

## Diagnóstico Automatizado de Auscultación Pulmonar Pediátrica

Jorge I. Lopez Perez<sup>1</sup>[0000-0001-7132-676X], Damián L. Taire<sup>2</sup> <sup>3</sup>[0000-0001-6505-1560] y  
Claudio Delrieux<sup>4</sup>[0000-0002-2727-8374]

<sup>1</sup> Universidad Nacional Del Sur, Instituto de Ciencias e Ingeniería de la Computación - ICIC (UNS-CONICET), Av. Alem 1253 (B8000CPB), Bahía Blanca, Buenos Aires, Argentina

<sup>2</sup> Instituto Patagónico de Ciencias Sociales y Humanas, Centro Nacional Patagónico (IPCSH), Consejo Nacional de Investigaciones Científicas y Técnicas, Puerto Madryn, Argentina

<sup>3</sup> Departamento de Neumonología Pediátrica, Hospital Zonal “Dr. Andrés R. Isola”, Puerto Madryn, Argentina

<sup>4</sup> 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

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 realizó 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 78% 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.

## 1 Introducción.

Las enfermedades respiratorias a lo largo de la vida son una carga global para los individuos, las familias, los servicios de atención médica y las sociedades. El impacto incluye costos directos de servicios de salud, costos sociales y de funcionamiento familiar, años de vida perdidos, pérdida de productividad por días de ausentismo escolar y laboral [1 2 3]. La auscultación pulmonar ha demostrado ser útil en el diagnóstico y seguimiento de diversas enfermedades [4 5]. Sonidos respiratorios adventicios (SRA), como roncus, sibilancias y crepitantes, se asocian con enfermedades comunes como asma, enfermedad pulmonar obstructiva crónica, enfermedad pulmonar intersticial, fibrosis quística (FQ), bronquiectasias no FQ, tuberculosis pulmonar, bronquiolitis y neumonía. Los hallazgos positivos durante la auscultación influyen en las decisiones clínicas tales como la frecuencia de prescripciones de antibióticos y derivaciones a atención especializada [6]. La mayoría de los casos de asma tienen su primera manifestación clínica durante los primeros años de vida cuando la obstrucción reversible de la vía aérea es difícil de medir [7, 8, 9].

Por lo tanto, mucho se basa en observaciones de segunda mano de parte de los padres y cuidadores y su reporte de síntomas, incluyendo sibilancias. Existe una falta de conocimiento de la relación entre sibilancias escuchadas con el estetoscopio y las que escuchan los padres/cuidadores [10, 11]. Los sonidos respiratorios (SR) adquiridos mediante auscultación están disponibles casi universalmente, son económicos y no invasivos, cómodos (no es necesario tolerar una mascarilla facial o un sello alrededor de una boquilla) y rentables, pueden repetirse tantas veces como sea necesario y requiere una cooperación mínima del paciente y capacitación de los profesionales de la salud. Sin embargo, la subjetividad de la auscultación es ampliamente reconocida, lo que ha llevado a una nueva era de desarrollos en técnicas computarizadas para la adquisición y análisis de sonidos respiratorios [12].

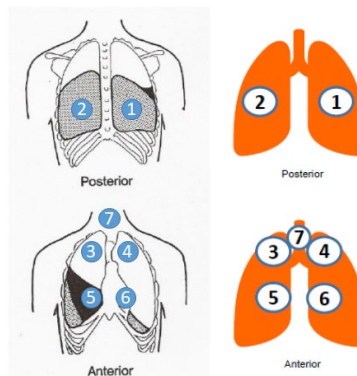
En las últimas décadas, la automatización de la clasificación de sonidos respiratorios anormales ha sido objeto de amplia investigación. En este sentido, se han recopilado y analizado métodos utilizados con mayor frecuencia en investigaciones [13]. La mayoría de estos métodos aún están en fase de desarrollo. Este proceso de desarrollo suele constar de dos etapas principales: la primera es el preprocesamiento de los audios obtenidos por especialistas: Esta etapa implica el uso de diversas técnicas para eliminar ruidos no deseados de las señales, con el fin de mejorar la calidad de los datos. La segunda es la extracción de características: Una vez procesados los datos, se extraen características necesarias para la detección o clasificación. Posteriormente, se aplican técnicas específicas sobre estas características extraídas. Entre las características más comúnmente empleadas en la literatura se incluyen los coeficientes cepstrales de frecuencia Mel (MFCC's) [14, 15], características espectrales, mediciones de energía, entropía y coeficientes de wavelet. Algunos algoritmos de aprendizaje automático implementados en la literatura están basados en reglas empíricas, tales como las máquinas de soporte vectorial (SVM), redes neuronales artificiales (ANN) [16], modelos de mezcla Gaussiana (GMM), regresión logística, entre otros. Gran parte del éxito en la clasificación de estos algoritmos depende del volumen de datos con el cual estos son alimentados, pero el reducido número de pacientes que se emplean en los estudios hace que esta tarea se

complice un poco más. Esto no significa que no se obtengan resultados satisfactorios en la clasificación implementando estos algoritmos ya que estos pueden diseñarse, ajustarse a la medida de los datos y sus características obtenidas de un número limitado de pacientes. No obstante, cuando el número de pacientes aumenta en grandes escalas pudiendo llegar incluso hasta los cientos, estas características aprendidas de conjuntos de datos pequeños tienden a no aplicarse de manera generalizada [17]. Por esta razón, el uso de redes neuronales convolucionales (CNN) pre entrenadas, se vuelve especialmente ventajoso al enfrentarse a conjuntos de datos de diversos tamaños. Las arquitecturas pre entrenadas, como la VGG-16 [18, 19], son ampliamente empleadas en este tipo de trabajos debido a su capacidad para adaptarse a diferentes escenarios. Esta elección estratégica no solo contribuye a reducir la carga computacional y los tiempos de entrenamiento, sino que también permite capitalizar el conocimiento previamente adquirido por la red a partir de conjuntos de datos extensos. Al utilizar una red pre entrenada como base, se puede comenzar con parámetros que ya tienen cierto grado de optimización, lo que agiliza el proceso de ajuste a los datos específicos del problema en cuestión [20, 21]. Esto conlleva una mejora significativa en la eficiencia y precisión del modelo, sin la necesidad de realizar un entrenamiento desde cero. Además, al evitar la necesidad de entrenar desde el principio, se reduce el riesgo de sobreajuste y se acelera el ciclo de desarrollo del modelo, lo que resulta en una implementación más rápida y efectiva en aplicaciones enfocadas al diagnóstico de patologías respiratorias permitiendo así a los especialistas poder generar diagnósticos más precisos. Este estudio se centra en la utilización de una red neuronal pre entrenada VGG-16 con el conjunto de datos de ImageNet, con el objetivo de extraer características MFCC's para determinar pacientes con o sin patologías respiratorias. Estas características se utilizan para clasificar los sonidos respiratorios, 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.

## 2 Metodología.

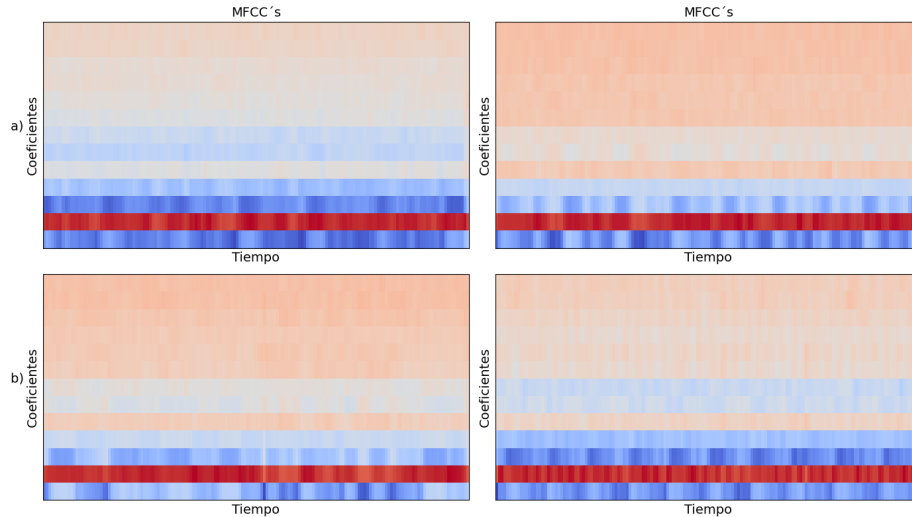
El equipo de investigación argentino lideró la creación de una base de datos de sonidos respiratorios, que comprende 342 grabaciones con una duración de 20 segundos cada una. Estas grabaciones se llevaron a cabo en 7 ubicaciones específicas de auscultación (cinco en el frente y dos en el dorso) como se puede observar en la Fig. 1, con la participación de 138 pacientes menores de 18 años, tanto asintomáticos como sintomáticos, incluyendo lactantes, niños y adolescentes, con diversas condiciones respiratorias. Los sitios 1 y 2 corresponden a los lóbulos inferiores derecho e izquierdo, respectivamente; los sitios 3 y 4 corresponden a los lóbulos superiores derecho e izquierdo, respectivamente; y los sitios 5 y 6 al lóbulo medio derecho y lingula, respectivamente; y el sitio 7 a la tráquea. Cada grabación contiene entre 5 y 12 ciclos respiratorios, lo que suma un total de 1.54 horas de grabaciones, abarcando un espectro de sonidos respiratorios como estridor, roncus, crepitantes y sibilancias, con algunas grabaciones que contienen múltiples tipos de sonidos. Se incluyeron tanto grabaciones de sonidos respiratorios sin

interferencias como aquellas con ruido, replicando así las condiciones en las que los especialistas trabajan en entornos clínicos. Los datos fueron recopilados en el consultorio externo de neumonología infantil del Hospital Zonal "Dr. Andrés R. Isola" en Puerto Madryn, Chubut, Argentina, durante los años 2018 (103 grabaciones), 2019 (234 grabaciones) y 2020 (5 grabaciones). 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 de los autores.

Cada archivo de audio fue etiquetado por expertos para garantizar la precisión de las etiquetas asignadas a los sonidos respiratorios. Se estructuraron los nombres de los archivos en cinco elementos separados por guiones bajos ("\_"), que incluyen el número de paciente, el índice de grabación, la ubicación torácica, el modo de adquisición (Secuencial/Canal Único (SC)), y el equipo de grabación (Shure MVL-3M Littmann Classic II Pediatric Stethoscope (ShuLittC2 PS)). Posteriormente, se dividieron en dos grupos según la presencia o ausencia de sibilancias (presencia=1, ausencia=0). Dado que los audios grabados en condiciones clínicas típicamente contienen ruido, se procedió a preprocesar los archivos de audio para garantizar la calidad de los datos. Primero, se resampearon los archivos de audio a 4Khz. Luego, se diseñó un filtro Butterworth pasa bandas de quinto orden con frecuencias de corte entre 100-400 Hz para eliminar los ruidos ambientales mientras se conservaban las señales de interés. Este proceso se llevó a cabo utilizando la librería de Python para manipulación de archivos de audio. Los MFCC's se destacan como herramientas efectivas para extraer características al analizar archivos de audio de auscultación, se puede observar un ejemplo de MFCC's tanto de un paciente que no presenta patología comparado con uno que si posee alguna condición **Fig. 2**. Aunque existen otras técnicas de extracción de características como la transformada de wavelet, el espectrograma y la representación en el dominio del tiempo, los MFCC's son particularmente adecuados para esta tarea. En este estudio, se emplearon los MFCC's, generando 13 coeficientes para cada archivo de audio. Estos coeficientes fueron transformados en imágenes para ser utilizados como entrada en una red pre entrenada VGG-16.



**Fig. 2.** Coeficientes MFCC's obtenidos para a) paciente que no presenta patología, b) paciente con patología respiratoria.

Se empleó la arquitectura CNN pre entrenada VGG-16, utilizando imágenes que representan los MFCC's como entrada de datos. Para el entrenamiento, se utilizó una laptop equipada con un procesador Intel(R) Core (TM) i7-9750H CPU @ 2.60GHz, con 16GB de RAM y una tarjeta gráfica RTX 2070 HQ. Inicialmente, el modelo VGG-16 pre entrenado se empleó en una tarea de clasificación binaria, donde la clase 0 representa a pacientes sin patología y la clase 1 a pacientes con patología respiratoria. Se conservó la arquitectura pre entrenada del modelo, sustituyendo únicamente la capa de salida para adaptarla a la nueva tarea. Posteriormente, el modelo se reentrenó con la nueva capa de salida, lo que resultó en un modelo afinado para procesar las representaciones de MFCC's. Para garantizar una evaluación adecuada y prevenir el sobreajuste, se dividió el conjunto de datos en conjuntos de entrenamiento (80%), prueba (20%) y validación (10%).

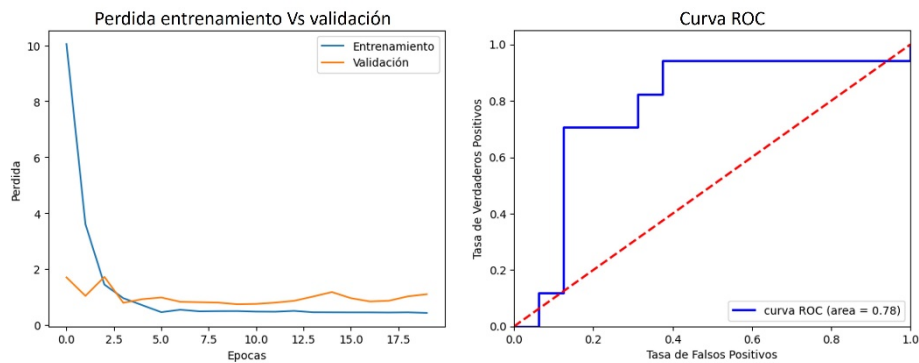
### 3 Resultados.

La implementación de una CNN pre entrenada VGG-16 para la clasificación de sonidos respiratorios ha arrojado resultados prometedores. Al alcanzar una exactitud del 78% en la clasificación de sonidos respiratorios en lactantes, niños y adolescentes, este modelo ha demostrado una capacidad significativa para distinguir entre diferentes patrones respiratorios, lo que sugiere su viabilidad en aplicaciones clínicas. La matriz de confusión en la **Tabla.1**, se puede observar que, del total del conjunto de prueba, se obtienen 11 verdaderos negativos (VN), 5 falsos positivos (FP), 3 falsos negativos (FN) y 14 verdaderos positivos (VP). Estos resultados destacan 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
<i>Real</i>	Clase 0	<b>VN</b>	<b>FP</b>
		11	5
	Clase 1	<b>FN</b>	<b>VP</b>
		3	14
		<i>Predicción</i>	

A pesar de que se han observado algunas confusiones entre clases específicas, la mayoría de las predicciones coinciden con las etiquetas verdaderas, lo que respalda la robustez del modelo. El análisis de la curva de pérdida junto con la ROC **Fig. 3** inicialmente indica que el modelo no sufre ni de sobreajuste ni de subajuste. No obstante, es importante destacar que el modelo alcanza un valor óptimo en menos de 20 épocas, lo que sugiere una eficiencia en el proceso de entrenamiento. Sin embargo, se debe ejercer cautela al interpretar esta curva, ya que la falta de datos podría estar influyendo en los resultados. El valor obtenido, 0.78, refleja un buen desempeño del modelo en la clasificación de los sonidos respiratorios, indicando una tasa aceptable de verdaderos positivos (sensibilidad) en comparación con la tasa de falsos positivos (1 - especificidad). En otras palabras, el modelo es capaz de distinguir eficazmente entre los sonidos respiratorios de pacientes con y sin patología respiratoria en la mayoría de los casos. Además, se realizaron análisis adicionales para comprender mejor el desempeño del modelo en diferentes subconjuntos de datos, como la edad de los pacientes, el tipo de afección respiratoria y la ubicación torácica de los sonidos auscultados. Estos análisis ofrecen información valiosa sobre la generalización del modelo en diversos contextos clínicos y pueden orientar futuras mejoras en su arquitectura.

**Fig. 3.** Graficas representativas del rendimiento del modelo.

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 po-

tencial 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 para el análisis del sonido pulmonar. Hasta la fecha de esta publicación, sólo tres estudios [23 24 25] han empleado MFCC's para dos tareas médicas distintas: la detección de sonidos anormales (ASD, por su sigla en inglés) y el reconocimiento de enfermedades respiratorias (RDR, por su sigla en inglés). Destacamos que nuestro conjunto de datos, centrado en lactantes, niños y adolescentes, es el primero en su clase en utilizar este método para extraer características al analizar archivos de audio de auscultación. Es crucial señalar que, aunque el valor de 78% sugiere un buen rendimiento del modelo, aún existe margen para mejorar su capacidad de discriminación. Estas mejoras podrían lograrse mediante ajustes en los parámetros del modelo, la optimización de la arquitectura de la red neuronal, o la incorporación de más datos de entrenamiento para fortalecer su capacidad predictiva en diversos escenarios clínicos.

#### **5 Conclusiones.**

Los resultados obtenidos en este estudio respaldan la eficacia de la implementación de la red neuronal convolucional (CNN) pre entrenada VGG-16 en la clasificación de sonidos respiratorios en lactantes, niños y adolescentes, con una precisión del 78%. Estos hallazgos son prometedores y sugieren que la aplicación de la inteligencia artificial en el análisis automatizado de sonidos respiratorios tiene el potencial de mejorar significativamente el diagnóstico y la gestión de enfermedades pulmonares en este grupo etario.

La capacidad del modelo para distinguir entre diferentes patrones respiratorios, incluso en presencia de ruido ambiental, destaca su robustez y su utilidad potencial en entornos clínicos. La matriz de confusión revela un buen rendimiento general del modelo, con algunas confusiones que podrían ser abordadas mediante una mayor optimización y refinamiento del algoritmo. Además, los análisis adicionales realizados para explorar el desempeño del modelo en diferentes subconjuntos de datos proporcionan información valiosa sobre su generalización en diferentes contextos clínicos. Estos hallazgos son fundamentales para comprender los posibles sesgos y limitaciones del modelo y pueden guiar futuras mejoras en su arquitectura y aplicación. En resumen, este estudio subraya el potencial de la inteligencia artificial, específicamente de las redes neuronales convolucionales pre entrenadas, en la mejora del diagnóstico y la gestión de enfermedades respiratorias a través del análisis automatizado de sonidos respiratorios.

Estas investigaciones tienen potenciales aplicaciones en el ámbito médico, particularmente en el diagnóstico y monitoreo remoto de pacientes pediátricos y adultos con enfermedades respiratorias, los usos futuros también pueden incluir la evaluación de habilidades y el examen de estudiantes y profesionales del personal sanitario.

**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.