

Estimación del volumen pulmonar a partir de bandas inductivas mediante modelos basados en redes neuronales

^aJosé R. Ortega, ^aFidel A. Ortega, ^aArturo Sotelo Orozco, ^aLeonardo Trujillo, ^aYazmin Maldonado, ^bJorge Tovar-Díaz.

^aTecnológico Nacional de México/IT Tijuana. Tijuana, Baja California, México, {jose.or, fidel.or, asotelo, leonardo.trujillo, yaz.maldonado}@tectijuana.edu.mx

^bUniversidad Autónoma de Baja California. Tijuana, Baja California, México, jorge.tovar.diaz@uabc.edu.mx

Resumen

El análisis de la función pulmonar es esencial para evaluar la salud respiratoria, siendo la espirometría el método estándar. Sin embargo, su uso continuo es limitado por la necesidad de equipos especializados. Las bandas inductivas ofrecen una alternativa no invasiva al medir la expansión torácica, aunque su conversión a valores espirométricos requiere procesamiento adicional. En este trabajo, se emplearon modelos de redes neuronales para transformar la señal de la banda inductiva en una señal equivalente a la de espirometría, con el objetivo de estimar el volumen pulmonar con alta precisión sin ajustes manuales. Se recopilaron 1180 mediciones en voluntarios sanos con cuatro pruebas respiratorias distintas. Las señales fueron procesadas mediante filtrado, derivación, integración y alineación temporal. Posteriormente, se entrenaron modelos basados en redes neuronales utilizando validación cruzada. Se evaluó el desempeño de los modelos, concluyendo que un modelo de red neuronal convolucional (CNN, por sus siglas en inglés) logró el mejor rendimiento. Los resultados mostraron que la integral de la espirometría tiene una alta correlación con la señal de la banda inductiva escalada, validando el enfoque propuesto. Como trabajo futuro, se plantea comparar el desempeño de redes neuronales con otros enfoques de aprendizaje automático, como modelos basados en árboles de decisión o modelos basados en distancias. Además, se explorarán implementaciones en hardware embebido para aplicaciones en monitoreo respiratorio en tiempo real.

Palabras clave — Banda inductiva, Capacidad pulmonar, Espirometría, Procesamiento de señales, Redes Neuronales.

Abstract

Pulmonary function analysis is essential for assessing respiratory health, with spirometry being the standard method. However, its continuous use is limited by the need for specialized equipment. Inductive bands offer a non-invasive alternative by measuring thoracic expansion, although their conversion to spirometric values requires additional processing. In this study, neural network models were employed to transform the inductive band signal into an equivalent spirometry signal, aiming to estimate lung volume with high precision without manual adjustments. A total of 1,180 measurements were collected from healthy volunteers across four different respiratory tests. The signals were

processed through filtering, differentiation, integration, and temporal alignment. Subsequently, neural network-based models were trained using cross-validation. Performance was evaluated, concluding that a convolutional neural network (CNN) achieved the best performance. The results showed that the integral of the spirometry signal exhibits a high correlation with the scaled inductive band signal, validating the proposed approach. As future work, the performance of neural networks will be compared with other machine learning approaches, such as decision tree-based models and distance-based models. Additionally, implementations in embedded hardware will be explored for real-time respiratory monitoring applications.

Keywords — Inductive band, Lung capacity, Neural networks, Signal processing, Spirometry.

1. INTRODUCCIÓN

La evaluación de la función pulmonar es fundamental en el diagnóstico y monitoreo de enfermedades respiratorias como la enfermedad pulmonar obstructiva crónica (EPOC), el asma y la fibrosis pulmonar. La espirometría es la herramienta estándar utilizada en la clínica para medir parámetros clave como la capacidad vital forzada (FVC, por sus siglas en inglés) y el volumen espiratorio forzado en un segundo (FEV1, por sus siglas en inglés). Sin embargo, su uso continuo y fuera del entorno clínico se ve limitado por la necesidad de equipos especializados, la calibración frecuente y la dependencia de la cooperación del paciente [1].

En diversos escenarios, como la monitorización remota de pacientes con enfermedades pulmonares o la evaluación del desempeño respiratorio en atletas, se requiere una alternativa que permita medir la función pulmonar de manera no invasiva. Como alternativa, las bandas inductivas registran los cambios en la expansión torácica a partir de variaciones en la inductancia, lo que las convierte en una opción prometedora para evaluar la función pulmonar de manera no invasiva. [2]. No obstante, la conversión de esta señal en datos equivalentes a los obtenidos por espirometría representa un desafío debido a diferencias en amplitud y frecuencia.

Este trabajo explora el uso de modelos de aprendizaje automático basados en redes neuronales para transformar la señal de la banda inductiva en una señal equivalente a la espirometría. Se plantea la hipótesis de que, mediante un adecuado procesamiento de señales y la implementación de modelos de aprendizaje automático sobre las señales inductivas, es posible estimar el volumen respiratorio con alta precisión. El resto del documento se organiza de la siguiente manera: en la Sección 2 se describe la metodología seguida, incluyendo la adquisición y el procesamiento de señales, la selección de modelos de aprendizaje automático y los resultados obtenidos. Finalmente, en la Sección 3, se discuten las conclusiones del estudio y se proponen líneas de trabajo futuro.

2. DESARROLLO METODOLÓGICO

En este trabajo se desarrolló una metodología para la construcción de una base de datos de señales respiratorias y su posterior análisis mediante modelos de aprendizaje automático. El proceso, representado en la Figura 1, inicia con la selección de voluntarios y la adquisición de datos mediante un espirómetro y una banda inductiva, registrando señales en distintas condiciones respiratorias. Una vez recolectadas, las señales son procesadas mediante filtrado, derivación, integración y alineación temporal.

Posteriormente, los datos preprocesados se emplean para el entrenamiento de modelos de aprendizaje automático basados en redes neuronales con el objetivo de transformar las señales de la banda inductiva en valores equivalentes a los de la espirometría. Finalmente, se selecciona el modelo con mejor desempeño y se valida su funcionamiento utilizando datos no empleados en el entrenamiento.

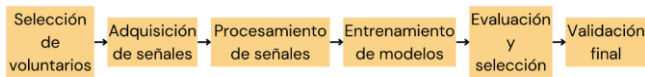


Fig. 1. Diagrama de bloques del proceso metodológico.

2.1 Construcción de la base de datos

Las mediciones obtenidas de los voluntarios se realizaron utilizando un espirómetro y una banda inductiva en cuatro tipos de pruebas respiratorias, cada una diseñada para evaluar la función pulmonar en diferentes condiciones (respiración en reposo, espiración forzada en reposo, espiración forzada después del esfuerzo, respiración después del esfuerzo). En total, se obtuvieron 1180 registros distribuidos de la siguiente manera:

- Prueba de respiración en reposo P1: respiración normal, sin esfuerzo previo, con el objetivo de obtener un patrón de ventilación en estado de reposo.
- Prueba de espiración forzada en reposo P2: exhalación máxima y rápida desde la capacidad pulmonar total hasta el volumen residual, evaluando la capacidad de expulsión de aire en ausencia de fatiga.
- Prueba de espiración forzada después del esfuerzo P3: similar a la Prueba P2, pero realizada después de hacer 2 minutos de ejercicio moderado, permitiendo analizar los cambios en la ventilación inducidos por el esfuerzo
- Prueba de respiración después del esfuerzo P4: respiración normal tras 2 minutos de actividad física moderada.

La Figura 2 presenta un ejemplo de medición obtenida en una Prueba P1, mientras que la Figura 3 muestra una medición correspondiente a una Prueba P2. Aunque las Pruebas P3 y P4 presentan diferencias en la frecuencia respiratoria debido a la actividad física realizada, sus patrones visuales son similares a los de las Pruebas P1 y P2.

2.2 Procesamiento de las señales

Las señales adquiridas presentan variabilidad en su amplitud y frecuencia debido a factores como el movimiento del sujeto, interferencias del entorno y la presencia de componentes no relevantes para el análisis respiratorio. En particular, las señales respiratorias de interés suelen encontrarse en un rango de frecuencias por debajo de los 5 Hz, mientras que el ruido y artefactos pueden abarcar un espectro más amplio.

Para mejorar la calidad de las señales y aislar las componentes de interés dentro del rango respiratorio, se aplicó un filtro Butterworth pasa banda de cuarto orden [3], configurado con una frecuencia de corte baja de 0.1 Hz y una frecuencia de corte alta de 5 Hz. La función de transferencia del filtro utilizado está dada por

$$H(z) = \frac{b_0 + b_1 z^{-1} + b_2 z^{-2} + b_3 z^{-3} + b_4 z^{-4}}{a_0 + a_1 z^{-1} + a_2 z^{-2} + a_3 z^{-3} + a_4 z^{-4}} \quad [1]$$

donde b_i son los coeficientes del numerador que definen la respuesta en frecuencia del filtro, a_i son los coeficientes del denominador, que determinan la estabilidad del sistema, z es la variable compleja en el dominio Z que representa la transformada Z de la señal y z^{-n} es el retardo de n muestras en el dominio discreto.

Una vez aplicado el filtro, la señal de flujo del espirómetro, que representa la tasa de cambio del volumen inhalado y exhalado durante la respiración, fue integrada para obtener la curva de volumen acumulado, fundamental en la evaluación de la capacidad respiratoria. Por otro lado, la señal de la banda inductiva, relacionada con el cambio en la expansión torácica, fue derivada temporalmente para generar una señal equivalente a la tasa de cambio de la expansión.

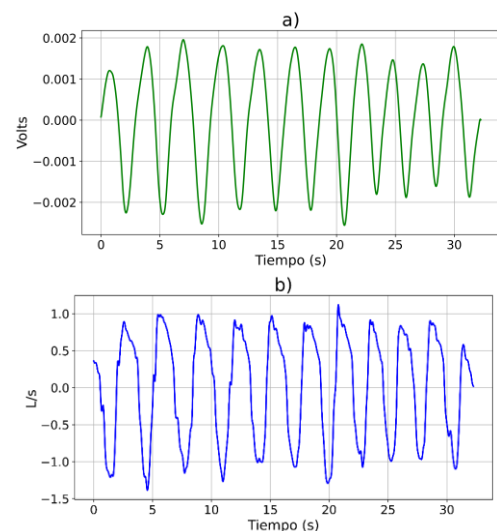


Fig. 2. Series de tiempo para la prueba de respiración normal, a) Señal de la banda inductiva, b) Señal de espirometría.

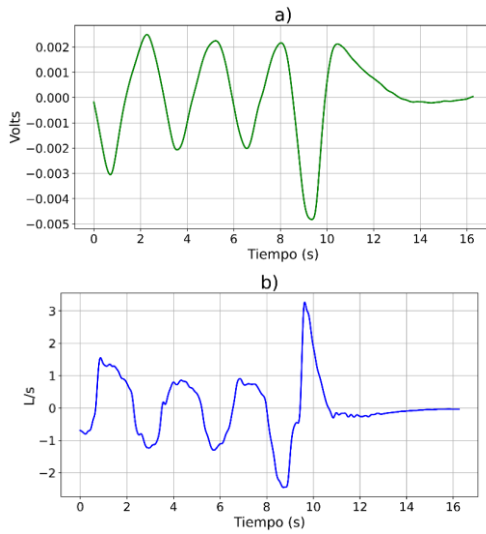


Fig. 3. Series de tiempo para la prueba de respiración forzada, a) Señal de la banda inductiva, b) Señal de espirometría.

Tras aplicar las operaciones de derivación e integración, se ajustó la escala vertical en la señal inductiva (en la señal original y en su derivada) para igualar la amplitud de ambas señales, facilitando su comparación. Posteriormente, se calculó el *lag* o retraso temporal óptimo entre ambas señales utilizando la ecuación de la correlación cruzada [4] dada por

$$R_{xy}[k] = \sum_{i=-\infty}^{\infty} x[i]y[i+k] \quad [2]$$

donde $R_{xy}[k]$ es la correlación cruzada en el desplazamiento k , $x[i]$ y $y[i]$ son las señales de entrada y k es el desplazamiento entre las señales.

Este *lag* se aplicó para corregir la posición de las señales en el tiempo, logrando una superposición precisa entre los picos y valles correspondientes. Las Figuras 4 y 5 muestran la comparación entre las series de tiempo procesadas para las Pruebas P1 y P2, mientras que los resultados para las Pruebas P3 y P4 muestran un comportamiento visualmente similar. Se observa que la derivada de la banda inductiva escalada y la señal de espirometría son muy similares, aunque con ciertas variaciones debido a que la derivación amplifica cambios pequeños en la señal. Por otro lado, la integral de la espirometría y la señal de la banda inductiva escalada muestran una mayor similitud, ya que la integración suaviza las fluctuaciones menores. Además, el error residual, representado en color rojo, se mantiene bajo en ambos casos, lo que confirma la similitud entre las señales después del procesamiento realizado.

Para analizar la similitud en términos de frecuencia entre las señales procesadas, se aplicó la transformada de Fourier discreta [5] definida como

$$X[k] = \sum_{i=0}^{N-1} x[i]e^{-j2\pi ki/N}, \quad k = 0, \dots, N-1 \quad [3]$$

donde N es el número total de muestras, $X[k]$ es el valor de la transformada en la frecuencia k y $x[n]$ es la señal en el dominio del tiempo.

Los espectros de frecuencia obtenidos mediante la Ecuación 3 se presentan en la Figura 6. En estos espectros se observan los cambios en las componentes de frecuencia de las señales de la banda inductiva y la espirometría, así como las modificaciones resultantes tras la derivación e integración aplicadas previamente. De manera visual, se observa que el espectro de la señal de la banda inductiva derivada se aproxima al espectro de la señal de espirometría. Por otro lado, el espectro de la integración de la señal de espirometría muestra una gran similitud con el de la banda inductiva, lo que refuerza la relación entre ambas señales.

En el procesamiento de señales, una estrategia común es la normalización a un rango fijo, como $[0,1]$, para reducir la variabilidad en amplitud y facilitar el entrenamiento de modelos de aprendizaje automático. Sin embargo, en este caso, la relación de amplitud entre la señal de la banda inductiva y la de espirometría variaba significativamente entre pruebas. En algunas mediciones, la diferencia de escala era de un factor de 100, mientras que en otras llegaba hasta 1000, lo que impedía definir un ajuste único para todas las señales. Por este motivo, se optó por mantener la escala original de las señales. No obstante, al no encontrarse una relación de escalamiento consistente entre ambas señales, no fue posible aplicar un ajuste directo de amplitud, lo que motivó la exploración de algoritmos basados en aprendizaje automático.

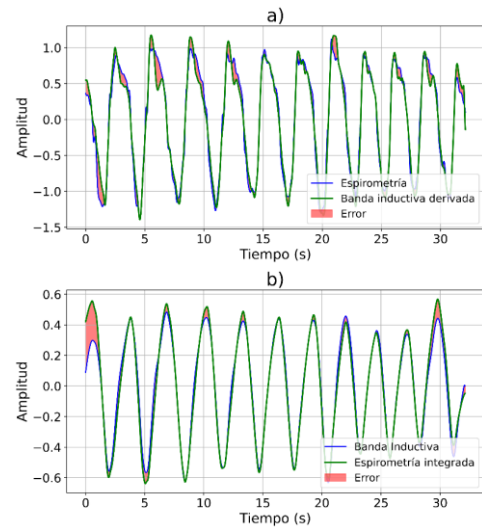


Fig. 4. Comparación entre series de tiempo procesadas para la prueba de respiración normal, a) Señal de espirometría y derivada de la banda inductiva escalada, b) Señal de la banda inductiva escalada e integral de la señal de espirometría.

2.3 Algoritmos de aprendizaje automático

Debido a la imposibilidad de determinar un escalamiento constante y que aún persistían errores que no podían corregirse manualmente, se optó por el uso de algoritmos de

aprendizaje automático basados en redes neuronales [6]. Estos modelos fueron seleccionados debido a su capacidad para capturar patrones complejos en series de tiempo.

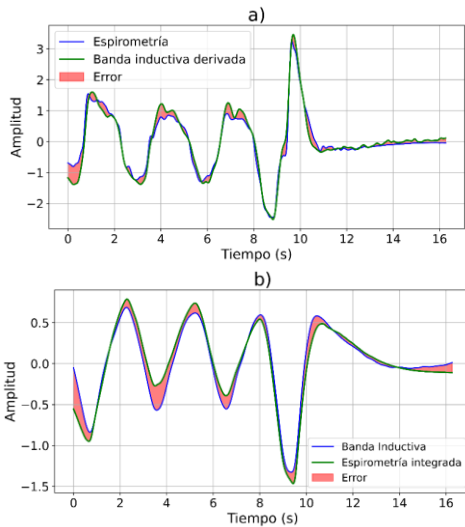


Fig. 5. Comparación entre series de tiempo procesadas para la prueba de respiración forzada, a) Señal de espirometría y derivada de la banda inductiva escalada, b) Señal de la banda inductiva escalada e integral de la señal de espirometría.

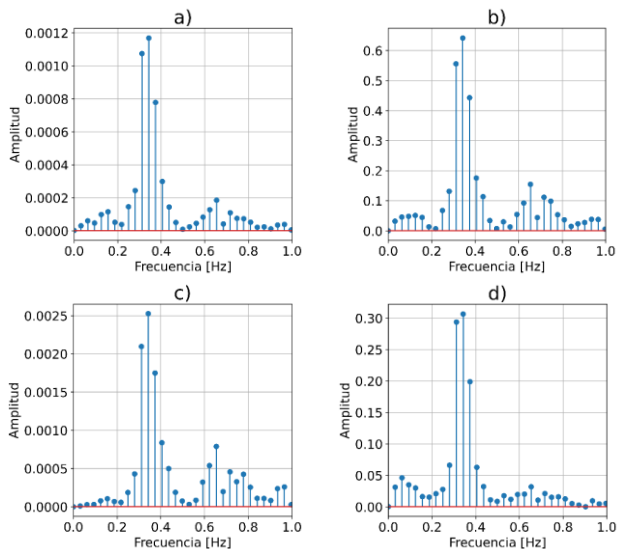


Fig. 6. Espectro de las series de tiempo para la Prueba P1, a) Banda inductiva, b) Espirometría, c) Derivada de la banda inductiva, d) Integral de la Espirometría.

El objetivo del modelo es transformar una señal de entrada en una señal de salida. Para ello, se estableció como entrada la señal de la banda inductiva sin escalar y como salida la integral de la espirometría. Esta elección se realizó debido a dos razones principales: evita la necesidad de calcular la derivada antes de la entrada del modelo y la integral de la espirometría representa volumen, una magnitud fundamental en la evaluación de la capacidad pulmonar [7]. Para la conversión de señales de la banda inductiva a señales equivalentes de espirometría, se exploraron distintas

arquitecturas de redes neuronales. Todas las arquitecturas fueron implementadas utilizando *TensorFlow/Keras*, excepto el modelo Transformer, que fue desarrollado en *PyTorch*. A continuación, se describen las principales arquitecturas utilizadas:

- Autocodificador: Aprende representaciones comprimidas de las señales, facilitando la reducción de ruido y la extracción de características clave [8].
- CNN: Detecta patrones locales en la señal respiratoria, permitiendo la identificación de estructuras relevantes en el tiempo [9].
- Autocodificador + CNN: Combina la capacidad de compresión del autocodificador con la detección eficiente de características mediante convoluciones.
- InceptionTime: Captura relaciones en múltiples escalas temporales, lo que permite un análisis más detallado de la dinámica de la respiración [10].
- Redes Neuronales Líquidas (LNN, por sus siglas en inglés): Modelo adaptable que capta dinámicas complejas y cambios en tiempo real [11].
- Memoria a Largo Corto Plazo (LSTM, por sus siglas en inglés): Modela secuencias largas y dependencias temporales [12].
- ResNet: Utiliza conexiones residuales para mejorar la extracción de características profundas sin perder información crítica [9].
- Transformer: Modelo para la captura de relaciones globales dentro de la serie temporal, permitiendo una modelación más robusta [12].
- Unidad Recurrente Gated (GRU, por sus siglas en inglés): Variante de LSTM, capaz de modelar dependencias temporales con menor costo computacional [13].

2.4 Entrenamiento de los modelos

Para la evaluación del modelo, se asignó el 5% de las señales como conjunto de prueba. Para el entrenamiento, se aplicó un ventaneo de las señales, dividiéndolas en secciones de 50 muestras con un desplazamiento de 25 muestras. Además, se implementó una validación cruzada con 19 pliegues [14], donde en cada iteración, un 5% de los datos fue utilizado para validación, mientras que el 90% restante se empleó para el entrenamiento del modelo. Como resultado, se utilizaron en promedio 122,274 ventanas en el proceso de entrenamiento.

2.5 Desempeño de los modelos

Durante el entrenamiento del modelo, en cada uno de los pliegues de la validación cruzada, se calcularon las métricas de error absoluto medio (MAE, por sus siglas en inglés), correlación de Spearman (ρ) y el coeficiente de determinación (R^2) [15]. Estas métricas permitieron evaluar el desempeño del modelo en términos de precisión, tendencia y ajuste a los valores reales. El MAE se calculó utilizando la siguiente expresión

$$MAE = \frac{1}{N} + \sum_{i=1}^N |y_i - \hat{y}_i| \quad [4]$$

donde y_i es el valor real de la muestra i y \hat{y}_i es el valor predicho en la muestra i . Por otro lado, la correlación de Spearman se definió como

$$\rho = 1 - \frac{6 \sum_{i=1}^N d_i^2}{N(N^2-1)} \quad [5]$$

donde d_i es la diferencia entre los rangos de los valores reales y predichos en la muestra i . Finalmente, la R^2 se calculó mediante

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \bar{y})^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad [6]$$

donde \bar{y} es el valor medio de las muestras.

En la Figura 7 se muestran los valores promedio de las métricas calculados a partir de los resultados obtenidos en la validación cruzada, para cada uno de los modelos. Basado en estos resultados, se construyó la Tabla 1, que muestra un resumen del mejor modelo (CNN), el peor (Autocodificador+CNN) y un modelo con resultados regulares (LSTM). Como se puede observar, CNN mostró el mejor desempeño general.

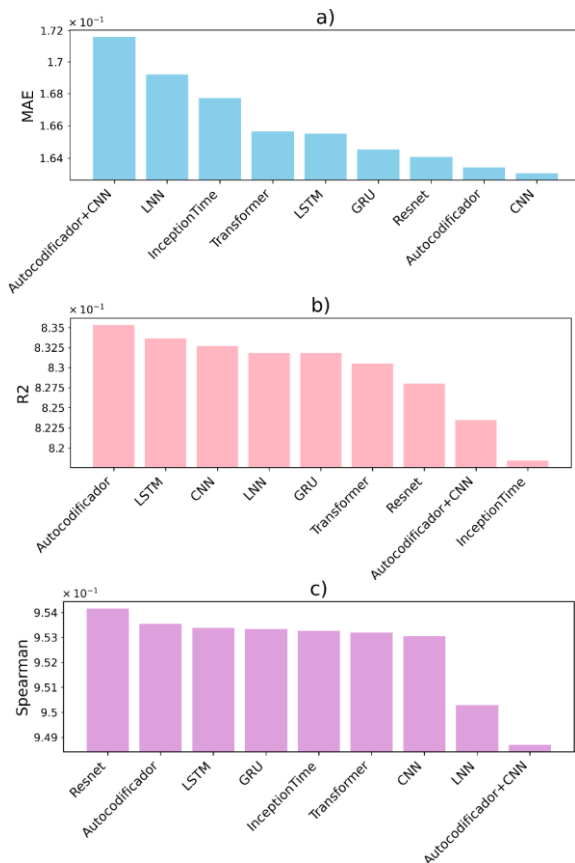


Fig. 7. Métricas de desempeño de los modelos entrenados, a) MAE, b) R^2 , c) Correlación de Spearman.

2.6 Predicciones utilizando el mejor modelo

Se evaluó el desempeño del modelo CNN utilizando el conjunto de prueba. En la Figura 8 se muestra un gráfico de dispersión que compara las predicciones obtenidas con los valores reales para las distintas pruebas. Se observa que los mejores resultados se obtuvieron en las Pruebas P1 y P4, cuyos puntos se encuentran más cerca de la diagonal. En contraste, las Pruebas P2 y P3 muestran una mayor dispersión, lo que sugiere un menor ajuste del modelo.

Esta diferencia podría estar relacionada con el espectro de las señales. Las Pruebas P2 y P3 implican una exhalación forzada donde los voluntarios expulsan el aire lo más rápido posible, por lo que contienen componentes de frecuencia más altas. Es posible que el modelo no haya capturado con precisión estos cambios rápidos en la señal, lo que explica el mayor error en estas pruebas. Por otro lado, la Figura 9 muestra el error absoluto por cada ventana utilizada en la generación de predicciones. De manera consistente con los resultados anteriores, se observa un mayor error en las Pruebas P2 y P3, mientras que la Prueba P4 muestra un error consistentemente bajo.

Tabla 1. Comparación de modelos.

Modelo	Métricas		
	MAE	Spearman	R^2
CNN	0.1631	0.9530	0.8327
LSTM	0.1655	0.9534	0.8336
Autocodificador+CNN	0.1716	0.9486	0.8234

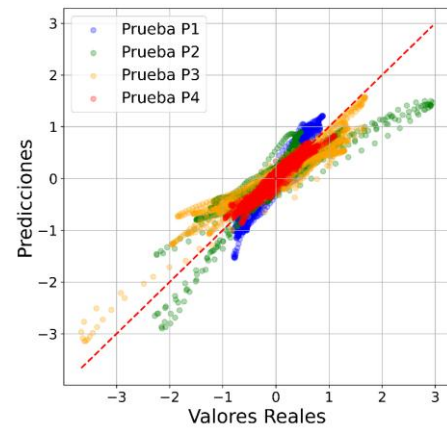


Fig. 8. Comparación de Predicciones vs Valores Reales para las distintas pruebas utilizando CNN.

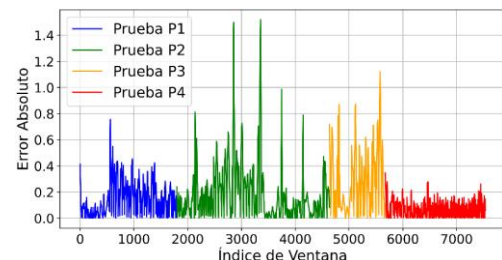


Fig. 9. Error absoluto por ventana para las distintas pruebas utilizando CNN.

3. CONCLUSIONES

En este trabajo, se abordó la conversión de señales obtenidas con una banda inductiva en datos equivalentes a los de la espirometría mediante el uso de aprendizaje automático. Se implementó un procesamiento de señales que incluyó filtrado, derivación, integración y alineación temporal, seguido de un entrenamiento con modelos basados en redes neuronales para regresión de series de tiempo. Los resultados obtenidos confirmaron que la integral de la espirometría presenta una mayor correspondencia con la señal de la banda inductiva escalada, lo que permite estimar el volumen pulmonar con mayor precisión.

Dado que el objetivo principal del modelo es la estimación del volumen pulmonar, se optó por utilizar la integral de la espirometría como variable de salida. En la práctica clínica, el flujo medido en la espirometría se utiliza principalmente para calcular el volumen pulmonar, ya que este último es la magnitud de mayor relevancia en la evaluación de la función respiratoria. Como el modelo aprende a predecir directamente una señal de volumen, esta estrategia podría ser más eficiente y alineada con los objetivos clínicos. En este sentido, modelar directamente la integral de la espirometría proporciona una representación fisiológicamente más relevante que depender de la estimación intermedia del flujo, eliminando la necesidad de realizar ajustes manuales o suposiciones sobre la relación entre ambas señales.

El modelo basado en CNN obtuvo el mejor desempeño general, logrando minimizar los errores y capturar con mayor precisión la estructura de las señales respiratorias. Sin embargo, se identificaron desafíos en la modelización de señales con componentes de alta frecuencia, particularmente en las pruebas de espiración forzada, lo que sugiere la necesidad de explorar arquitecturas más especializadas en la captura de transiciones rápidas.

Como parte del trabajo futuro, se consideran las siguientes mejoras y aplicaciones:

- Optimización del modelo para mejorar la precisión en señales de alta frecuencia.
- Implementación en hardware embebido que permita el procesamiento en tiempo real en dispositivos vestibles.
- Validación en poblaciones más amplias, incluyendo sujetos con patologías respiratorias, para evaluar la robustez del modelo.
- Comparación de los resultados obtenidos con redes neuronales frente a otros enfoques, como modelos basados en árboles de decisión o modelos basados en distancias, con el fin de evaluar si estos métodos pueden ofrecer ventajas en términos de desempeño.

4. REFERENCIAS

- [1] J.-C. Vázquez-García, J. Salas-Hernández, R. Pérez Padilla, y M. Montes de Oca, "Respiratory Health in Latin America: Number of Specialists and Human Resources Training," *Archivos de Bronconeumología (English Edition)*, vol. 50, no. 1, pp. 34-39, 2014. DOI: 10.1016/j.arbr.2013.12.003.
- [2] J. M. Marín Trigo, "Pletismografía inductiva en la monitorización respiratoria," *Archivos de Bronconeumología*, vol. 24, no. 2, pp. 78-80, 1988. doi: 10.1016/s0300-2896(15)31861-5.
- [3] S. Butterworth, "On the theory of filter amplifiers," *Exp. Wireless & Wireless Eng.*, vol. 7, no. 10, pp. 536-541, Oct. 1930.
- [4] C. C. Botero Suárez, E. T. Fonoff, G. Ballester, and F. J. Ramírez-Fernández, "Previsión de crisis epilépticas usando transformada wavelet y correlación cruzada," *Rev. Ing. Univ. Medellín*, vol. 6, no. 11, pp. 223-232, Jul.-Dec. 2007.
- [5] A. V. Oppenheim and A. S. Willsky, *Signals and Systems*, 2nd ed. Upper Saddle River, NJ, USA: Prentice Hall, 1996.
- [6] N. Sáenz Bajo and M. Álvaro Ballesteros, "Redes neuronales: concepto, aplicaciones y utilidad en medicina," *Aten. Primaria*, vol. 30, no. 2, pp. 119-120, 2002.
- [7] D. Rivero-Yeverino, "Espirometría: conceptos básicos," *Rev. Alerg. Méx.*, vol. 66, no. 1, pp. 76-84, 2019. doi: 10.29262/ram.v66i1.53.
- [8] P. Li, Y. Pei, and J. Li, "A comprehensive survey on design and application of autoencoder in deep learning," *Appl. Soft Comput.*, vol. 138, p. 110176, May 2023. doi: 10.1016/j.asoc.2023.110176.
- [9] H. I. Fawaz, G. Forestier, J. Weber, L. Idoumghar, and P.-A. Muller, "Deep learning for time series classification: a review," *Data Min. Knowl. Discov.*, vol. 33, pp. 917-963, 2019. doi: 10.1007/s10618-019-00619-1.
- [10] H. I. Fawaz et al., "InceptionTime: Finding AlexNet for time series classification," *Data Min. Knowl. Discov.*, vol. 34, pp. 1936-1962, 2020. doi: 10.1007/s10618-020-00710-y.
- [11] R. Hasani, M. Lechner, A. Amini, D. Rus, and R. Grosu, "Liquid Time-constant Networks," in *Proc. AAAI Conf. Artif. Intell.*, vol. 35, no. 9, pp. 7657-7666, 2021. [Online]. Available: arXiv:2006.04439.
- [12] G. Ammann, "Using LSTMs and Transformers to Forecast Short-term Residential Energy Consumption," M.Sc. thesis, School of Humanities and Digital Sciences, Tilburg Univ., Netherlands, Jun. 2022.
- [13] K. Cho, B. van Merriënboer, C. Gulcehre, F. Bougares, H. Schwenk, D. Bahdanau, and Y. Bengio, "Learning phrase representations using RNN Encoder-Decoder for statistical machine translation," *arXiv preprint, arXiv:1406.1078*, Sep. 2014.
- [14] L. Laura-Ochoa, "Evaluation of Classification Algorithms using Cross Validation," in *Proc. 17th LACCEI Int. Multi-Conf. Eng. Educ. Technol.*, Jamaica, Jul. 2019. doi: 10.18687/LACCEI2019.1.1.471.
- [15] O. Rainio, J. Teuho, and R. Klén, "Evaluation metrics and statistical tests for machine learning," *Sci. Rep.*, vol. 14, no. 6086, 2024. doi: 10.1038/s41598-024-56706-x.