

Fig. 2. Etapas de preprocesamiento.

2.3 Extracción de características

Los MFCC´s se destacan como una de las herramientas más eficaces para la extracción de características en señales de audio de auscultación. Esta técnica permite representar el espectro de audio de manera similar a cómo el oído humano percibe las frecuencias, lo cual es especialmente relevante para detectar sibilancias, ya que estas corresponden a sonidos de alta frecuencia con un perfil tonal específico. En la **Fig. 3** se muestra un ejemplo comparativo de los MFCC´s obtenidos de un paciente sin patología y otro con alguna condición respiratoria.

Si bien existen otros métodos de extracción de características, como la transformada de wavelet, el espectrograma de potencia y la representación en el dominio del tiempo, los MFCC´s son particularmente adecuados para esta tarea, ya que dividen las frecuencias de manera lineal en las bajas frecuencias y logarítmica en las altas, lo que permite una mejor detección de las variaciones espectrales características de los sonidos respiratorios anómalos. Este enfoque resulta especialmente ventajoso cuando se trabaja con señales que contienen componentes de alta frecuencia, como las sibilancias. Es así como se extrajeron 13 coeficientes MFCC por cada archivo de audio. Estos coeficientes representan las principales características espectrales de la señal, donde los valores más bajos corresponden a las frecuencias fundamentales, mientras que los coeficientes superiores contienen información detallada sobre los armónicos y los componentes de menor energía. Una vez calculados, los coeficientes se normalizaron y se transformaron en imágenes mediante interpolación de vecinos más cercanos, generando una representación visual de tamaño 224x224 píxeles. Esta conversión fue realizada con el objetivo de utilizar las imágenes como entrada en una red neuronal convolucional preentrenada (VGG-16), la cual ha demostrado un buen desempeño en tareas de clasificación de imágenes. Además, se aplicó una ventana de tamaño 2048 muestras para calcular la Transformada Rápida de Fourier (FFT), lo que permitió obtener un espectrograma de alta resolución. El desplazamiento entre ventanas sucesivas, determinado por el parámetro tamaño de salto, fue configurado en 512 muestras. Este valor fue seleccionado para lograr un equilibrio entre una adecuada resolución temporal y el costo computacional. Al utilizar un desplazamiento entre ventanas intermedio, se asegura que las características espectrales esenciales de las señales se capturen con precisión sin generar una carga computacional excesiva.

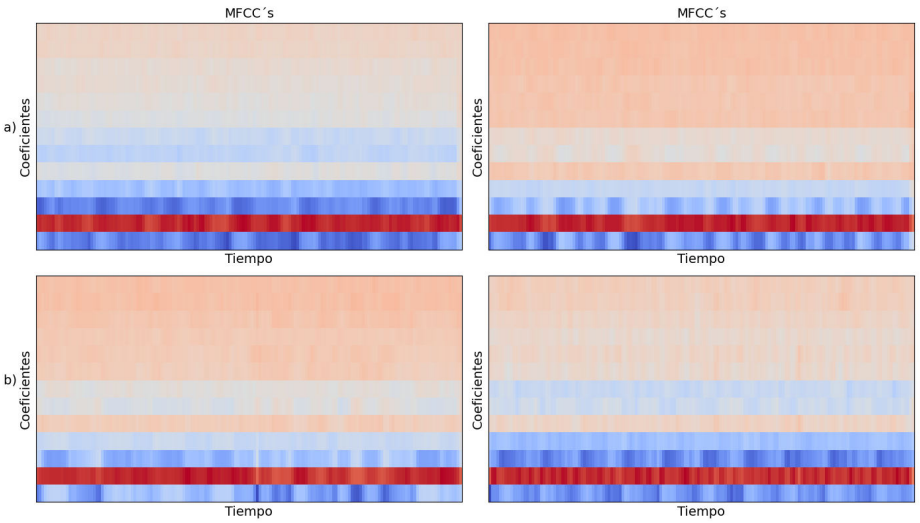

Fig. 3. Coeficientes MFCC´s obtenidos para a) paciente que no presenta patología, b) paciente con patología respiratoria.

2.4 Modelo de entrenamiento

Se empleó la arquitectura CNN preentrenada VGG-16 con el conjunto de datos "ImageNet", la cual posee un total de 17,934,401 parámetros, de los cuales 14,714,688 son no entrenables y 3,219,713 son parámetros de entrenamiento. Este modelo destaca por su habilidad para capturar características significativas y detalladas de las imágenes gracias a su arquitectura con múltiples capas convolucionales. La red VGG-16 resulta especialmente eficiente en el reconocimiento de patrones complejos y en la extracción de características a distintos niveles de abstracción, lo que la convierte en una opción adecuada para tareas de clasificación de imágenes. Para adecuar el modelo a los requerimientos específicos de este estudio, se llevaron a cabo diversas modificaciones en las capas finales de la red. En lugar de utilizar las capas de salida predeterminadas del modelo VGG-16, se diseñó e implementó una nueva estructura de red con el objetivo de optimizar el rendimiento en la tarea específica de clasificación binaria. Estas modificaciones permitieron ajustar el modelo a las características propias del conjunto de datos empleado, logrando una mejor adaptación a la naturaleza de las señales analizadas. La estructura resultante se muestra en la **Fig. 4**, donde se observa el diseño final de las capas adaptadas para esta tarea.

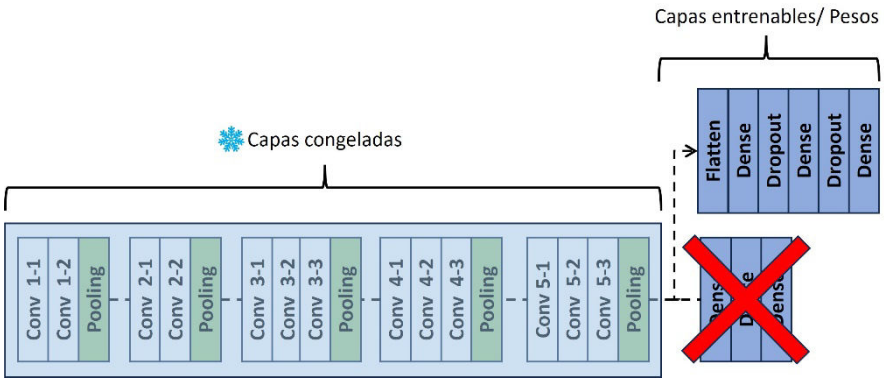

Fig. 4. Modelo VGG-16 adaptado.

Las modificaciones realizadas en la arquitectura incluyen varias capas densas y de regularización diseñadas específicamente para mejorar el rendimiento del modelo en la tarea de clasificación binaria. El flujo de la red comienza con la entrada del modelo preentrenado VGG-16, seguida por una serie de capas adicionales destinadas a refinar las características extraídas por las capas convolucionales. A continuación, se describen las modificaciones clave:

1. **Capa Flatten**: Se incluyó una capa de aplanamiento (Flatten) que convierte el volumen tridimensional de salida del modelo VGG-16 en un vector unidimensional. Este paso es esencial para conectar las capas convolucionales con las capas densas posteriores.

2. **Capa Densa de 128 Neuronas**: Se añadió una capa densa con 128 neuronas y función de activación ReLU, cuyo objetivo es aprender representaciones de mayor nivel a partir de las características extraídas por el modelo VGG-16. Esta capa ayuda a capturar patrones complejos, mejorando la capacidad de discriminación del modelo.
3. **Capa Dropout al 50%**: Para reducir el riesgo de sobreajuste, se implementó una capa de Dropout con una tasa del 50%. Esta técnica desactiva aleatoriamente la mitad de las neuronas durante cada iteración de entrenamiento, lo que mejora la capacidad del modelo para generalizar a nuevos datos.
4. **Capa Densa de 64 Neuronas**: Posteriormente, se agregó una segunda capa densa con 64 neuronas y función de activación ReLU. Esta capa proporciona un refinamiento adicional de las características aprendidas, permitiendo al modelo enfocarse en detalles específicos que contribuyen a una mejor clasificación.
5. **Capa Dropout Adicional**: Se incorporó una segunda capa de Dropout al 50% después de la capa de 64 neuronas para reforzar aún más la capacidad de generalización del modelo.
6. **Capa de Salida con Activación Sigmoide**: Finalmente, se añadió una capa densa de salida con una sola neurona y función de activación sigmoide. Esta configuración es adecuada para la clasificación binaria, ya que la función sigmoide transforma la salida del modelo en un valor de probabilidad, lo que facilita la interpretación del resultado como la probabilidad de que una entrada pertenezca a una de las dos clases.

La secuencia de capas descrita en la **Fig. 4**, garantiza que el modelo sea capaz de aprender representaciones profundas y relevantes de las señales de entrada, optimizando el rendimiento en la tarea de clasificación binaria específica. Se empleó el modelo VGG-16 preentrenado, aprovechando su capacidad para extraer características generales de manera eficiente. Las capas añadidas fueron diseñadas para adaptar el modelo a las particularidades del conjunto de datos de este estudio, lo que permitió afinar su desempeño en la tarea de clasificación de pacientes según la presencia o ausencia de patologías respiratorias. Para el entrenamiento, se utilizó una laptop equipada con un procesador Intel(R) Core (TM) i7-9750H CPU @ 2.60GHz, 16GB de RAM y una tarjeta gráfica RTX 2070 HQ. El modelo VGG-16 se empleó inicialmente en una tarea de clasificación binaria, donde la clase 0 corresponde a pacientes sin patología y la clase 1 a pacientes con patología respiratoria. Se mantuvo la arquitectura convolucional preentrenada, realizando modificaciones únicamente en las capas finales, específicamente en la capa de salida, para ajustarla a la nueva tarea de clasificación. Tras estas modificaciones, el modelo se reentrenó utilizando las representaciones de los MFCCs generadas a partir de las señales de audio de auscultación. Este proceso resultó en un modelo ajustado y optimizado para la detección de patrones relevantes en las representaciones paramétricas de audio. Con el fin de asegurar una evaluación precisa y evitar el sobreajuste, el conjunto de datos se dividió en tres subconjuntos: entrenamiento

(80%), validación (10%) y prueba (20%). Esta estrategia permitió monitorear el desempeño del modelo en cada etapa del proceso y garantizar una generalización adecuada a nuevos datos.

3 Resultados

La implementación de una CNN preentrenada VGG-16 para la clasificación de sonidos respiratorios ha arrojado resultados prometedores. El modelo alcanzó una exactitud del 79% en la clasificación de sonidos respiratorios en lactantes, niños y adolescentes, demostrando una capacidad significativa para distinguir entre diferentes patrones respiratorios, lo que sugiere su viabilidad en aplicaciones clínicas. En la matriz de confusión (**Tabla 1**), se observa que, del total del conjunto de prueba, se obtuvieron 13 verdaderos negativos (VN), 3 falsos positivos (FP), 4 falsos negativos (FN) y 14 verdaderos positivos (VP). Estos resultados resaltan la eficacia del modelo en la clasificación de las distintas categorías de sonidos respiratorios.

Tabla 1. Matriz de Confusión del Modelo Clasificador en el Conjunto de Pruebas.

		Clase 0	Clase 1
Real	Clase 0	**VN** 13	**FP** 3
	Clase 1	**FN** 4	**VP** 13
		Predicción	

A pesar de ciertas confusiones observadas entre clases específicas, la mayoría de las predicciones coinciden con las etiquetas reales, lo que respalda la robustez del modelo. El análisis de las curvas de pérdida y precisión (**Fig. 5**) indica que el modelo no presenta signos evidentes de sobreajuste ni de subajuste. No obstante, se observó que el modelo alcanza un rendimiento óptimo entre las 25 y 30 épocas, lo que sugiere un proceso de entrenamiento eficiente. Sin embargo, es importante interpretar estas curvas con cautela, ya que la cantidad limitada de datos disponibles podría estar afectando los resultados. En términos de métricas de clasificación, la precisión para la Clase 0 fue de 76%, mientras que para la Clase 1 alcanzó un 81%. La precisión mide la proporción de verdaderos positivos entre el total de predicciones positivas realizadas por el modelo, indicando una moderada capacidad para minimizar los falsos positivos. El recall o sensibilidad fue de 81% para la Clase 0 y 76% para la Clase 1, lo que refleja la capacidad del modelo para detectar la mayoría de las instancias positivas, aunque con algunas limitaciones en ambas clases. El F1-score, que representa una medida equilibrada entre precisión y recall, fue de 0.79 para ambas clases, lo que indica un buen balance general en el desempeño del modelo. El área bajo la curva ROC (AUC-ROC) obtenida fue de 0.79 (**Fig. 6**), lo que sugiere un buen rendimiento en la clasificación de sonidos respiratorios, al mostrar una tasa aceptable de verdaderos positivos frente a la tasa de falsos

positivos. En otras palabras, el modelo es capaz de distinguir adecuadamente entre los sonidos respiratorios de pacientes con y sin patología en la mayoría de los casos.

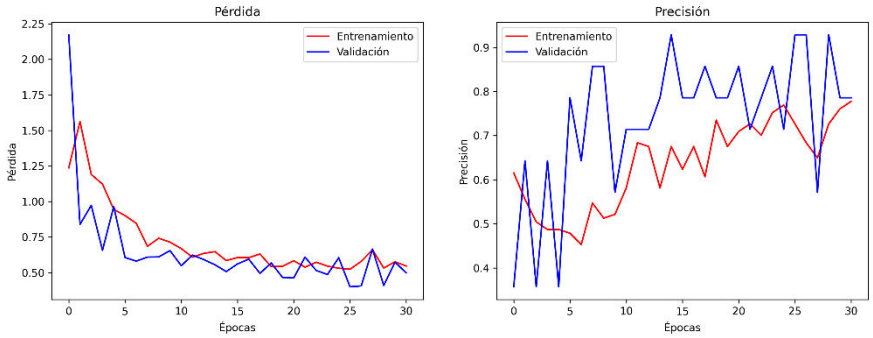

Fig. 5. Graficas representativas del rendimiento del modelo.

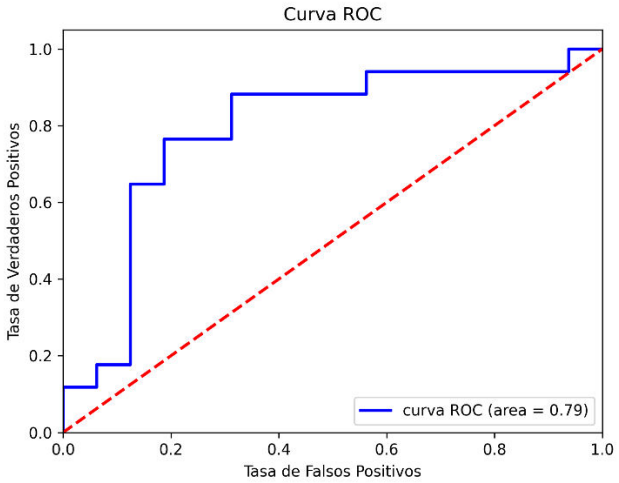

Fig. 6. Curva ROC utilizada para evaluar el rendimiento del modelo en la clasificación de sibilancias.

Finalmente, se realizaron análisis adicionales para evaluar el desempeño del modelo en distintos subconjuntos de datos, considerando factores como la edad de los pacientes, el tipo de afección respiratoria y la ubicación torácica de los sonidos auscultados. Estos análisis proporcionan información relevante sobre la capacidad del modelo para generalizar en diversos contextos clínicos y pueden ser útiles para orientar mejoras futuras en su arquitectura. En conjunto, estos resultados sugieren que la aplicación de la red neuronal convolucional VGG-16 pre entrenada en la clasificación de sonidos respiratorios tiene el potencial de mejorar la precisión y la eficiencia en el diagnóstico de

enfermedades pulmonares, lo que podría tener un impacto significativo en la práctica clínica y la atención al paciente. Sin embargo, se requieren estudios adicionales para validar y optimizar aún más el modelo en diferentes poblaciones y entornos clínicos.

4 Discusión

Recientemente, Huang y colaboradores [22] describieron métodos de aprendizaje profundo aplicados al análisis de sonidos pulmonares. Hasta la fecha de esta publicación, solo tres estudios [23, 24, 25] han empleado MFCCs en dos tareas médicas específicas: la detección de sonidos anormales (ASD, por su sigla en inglés) y el reconocimiento de enfermedades respiratorias (RDR, por su sigla en inglés). Estos estudios han explorado enfoques diversos para la clasificación de sonidos respiratorios, evidenciando que el uso de MFCCs es una técnica prometedora para la extracción de características relevantes en señales acústicas. Destacamos que nuestro conjunto de datos, enfocado en lactantes, niños y adolescentes, es el primero de su tipo en implementar este método para la extracción de características en el análisis de audios de auscultación, lo que contribuye de manera significativa al avance en el campo del análisis acústico de sonidos respiratorios pediátricos. Si bien el valor de exactitud del 79% sugiere un rendimiento satisfactorio del modelo, aún existe margen de mejora en su capacidad de discriminación, especialmente en escenarios donde los sonidos respiratorios pueden ser muy similares entre clases. Estas mejoras podrían lograrse mediante la optimización de los hiperparámetros del modelo, como la tasa de aprendizaje y el número de neuronas en las capas ocultas, así como mediante el ajuste de la arquitectura de la red neuronal, por ejemplo, añadiendo más capas convolucionales o utilizando mecanismos de atención que permitan al modelo centrarse en patrones acústicos específicos. Además, la incorporación de un mayor volumen de datos de entrenamiento, incluyendo sonidos respiratorios de diferentes entornos y equipos de grabación, podría contribuir a una mejor generalización del modelo en condiciones clínicas reales.

Este tipo de investigaciones tiene importantes aplicaciones en el ámbito médico, particularmente en el diagnóstico y monitoreo remoto de pacientes pediátricos y adultos con enfermedades respiratorias, lo que podría facilitar la detección temprana de patologías respiratorias en regiones con acceso limitado a especialistas. Asimismo, futuras aplicaciones pueden incluir el uso del modelo como herramienta educativa, permitiendo la evaluación de competencias y el examen de habilidades de estudiantes y profesionales de la salud en la auscultación pulmonar, con el fin de estandarizar y mejorar la calidad de la atención médica relacionada con enfermedades respiratorias.

Finalmente, dado el desafío inherente en la clasificación de sonidos respiratorios y las limitaciones impuestas por la cantidad de datos, es fundamental continuar desarrollando modelos más robustos que no solo mejoren la exactitud global, sino que también reduzcan las tasas de falsos positivos y falsos negativos, garantizando un mejor desempeño en escenarios clínicos diversos. Este esfuerzo podría incluir el uso de técnicas avanzadas de aumento de datos, como la generación sintética de sonidos respiratorios mediante redes generativas adversarias (GANs), así como la combinación de múltiples

representaciones acústicas, como MFCCs, espectrogramas y características cepstrales diferenciales, lo que permitiría al modelo capturar una mayor cantidad de información relevante para la clasificación.

5 Conclusiones

Los resultados obtenidos en este estudio respaldan la eficacia de la implementación de la red neuronal convolucional (CNN) preentrenada VGG-16 en la clasificación de sonidos respiratorios en lactantes, niños y adolescentes, logrando una precisión del 79%. Estos hallazgos son alentadores y sugieren que el uso de inteligencia artificial en el análisis automatizado de sonidos respiratorios tiene el potencial de mejorar significativamente tanto el diagnóstico como la gestión de enfermedades pulmonares en este grupo etario. La capacidad del modelo para diferenciar entre diversos patrones respiratorios, incluso en presencia de ruido ambiental, resalta su robustez y su viabilidad en entornos clínicos reales. La matriz de confusión muestra un desempeño general aceptable del modelo, aunque con ciertas confusiones entre clases que podrían reducirse mediante una mayor optimización de los hiperparámetros y ajustes adicionales en el algoritmo. Asimismo, los análisis complementarios realizados para evaluar el desempeño del modelo en diferentes subconjuntos de datos aportan información clave sobre su capacidad de generalización en múltiples escenarios clínicos. Estos hallazgos son esenciales para identificar posibles sesgos y limitaciones del modelo, proporcionando una base sólida para guiar futuras mejoras en su arquitectura y su implementación.

En este sentido, las investigaciones futuras podrían centrarse en ampliar los conjuntos de datos utilizados, incorporando muestras provenientes de distintos dispositivos de auscultación y en diferentes condiciones clínicas. Esto no solo mejoraría la generalización del modelo, sino que también fortalecería su aplicabilidad en situaciones más diversas. La integración de nuevas características fisiológicas, como información demográfica y antecedentes médicos, podría contribuir a aumentar aún más la precisión y la sensibilidad del modelo en la detección de enfermedades respiratorias. Este estudio subraya el notable potencial de la inteligencia artificial, y en particular de las redes neuronales convolucionales preentrenadas, para optimizar el diagnóstico y el manejo de enfermedades respiratorias mediante el análisis automatizado de sonidos respiratorios. Las aplicaciones de esta tecnología incluyen, además del diagnóstico clínico, el monitoreo remoto de pacientes, lo que resulta especialmente valioso en áreas con acceso limitado a profesionales de la salud. Adicionalmente, se vislumbran futuros usos en la formación médica, permitiendo la evaluación y mejora de las habilidades de auscultación de estudiantes y profesionales del sector sanitario.

En conclusión, aunque los resultados obtenidos son prometedores, es necesario continuar con investigaciones adicionales para validar el modelo en escenarios clínicos más variados y ampliar el conjunto de datos disponible. Esto garantizaría una mejor generalización del modelo y permitiría reducir las confusiones restantes. La incorporación de herramientas de inteligencia artificial en la práctica clínica diaria representa una vía prometedora para mejorar el diagnóstico precoz de enfermedades respiratorias, facilitando una intervención temprana y orientada a los síntomas, lo que podría resultar en tratamientos más rápidos y eficaces.

Agradecimientos. Nos gustaría agradecer a todos los participantes en el primer estudio de Puerto Madryn, así como al personal del hospital, para hacer posible esta investigación.

Referencias

1. World Health Organization. The top 10 causes of death. (2017).
2. Landau LI, Taussig LM: Early childhood origins and Economic impact of respiratory disease throughout life. En: Pediatric Respiratory Medicine (Second Edition), editado por Taussig LM, Landau LI, 1-8. Mosby, Philadelphia (2008).
3. Gibson GJ, Loddenkemper R, Lundbäck B, Sibille Y: Respiratory health and disease in Europe: the new European Lung White Book. Eur Respir J. 42, 559–563 (2013).
4. Marques A, Oliveira A, Jácome C: Computerized adventitious respiratory sounds as outcome measures for respiratory therapy: a systematic review. Respir Care. 59, 765–776 (2014).
5. Marques A, Jácome C: Breath sounds from basic science to clinical practice, editado por Priftis KN, Hadjileontiadis LJ, Everard ML, 291-304. Springer, Switzerland (2018).
6. Aviles-Solis JC, Jácome C, Davidsen A, Einarsen R, Vanbelle S, Pasterkamp H, Melbye H: Prevalence and clinical associations of wheezes and crackles in the general population: the Tromsø study. BMC Pulm Med. 19, 173 (2019).
7. Saglani S, Payne DN, Zhu J, et al: Early detection of airway wall remodeling and eosinophilic inflammation in preschool wheezers. Am J Respir Crit Care Med. 176, 858–864 (2007).
8. Saglani S, Malmstrom K, Pelkonen AS, et al: Airway remodeling and inflammation in symptomatic infants with reversible airflow obstruction. Am J Respir Crit Care Med. 171, 722–727 (2005).
9. Brand PL, Baraldi E, Bisgaard H, et al: Definition, assessment and treatment of wheezing disorders in preschool children: an evidence-based approach. Eur Respir J. 32, 1096–1110 (2008).
10. Kelada L, Molloy CJ, Hibbert P, Wiles LK, Gardner C, Klineberg E, Braithwaite J, Jaffe A: Child and caregiver experiences and perceptions of asthma self-management. NPJ Prim Care Respir Med. 31, 42 (2021).
11. Zhjeqi V, Kundi M, Shahini M, Ahmetaj H, Ahmetaj L, Krasniqi S: Correlation between parents and child's version of the child health survey for asthma questionnaire. Eur Clin Respir J. 10, 2194165 (2023).
12. Rocha B M, Filos D, Mendes L, Vogiatzis I, Perantoni E, Kaimakamis E, Natsiavas P, Oliveira A, Jácome C, Marques A, Paiva RP, Chouvarda I, Carvalho P, & Maglaveras N: A Respiratory Sound Database for the Development of Automated Classification. (2017).
13. Pramono RXA, Bowyer S, Rodriguez-Villegas E: Automatic adventitious respiratory sound analysis: A systematic review. PLoS One. 12, e0177926 (2017).
14. Liu GK: Evaluating Gammatone Frequency Cepstral Coefficients with Neural Networks for Emotion Recognition from Speech. Informe de investigación. Ravenwood High School, Brentwood, TN 37027 (2018).
15. Wei H, Chan C, Choy C, Pun P: An efficient MFCC extraction method in speech recognition. En: Circuits and Systems (2006).
16. Acharya J, & Basu A: Deep Neural Network for Respiratory Sound Classification in Wearable Devices Enabled by Patient Specific Model Tuning. IEEE Transactions on Biomedical Engineering (2020).

17. Chamberlain D, Kodgule R, Ganelin D, Miglani V, Fletcher RR: Application of Semi-Supervised Deep Learning to Lung Sound Analysis. 38th Annu Int Conf IEEE Eng Med Biol Soc; 2016;804–7.
18. Simonyan K, & Zisserman A: Very Deep Convolutional Networks for Large-Scale Image Recognition (2015).
19. Xia T, Han J, & Mascolo C: Exploring machine learning for audio-based respiratory condition screening: A concise review of databases, methods, and open issues. Department of Computer Science and Technology, University of Cambridge, 15 JJ Thomson Avenue, Cambridge CB3 0FD, Reino Unido (2022).
20. Kochetov K, Putin E, Azizov S, Skorobogatov I, & Filchenkov A: Wheeze Detection Using Convolutional Neural Networks (2017).
21. Chang G-C, Lai Y-F: Performance evaluation and enhancement of lung sound recognition system in two real noisy environments. Comput Methods Prog Biomed. 97(2):141–150 (2010).
22. Huang DM, Huang J, Qiao K, et al. Deep learning-based lung sound analysis for intelligent stethoscope. Military Med Res 10, 44 (2023). https://doi.org/10.1186/s40779-023-00479-3.
23. Messner E, Fediuk M, Swatek P, Scheidl S, Smolle-Juttner FM, et al. Crackle and breathing phase detection in lung sounds with deep bidirectional gated recurrent neural networks. In: 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). Honolulu, HI, USA; 2018. p. 356–9.
24. Sengupta N, Sahidullah M, Saha G. Lung sound classification using cepstral-based statistical features. Comput Biol Med. 2016;75: 118–29.
25. Perna D, Tagarelli A. Deep auscultation: predicting respiratory anomalies and diseases via recurrent neural networks. En: 2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS). Córdoba, Spain; 2019. p. 50–5.